Evaluation of an Auditory Localization Training System for Use in Portable Configurations: Variables, Metrics and Protocol

Kara Meghan Cave

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Industrial and Systems Engineering

John G. Casali, Chair
Kristen L. Casto
Joseph L. Gabbard
Brian M. Kleiner
Kichol Lee

December 10, 2019
Blacksburg, Virginia

Keywords: Auditory Localization, Auditory Situation Awareness, Auditory Training,
Hearing Protection, Noise Exposure

Evaluation of an Auditory Localization Training System for Use in Portable Configurations: Variables, Metrics and Protocol

Kara M. Cave

ABSTRACT

Hearing protection can mitigate the harmful effects of noise, but for Service Members these devices can also obscure auditory situation awareness cues. Tactical Communication and Protective Systems (TCAPS) can restore critical cues through electronic circuitry with varying effects on localization. Evidenced by past research, sound localization accuracy can improve with training. The investigator hypothesized that training with a broadband stimulus and reducing the number of presentations would result in training transfer. Additionally, training transfer would occur with implementation of more user-engaged training strategies. The purpose of the experiments described in this study was to develop an optimized auditory azimuth-training protocol for use in a field-validated portable training system sensitive to differences among different TCAPS.

A series of indoor experiments aimed to shorten and optimize a pre-existing auditory localization training protocol. Sixty-four normal-hearing participants underwent localization training. The goal of training optimization included the following objectives: 1) evaluate the effects of reducing stimulus presentations; 2) evaluate the effects of training with a broadband stimulus (but testing on untrained military-relevant stimuli); and 3) evaluate performance differences according to training strategies.

Twenty-four (12 trained and 12 untrained) normal-hearing listeners participated in the field-validation experiment. The experiment evaluated localization training transfer from the indoor portable system to live-fire blanks in field. While training conducted on the portable system was predicted to transfer to the field, differences emerged between an in-the-ear and over-the-ear TCAPS. Three of four untrained stimuli showed evidence of training transfer. Shortening the training protocol also resulted in training transfer, but manipulating training strategies did not. A comparison of changes in localization scores from the indoor pretest to the field posttest demonstrated significant differences among listening conditions. Training improved accuracy and response time for the open ear and one of two TCAPS. Posttest differences between the two TCAPS were not statistically significant.

Despite training, localization with TCAPS never matched the open ear. The portable apparatus employed in this study offers a means to evaluate the effects of TCAPS on localization. Equipped with a known effect on localization, TCAPS users can render informed decisions on the benefits or risk associated with certain devices.

Evaluation of an Auditory Localization Training System for Use in Portable Configurations: Variables, Metrics and Protocol

Kara M. Cave

GENERAL AUDIENCE ABSTRACT

Hearing protection can mitigate the harmful effects of noise, but for Service Members these devices can obscure auditory situation awareness cues. Certain powered hearing protection can restore critical cues through electronic circuitry with varying effects on localization. Evidenced by past research, sound localization accuracy can improve with training. The investigator hypothesized that training with a broadband stimulus and reducing the number of presentations would result in auditory learning. Additionally, implementing more user-engaged training strategies would demonstrate more auditory learning.

The purpose of the experiments described in this study was to develop an optimized auditory azimuth-training protocol for use in a field-validated training system sensitive to differences among active hearing protection.

A series of indoor experiments aimed to shorten and optimize a pre-existing auditory localization training protocol. Sixty-four normal-hearing participants underwent localization training. The goal of training optimization included the following objectives: 1) evaluate the effects of reducing stimulus presentations; 2) evaluate the effects of training with a broadband stimulus (but testing on untrained military-relevant stimuli); and 3) evaluate performance differences in localization performance according to training strategies.

In the field-validation study, 12 trained and 12 untrained normal-hearing listeners participated. The experiment evaluated localization learning from the indoor portable training system to live-fire blanks in a field. Training conducted on the portable system was predicted to transfer to the field, but differences would emerge between an in-the-ear and an over-the-ear TCAPS. Three of four untrained stimuli showed evidence of localization learning. Shortening the protocol also resulted in localization learning, but manipulating training strategies did not. A comparison of changes in localization scores from the indoor pretest to the field posttest demonstrated significant differences among listening conditions. Training improved performance for the open ear and one of two active hearing protectors. Posttest differences between the two devices were not significant.

Despite training, performance with hearing protection never equaled the open ear. The portable apparatus employed in this study offers a means to evaluate the effects of hearing protection on localization. Knowing the effects of hearing protection on localization apprises users of the benefits and/or risk associated with the use of certain devices.

ACKNOWLEDGMENTS

I wish to extend my deepest gratitude to Dr. John G. Casali for his belief in my abilities as a researcher and professional. He ensured the highest level of preparedness for a career in research and for that, I am grateful. With great pleasure, I also wish to extend my gratitude to the committee members Colonel (Dr.) Kristen L. Casto, Dr. Joseph L. Gabbard, Dr. Brian M. Kleiner, and Dr. Kichol Lee. The committee remained steadfast in their encouragement and challenged me to think more critically.

The combined programming, hardware design and experimental design support of Dr. Lee and Lieutenant Colonel Brandon Thompson made the series of experiments possible. Their patience and support were invaluable. They helped me realize that truly impactful research is a collaborative effort.

The experiments discussed herein were supported in part by funding from the Office of Naval Research, Noise-Induced Hearing Loss Program, with contract monitors Kurt Yankankas and Kristy Hentchel. The view expressed herein are those of the author, and do not represent those of any branch of the United States military.

My family and friends deserve a special recognition for their years of love and support that pushed me to the finish line. The person who sacrificed the most in this process was my husband, Josh, who made sure “headquarters” ran smoothly and ensured Max and Molly flourished throughout this process. I have a truly exceptional tribe to thank.

PREFACE

To meet the objective of optimizing and field-validating an auditory localization training protocol and system, experiments were conducted in three phases. Given the breadth of this objective, the experiments were covered in two dissertations. This dissertation covered Phase I and sought to improve the training delivered in an auditory localization training protocol. LTC Brandon Thompson (U.S. Army) conducted Phase II whereby the improved training protocol was employed in an experiment to validate a laboratory-grade training system against a portable one. Phase III was the combined effort of both authors of Phase I and Phase II. As such, Phase III in this document is duplicative in this dissertation and that of LTC Thompson’s, and is included with the knowledge and assent of the faculty who comprise both students’ advisory committees.

TABLE OF CONTENTS

ABSTRACT ii

GENERAL AUDIENCE ABSTRACT iv

ACKNOWLEDGMENTS vi

PREFACE vii

TABLE OF CONTENTS viii

GLOSSARY AND ABBREVIATIONS iv

LIST OF FIGURES viii

LIST OF TABLES xxv

INTRODUCTION 1

- Purpose 6

BACKGROUND 8

- Hearing in military environments 8
- Military fitness-for-duty 8
- Hearing loss and auditory situation awareness 10
- Hearing loss and noise 12
- Human hearing as a combat multiplier 13
- Auditory situation awareness defined 14
- Auditory anatomy and physiology 15
  - Outer Ear 15
  - Middle Ear 18
  - Inner ear 19

Auditory localization 19

  Sensation and Perception of Localization 19

  Sound characteristics 24

  Head movements 30

  Distance judgements 30

  Signal duration effects 31

  Intensity effects on localization 31

  Sound/listener movement and Doppler effect 32

  Sound source movement 32

  Environmental effects on localization 33

Hearing Loss and Localization 35

Hearing in Noise 36

  Physiological Effects of Noise on the Cochlea 36

  Non-Auditory Effects of Noise 40

Hearing Protection Devices and Effects on Perception 40

Effects of Hearing Protection on Localization 41

Auditory Learning 45

  Plasticity and Adaptation of the Auditory System 45

  Auditory Training 47

    Auditory Training as a Clinical Rehabilitative Tool 52

    General Training Methodologies 53

Objectives 57

  Summarized Objectives 60

Hypotheses 61

PHASE I: METHODOLOGY 61

Pilot Experiment 1 61

Pilot Experiment 2 64

Main experiment 65

PHASE I: EXPERIMENTAL DESIGN 67

Pilot 1 67

Pilot 2 68

Main experiment 69

Participants 70

Apparatus 71

PHASE I: EXPERIMENTAL PROCEDURES 73

Participant Recruitment 73

Pilots 1 and 2 75

Main Experiment 79

PHASE I: RESULTS 81

Data Reduction Pilots 1 and 2 81

Results for Pilots 1 and 2 84

Pilot 1 and 2 Discussion 90

Pilot 1 and 2 Conclusions 93

Main Experiment Data Reduction for Objective Performance 93

Main Experiment Results for Objective Performance 98

Main Experiment Data Reduction by Location 131

PHASE I: MAIN EXPERIMENT OBJECTIVE MEASURES DISCUSSION AND CONCLUSIONS 136

Objective Measures Discussion... 136

Objective Measures Conclusions... 139

Main Experiment Subjective Performance Data Reduction 140

Main Experiment Subjective Performance Results... 142

Main Experiment Subjective Measures Discussion... 169

Main Experiment Subjective Measures Conclusions 172

Implications from Phase I 172

PHASE III: IN-FIELD INVESTIGATION OF TRANSFER-OF-TRAINING................. 177

Phase III: Objectives 177

Phase III: Methodology 177

Phase III: Experimental Design... 179

Independent Variables (IVs).. 181

Independent Variable – Group.. 181

Open ear.. 182

In-the-ear TCAPS 183

Over-the-Ear TCAPS 186

Independent Variable – Stage of Training 188

Dependent Measures.. 190

Localization accuracy 190

Response time.. 194

            Subjective ratings 195
        Participants 196
    Apparatus 198
        In-Office: PALAT System 198
        In-Field Site 204
            Outdoor Azimuthal Gunshot Presentation System 207
        Outdoor Auditory Localization Data Capture System 211
    Phase III: Experimental Procedures 213
        Recruitment and screening 213
        In-office pretest 213
        Training Session (Experimental group only) 219
        In-field posttest 222
    Phase III: Results 225
        Outlier Analysis 225
        Objective performance 227
            Analysis technique overview 227
            Results: Evaluation of transfer-of-training effects from the in-lab to in-field localization performance 230

Table 85. Mauchly's Test of Sphericity and MANOVA results evaluating the effects of training group, listening condition, and stage of training on absolute score, ballpark score, and response time (bolded text indicates a significant test result at the α=0.05 significance level). 231

            Group Main Effect: Post hoc test for Ballpark Correct Score 233

Listening Condition Main Effect: Post hoc test for Absolute Correct Score 233

Listening Condition Main Effect: Post hoc test for Ballpark Correct Score 234

Listening Condition Main Effect: Post hoc test for Response Time 235

Stage of Training Main Effect: Post hoc test for Response Time 236

Stage of Training x Group Interaction: Post hoc test for Absolute Correct Score. 237

Listening Condition x Stage of Training x Group Interaction: Post hoc test for Absolute Correct Score 239

Front-back Reversal Errors 244

Regression Analysis 246

Subjective ratings 250

Question 1. Perceived Confidence 251

Question 2. Perceived Accuracy 257

Question 3. Perceived Difficulty 263

Question 4. Perceived Reaction Time 268

Question 5. Perceived Comfort 273

Question 6. Likelihood of Wearing Device during a Sound Localization Task 279

Question 7. Degree of Preparedness as Result of PALAT Training 284

Conclusions: In-Field Investigation of Transfer-of-Training 285

Listening Condition Conclusions 286

Training Effect Conclusions 287

Localization performance wearing TCAPS 289

Front-Back Errors 290

Response Time 291

Signal Duration and Head Movement 292

Signal Effect on Localization 294

IMPLICATIONS OF THE RESULTS 306

Limitations of the Research 306

Results explained as a possible function of TCAPS design variables 308

Recommendations for Efficiency and Effectiveness in Training with the Portable Auditory Localization Training System (PALAT) 310

Recommendations for use of the PALAT in various room environments.................. 310

Recommendations for use of the PALAT System as setup and deployed by a trainee 312

Recommendations for use of the PALAT System given time likely available and time likely required 313

Implications for military implementation 314

Relevance to TCAPS design, selection, and procurement... 314

Relevance to ground combat service member duties and mission 316

Implications for NIHL reduction 318

FINAL CONCLUSIONS... 319

REFERENCES.. 321

Appendix A. Email to Subject Matter Experts 338

Appendix B. Participant Recruitment Flyer 340

Appendix C. Phase I Informed Consent 341

Appendix D. Demographic Form 346

Appendix E. Participant Audiogram 347

Appendix F. Pilot 1 Instructions 348

Appendix G. Example of Experimenter Checklist for Pilots 1 and 2 352

Appendix H. Example of Experimenter Checklist for the Main Experiment 353

Appendix I. Questionnaires for the Main Experiment 356

Appendix J. Experimenter Script for the Main Experiment 368

Appendix K. Phase III Questionnaire 379

Appendix L. Phase III Screening Form 382

Appendix M. Remote Firing Device Wiring Diagram 383

Appendix N. Phase III Participant Flyer 384

Appendix O. Phase III Informed Consent 385

Appendix P. Phase III Participant Instructions 396

Appendix Q. Figures of statistically non-significant findings included in qualitative analysis. 403

GLOSSARY AND ABBREVIATIONS

Acoustic beats: Two sound waves interacting to increase the amplitude at certain points in the phase and attenuate the signal at others, creating a perception of a sound fluctuating in loudness.

Active hearing protection: Hearing protection that uses a uses analog circuitry, digital signal processing, or both to prevent hazardous exposures and process other auditory input (Casali, 2010).

Active noise reduction (ANR): Refers to noise reduction strategies through active, or powered, components. Active noise cancellation is a type of ANR that introduces a signal equal to the noise in frequency and amplitude, but opposite in phase of the detected one, summing the total energy of a waveform to zero (Casali, 2010).

Auditory situation awareness (ASA): The ability, through auditory input, to detect, recognize, and localize sound-related events and to anticipate the status of these events in the future (Hajicek, Myrent, Li, Barker, & Coyne, 2010; Endsley M. R., 1988).

Augmented hearing protection: Hearing protection that employs either dynamic mechanical components or electronics to provide protection and access to certain auditory information not otherwise accessible with conventional hearing protection (Casali, 2010).

Circumaural hearing protection: Hearing protection that encloses the outer ear, thereby covering the pinnae.

Decibel (dB): A ratio of measured pressure to a reference pressure, converted into a logarithmic scale.

Decibel hearing level (dB HL): The terminology used to express hearing threshold levels where 0 dB HL represents the softest sound a normal adult listener can hear 50% of the time. The dB HL scale uses a minimum audibility curve according to dB sound pressure level and sets all of the minimum values to 0 dB HL.

Decibel sensation level (dB SL): Using the listener’s threshold as a reference, this is the decibel level above threshold at which a signal is presented.

Department of Defense (DoD): A department within the executive branch of government responsible for maintaining and overseeing all branches of the United States military to deter war and ensure national security.

Department of the United States Army (DA): One of the military service branches under the DoD.

Detection, Recognition/Identification, Localization, and Communication (DRILCOM): A Virginia Tech Auditory Systems Laboratory-developed test battery designed to assess the components of auditory situation awareness (Casali & Lee, 2017).

Doppler effect: The change in frequency that occurs when a sound approaches, passes, and then recedes from a listener. The change in frequency is due to the compressed sound waves as the sound source moves towards the listener, that subsequently moves into a rarefaction phase as it travels away from the listener. The compressed sound waves result in an increase in frequency, and likewise, the rarefaction results in lower frequency content of the signal.

Gruntworks: A U.S. Marine Corps organization tasked with research, design, acquisition, and equipping Marine rifle squads.

Head related transfer function: The changes in sound parameters of amplitude, frequency, and phase that occurs at the opening of the ear canal relative to the aforementioned sound parameters from the signal source (Rash, Russo, Letowski, & Schmeisser, 2009). The changes at the ear canal are the result of spectral shaping of the torso, head, and outer ear(s). The spectral shaping that occurs at each ear canal is generally accepted to be different, especially when the sound source does not occur at the midline of the head.

Hertz (Hz): A unit measurement for sound frequency, or wave cycles completed per second.

Improvised explosive device (IED): Bombs constructed by makeshift means often used as roadside bombs against U.S. forces in current conflicts.

Insert-style hearing protection: Earplugs that are inserted into the ear canal to form seal that serves as an acoustical barrier to sound entering the ear canal.

Interaural level different (ILD): The difference in intensity level between a single sound source.

Interaural time difference (ITD): The perceived difference in temporal components of a single sound source. The phase, or timing difference, is due to a sound the sound traveling different distances that corresponds to the different position of the two ears.

Learning unit (LU): A complete block of auditory localization training that when combined with a preset number of iterations of the training block comprises a complete localization training protocol for each participant.

Level-dependent hearing protection: Also known as non-linear hearing protection, provides differing levels of sound dampening according to the sound level. At certain levels of turbulent airflow, these devices employ barriers to acoustical energy.

Localization blur: The smallest change in a sound parameter or parameters, including a correlated signal that can affect the original signal, that changes the location of the auditory signal (Blauert, 1997).

Minimum audible angle (MAA): The smallest angular displacement of an auditory event, given a stationary presentation, that can be detected.

Minimum audible movement angle (MAMA): The smallest detectable angular displacement of a sound moving at a constant velocity (Letowski & Letowski, 2012).

National Institute for Occupational Safety and Health (NIOSH): A research agency developed by the Occupational Safety and Health Administration tasked with researching and recommending worker safety and practices (National Institute for Occupational Safety and Health, 2018).

Occupational Health and Safety Administration (OSHA): Federal enforcement agency responsible for upholding the Occupational Safety and Health Act to ensure workplace safety.

Program Executive Office (PEO) Soldier: Organization within the U.S. Army responsible for acquisition and fielding of equipment for Soldiers (U.S. Army, 2019).

Signal-to-noise-ratio (S/N): The ratio of a signal relative to the level of background noise.

Standing wave: When waves from one source interfere with sound from another and create another sound event. In some cases, the sound event can consist of wave cancellation.

Tactical Communication and Protective Systems (TCAPS): Powered, non-linear, hearing protection with pass-through capabilities. These systems incorporate a communication device that enable a Service Member to maintain remote communication while monitoring his or her environment.

Tactical Communication and Protective Systems-Lite (TCAPS-Lite): Powered, non-linear hearing protection with pass-through capabilities that does not interface directly with communication equipment.

Transfer-of-training: Occurs when the trainee recognizes when to apply a certain skillset outside of the training environment and to other applications other than those learned while in training (Hays & Singer, 1989).

Tympanic membrane (TM): The membranous boundary lying between the medial-most portion of the outer and the lateral-most portion of middle ear. The TM assists in overcoming the impedance mismatch between the outer ear and inner ear by increasing energy transmitted to the inner ear (Pickles, 1988).

Veterans Administration (VA): a federal agency also known as the U.S. Department of Veterans affairs that provides healthcare and other benefits to qualified Veterans and their dependents (U.S. Department of Veterans Affairs, 2018).

Unity gain: The gain setting in an active hearing protector where ambient sounds are represented at the same level under the protector as they occur in the environment (Casali & Robinette, 2015).

Virginia Tech Auditory Systems Laboratory (VT-ASL): Acoustical and auditory research and evaluation facility designed to perform laboratory and in-field-related investigations (Virginia Tech, 2019).

LIST OF FIGURES

Figure 1. Localization accuracy across TCAPS devices using the DRILCOM test battery. Adapted from Casali & Lee (2016), slide 39. 6

Figure 2. Schematic of auditory situation awareness adapted from Casali & Clasing, 2013, p. 11. Copyright 2012 by John G. Casali. 15

Figure 3. Cross-section of the outer, middle, and inner ear. Adapted from Hansen (2014), p. 469, Figure 8-35. 17

Figure 4. General description of sound orientation relative to the head. Adapted from Kendall, 1995, p. 26. 20

Figure 5. Depiction of azimuth, elevation and distance relative to the head. Adapted from Kendall, 1995, p. 26. 21

Figure 6. Depiction of interaural time difference (left) and interaural level difference (right). Adapted from Kapralos, Milios, & Jenkin, 2008, p. 528. 21

Figure 7AB. Outer ear spectral shaping according angle of sound incidence. Adapted from Shaw, 1974, p. 461-462. 23

Figure 8. Minimum audible angle across stimulus frequency (sinusoids) for different azimuths. Adapted from Mills, 1958, p. 240. 25

Figure 9. Illustration of the cone of confusion. Adapted from Kapralos et al., 2008, p. 529. 26

Figure 10. Spectral filtering performed by the pinna to shape elevation cues. Figures *A* and *B* adapted from Duda, 2011. Figure *C* adapted from K. Craig (2001). 27

Figure 11. Localization blur in the horizontal axis using white noise stimuli. Adapted from Blauert, 1997, p. 40. 29

Figure 12. Localization blur in the vertical plane using white noise stimuli. Adapted from Blauert, 1997, p. 44. 29

Figure 13. Illustration of upward spread of masking for a 410 Hz signal presented at 40, 60 and 80 dB. Adapted from Tufts & Casali (2017), Figure 14.2. 39

Figure 14. Pre-training and post training mean absolute localization error for the open ear and TCAPS with which the participant conducted training (referred to in this Figure as Training EMHP). Adapted from Casali & Robinette (2015), Figure 2, p. 5. 50

Figure 15. Training effect demonstrated using a training protocol delivered through the DRILCOM test battery and apparatus. Device “A” is the TEA Invisio and “B” is the prototype device. Adapted from Casali & Lee (2016), slide 47. 52

Figure 16. Habit strength index as a function of delay in feedback. Adapted from Wolfle (1951), p. 1268.. 56

Figure 17. DRILCOM testing facility and participant interface demonstrating the 12 response location options and system-generated feedback for an incorrect response. Adapted from Casali & Lee (2019), Figure 1, pS67 and Figure 3, pS69... 59

Figure 18. 1/3-octave band spectral content of the dissonant training signal and untrained stimuli as used in both pilot experiments and main experiment containing 104, 295, 450, 737, 2967, 4959, 7025 and 7880 Hz. 64

Figure 19. Experimental design schematic for Pilot 1. 68

Figure 20. Experimental design schematic for Pilot 2. 69

Figure 21. Experimental design schematic for the main experiment... 70

Figure 22. DRILCOM testing apparatus. Adapted from Casali & Lee (2015), Figure 4, p.27. ... 72

Figure 23. Azimuth speaker array. Adapted from Casali & Lee (2015), Figure 11, p.27. 72

Figure 24. Participant response display showing 12 sequential options. Adapted from Lee & Casali (2019), slide 12. 77

Figure 25. Participant response display for showing feedback for correct (left) and incorrect (right) response. Adapted from Lee & Casali (2019), slides 14 and 13, respectively. 77

Figure 26. Participant response display showing 24 options with only 12 active speakers. Adapted from Lee and Casali (2019), slide 15. 79

Figure 27. Mean score for each presentation condition at each training stage for Pilots 1 and 2. 83

Figure 28. Mean *response time* and *absolute score* at each stage of training plotted according to training strategy. 102

Figure 29. Mean *response time* and *absolute score* at each stage of training plotted according for the adaptive training strategy. 103

Figure 30. Mean *response time* and *absolute score* at each stage of training plotted according for the choose training strategy. 104

Figure 31. Mean *response time* and *absolute score* at each stage of training plotted according for the adapt + choose training strategy. 105

Figure 32. Mean *response time* and *absolute score* at each stage of training plotted according for the DRILCOM training strategy. 106

Figure 33. *Changes in absolute score* versus *response time* from pretest to LU8 with the regression line and equation plotted. 117

Figure 34. *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted. 118

Figure 35. *Changes in absolute score* versus *response time* from **LU4 to LU8** with the regression line and equation plotted. 119

Figure 36. *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the adaptive condition. 122

Figure 37. *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the choose condition....................................... 122

Figure 38. *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the adaptive + choose condition...................... 123

Figure 39. *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the DRILCOM condition. 123

Figure 40. *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted for the adaptive condition. 126

Figure 41. *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted for the choose condition....................................... 126

Figure 42. *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted for the adaptive + choose condition...................... 127

Figure 43. *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted for the DRILCOM condition. 127

Figure 44. *Changes in absolute score* versus *response time* from **LU4 to LU8** with the regression line and equation plotted for the adaptive condition.. 130

Figure 45. *Changes in absolute score* versus *response time* from **LU4 to LU8** with the regression line and equation plotted for the choose condition. 130

Figure 46. *Changes in absolute score* versus *response time* from **LU4 to LU8** with the regression line and equation plotted for the adaptive + choose condition. 131

Figure 47. *Changes in absolute score* versus *response time* from **LU4 to LU8** with the regression line and equation plotted for the DRILCOM condition... 131

Figure 48. Mean percent accuracy at LU4 for each speaker location for the choose subunit compared to the user-selected practice subunit. 132

Figure 49. Frequency of loudspeaker selection compared to mean percent accuracy at each speaker location for the user-selected practice subunit.. 133

Figure 50. Mean percent accuracy at LU4 for each speaker location....................................... 133

Figure 51. Mean ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 1, Perceived Confidence. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test. 144

Figure 52. Mean changes in ratings from LU1 and LU8 (LU8-LU1) with confidence intervals for each training strategy for Question 1, Perceived Confidence. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test.. 145

Figure 53. Mean ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 2, Perceived Difficulty. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test. 147

Figure 54. Mean changes in ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 2, Perceived Difficulty. Mean ratings with a different upper-case

letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test. 148

Figure 55. Mean response ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 3, Perceived Ability. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test. 150

Figure 56. Mean changes in difficulty ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 3, Perceived Ability. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test. 151

Figure 57. Mean response ratings of usefulness at LU1 and LU8 with confidence intervals for each training strategy for Question 4, Perceived Usefulness. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test. 153

Figure 58. Mean changes in usefulness ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 4, Perceived Usefulness. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test. 154

Figure 59. Mean response ratings at LU1 and LU8 with confidence intervals for each training strategy, Question 5, Confidence in Abilities. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at *p*<0.05 using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at *p*<0.05 using a Kruskal-Wallis test. 156

Figure 60. Mean changes in ratings from LU1 to LU8 with confidence intervals for each training strategy for Question 5, Confidence in Abilities. Mean ratings with a different upper-case letter between Training Condition are significantly different at *p*<0.05 using a Kruskal-Wallis test.. 157

Figure 61. Mean response ratings for each training strategy for Question 6, Ability to Localize from Before to After Training. Mean Training Condition ratings with a different upper-case letter within the LU1 and within the LU8 graphs are significantly different at *p*<0.05 using a Kruskal-Wallis test. 158

Figure 62. Mean response ratings with confidence intervals for each training strategy for Question 7, Response Time from Before to After Training. Mean Training Condition ratings with a different upper-case letter within the LU1 and within the LU8 graphs are significantly different at *p*<0.008 using Mann-Whitney *U* tests following a Kruskal-Wallis test.. 159

Figure 63. Mean response ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 8, Appropriateness of Time Allotted. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at *p*<0.05 using a Wilcoxon test. Mean Training Condition ratings

with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test. 161

Figure 64. Mean changes in ratings from LU1 to LU8 with confidence intervals for each training strategy Question 8, Appropriateness of Time Allotted. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test.. 162

Figure 65. Mean ratings and confidence intervals for the **sequentia**l subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C. 163

Figure 66. Mean ratings and confidence intervals for the **random** subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C. 164

Figure 67. Mean ratings and confidence intervals for the **test** subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C. 165

Figure 68. Mean ratings and confidence intervals for the **adaptive** subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C. 166

Figure 69. Mean ratings and confidence intervals for the **choose** subunit separated by training condition and training stage. 167

Figure 70. Percentage of loss of interest reported by subunit for LU1. 168

Figure 71. Percentage of loss of interest reported by subunit for LU8. 168

Figure 72. Phase III experimental design order. 179

Figure 73. Experimental design for Phase III, with independent variables, experimental order, participant assignment, and dependent measures listed. 180

Figure 74. 3M™ PELTOR™ TEP-100 electronic earplug-style TCAPS device...................... 184

Figure 75. Mean frequency response of 70 dBA pink noise measured using KEMAR under open ear and TEP-100 devices on normal volume setting by 1/3 octave-band frequencies. .. 186

Figure 76. 3M™ PELTOR™ ComTac™ III electronic earmuff-style TCAPS device............. 187

Figure 77. Mean frequency response of 70 dBA pink noise measured using KEMAR under open ear and ComTac™ III devices at high volume setting by 1/3 octave-band frequency... 188

Figure 78. 1/3 octave-band spectral content of pretest Dissonant training signal (played on PALAT system) and posttest .22 caliber blank gunshot by 1/3 octave-band frequency. Recorded at the participant's ear in office environment (for PALAT) or at outdoor field site (for transfer-of-training test). Overall sound pressure level of 70 dBA for both signals. 190

Figure 79. Participant pretest and posttest screen displaying 24 response options (black circles) and 12-signal locations (black circles with yellow numbers). 191

Figure 80. Absolute correct response (arrow) and ballpark correct response region (grey shaded region) when the signal emanates from the one o'clock position................................. 192

Figure 81. Front and back regions (shaded regions) depicting the range of signal locations and response locations for possible front-back reversal errors.. 194

Figure 82. Example of semantic differential rating scale. 195

Figure 83. Portable audiometric booth (left) co-located with the PALAT system in the office environment used for training and pretesting. 198

Figure 84. PALAT system apparatus located in a semi-reverberant office room at Virginia Tech. 202

Figure 85. 1/3-octave band spectral content of PALAT system pink noise (green dashed line) and dissonant signal (blue solid line) in 1/3 octave-band frequency. Eight pure tones comprising the dissonant tone are labeled above the respective frequency........ 202

Figure 86. Example CSV file output of PALAT system. 204

Figure 87. Aerial view of field site layout with 12-remote firing device positions, located around the participant........ 205

Figure 88. Mean 1/3-octave band spectral content of .22 caliber blank gunshot from all 12 firing device positions (dashed green line) versus ambient noise floor (dotted purple line) measured in 1/3 octave-band frequency at the participant center head position. 206

Figure 89. Field site panoramic picture (from left at 7 o'clock to right at 6 o'clock), with clock face positions identified. Only the 12 o'clock position actually included a sign during the conduct of the field experiment. 207

Figure 90. Participant view of computer tablet used to operate the field posttest and 12 o'clock reference sign to orient participant. 207

Figure 91. Remote firing device design concept sketch and final product. 208

Figure 92. One remote firing device containing three remote firing mechanisms mounted on a steel u-post located in the wooded forest at the field localization site (Left: Front view of remote firing device, Right: Profile (side) view as seen from direction of the participant), which reduced the visual signature. 209

Figure 93. Remote firing device control box. 210

Figure 94. Remote firing device block diagram with all major components........ 211

Figure 95. Participant-controlled computer tablet placed on music stand at the center of the field experiment site. 212

Figure 96. PALAT system initialization screen. 214

Figure 97. Participant operating the PALAT system. 215

Figure 98. PALAT system main menu screen on computer tablet. 216

Figure 99. PALAT system sequential training screen on computer tablet. 217

Figure 100. PALAT system test screen on computer tablet. 218

Figure 101. Participant interface on the computer tablet demonstrating the 24 response location options and system-generated feedback for an absolute correct response. 221

Figure 102. Participant interface on the computer tablet demonstrating feedback provided for an incorrect response. 221

Figure 103. Example of the user-select display interface on the computer tablet. 222

Figure 104. Field posttest site layout with participant standing in center of remote firing devices facing 12 o'clock target position and two investigators seated behind participant between the 6 o'clock and 7 o'clock positions. 223

Figure 105. Absolute correct score on pretest under TEP-100 listening condition for all participants. Values ordered and plotted from lowest to highest score by group. 226

Figure 106. Response time on posttest under open ear listening condition for all participants. Values ordered and plotted from lowest to highest response time by group. 226

Figure 107. Mean ballpark correct scores with 95% confidence intervals plotted. Different letters indicate significant differences at $p \leq 0.10$. 233

Figure 108. Mean absolute correct scores for each listening condition with 95% confidence intervals plotted. Different letters indicate significant differences at $p \leq 0.10$. 234

Figure 109. Mean ballpark correct scores for each listening condition with 95% confidence intervals plotted. Different letters indicate significant differences at *p*≤0.10. 235

Figure 110. Mean response times for each listening condition with 95% confidence intervals plotted. Different letters indicate significant differences at *p*≤0.10. 236

Figure 111. Mean response times from pretest to posttest. Different letters indicate significant differences at *p*≤0.10. 237

Figure 112. Mean absolute correct scores for each group at pretest and posttest. Different letters indicate significant differences at *p*≤0.10. 238

Figure 113. Mean absolute correct scores for each training group comparing pretest and posttest performance. Different letters indicate significant differences at *p*≤0.10. 239

Figure 114. Mean absolute correct scores for the trained versus untrained groups at pretest for each listening condition. Different letters indicate significant differences at *p*≤0.10. ... 241

Figure 115. Mean absolute correct scores at posttest for each listening condition comparing the trained and untrained groups. Different letters indicate significant differences at *p*≤0.10. 241

Figure 116. Mean absolute correct scores for each listening condition for the **untrained** group **comparing pretest and posttest** performance. Different letters indicate significant differences at *p*≤0.10. 243

Figure 117. Mean absolute correct scores for each listening condition for the **trained** group **comparing pretest and posttest** performance. Different letters indicate significant differences at *p*≤0.10. 243

Figure 118. Mean front-back reversal errors using the 120-degree arc criterion for each listening condition. Different letters indicate significant differences at *p*≤0.10. 246

Figure 119. Mean absolute correct score on pretest and posttest for trained group. Regression line and equation plotted. 248

Figure 120. Slopes from pretest to posttest for mean absolute correct score for open ear by group. 249

Figure 121. Slopes from pretest to posttest for mean absolute correct score for open ear by group. 250

Figure 122. Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 1, Perceived Confidence. 252

Figure 123. Mean ratings for each group at pretest for Question 1, Perceived Confidence. 256

Figure 124. Mean ratings for each group at posttest for Question 1, Perceived Confidence. 257

Figure 125. Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 2, Perceived Accuracy. 258

Figure 126. Mann-Whitney *U* results comparing trained and untrained ratings at pretest in the open ear condition for Question 2, Perceived Accuracy 260

Figure 127. Mean ratings for each group at pretest for Question 2, Perceived Accuracy. 261

Figure 128. Mean ratings for each group at posttest for Question 2, Perceived Accuracy. 262

Figure 129. Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 3, Perceived Difficulty 264

Figure 130. Mean ratings for each group at pretest for Question 3, Perceived Difficulty 266

Figure 131. Mean ratings for each group at posttest for Question 3, Perceived Difficulty. 267

Figure 132. Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 4, Perceived Reaction Time 269

Figure 133. Mean ratings for each group and listening condition at pretest for Question 4, Perceived Reaction Time. 272

Figure 134. Mean ratings of perceived reaction time for each group and listening condition at posttest for Question 4, Perceived Reaction Time. 273

Figure 135. Plotted means and 95% confidence intervals at pretest and posttest for each group and listening condition for Question 5, Perceived Comfort. 275

Figure 136. Mean ratings of perceived comfort for each group and listening condition at pretest for Question 5, Perceived Comfort. 277

Figure 137. Mean ratings of perceived comfort for each group and listening condition at pretest for Question 5, Perceived Comfort. 278

Figure 138. Plotted means and 95% confidence intervals for likelihood to maintain the same listening condition ratings at pretest and posttest for each group and listening condition for Question 6, Likelihood of Wearing Device. 280

Figure 139. Mean ratings of likelihood of wearing device given similar conditions for each group and listening condition at pretest for Question 6, Likelihood of Wearing Device. 282

Figure 140. Mean ratings of likelihood of wearing device given similar conditions for each group and listening condition at posttest for Question 6, Likelihood of Wearing Device........ 283

Figure 141. Plotted means and 95% confidence intervals for perceived degree of preparedness for each listening condition ratings at posttest for the trained group for Question 7, Degree of Preparedness.. 285

Figure 142. Radial plots of mean absolute correct accuracy percentage for each listening condition during pretest and posttest by group, trained (solid line) and untrained (dashed line). 288

Figure 143. Level dependent function of the TEP-100 measured using a broadband pink noise signal (Stegar et al., 2019). 296

Figure 144. Frequency response radial plots of mean sound pressure level at a single 1/3 octave-band frequency. Measurements recorded using KEMAR manikin with open ear (dashed blue lines) and TEP-100 (solid orange line) in right ear. Dissonant tone set at 55 dBA (left column) and 80 dBA (right column). 298

Figure 145. Input-output curve across frequencies using a pure-tone sweep at 60 and 90 dB SPL for the TEP-100 in the unity gain setting. Data was obtained with an Audioscan Verifit system. 310

Figure 146. Mann-Whitney U results comparing mean ratings of confidence for each listening condition at pretest for trained versus untrained groups for Question 1, Perceived Confidence. 403

Figure 147. Mann-Whitney U results comparing mean ratings of confidence for each listening condition at posttest for trained versus untrained groups for Question 1, Perceived Confidence. 403

Figure 148. Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest in the TEP-100 and ComTac™ III conditions for Question 2, Perceived Accuracy. 404

Figure 149. Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest for all listening condition for Question 3, Perceived Difficulty....... 404

Figure 150. Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at posttest for all listening condition for Question 3, Perceived Difficulty. 405

Figure 151. Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest and posttest collapsed across listening conditions for Question 4, Perceived Reaction Time. 405

Figure 152. Mann-Whitney *U* results comparing trained and untrained ratings of perceived reaction time at pretest for all listening condition for Question 4, Perceived Reaction Time. 406

Figure 153. Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest for all listening condition for Question 4, Perceived Reaction Time. 406

Figure 154. Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at posttest for all listening condition for Question 4, Perceived Reaction Time. 407

Figure 155. Mann-Whitney U results comparing mean ratings of confidence for each listening condition at pretest for trained versus untrained groups for Question 5, Perceived Comfort. 407

Figure 156. Mann-Whitney U results comparing mean ratings of confidence for each listening condition at posttest for trained versus untrained groups for Question 5, Perceived Comfort. 407

Figure 157. Mann-Whitney U results comparing mean ratings of likelihood of wearing device for each listening condition at pretest for trained versus untrained groups for Question 6, Likelihood of Wearing Device 408

Figure 158. Mann-Whitney U results comparing mean ratings of likelihood of wearing device for each listening condition at posttest for trained versus untrained groups for Question 6, Likelihood of Wearing Device.. 408

LIST OF TABLES

Table 1. Hearing Profile Criteria for the U.S. Air Force, Army, and Navy according to Service standards 10

Table 2. Localization blur according to stimulus time with sound presented at the midline. 24

Table 3. Subunit composition for each training condition in the main experiment. The sequential subunit was comprised of four presentations from each location for LU1 and only one from each location for LUs 2-8. 66

Table 4. Mean pure-tone hearing level thresholds (dBHL) by group study and group. 75

Table 5. Slope comparisons for each presentation condition and Pilot 2, evaluating differences in the change in absolute score over LUs between conditions. 86

Table 6. One-way ANOVA for the AK-47 untrained stimulus across presentation conditions... 87

Table 7. One-way ANOVA for the Apache untrained stimulus across presentation condition. .. 87

Table 8. One-way ANOVA for the Arabic untrained stimulus across presentation condition..... 87

Table 9. One-way ANOVA for the whistle untrained stimulus across presentation condition.... 88

Table 10. One-way, within-subjects ANOVA for the two-presentation condition comparing absolute performance across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 88

Table 11. Pairwise comparison by Bonferroni *t*-test for the two-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 88

Table 12. One-way, within-subjects ANOVA for the three-presentation condition comparing absolute performance across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 88

Table 13. Pairwise comparison by Bonferroni *t*-test for the two-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 89

Table 14. One-way, within-subjects ANOVA for the four-presentation condition comparing absolute score across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 89

Table 15. Pairwise comparison by Bonferroni *t*-test for the four-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 89

Table 16. One-way, within-subjects ANOVA for Pilot 2 (no sequential training in LU2-LU8) comparing absolute score across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 90

Table 17. Pairwise comparisons by Bonferroni *t*-test for the four-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level). 90

Table 18. Mixed-factor ANOVA results analyzing the effect of training strategy and LU on *absolute score* at **pretest and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 99

Table 19. Mixed factor ANOVA table analyzing the effect of training strategy and LU on *absolute score* at **pretest and LU4** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 99

Table 20. Mixed factor ANOVA table analyzing the effect of training strategy and LU on *absolute score* at **LU4 and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 101

Table 21. Mixed factor ANOVA table analyzing the effect of training strategy and LU on *response time* at **pretest and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 101

Table 22. Mixed factor ANOVA table analyzing the effect of training strategy and LU on *response time* at **pretest and LU4** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 107

Table 23. Mixed factor ANOVA table analyzing the effect of training strategy and LU on *response time* at **LU4 and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 107

Table 24. One-way ANOVA results for measuring effect of training strategy on the *absolute difference score* from **LU8 compared to pretest**. 108

Table 25. Descriptive statistics for the *absolute difference score* **from LU8 compared to pretest.** 108

Table 26. One-way ANOVA results for measuring effect of training strategy on the *absolute difference score* from **LU4 compared to pretest**. 108

Table 27. Descriptive statistics for using the *absolute difference score* from **LU4 compared to pretest**. 109

Table 28. One-way ANOVA results for measuring effect of training strategy on the *absolute difference score* from **LU8 compared to LU4**. 109

Table 29. Descriptive statistics for the *absolute difference score* from **LU4 compared to pretest.** 109

Table 30. One-way ANOVA results for measuring effect of training strategy on the *ballpark difference score* from **LU8 compared to pretest**. 110

Table 31. Descriptive statistics for the *ballpark difference score* from **LU8 compared to pretest.** 110

Table 32. One-way ANOVA results for measuring effect of training strategy on the *ballpark difference score* for **LU4 compared to pretest.** 110

Table 33. Descriptive statistics for the *ballpark difference score* from **LU4 compared to pretest.** 111

Table 34. ANOVA results for measuring effect of training strategy on ballpark score using the *ballpark difference score* from **LU8 compared to LU4**. 111

Table 35. Descriptive statistics for the *ballpark difference score* from LU8 compared to LU4. 111

Table 36. One-way ANOVA results for measuring effect of training strategy *response time difference score* from **pretest to LU8**. 112

Table 37. Descriptive statistics for the *response time difference score* from **pretest to LU8**. .. 112

Table 38. One-way ANOVA results for measuring effect of training strategy *response time difference score* from **pretest to LU4.** 112

Table 39. Descriptive statistics for the *response time difference score* from **pretest to LU4.** .. 113

Table 40. One-way ANOVA results for measuring effect of training strategy *response time difference score* from **LU4 to LU8.**... 113

Table 41. Descriptive statistics for the *response time difference score* from LU4 to LU8. 113

Table 42. One-way ANOVA results for measuring effect of training strategy on mean _slope_ from **pretest to LU8**. 114

Table 43. Descriptive statistics for _slope_ measured from **pretest to LU8** for each training strategy. 114

Table 44. One-way ANOVA results for measuring effect of training strategy on _slope_ from **pretest to LU4** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 114

Table 45. Descriptive statistics for _slope_ measured from pretest to **LU4 for each training strategy** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 114

Table 46. Tukey HSD pairwise comparisons for each training strategy using _slope_ from **pretest to LU4** as the dependent measure (bolded text in the table indicates a significant test result at the α=0.05 significance level). 115

Table 47. One-way ANOVA table results for measuring effect of training strategy on _slope_ from **LU4 to LU8**. 115

Table 48. Descriptive statistics for _slope_ measured **from LU4 to LU8** for each training strategy. 116

Table 49. ANOVA table and parameter estimates for linear regression for _change in response time given a change in absolute score_ measured from **pretest to LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 116

Table 50. ANOVA table and parameter estimates for linear regression for _change in response time_ given a _change in absolute score_ measured from **pretest to LU4**. 117

Table 51. ANOVA table and parameter estimates for linear regression for *changes in absolute score* versus *response time* measured from **LU4 to LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level). 119

Table 52. ANOVA results and linear regression parameter estimates for each training strategy analyzing the *changes in absolute score* versus *response time* from **pretest to LU8** (bolded text in the table indicates a significant test result at the the α=0.05 significance level). 121

Table 53. ANOVA results and linear regression parameter estimates for each training strategy analyzing the *change in response time* given a *change in absolute score* measured from **pretest to LU4** (bolded text in the table indicates a significant test result at the the α=0.05 significance level). 125

Table 54. ANOVA results and linear regression parameter estimates for each training strategy analyzing the *change in response time* given a *change in absolute score* measured from LU4 to LU8 (bolded text in the table indicates a significant test result at the α=0.05 significance level). 129

Table 55. Spearman correlations using Z scores for percentage of trials a speaker location was chosen and the percent correct performance at that location. 135

Table 56. Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy for Question 1, Perceived Confidence (bolded text in the table indicates a significant test result at $p<0.05$.) 143

Table 57. Kruskal-Wallis results and mean ranks comparing training strategies at LU1 and LU8 for Question 1, Perceived Confidence (bolded text in the table indicates a significant test result at $p<0.05$). 145

Table 58. Kruskal-Wallis mean ranks results comparing ratings among training strategies using the difference score (LU8-LU1) for Question 1, Perceived Confidence. (bolded text in the table indicates a significant test result at *p*<0.05). 145

Table 59. Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy for Question 2, Perceived Difficulty (bolded text in the table indicates a significant test result at *p*<0.05.) 146

Table 60. Kruskal-Wallis results comparing all training strategies at LU1 and LU8 for Question 2, Perceived Difficulty (bolded text in the table indicates a significant test result at *p*<0.05). 148

Table 61. Kruskal-Wallis mean ranks results comparing ratings for training strategies using the difference score in ratings (LU8-LU1) for Question 2, Perceived Difficulty (bolded text in the table indicates a significant test result at *p*<0.05). 148

Table 62. Wilcoxon results comparing ability ratings for LU1 versus LU8 for each training strategy group for Question 3, Perceived Ability (bolded text in the table indicates a significant test result at *p*<0.05.) 149

Table 63. Kruskal-Wallis results comparing ratings of ability for all training strategies at LU1 and LU8 for Question 3, Perceived Ability (bolded text in the table indicates a significant test result at *p*<0.05). 150

Table 64. Kruskal-Wallis mean ranks results comparing training strategies using the difference score in ratings (LU8-LU1) for Question 3, Perceived Ability (bolded text in the table indicates a significant test result at *p*<0.05). 151

Table 65. Wilcoxon results comparing confidence ratings for LU1 versus LU8 for each training strategy group for Question 4, Perceived Usefulness (bolded text in the table indicates a significant test result at *p*<0.05). 152

Table 66. Kruskal-Wallis results comparing ratings of usefulness for all training strategies at LU1 and LU8 for Question 4, Perceived Usefulness (bolded text in the table indicates a significant test result at *p*<0.05). 153

Table 67. Kruskal-Wallis results comparing ratings of usefulness for all training strategies at LU1 and LU8 for Question 4, Perceived Usefulness (bolded text in the table indicates a significant test result at *p*<0.05). 154

Table 68. Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy group for Question 5, Confidence in Abilities (bolded text in the table indicates a significant test result at *p*<0.05). 155

Table 69. Kruskal-Wallis results comparing ratings for all training strategies at LU1 and LU8 for Question 5, Confidence in Abilities (bolded text in the table indicates a significant test result at *p*<0.05). 156

Table 70. Kruskal-Wallis mean ranks results comparing ratings among training strategies using the difference score in ratings (LU8-LU1) for Question 5, Confidence in Abilities (bolded text in the table indicates a significant test result at *p*<0.05). 157

Table 71. Table 70. Mean ranks for perceived ability to localize from before to after the training for Question 6, Ability to Localize from Before to After Training (bolded text in the table indicates a significant test result at *p*<0.05). 158

Table 72. Mann-Whitney U pairwise comparisons for perceived response time ratings from after LU1 to after the training for Question 7, Response Time from Before to After Training (bolded text in the table indicates a significant test result at *p*<0.008). 159

Table 73. Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy group for Question 8, Appropriateness of Time Allotted (bolded text in the table indicates a significant test result at *p*<0.05). 160

Table 74.Kruskal-Wallis results comparing ratings of usefulness for all training strategies at LU1 and LU8 for Question 8, Appropriateness of Time Allotted (bolded text in the table indicates a significant test result at *p*<0.05). 161

Table 75. Kruskal-Wallis mean ranks results comparing ratings among training strategies using the difference score in ratings (LU8-LU1) for Question 8, Appropriateness of Time Allotted (bolded text in the table indicates a significant test result at *p*<0.05).............. 162

INTRODUCTION

Workplace noise and subsequent auditory injury remains pervasive globally. Occupational noise exposure accounts for 16% of hearing loss worldwide (Nelson, Nelson, Concha-Barrientos, & Fingerhut, 2005). The deleterious effects of occupational noise exposure can obviously impact workers' ability to participate in the workforce and degrade their quality of life. Epidemic rates of hearing loss within the U.S. serve as no exception to global trends. Over 28 million Americans suffer from hearing loss, ranking the third most common chronic health condition in adults (U.S. Department of Veterans Affairs [VA], 2016; National Institute for Occupational Safety and Health [NIOSH], 2019). With 22 million workers exposed to hazardous noise, the influence of noise on prevalence of hearing loss is obvious (Tak, Davis, & Calvert, 2009). Exacerbating the effects of noise is a lack of adherence to hearing protection requirements. When engineering or administrative controls are infeasible or ineffective, hearing protection is required (Occupational Safety and Health Administration [OSHA], 1971). However, approximately 34% of U.S. noise-exposed workers report never wearing hearing protection (Tak et al., 2009). Therefore, an obvious need exists to improve strategies for hearing protection compliance.

Of note, estimates for hearing loss among U.S. workers exclude Military Service Members. From 2003-2012, the prevalence of noise-induced hearing loss in the civilian sector was 12.94% (Matterson, Bushnell, Themann, & Morata, 2016). By comparison, in 2016 the prevalence rates of significant hearing loss among just the U.S. Army Service Members was 24% (DOEHRS-DR, 2016), nearly double that of the civilian sector. The long-term effects of noise exposure in the military are evidenced by the number of Veterans receiving compensation for hearing loss and tinnitus. For fiscal year 2016, 2.8 million Veterans received financial compensation for hearing loss and tinnitus, the two most commonly awarded disabilities (VA, 2017). The VA does not currently report spending for service-connected hearing loss and tinnitus according to percent disability ratings, as compensation is awarded in context of whole-body disability. Therefore, ratings for conditions are not directly summative. Given that 1.7 million Veterans receive a 10% disability rating for tinnitus or hearing loss, and the average annual compensation of $1,611.00 for those with a 10% rating, estimates far exceed the disability payments civilian workforce of $242.4 million annually (VA, 2017; NIOSH, 2012).

Of more immediate concern than the financial burden of hearing loss in Service Members are the consequences of hearing loss that adversely affect the individual. Service Members must maintain a certain level of physical readiness to be retained (Department of Defense [DoD], 2010). Upholding auditory fitness-for-duty standards better ensures that Service Members are capable of a certain level of auditory awareness in order to respond to mission-critical information. The ability to auditorily detect, recognize, and localize sound-related events and to anticipate the status of these events in the future is known as auditory situation awareness (ASA) (Hajicek, Myrent, Li, Barker, & Coyne, 2010; Endsley M. R., 1988). Service Members often face a decision to wear hearing protection that can attenuate critical cues or remain unprotected, suffering the consequences of temporary and/or permanent hearing loss. Of concern, hearing loss and any decrement to the fidelity of an auditory signal can impede threat detection, recognition, and localization (Casali & Lee, 2017). Therefore, both hearing loss and hearing protection use can present risks to ASA.

Augmented hearing protection seeks to address the simultaneous needs of maintaining ASA and noise protection. Conventional hearing protection generally refers to devices that serve as a physical barrier to sound entering the ear canal. These devices employ neither dynamic mechanical components nor electronics to provide protection (Casali, 2010). Level-dependent, or non-linear, hearing protection provides differing levels of attenuation depending on the input sound level. Non-linear devices employ barriers to acoustical energy that are typically activated at certain levels of turbulent airflow. These barriers do not involve moving parts, as in a valve, that can result in a delayed reaction to fast-occurring sounds, such as gunfire (Casali, 2010). Active, or battery-powered, hearing protection uses analog circuitry, digital signal processing, or both to prevent hazardous exposures (Casali, 2010). Active noise reduction (ANR) is a processing strategy that introduces a signal in the opposite phase but same amplitude as the detected one, summing the total energy of a waveform to zero (Casali, 2010). ANR can also mean that a detected signal is relatively constant, as opposed to speech that fluctuates, and the circuitry attenuates the constant signals. Electronically modulated sound transmission is a type of hearing protection strategy that at least partially overcomes the passive protection of a device (Casali, 2010). Sound is transmitted from an externally-mounted microphone, through the passive protector, to an output-limiting amplifier, and into a loudspeaker that emits a signal into the ear canal (Casali, 2010). Generally, the device circuitry either enables sound to "pass through" and amplifies certain signal frequencies while limiting other frequencies as well as high-level transient sounds. The DoD commonly refers to active hearing protection as Tactical Communication and Protective Systems (TCAPS), as long as the device includes a radio communications feature. If the radio feature is not included, the device is sometimes referred to as "TCAPS-Lite". TCAPS are often designed to interface with other military communication equipment, such as radios and ground vehicles or aircraft-mounted communication systems.

The effects of hearing protection devices on situation awareness have been studied extensively within the Virginia Tech Auditory Systems Laboratory (VT-ASL). Certain types of hearing protection, including passive and augmented systems, impeded detection and subsequent recognition of certain sounds (Alali & Casali, 2012; Clasing & Casali, 2014). In regards to detection, devices with higher attenuation ratings were correlated with shorter detection distances for ground vehicle reverse (back-up) alarms (Alali & Casali, 2012). In regards to localization, listeners' ability to localize backup alarms in hazardous listening environments remained intact with certain passive hearing protection (Alali & Casali, 2011). However, in a weapon-related localization task, performance was significantly worse with augmented protective devices compared to the unprotected ear in varying levels of noise (Talcott, Casali, Keady, & Killion, 2012). Therefore, hearing protection can have differing effects on localization according to its design, ambient noise levels, and content of the signal (Alali & Casali, 2011). Of note, ergonomic compatibility with combat-related tasks was identified as paramount to confidence and subsequent use (Casali, Ahroon, & Lancaster, 2009). Promisingly, localization accuracy using certain TCAPs showed improvement with training to achieve performance comparable to that of the open ear (Casali & Robinette, 2015). However, the authors found that training benefits did not cross over between circumaural (earmuff style) and insert (earplug-style) devices. Furthermore, Lee & Casali (2017), demonstrated that certain electronically-modulated sound transmission devices are not amenable to training benefits for localization (i.e. trainee performance with the device did not approach that of the open ear). Thus, proper design and training of TCAPs play a role in maximizing the benefits for protection and ASA.

Devices that incorporate circuitry or mechanics that enable access to certain localization cues show promise in enabling listeners to recalibrate altered cues to improve localization (Casali & Robinette, 2015; Hofman, Van Riswick, & Van Opstal, 1998). Improving access to certain cues that contribute to ASA may decrease workload and improve performance. For example, Casali, Lancaster, Valimont, and Gauger (2007) showed that when headsets with active noise reduction (ANR) were used by pilots in civilian aircraft simulators, speech intelligibility increased while ratings of mental workload and hazardous noise exposure decreased. For those with hearing loss, use of passive hearing protection significantly decreased word recognition in noise (Giguère, Laroche, & Vaillancourt, 2013). However, use of electronic level-dependent hearing protection showed significantly better word recognition compared to the passive protection condition (Giguère et al., 2013). Moreover, hearing-impaired listeners demonstrated performance better than the open-ear in some cases (Giguère et al., 2013). Casto & Casali (2013) also supported the use of augmented communication devices for pilots with hearing loss. For those with hearing loss, speech intelligibility decreased when these pilots used conventional passive headsets compared to ANR-based headsets. Those without significant hearing loss did not suffer these same detriments to speech intelligibility when using passive protection. Furthermore, evidence also supported that increasing ASA cues promoted efficacy of ground combat-related tasks. Sheffield, Brungart, Tufts, and Ness (2015) reported that cadets in simulated ground-combat tasks with reduced audibility, via wearing of a hearing loss simulator, participated more cautiously and remained "hidden" to maintain survivability at the cost of lethality, or efficacy. Thus, restoring access to ASA cues could potentially increase confidence as well as ability to detect and engage targets. Therefore, improving access to ASA cues could facilitate compliance with hearing protection.

To standardize testing of auditory situation awareness, VT-ASL developed a battery of tests to assess the components of ASA known as Detection, Recognition/Identification, Localization, and Communication (DRILCOM) (Casali & Lee, 2017). The test battery has previously demonstrated its sensitivity to differences among TCAPs on various aspects of ASA

(Casali & Robinette, 2015; Casali & Lee, 2016; Lee & Casali, 2017). Results of localization accuracy within ± 15° are displayed in Figure 1, illustrating differences in localization accuracy detected by DRILCOM among TCAPS.

Therefore, the instrumentation to test different aspects of ASA exists. This same instrumentation shows potential as a means to deliver training to improve localization accuracy, as demonstrated by Casali & Robinette (2015).

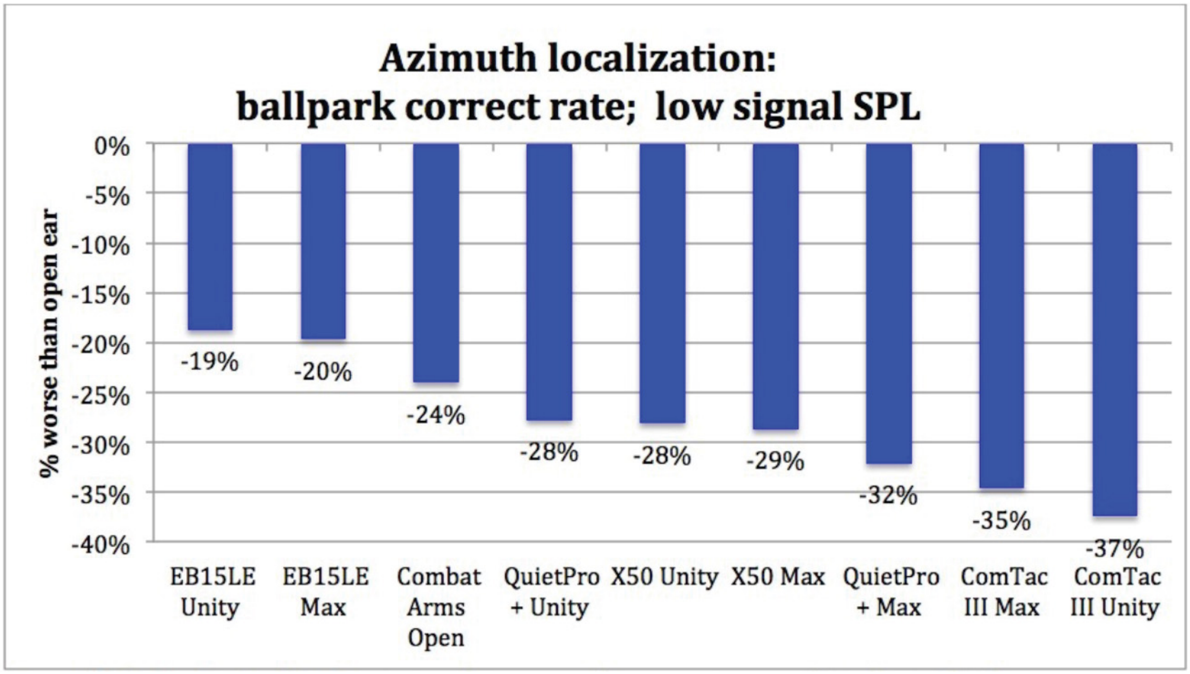


Figure . Localization accuracy across TCAPS devices using the DRILCOM test battery. Adapted from Casali & Lee (2016), slide 39.

Purpose

The Office of Naval Research (ONR) provided a contract to develop and validate a horizontal auditory localization training regimen that will be incorporated into a validated portable system. The goal of the training regimen and portable system was to improve auditory localization with and without TCAPS. The focus of this research was the development of stimuli and a training regimen. An optimized training protocol was incorporated into a portable system design project that was covered by LTC Brandon Thompson in another Ph.D. dissertation. A field validation experiment was conducted to evaluate the existence of transfer-of-training from the lab to a realistic field environment and are reported LTC Thompson's dissertation and the one described herein.

BACKGROUND

Hearing in military environments

Noise exposure remains ubiquitous in the U.S. Military. Service Members encounter considerable challenges functioning in environments where noise can mask speech and critical environmental cues, create temporary and permanent hearing loss, and create other negative physiological effects (Casali & Tufts, 2020). Additionally, Service Members must be able to rely on auditory information in reduced visibility environments, such as in darkness and, oftentimes under increased load on the working memory (i.e. driving and engaging targets) (Casali & Tufts, 2020). Military training environments often involve predictable sources of hazardous noise exposure from munitions, vehicle noise, and aircraft noise. Despite the predictable nature of many of these exposures, compared to nonveterans, Veterans are 30% more likely to have severe hearing impairment (Groenwold, Tak, & Matterson, 2011). Combat environments present an even greater challenge of having predictable noise sources encountered in training as well as unpredictable ones from enemy sources. In the midst of these varied noise sources, Service Members must maintain the ability to detect, recognize, localize, react, and communicate (Tufts, Vasil, & Briggs, 2009). Consequently, Veterans with combat deployments from 2001-2010 showed a four times greater likelihood of having severe hearing impairment than nonveteran counterparts (Groenewold et al., 2011). Implicated in many of these combat-related injuries are blast exposures, which can manifest as comorbid conditions of hearing loss, tinnitus, vestibular disorders, speech-language impairments, impaired cognition, and other latent sequelae resulting from traumatic brain injury (Cave, Cornish, & Chandler, 2007; Helfer et al., 2011).

Military fitness-for-duty

Auditory fitness-for-duty requirements assist in ensuring U.S. Military Service Members have sufficient capabilities to maintain ASA. Obviously, detection remains paramount to all other aspects of ASA since sounds can only be recognized, localized and reacted to if they are first detected (Tufts et al., 2009). Therefore, the military employs minimum audiometric pure tone threshold standards as thresholds are relatively good at predicting detection ability in quiet (i.e. ability to hear alarms and signals) (Kamm, Dirks, & Bell, 1985). For enlistment, appointment, or induction into the U.S. Military, Service Members are not to have average thresholds in each ear greater than 30 decibels hearing level (dB HL) with no individual level greater than 35 dB at 500, 1000, and 2000 Hz; the threshold at 3000 Hz cannot exceed 45 dB; and the threshold at 4000 Hz cannot exceed 55 dB (DoD, 2010). In the U.S. Army, in order to determine fitness-for-duty status, a medical profiling system rates assignment limitations based on certain injuries and illnesses of a permanent or temporary nature. In the hearing category of the profiling system, known as “H”, Soldiers are assigned a category H-1 through H-3 based on their pure tone thresholds (Department of the United States Army [DA], 2017). Those with an H-4 profile meet neither the H-3 pure tone requirements nor word recognition requirements with the use of a hearing aid. A rating of H-1 is consistent with no limitations on duty due to hearing thresholds with the exception of special duties and schools that have more stringent requirements (i.e. aviation and Special Forces). An H-2 is consistent with minor limitations on duty due to his/her hearing loss. An H-3 is consistent with significant duty limitations that requires a board to determine fitness for his or her job and/or fitness for service, and an H-4 involves a drastic duty limitation. When a Soldier is issued an H-3 or H-4 profile, a Speech Recognition in Noise Test (SPRINT) is administered (DA, 2017). The test evaluates the Soldier’s ability to hear Northwestern University Auditory Test No. 6, monosyllabic words delivered at 50 dB HL at a 9 dB signal-to-noise ratio to further determine duty limitations based on auditory performance (Wilson & Cates, 2008). Hearing profiling standards across the Services are listed in Table 1.

Table . Hearing Profile Criteria for the U.S. Air Force, Army, and Navy according to Service standards

	Air Force AFI 48-123	Army AR 40-501	Navy TM 620.51.99-2
H-1	Unaided hearing loss in either ear with no single value greater than: 25 dB at 500 1000 2000 Hz, 35 at 3000 Hz, 45 at 4000 Hz, and 45 at 6000 Hz	Audiometer average level for each ear not more than 25 dB at 500, 1000, 2000 Hz with no individual level greater then 30 dB. Not over 45 dB at 4000 Hz.	Unaided hearing loss in either ear with no single value greater than: 25 dB at 500 1000 2000 Hz, 35 at 3000 Hz, 45 at 4000 Hz, and 45 at 6000 Hz
H-2	Unaided hearing loss in either ear with no single value greater than: 35 dB at 500 1000 2000 Hz, 45 at 3000 Hz, and 55 at 4000 Hz; no requirement for 6000 Hz	Audiometer average level for each ear at 500, 1000, 2000 Hz, or not more than 30 dB, with no individual level greater than 35 dB at these frequencies, and level not more than 55 dB at 4000 Hz; or audiometer level 30 dB at 500 Hz, 25 dB at 1000 and 2000 Hz, and 35 dB at 4000 Hz in better ear. (Poorer ear may be deaf.)	Unaided hearing loss in either ear with no single value greater than: 35 dB at 500 1000 2000 Hz, 45 at 3000 Hz, and 55 at 4000 Hz; no requirement for 6000 Hz
H-3	Any loss that exceeds the values noted above, but does not qualify for H-4.	Speech reception threshold in best ear not greater than 30 dB HL, measured with or without hearing aid; or acute or chronic ear disease.	Any loss that exceeds the values noted in the above definition.
H-4	Hearing loss sufficient to preclude safe and effective performance of duty, regardless of level of pure tone hearing loss, and despite use of hearing aids.	Functional level below H3.	Hearing loss sufficient to preclude safe and effective performance of duty, regardless of degree of pure tone hearing loss, or unknown hearing loss values. The H-4 profile indicates an incomplete follow-up or a requirement for a Medical Evaluation Board.

Hearing loss and auditory situation awareness

Price, Kalb, and Garinther (1989) illustrated the effects of hearing loss on Soldiers' abilities to detect environmental cues. The researchers simulated hearing loss and applied models that accounted for loss of audibility, but did not account for distortion that can also occur in sensorineural hearing loss (Price et al., 1989). In other words, the models may have underestimated the effects of hearing loss on detection. Nonetheless, their results illustrated the deleterious effects of reduced audibility on ASA. Those with hearing within H-1 limits could hear footfalls in leaves for a 100 meters (m) away versus 0.6 m away for those with simulated hearing loss within H-3 limits (Price et al., 1989). The authors calculated that the detection distance for those H-1 hearing corresponded to two minutes of reaction time compared to those with H-3 hearing who had almost none.

Peters & Garinther (1990) quantified the effects of reduced speech intelligibility on tank crew performance through tasks involving detection and recognition. In a tank simulator, the researchers examined experienced crewmember performance on measures of response time, mission error rates, and gunner accuracy. The greatest effects of reduced speech audibility were evidenced by longer response times to initially identify the target (Peters & Garinther, 1990). Increased detection time proportionally increased all other mission completion times (Peters & Garinther, 1990). Reduced speech intelligibility resulted in reduction of mission completion, lethality, and survivability (Peters & Garinther, 1990). The frequency of wrong targets fired upon and communication errors also increased significantly with decreased speech intelligibility (Peters & Garinther, 1990). Casto & Casali (2013) further demonstrated that reduced speech intelligibility can have an adversely synergistic effect on performance when workload is high. While reaction time was not directly measured, Blackhawk helicopter pilots requested significantly more repeat messages from the air traffic control tower when signal quality was poor and flight-imposed workload was high.

Studies investigating the impact of hearing loss on ASA illustrate the need to maintain auditory fitness-for-duty standards for certain military duties. Even before localization can occur in the ASA task system, hearing loss can prevent detection and recognition. Sheffield et al. (2015) measured the effects of simulated hearing loss on combat-integrated tasks incorporated into close quarters combat scenarios. Results showed that the number of opponents eliminated, or combat effectiveness, decreased significantly with severe levels of hearing loss. Survivability, or time until elimination, was not affected by hearing loss. The researchers observed, and some participants anecdotally noted, that those with more severe hearing loss maintained more defensive postures, such as hiding. These behaviors may increase the chances for survivability, but at the cost of lethality. The results clearly supported the adverse effects of hearing loss on signal detection, and the operator's awareness of reduced audibility.

Hearing loss and noise

Hearing loss resulting from damage to the sensory cells, known as cochlear hearing loss, creates broader filters on the sensory epithelium of the inner ear. In a healthy ear, frequency-specific filtering occurs whereby sensory cells selectively respond according to the intensity, frequency, and phase of the signal. This filtering in the cochlea enables differentiated processing of certain types of acoustic stimuli. Contrariwise, in a damaged cochlea, more signal intensity is required to stimulate the cochlea and does so with broader filtering "banks" (Pickles, 1988) Therefore, in those with noise-induced hearing loss, the signal requires higher intensity levels to stimulate the sensory cells and is more likely to be processed within the same channel as noise.

Hearing loss-related distortion can render those with hearing loss more vulnerable to the masking effects of noise (Tufts et al., 2009). As hearing loss increases, better signal-to-noise ratios are required for detection, especially when auditory pure tone thresholds exceed 30 dB HL (Hétu, Getty, & Quoc, 1995; Tufts et al., 2009). Therefore, detection in quiet does not directly predict detection in noise. As such, pure tone thresholds recorded on an audiogram do not necessarily reflect the underlying status of the damage to the auditory mechanism (Donahue & Ohlin, 1993; Tufts et al., 2009). As previously mentioned, hearing loss not only results in decreased detection, but also distortion in perception of sounds. Walden and colleagues (1981) reported that even when audibility is ensured in those with a unilateral hearing loss, the ear with normal hearing performed significantly better in tasks of speech perception. These results support that hearing loss not only reduces detection of sounds, but hearing loss also creates distortion that cannot be restored via audibility.

Human hearing as a combat multiplier

The current practice of assessing hearing via pure tone threshold does not thoroughly evaluate the aspects of audition required to discern critical information. Functional hearing abilities describe the ability to communicate as a result of information collected from the acoustic environment through detection, recognition, and localization (Soli, 2003; Tufts et al., 2009). Intact functional hearing is a combat multiplier (Donahue & Ohlin, 1993). Ground Soldiers employ functional hearing to discern enemy versus friendly fire, locate snipers, locate patrol members, identify vehicles, and determine key nuances in their environment (Donahue & Ohlin, 1993). The documentary *Korengal* catalogued combat experiences of Soldiers assigned to securing the Korengal Valley, Afghanistan in 2010 (Quested, 2014). The following excerpts from interviews with ground Soldiers of various ranks during this deployment illustrate functional hearing in combat environments:

PV2 Misha Pemble-Belkin: First thing you hear when you get ambushed or you get in a firefight or whatever it is, the first thing you hear is just a loud crack.
Sergeant Brendan O'Byrne: The bullets passing by your head- the snaps.
Specialist Kyle Steiner: You hear that snap and your first…exactly how we are trained, the snap is the first instinct to is to just get behind something.
O'Byrne: That's exactly what you do. You get cover.
Sergeant First Class Mark Patterson: Then you find out by sound and distance "where's this coming from"…
First Sergeant LaMonta Caldwell: One of the things that you learn about is that you may not can see it, but you can hear it. That's called 'tactical awareness'. We [are] able to pick up the different sounds…

Patterson: You have to name the terrain features around you. You can't just say 'that green hill over there', is it coming from Honcho Hill, is it coming from 17-05? Is it coming from Spartan Spur?
Pemble-Belkin: Everybody just starts shooting in that direction.
Steiner: As a team leader, my second instinct is to find out where everybody else is. We return fire and I immediately start checking my guys…
Pemble-Belkin:…as soon as I hear a crack, I'm on the radio saying like 'troops in contact, we're taking fire from the east. There's probably five of them out there.'
Patterson: Then you have the squad leaders bouncing around from each position where the Soldiers are firing from trying to give me a situation report- give me a status. "Is the guy still there? Has he picked up and move?" I'm trying to paint a picture to the company commander because he is talking to higher trying to gather all these other assets.

Auditory situation awareness defined

Hearing enables Soldiers to quickly react, engage, and communicate in combat scenarios. The *Korengal* excerpt vividly illustrates Endsley's (1988) concept of situation awareness. In this model, through his or her senses, the decision-maker perceives an element and its germane details in the environment (Endsley, 1998). The decision-maker then connects incoming knowledge with pre-existing knowledge of relevant concepts to synthesize an overall picture, or gestalt, of the environment (Endsley, 1998). The individual can then attach meaning to the implications of certain elements and events (Endsley, 1998). Knowing how these elements and events will affect future events is the final stage of situation awareness (Endsley, 1998). Casali & Clasing (2013) provide a schematic in Figure 2 to demonstrate the processing steps in ASA.

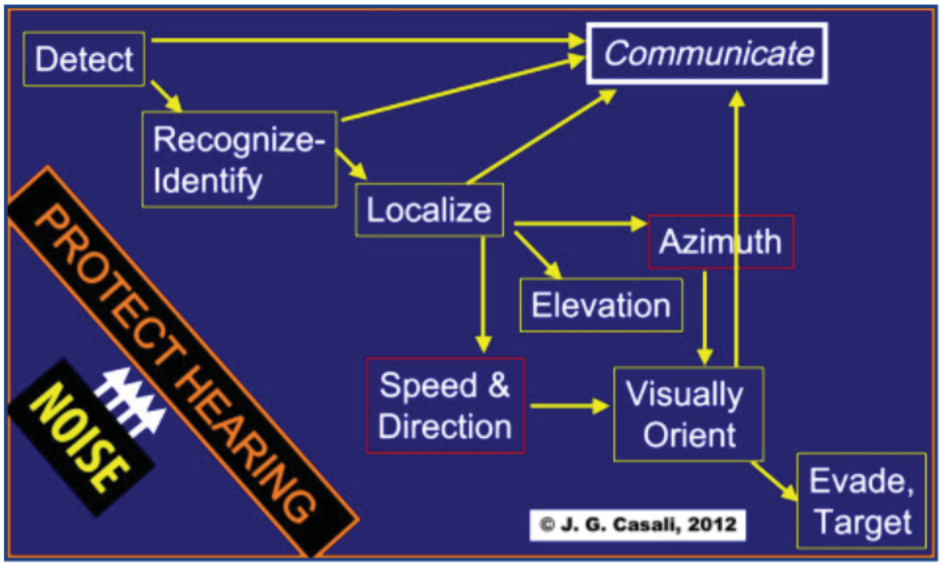


Figure . Schematic of auditory situation awareness adapted from Casali & Clasing, 2013, p. 11. Copyright 2012 by John G. Casali.

In a simulated ground combat game, those with degraded access to localization cues won the game substantially less often than those with severe hearing loss (Brungart & Sheffield, 2016). Contrary to those with hearing loss who knew to adapt a strategy to compensate for their handicap (i.e. passive hiding behaviors), a new strategy was not adopted by those with impaired localization imposed by the simulator. Therefore, those with degraded localization did not notice the absence of localization cues. Unlike degraded cues to detection, the authors found that degraded localization cues can have insidious and largely unnoticed effects on performance.

Auditory anatomy and physiology

Outer Ear

The outer ear, Figure 3, consists of an external flange, known as the pinna and the auditory canal. In general, the outer ear serves as a funneling device and resonator that emphasizes cues for transmission into the cochlea and enables sound localization (Pickles, 1988). The pinna's angling on the skull enables funneling of sound towards the tympanic membrane

(TM). Along with the angling, the contouring of the pinna shapes the collected sound and creates an “imprint” by filtering the acoustic signal with directionally-associated cues. The paired pinnae are slightly asymmetrical both in shape and position on the skull, similar in concept to how a right foot can be a slightly different size than a left one (Maroonroge, Emanuel, & Letowski, 2009). The asymmetry and dual receipt of acoustic information enables localization, and “customization” of localization cues among individuals (Maroonroge et al., 2009). The difference between the sound arriving at the ear and the influence of the individual’s head and pinna that results in directional-dependent filtering is known as the head related transfer function (HRTF) (Moore, Space perception, 1997).

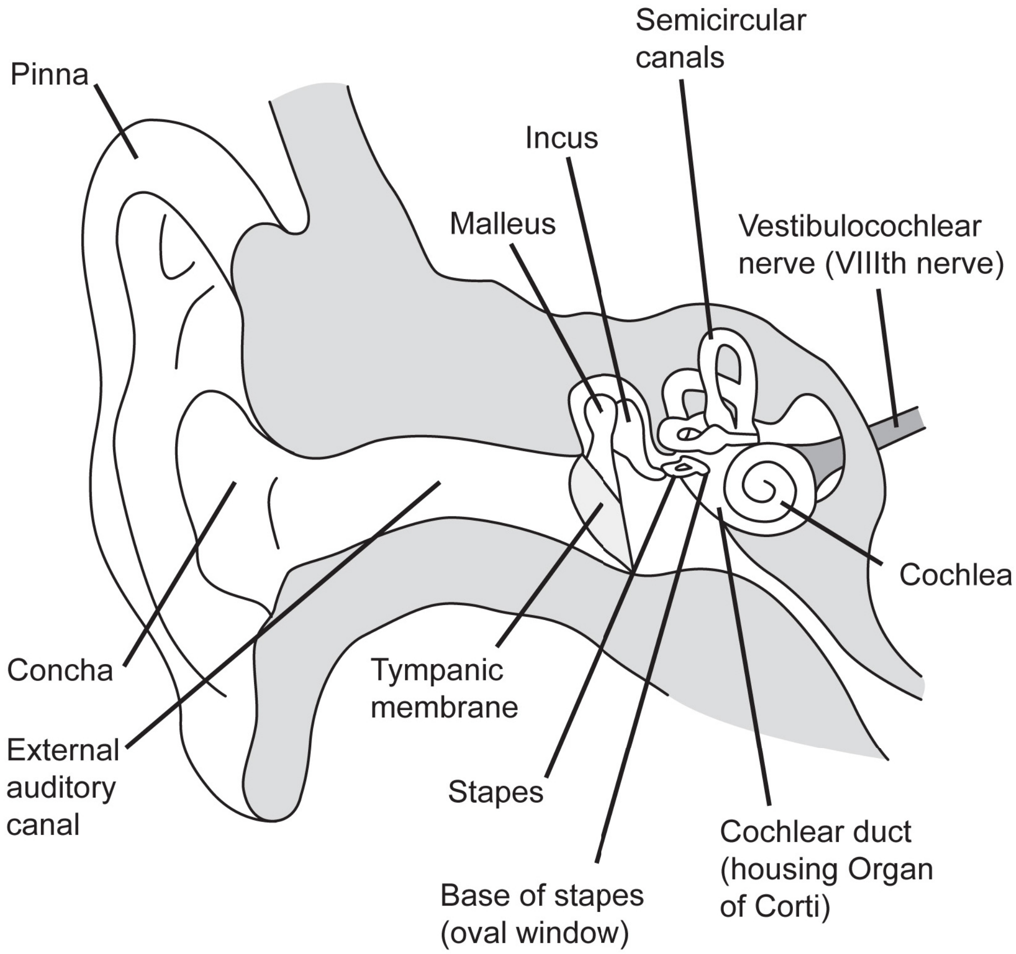


Figure . Cross-section of the outer, middle, and inner ear. Adapted from Hansen (2014), p. 469, Figure 8-35.

Upon filtering by the head and pinnae, sound is then funneled into the ear canal through a bowl-like depression in the pinna known as the concha (Maroonroge et al., 2009). The wider opening (cavum) of the concha into entrance of the canal, with a peak resonance of about 5500 Hz, further enables collection of sound towards the TM while protecting the TM from adverse environmental factors (Maroonroge et al., 2009; Pickles, 1988). The canal serves as an approximately 25 mm closed-tube resonator to enhance certain frequencies, especially those important to the perception of speech (Maroonroge et al., 2009; Pickles, 1988). Frequencies between 2000-7000 Hz are amplified, with a peak gain in the canal at 2000 Hz of between 15-20 dB (Pickles, 1998). The combined effect of the concha and canal results in a peak resonance occurring at 2500 Hz (Pickles, 1988). Therefore, the head, each pinna and comparison of cues thereof, and canal provide directional-dependent information.

Middle Ear

The boundaries for the middle ear are marked by the TM and the oval window at the beginning and end, respectively. The middle ear space is approximately 2 cm^3 and houses the three middle ear bones, malleus, incus, and stapes, and their supporting muscles and ligaments (Pickles, 1988). Together, the outer and middle ear overcome the energy lost when sound is transmitted through air to the liquid medium of the cochlea. The damping resulting from transmission from air to fluid results in approximately 30 dB of attenuation when it travels from air to liquid (Rappaport & Provencal, 2002). The middle ear employs three mechanisms to overcome the impedance mismatch. The first mechanism is the increase in force created by the same amount of energy transmitted from a larger surface area of the mallear attachment to the TM compared to smaller surface area of the stapedial attachment to the oval window. Areal mismatch results in 35 times more pressure at the stapes than the TM (Pickles, 1988). The second mechanism results from a lever action between the longer malleus and shorter incus, increasing pressure by 1.32 times (Pickles, 1988). Third, the TM moves in a buckling manner, but the malleus does not, which creates a four-fold increase in pressure (Pickles, 1988). The Eustachian tube provides aeration to the middle ear space and serves as a pressure-equalization mechanism.

Inner ear

At the inner ear, or cochlea, sound is converted from mechanical energy into an electrochemical message. As sound is transmitted through the footplate of the stapes, the oval window responds accordingly by moving the cochlear fluid. Housed within the cochlea is the organ of Corti that serves as the sensory receptor to sound. At the base of the organ of Corti are receptor cells, known as outer hair cells. The outer hair cells line the basilar member and are stimulated in response to fluid moving inside the cochlea. The organ of Corti also contains inner hair cells that serve as the link to transmit signals from outer hair stimulation to the auditory, VIIIth, nerve. Hair cells are organized in a frequency-specific fashion along the basilar member, and therefore selectively respond according to frequency. Hair cells at the base of the cochlea generate the greatest response to higher frequency information, whereas those at the apex generate largest responses for lower frequency signals (Gelfand, 1998). Several mechanisms exist for encoding signal amplitude within the cochlea, but in general, the response of the outer hair cells increases non-linearly with increases in intensity (Pickles, 1988). The motility of the outer hair cells creates amplification of lower-intensity signals and damping of higher-intensity ones (Pickles, 1988).

Auditory localization

Sensation and Perception of Localization

Sound location in the horizontal plane, or azimuth, is determined by signal differences between the right and left ears. A schematic of localization labeling conventions is provided in Figures 4 and 5. The two main binaural cues used for azimuth localization are interaural time difference (ITD) and interaural level difference (ILD) pictured in Figure 6 (Moore, 1997).

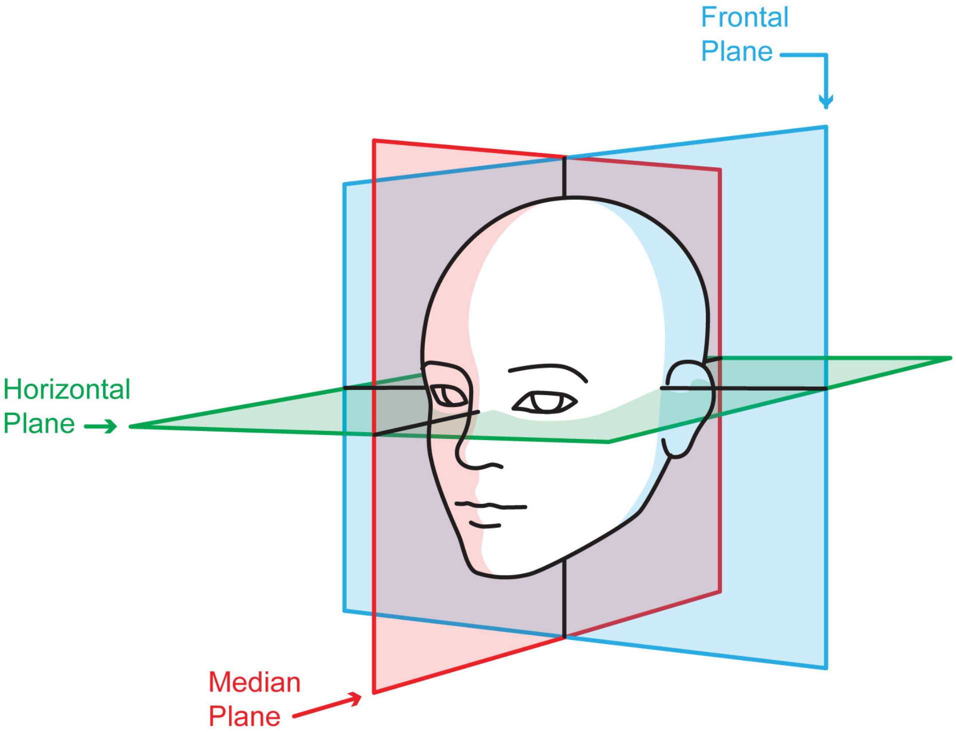


Figure . General description of sound orientation relative to the head. Adapted from Kendall, 1995, p. 26.

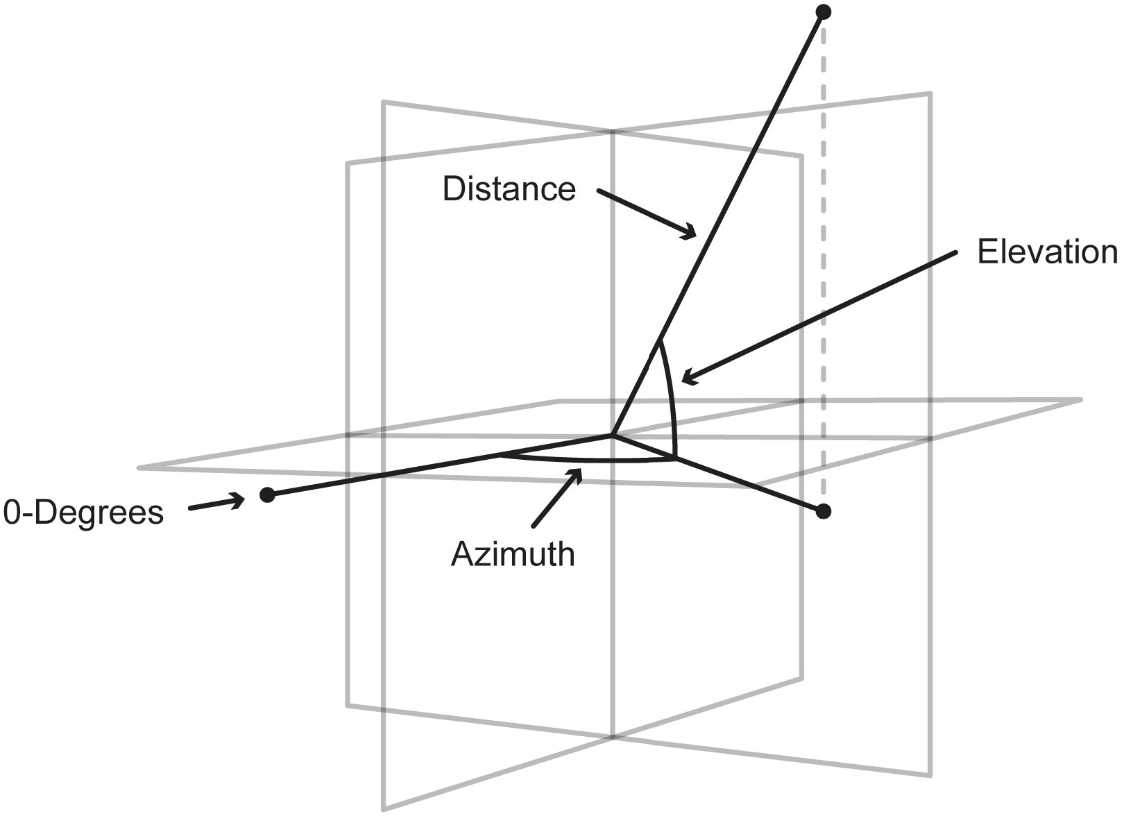


Figure . Depiction of azimuth, elevation and distance relative to the head. Adapted from Kendall, 1995, p. 26.

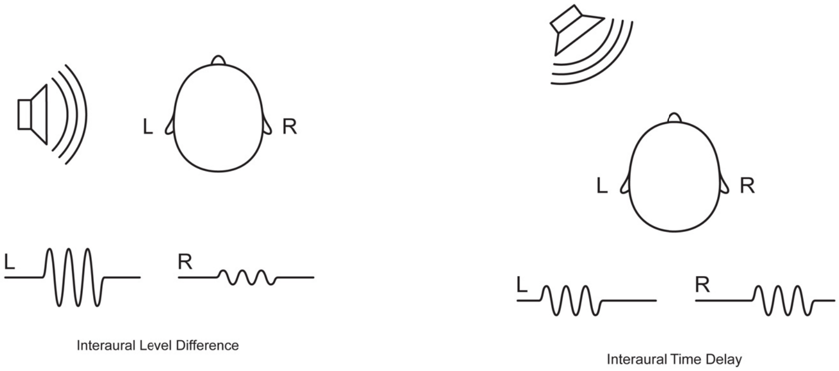
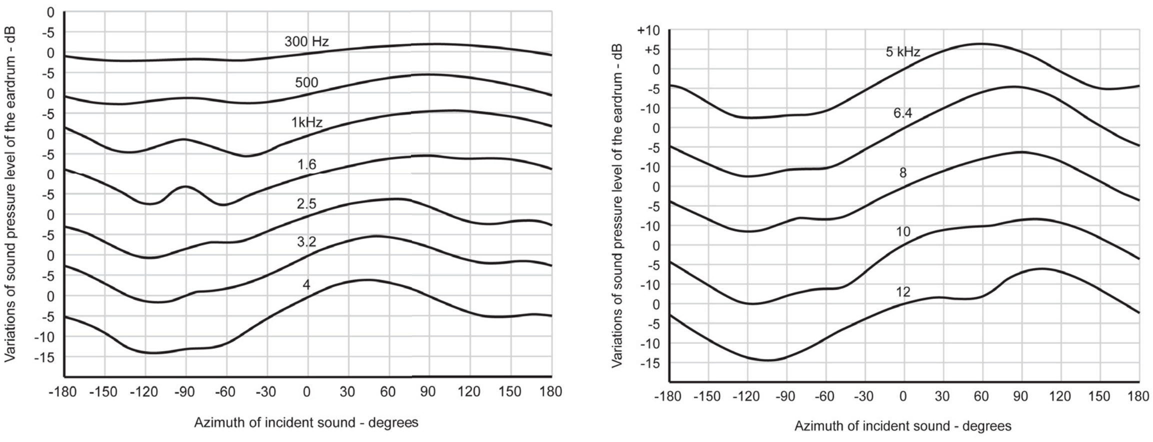


Figure . Depiction of interaural time difference (left) and interaural level difference (right). Adapted from Kapralos, Milios, & Jenkin, 2008, p. 528.

ITD cues predominate for localization of sounds below 1500 Hz as the wavelength of the sound must exceed the diameter of the head (Blauert, 1997). Essentially, wavelengths must be able to “bend around” the head to timing cues accessible (Kapralos et al., 2008). The head, torso and pinnae serve to filter sound such that some components of the sound are deflected while others are amplified (Emanuel, Maroonroge, & Letowski, 2009). When sound is located at the midline of the head, i.e., on the median plane, approximately the same signal arrives simultaneously to both ears (Emmanuel et al., 2009). However, if sound is off-center, the near ear will receive the sound before the far ear and will be reflected in the time for signal onset as well as the phase of the signal (Emmanuel et al., 2009). The perceived difference in temporal components are ITDs.

ILDs occur when the near ear receives a more intense signal. A shadowing effect from the head and torso deflects components of the signal spectrum arriving at the far ear (Emmanuel et al., 2009). Because of the longer wavelengths of sounds below 1500 Hz, the diffraction and subsequent attenuation do not generally exceed 5 dB (Kapralos et al., 2008; Wightman & Kistler, 2003). ILD cues are dominated by higher frequencies with frequencies above 2000-3000 Hz providing the most information (Moore, 1997).

When sound occurs in the frontal plane, the highly contoured surface of the pinnae and successive funneling into the ear canal resonates higher frequencies (Emanuel et al., 2009). This contouring creates spectral changes in the signal even with small changes in sound location. In particular, when the sound source is behind the listener, the pinna deflects sound, especially in the 3000-6000 Hz range (Pickles, 1988). Due to pinnae deflection of sound waves in the 3000-6000 Hz region, intensity changes are the most salient in this region as the sound source moves in the horizontal plane (Shaw, 1974). The concha provides filtering in higher frequency regions than the pinna. According to the azimuth location, attenuation occurs in the 10,000-12,000 Hz region due to reflections from the concha (Pickles, 1988). These reflections can result in phase cancellation (Pickles, 1988). The frequency of the cancellation, or dip, is highly directionally dependent, as shown in Figure 7. Therefore, discerning front from equal but opposite positions in the anterior versus posterior plane (i.e. 0 ° vs 180 °) is largely determined by the shape of the pinnae, a monaural cue (Scharine, Cave, & Letowski, 2009). However front-back localization also uses pinnae filtering as a binaural cue. A slight asymmetry in the orientation, size and shape of the pinnae enables access to some binaural cues (Scharine et al., 2009). Therefore, the pinnae contribute to directionally-dependent sound modulation, with spectral information changing considerably in the higher frequencies to enable front-back distinctions.



A. *B.*

Figure AB. Outer ear spectral shaping according angle of sound incidence. Adapted from Shaw, 1974, p. 461-462.

Sound characteristics

In general, the more spectral information incorporated into a signal, the smaller the minimum detectable change in azimuth, or minimum audible angle (MAA). Localization blur is defined as the smallest change in a sound parameter or parameters, including a correlated signal that can affect the original signal, that changes the location of the auditory signal (Blauert, 1997). In other words, localization blur is the precision with which a sound location can be identified (Letowski & Letowski, 2012). Broader bandwidth stimuli afford the listener access to more intensity and frequency cues, as shown in Table 2, adapted from Table 1 in Robinette, 2012 and Table 2.1 in Blauert, 1997, p. 39. The range in values are due to different experimental methodology. While broader band stimuli provide access to lower frequency ITDs and higher frequency ILDs, broader bandwidth stimuli must contain common information in the higher frequencies presented to each ear for this rule to hold true (Moore, Space perception, 1997).

Table . Localization blur according to stimulus time with sound presented at the midline.

Signal	Localization blur
Impulse (click or click train)	0.75-2°
Narrowband noise	1.4-2.8°
Speech	0.9-1.5°
Broadband noise	3.2°
Sinusoids	1-4.4°

Although broader band stimuli provide access to lower frequency ITDs and higher frequency ITDs, broader bandwidth stimuli must contain common information in the higher frequencies presented to each ear for this rule to hold true (Moore, 1997). As such, complex signals may provide each ear with disparate, and thus ambiguous, high-frequency information.

Figure 8 demonstrates that the ear is most accurate at localizing pure tones below 1000 Hz and from about 3000-6000 Hz, which corresponds to ILDs and ITDs. The figure also demonstrates that MAA increases as deviation from the midline increases.

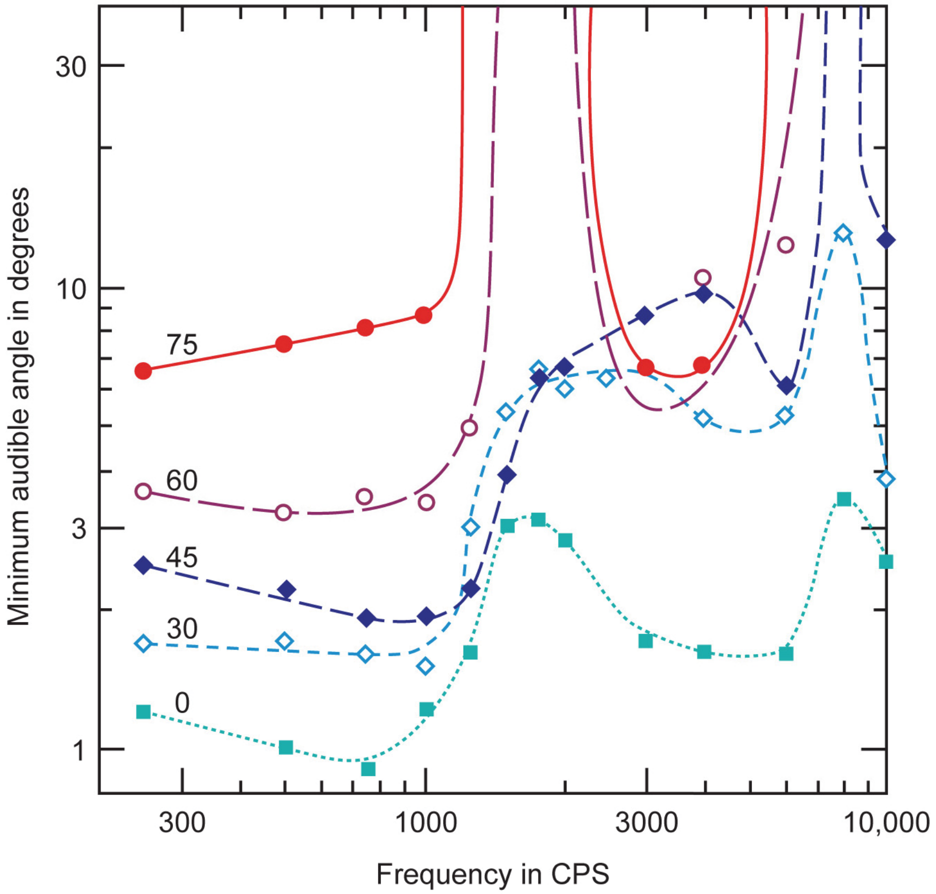


Figure . Minimum audible angle across stimulus frequency (sinusoids) for different azimuths. Adapted from Mills, 1958, p. 240.

Elevation perception, known as zenith, describes the angle at which a sound occurs in the vertical plane (Scharine et al., 2009). Unlike azimuth perception, elevation perception uses minimal, if any, binaural cues. Within a conical area, see Figure 9, any source occurring on the surface of the cone will render the same binaural cues (Kapralos et al., 2008; Scharine et al., 2009). Therefore, the contouring of the pinnae provides spectral filtering as shown in Figure 10A-C. The attenuation and resonance provided by the pinna provides cues used to discern elevation differences in the frontal plane and front versus back stimuli. Discrepancies between cues provided by minor asymmetries of the pinnae location can also furnish some localization cues.

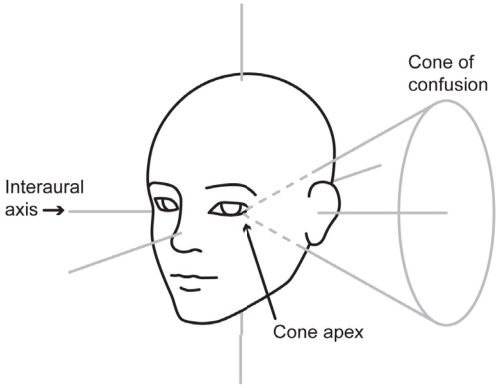


Figure . Illustration of the cone of confusion. Adapted from Kapralos et al., 2008, p. 529.

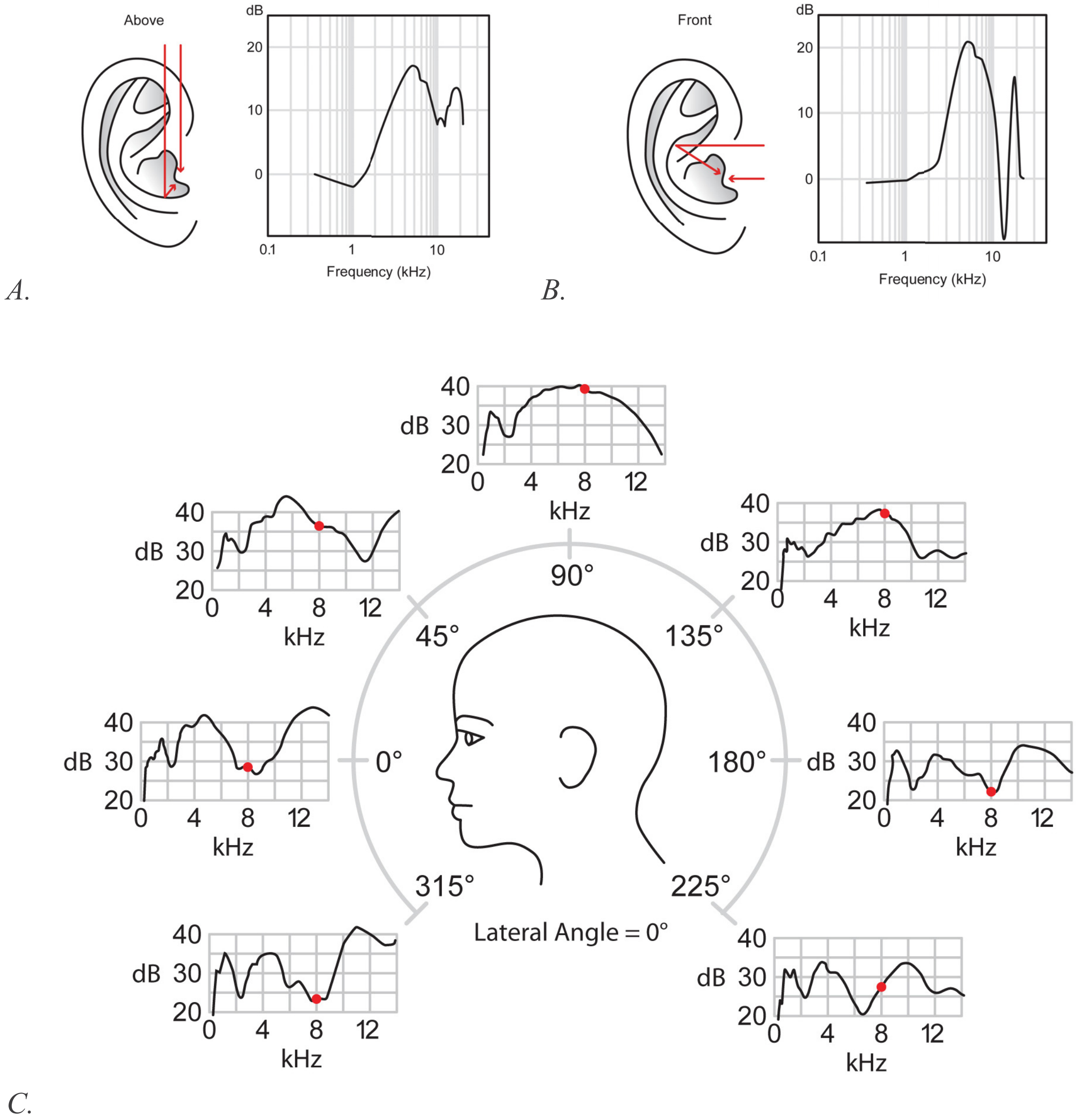


Figure . Spectral filtering performed by the pinna to shape elevation cues. Figures *A* and *B* adapted from Duda, 2011. Figure *C* adapted from K. Craig (2001).

Resolution in localization accuracy depends on location of the sounds as a function of availability of monaural and binaural cues (Moore, 2014). In the horizontal, frontal plane, the minimum detectable difference in angle can be as small as 1° (Kapralos et al., 2008; Kendall, 1995; Blauert, 1997). Specifically, in front of the listener, or 0° azimuth, accuracy can be 4° for sounds between 200 and 4000 Hz (Kapralos et al., 2008). For a 500 Hz tone, detection can be as exact as 1° (Kapralos et al., 2008). The minimum audible angle (MAA) increases as the sound source approaches 75° (sides) to 10° and reduces to 6° behind the listener (Blauert, 1997). The smallest MAA in the vertical plane is ± 9° directly in front of the listener, or 0° in lateral and vertical plane, but MAA increases to ±22° directly above the head, or 90° zenith (Blauert, 1997). MAA improves as elevation decreases behind the listener, reaching ±15° behind the listener (Blauert, 1997). Mechanisms for elevation perception are not as clearly understood as those employed in lateral localization. However, spectral shaping that creates peaks and notches in the sound spectra serve as key cues in vertical perception (Shaw, 1974). As Figure 10 demonstrates, sounds emanating from the frontal plane result in larger peaks and notches in the transfer functions compared to those above the listener. Given the more distinct information afforded by greater peaks and valleys in signals originating from the front, these signals are easier for the listener to discern than signals incorporating more similar information. Therefore, lower frequency stimuli at midline are easiest to perceive in azimuth while higher frequency stimuli in the frontal plane are easiest to localize in zenith.

Overall, localization in the horizontal plane is more accurate than in the vertical plane, and accuracy depends on location and spectral content. Horizontal and vertical accuracy blurs for broadband (white noise) stimuli are provided in Figures 11 and 12, respectively.

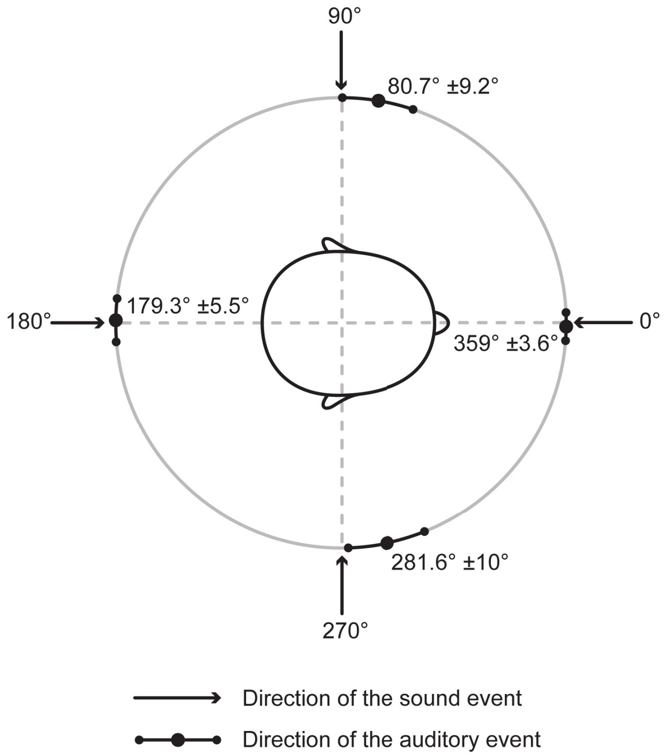


Figure . Localization blur in the horizontal axis using white noise stimuli. Adapted from Blauert, 1997, p. 40.

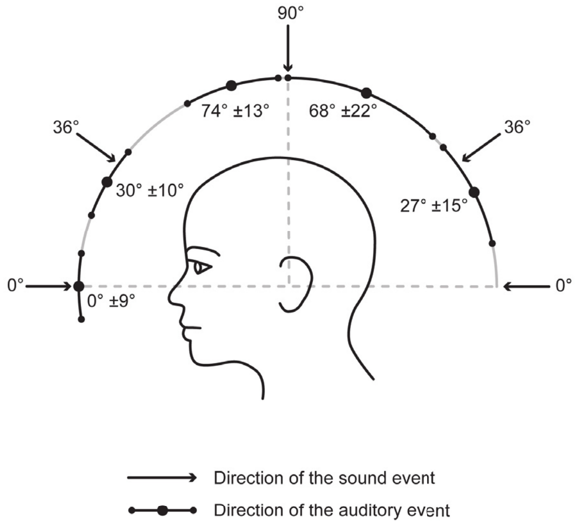


Figure . Localization blur in the vertical plane using white noise stimuli. Adapted from Blauert, 1997, p. 44.

Head movements

In many cases, humans can localize sounds with dynamic information afforded by head movements. Head movements can resolve ambiguities, especially front vs back discrimination by enabling greater access to ILD, ITD, and spectral cues (Scharine et al., 2009; Moore,1997). A head movement can displace an auditory image from a difficult-to-localize position to a location that can more easily employ binaural cues. For example, turning the head 30° may move the sound out of the cone of confusion. Likewise, if the head moves and the percept remains stable, the listener determines that the sound must be in the vertical plane, either directly above or below (Moore, 1997). Head movements can be employed to improve localization for signals of at least 600-800 ms in duration (Letowski & Letowski, 2012). Head movements require integrating information over time and is one of the longer auditory processes (Letowski & Letowski, 2012). Of note, head movements do not just alter ITD and ILDs, both head and source movement convey different spectral information, as illustrated in Figures 7 and 10. Therefore, head movements help resolve ambiguity in location by potentially providing more ILD, ITD, and monaural cues.

Distance judgements

Like localization judgements, auditory distance judgements also use ILDs, ITDs, spectral cues, and motion cues. However, listeners demonstrate far less accuracy in distance estimations compared to localization (Zahorik, Brungart, & Bronkhorst, 2005). For sounds less than 1 m, listeners tend to underestimate distances, whereas farther distance estimations tend to be overestimated (Zahorik et al., 2005). Intensity provides the most salient cue as decreases in intensity are associated with father source distances. Most likely, listeners employ the inverse square law where a doubling of distance results in a decrease in 6 dB (Zahorik et al., 2005). Of note, the inverse square law does not hold true in reverberant environments where reflections from the original sound source reinforce the signal intensity. Spectral content of a sound also contributes to the distance cues. Absorption from the air and other environmental factors reduce higher frequency information. As distance from the source increases, signals can be further altered since reverberation can increase, thus changing the spectral content (Zahorik et al., 2005).

Signal duration effects

Other monaural cues, such as signal duration, can also provide cues for localization. Rapid stimuli onset, particularly those less than 100 ms, results in “spectral splatter,” (i.e., increasing the frequency content of a signal) (Rakerd & Hartman, 1986). While shorter onsets provide important spectral cues, longer duration signals can better depict a spatial image (Letowski & Letowski, 2012). Pollack & Rose (1967) found that without head movement, localization accuracy increased from 10° for a 3 ms signal duration to 2° for a 1 s signal. The discussion of temporal characteristics of a sound warrants mention of the precedence effect. Listeners tend to localize based on properties of the earliest arriving waveform while disregarding later information. Specifically, reverberations greater than 1-20 ms and less than 10 dB of the primary sound are disregarded in localization perception (Letowski & Letowski, 2005).

Intensity effects on localization

In addition to ILDs, the intensity of the source signal can aid in localization. Sabin, Macpherson, and Middlebrooks (2005) found that azimuth localization of broadband noise improves rapidly up to approximately 10 dB sensation level (SL) and asymptotes at around 20 dB SL. Vertical localization reached asymptote at around 30 dB SL (Sabin et al., 2005). Due to the HRTF, listeners are not equally sensitive to sound at different azimuths. Specifically, results demonstrated that the best (lowest) thresholds were obtained at ± 45° and the worst (highest) thresholds were obtained at 180° (Sabin et al., 2005). Optimal localization is related to audibility of a full range of spectral cues, as the ear is not equally sensitive across frequencies (Sabin et al., 2005). For example, if frequencies above 4000 Hz are not audible, the listener will place more emphasis on ITD that is associated with frequencies below 1500 Hz. However, Scharine & Letowski (2005) reported that if cues are audible and not masked by other sounds, increases in intensity do not result in improved horizontal localization accuracy. Conversely, localization accuracy in the vertical plane exhibits a non-monotonic function where accuracy increases from about 20-40 dB of signal intensity and then decreases from about 55-65 dB (Vliegen & Opstal, 2004).

Sound/listener movement and Doppler effect

Movement of the source or listener can change the frequency content of the signal and subsequent auditory perception. Specifically, the Doppler effect occurs when the frequency of a sound changes as it moves closer, passes in front of and then moves away from the listener. As a sound approaches the listener, its wavetrain compresses (Letowski & Letowski, 2012). When the sound retreats from the listener the wave train becomes more expanded (Letowski & Letowski, 2012). The perceived frequency increases corresponding to the increased rate of movement towards the listener. Similarly, retreating sounds are perceived as lower in frequency. Additionally, as sounds move toward the listener and retreat, intensity increases and decreases, respectively. Therefore, perceived changes in frequency and amplitude due to the Doppler effect provide important cues to distance of a moving source.

Sound source movement

In addition to using the monaural cues of Doppler shift and intensity changes, sound source movement perception employs interaural differences. The minimum audible movement angle (MAMA) is the unit of measurement used to describe the smallest detectable angular displacement of a sound moving at a constant velocity (Letowski & Letowski, 2012). Essentially, a MAMA is the perception of how much a source moved. The MAMA exceeds that of the minimum audible angle (MAA) regardless of direction in a horizontal plane for the same sound having the same reference position (Chandler & Grantham, 1992). As with MAA, the MAMA accuracy decreases from midline. In particular, MAMA decreases from 5° at midline (on the medial plane) to 30° as the sound moves laterally to ± 90° (on the frontal plane) (Grantham, 1986). Chandler & Grantham (1992) investigated the effects of velocity, frequency, and bandwidth on MAMAs. The authors reported that increasing velocity increased MAMAs linearly above 10°/s. The results suggest a prolonged minimum integration time as detection times varied from 336 ms to 1116 ms with increasing velocity. Frequency manipulation showed the highest threshold at 3000 Hz and the lowest thresholds were at 500, 1000 and 5000 Hz. Similar to increased signal bandwidth in MAA experiments, increased bandwidth decreased MAMA thresholds. Therefore, MAMAs vary according to velocity, bandwidth, and angle of incidence (Chandler & Grantham, 1992). Furthermore, with higher velocity, sounds generated shorter duration signals resulting in larger MAMAs.

Environmental effects on localization

Elements in the environment, such as ambient noise, can interfere with cues normally used in localization. The extent to which noise interferes with localization depends on noise intensity, signal-to-noise ratio, noise spectrum, spectrum of the signal, and meaning of the signal (i.e. the listener perceives an alarm as an alarming stimulus). Alali & Casali (2011) examined the effects of differing noise levels and hearing protection on alarm localization performance. Specifically, the authors evaluated performance in 60 dBA and 90 dBA of pink noise in an open ear and several hearing protection conditions. The 90 dBA condition had a significant and adverse effect on localization with mean absolute error increasing from 17.5° at 60 dBA to 35.5° at 90 dBA. Front-back reversals increased from 7.2% at 60 dBA to 17.3% at 90 dBA. Of note, the authors employed 90 dBA of noise, the threshold for implementation of engineering noise controls according to OSHA (Alali & Casali, 2011). At the OSHA action level, localization performance was half as good or worse compared to the "quiet" condition of 60 dBA on the measures of left-right confusions, front-back reversals, and absolute deviation (Alali & Casali, 2011). One underlying cause for the adverse effects of noise on performance is the disruption of ITDs. Recalling that onset of sound provides an ITD cue, noise can mask this important temporal cue (Rakerd & Hartman, 1986). Scharine & Letowski (2005) reported that based on their review of Abouchacra, Emanuel, Blood, and Letowski (1998) and Letowski, Mermagen, and Abouchacra (2004), a signal-to-noise ratio (S/N) of -7 to -4 dB should be maintained in order to achieve 50% accuracy. Accordingly, the authors recommended a sensation level of at least 9 dB according to their review of studies by Smith-Abouchacra (1993) and Abouchacra & Letowski (2001) to reach a similar level of performance. Reflections in reverberant environments can serve as maskers to spectral cues. Reverberations can mix with the original signal to reduce ILD cues and create peak energy that provides false ITD cues (Scharine & Letowski, 2005). However, when the reflected energy emanates from the same area as the first wave front, thus eliciting the precedence effect, signal localization can improve (Rakerd & Harman, 1985). Therefore, noise, either ambient or subsequent to reverberations, can interfere with the interaural differences and mask spectral cues employed in localization.

Hearing Loss and Localization

The link between hearing loss and impaired localization remains obvious due to the loss of detection of cues. Surprisingly, unilateral and even asymmetrical losses can be more detrimental to localization accuracy than symmetrical losses (Humes, Allen, & Bess, 1980; Letowski & Letowski, 2012; Viehweg & Campbell, 1960). A unilateral loss disrupts the predominant cues of ILD and ITD comparisons needed to localize in a horizontal plane. Viehweg and Campbell (1960) reported that comparison cues were more important than high-frequency audibility in the better-hearing ear. The authors also supported that the amount of high-frequency loss did not predict performance, but degree of loss in the worse ear did (Viehweg & Campbell, 1960). In this study, localization errors increased as sounds moved toward the midline and away from the better ear (Viehweg & Campbell, 1960). The worst performance in unilaterally-impaired listeners occurred when the sound occurred behind the listener and toward the poorer ear (Viehweg & Campbell, 1960). The findings of this study suggested that when interaural cues are disrupted, the better hearing ear relied on spectral cues. According to Humes et al. (1980), less errors occur when using a low-frequency signal versus a high frequency one. Those with unilateral losses still had significantly more problems hearing in background noise when using a low-frequency signal (Humes et al., 1980).

Symmetrical hearing loss is known to be less detrimental to localization ability than unilateral loss: however, problems with impaired detection, or audibility, are strongly associated with poorer localization (Letowski & Letowski, 2012; Lorenzi, Gatehouse, & Lever, 1999; Noble, Byrne, & Lepage, 1994). Hearing loss configuration, type, severity, and listening experience can contribute to localization perception. Performance at the midline for hearing impaired listeners showed little difference than for normal-hearing listeners: however, performance degraded significantly more for hearing-impaired listeners in ± 90° azimuth, especially at reduced signal-to-noise ratios (Lorenzi et al., 1999). When stimuli were low-pass filtered for normal hearing listeners to match the losses of the hearing-impaired subjects, hearing-impaired listeners still performed worse. Results are suggestive of a distortion component in addition to loss of audibility that disrupts localization accuracy (Lorenzi et al., 1999; Noble et al., 1994).

Noble et al. (1994) further demonstrated that in symmetrical losses, the degree to which localization accuracy is affected depends on degree of loss, frequency region of loss, and signal level. Those with conductive losses performed poorer than those with sensorineural losses when matched on severity. Worse performance in those with conductive losses was most likely related to the loss of low-frequency, and thus, ITD cues. Of note, localization within the frontal horizontal plane remained intact as long as a severe loss was not present at 250-1000 Hz or 2000-4000 Hz. These results suggest that hearing-impaired listeners weigh ILDs or ITDs more heavily depending on the frequency region of the loss. Thus, while hearing-impaired listeners experience distorted localization cues, residual hearing within ITD and ILD bands may render frontal azimuth localization intact.

Hearing in Noise

Physiological Effects of Noise on the Cochlea

The pervasiveness of noise in U.S. Military environments is evidenced by the 2.8 million VA beneficiaries receiving compensation for auditory injuries and illnesses (VA, 2016). Troops trained for direct-combat related missions are at particular risk for hearing loss. As such, this population demonstrated a prevalence rate of 41.2 per 1,000 person-years with newly-diagnosed noise-induced injury (Helfer, 2011). The prevalence of noise-induced hearing loss stems from exposure to two common types of noise: impulse and steady-state noise. Impulse noise is noise that lasts less than 1 second, and has a decay time from either 20 dB or 10% of the highest peak is between 10 to 300 msec (DA, 2015). Impulse noise in excess of 140 peak SPL (pSPL) is considered hazardous and can cause physical damage to hair cells and supporting structures (Cave et al., 2007). Small arms fire, mortar rounds, tank-mounted weapons, and improvised explosive devices (IEDs) are all examples of impulse noise sources that exceed 140 dB pSPL (DA, 2015). Just one exposure to hazardous noise at this level can create permanent damage. Metabolic changes associated with the inflammatory process in response to cellular injury can occur long after the exposure ceases in animal models, resulting in permanent hearing loss (Kujawa & Liberman, 2009). OSHA also requires, and NIOSH recommends, a 140 dB pSPL limit to impulse noise exposures (NIOSH, 2014). NIOSH (2014) recommends the use of double hearing protection with fit attenuation testing when impulse sound levels exceed 150 dB pSPL. Steady-state noise is noise that lasts one second or greater and can result in a more gradual onset of hearing loss with repeated exposures (DA, 2015). Examples of excessive steady-state exposures include High Mobility Multipurpose Wheeled Vehicle (HMMWVs), tracked vehicles, and aircraft. According to NIOSH (1998), the eight-hour time weighted average damage risk criteria for noise-induced hearing loss (NIHL) is 85 dBA. However, 85 dBA does not necessarily mean no one will suffer hearing loss from exposures at this level. This action level means that 8% of people will experience auditory problems after 40 years of exposure (Fligor, 2009).

Both impulse and steady-state noise exposures can result in immediate, temporary hearing loss that can degrade situation awareness. This type of noise can also lead to long-term neural loss that may not manifest until years later (Kujawa & Liberman, 2009; Maison, Usubuchi, & Liberman, 2013). The clinical presentation of noise-induced hearing injury often includes tinnitus, hyperacusis, and high-frequency sensorineural hearing loss (Yong & Wang, 2015).

Noise can have detrimental effects on perception well below background noise levels that result in long-term pathophysiological processes in the cochlea. The masking effects of ambient noise within the cochlea is well documented. The ear processes sound according to three main parameters: intensity, frequency and duration. Intensity is encoded through increased hair cell activation rate, the area of the cochlea that responds, and the time between neural discharges (Pickles, 1988). Frequency is encoded through a combination of location (place) on the cochlea and timing cues (Pickles, 1988). Place on the cochlea refers to the spatially-organized highly tuned cochlear filters (Moore, 1997). These filters respond maximally to certain stimulus frequencies known as characteristic frequencies (CFs) (Moore, 1997). In order to encode timing, the discharge (synchronization) of neurons in the CF region are related to the repetition rate of the waveform (Moore, 1997). Temporal characteristics of sounds are encoded according to neural discharges synchronizing with the phase structure of the signals (Henry & Heinz, 2012). When noise reaches 80-90 dB, the finely-tuned filters in the cochlea broaden. The broader excitation along the cochlea creates distorted spectral information that subsequently distorts neural transmission (Casali & Tufts, 2020).

The distortion in the cochlea resulting from overloading regions corresponding to certain CFs can be attributable mostly to upward spread of masking. Instead of producing a sharp-peaked waveform on the basilar membrane, the waveform becomes rounded by exciting adjacent areas (Casali & Tufts, 2020). The cochlear region below the masker, which encodes higher frequencies, is most affected by the spread, as shown in Figure 13. The more intense the sound, the greater the spread of masking. Hawkins & Stevens (1950), as cited in Casali & Tufts, 2020, found that the degree to which the threshold of detection is elevated is directly proportional to the level of making noise regardless of frequency of the signal. For example, a 10 dB increase in masking will increase the signal threshold by 10 dB (Casali & Tufts, 2020). Therefore, noise mitigation to improve signal detection involves any or all of the following: attenuating the masker, increasing signals levels above those of the level the masker, but below 80-90 dB, or attenuating the signal and noise through hearing protection. However, the overall level transmitted through the protector should be below 80-90 dB to prevent cochlear distortion and hearing loss.

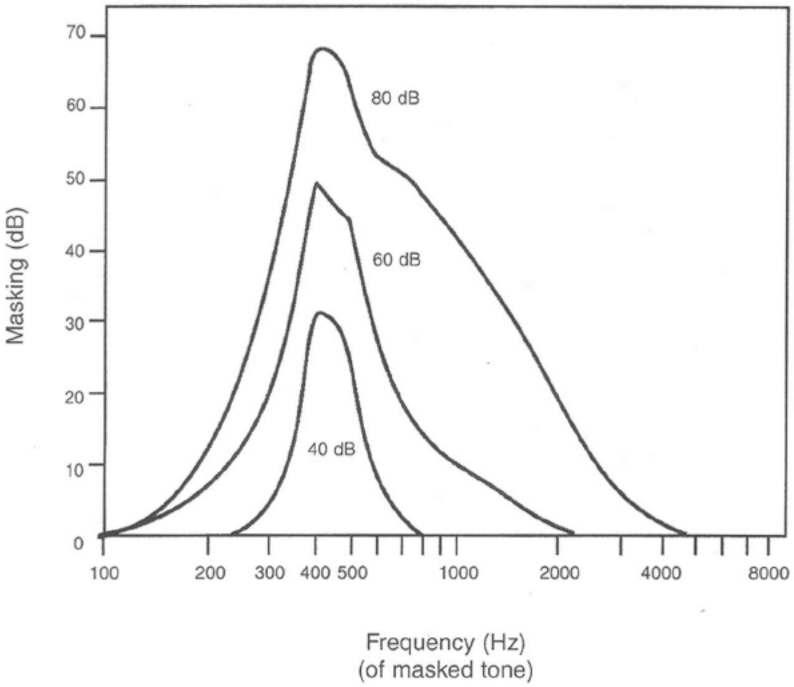


Figure . Illustration of upward spread of masking for a 410 Hz signal presented at 40, 60 and 80 dB. Adapted from Casali & Tufts, 2020 & Casali (2017), Figure 14.2.

Those with hearing loss, either temporary or permanent, are particularly vulnerable to the effects of noise. As Henry & Heinz (2012) explain, hair cell damage widens the highly selective filtering that otherwise separates signal from noise of different frequencies on the basilar membrane. Therefore, wider cochlear filters in those with hearing loss can render the signal and noise as indistinguishable due to processing within the same filter. While those with hearing loss may detect a sound in quiet at the same threshold as those with normal hearing, they are much more vulnerable to the masking effects of noise.

Non-Auditory Effects of Noise

While improving signal-to-noise ratio can improve audibility, audibility does not necessarily ensure listeners will attend to a signal. For example, Alali & Casali (2012) proposed that back-up alarms containing broader spectral information than traditional tonal alarms should be easier to detect. However, workers did not perceive broader band alarms as salient and meaningful, and therefore, did not detect these alarms as readily as the tonal ones. Non-auditory effects of noise can create stressors that can cause inattention to important environmental cues (Smith, 1991). Smith (1991) explains that the activation of certain cerebral areas increases production of the neurotransmitters catecholamine and cortisol. Higher levels of these neurotransmitters can lead to adverse changes in mood, emotion, and sleep disruption. An avowed link between fatigue and distractibility exists. Furthermore, catecholamines, a known marker of stress, have been measured at elevated levels in those exposed to noise (Smith, 1991). Therefore, the non-physiological adverse effects of noise can impact signal detection.

Hearing Protection Devices and Effects on Perception

Hearing protection attenuates noise and prevents cochlear distortion and subsequent damage, but the effects on ASA is varied. Certain types of hearing protection can improve speech understanding. For example, when headsets with active noise reduction (ANR) were used by pilots in civilian aircraft simulators, flight control performance increased (Casali et al., 2007). ANR headset use was attributed with the improvement due to decreased ratings of mental workload, increased speech intelligibility, and decreased hazardous noise exposure levels.

Giguère et al. (2013) investigated the effects of electronic level-dependent and passive hearing protection on speech recognition in noise in those with and without hearing loss. Compared to the open ear, normal hearing listeners showed little to no change in speech recognition when using passive hearing protection in hazardous noise. In those with hearing loss, use of passive hearing protection significantly decreased word recognition in noise. However, when hearing-impaired participants used electronic level-dependent hearing protection, speech recognition improved significantly compared to passive protection performance. For some hearing-impaired participants, electronic hearing protection use improved speech recognition scores compared to the open ear condition. Therefore, hearing protection with noise reduction strategies beyond conventional passive protection demonstrate promise as way to restore speech audibility needed for detection and recognition, especially in users with hearing loss.

Effects of Hearing Protection on Localization

While speech detection and recognition may improve with the use of certain types of hearing protection, wear of these devices can obscure localization cues. The most common hearing protection-related localization error, and the most perhaps detrimental, is front-back reversals (Abel, Boyne, & Roesler-Mulroney, 2009; Brown, Beemer, Greene, Argo, Meegan, & Tollin, 2015; Vause & Grantham, 1999; Zimpfer & Sarafian, 2014). Use of earplugs in combination with earmuffs can have particularly detrimental effects. Use of these combined passive protectors can result in the same localization performance as when only visual cues are available (Simpson, Bolia, McKinley, & Brungart, 2005). Powered earmuffs with pass-through capabilities consistently demonstrated the largest localization errors when compared to other hearing protection, both powered and conventional (Alali & Casali, 2011; Brown et al., 2015; Talcott, et al. 2012; Zimpfer & Sarafian, 2014). Hajicek et al. (2010) reported that personal protective gear that covers the ear, such as Kevlar ballistic helmets, can be expected to result in localization errors greater or equal to 30%. In comparison, bareheaded, open ear azimuth errors hover between of 3.6-10% (Blauert, 1997).

The results of Alali & Casali (2011) specifically illustrate localization differences among different hearing protector use amidst noise. The researchers found that powered circumaural hearing protection offered no improvement in localization compared to non-powered earplugs, including level-dependent and flat-attenuation plugs. However, a flat-attenuating earplug did demonstrate an advantage over the dichotic, or separate microphones for each ear, powered earmuff and foam earplug (highest noise reduction rating [NRR]) in 90 dBA versus 60 dBA of noise. No significant difference existed in flat attenuating earplug performance when compared to that of preformed earplugs or passive muffs. Front-back reversals were significantly less when using preformed earplugs versus electronic muffs. The study also examined localization using a diotic earmuff, where one microphone input feeds into each ear. Lack of disparate interaural cues using the diotic earmuff resulted in the worst localization performance in all conditions. Overall, localization accuracy decreased when noise increased from 60 to 90 dBA regardless of hearing protection device (HPD), and this same effect was observed in the open ear.

While Alali & Casali's (2011) study examined the effect of localization using alarms that employ more tonal stimuli, Talcott et al. (2012) examined the effects of HPDs on blank gunshots. Specifically, the researchers examined hearing protection performance using two gunshot bursts in quiet and amidst 82 dBA of heavy military vehicle noise (Talcott et al., 2012). Compared to a back-up alarm, the gunshot spectrum contained more broadband information. Performance was significantly better in the open ear than any of the HPD conditions. HPD conditions included a powered sound transmission circumaural earmuff, known as the Peltor Comtac™ II using full gain, the 3M Single-Ended Combat Arms™ non-linear earplug, and two powered earplugs. One of the powered earplugs used was the EB1-LE BlastPLG® in the "Lo" position that is close to acoustically transparency for inputs ranging from 0-115 dB. The other powered earplug used was the EB15-LE BlastPLG® also in the "Lo" position that acts a compression circuit. This circuit provides acoustic transparency for inputs below 60 dB, attenuates input of 85-155 dB by 15 dB, and passively attenuates for impulse noise from 120-180 dB. No significant difference existed among the HPDs, except performance with the circumaural earmuff was significantly worse than the other HPDs. Unlike Alali and& Casali's (2011) study, no main effect for noise existed, perhaps due to the fact that the two-shot stimuli were suprathreshold to all subjects in noise.

Similar to Alali & Casali (2011) and Talcott et al. (2012) results, Vause & Grantham's (1999) showed an increase in localization errors using a foam and flat attenuation earplug compared to open ear performance. The authors used an M16 cocking as the experimental stimuli. The increase in localization errors were mostly attributable to the increase in front-back reversals. Interestingly, when the stimulus was attenuated and tested in the open ear to match the predicted attenuation of the flat attenuating plug, errors were significantly worse in the frontal plane, suggestive of interruption of ITD through attenuation of the lower frequencies. Conversely, no significant increase in front-back reversals occurred for the open ear attenuated signal compared to the standard signal with the flat-attenuation earplug. The results suggested preservation of high-frequency information when using the flat attenuation earplug (Vause & Grantham, 1999).

While active earplugs show potential for restoring audibility and localization cues, the results vary depending on the particular device. For example, Casali & Lee (2016) demonstrated no significant difference in absolute localization between the open ear and the EB15-LE BlastPLG® electronic earplug in low and high ambient noise. Zimpfer and Sarafian (2014) found that two types of unnamed active earplugs were associated with significantly more front-back confusions than the open ear. However, the same study showed no significant difference in front-back confusions between the active and non-linear passive earplugs. When comparing active earplugs and active earmuffs, the performance with the earplugs showed significantly less front-back reversals. The researchers reported that non-linear passive earplugs showed the least amount of impact on localization performance. The researchers also conducted HRTF measurements with and without all of the HPDs. Measurements revealed that circumaural earmuffs disrupted cues relating to ILDs in the front-right (45°) and back-right (135°). Non-linear passive sound transmission earplugs and an active earplug that uses compression and ANR to control noise, effected spectral cues less than a proprietary powered earplug and the powered circumaural muff that uses peak-clipping. The authors speculated that the better performance of the passive non-linear earplugs and one of the powered earplugs was due to the shorter distance between the sound inlet and canal entrance. The implication of this design is better preservation of interaural phase differences and pinnae effects. Similarly, Brown et al. (2015) analyzed localization performance and compared errors against transfer functions. The researchers compared performance among the following hearing protection: a commercially available non-linear electronic sound transmission earplug, the EB15-LE BlastPLG®, a prototype of an active earplug similar to the EB15-LE BlastPLG® but filling more of the concha, and a prototype of a lower profile passive non-linear earplug, known as "ShotShields". Results showed that the lower-profile, non-linear earplug was associated with the least front-back confusions and the smallest acoustic distortion. This device also had the least concha-occlusive design, supporting that preserving high-frequency concha cues could minimize localization errors.

Auditory Learning

Plasticity and Adaptation of the Auditory System

While hearing protection can disrupt cues used in localization, evidence supports that listeners can relearn to localize with altered cues (Hofman, Van Riswick, & Van Opstal, 1998). Hofman, Van Riswick, and Opstal (1998), demonstrated that when custom, concha-filling molded earplugs were inserted into the pinna to disrupt spectral cues, elevation localization accuracy was immediately disrupted. Pinna filter functions were obtained to reveal large disruptions by the molds in the 6000-10000 Hz region, the region employed in localization, as shown in Figure 7, and especially those for elevation localization, as demonstrated in Figure 10. However, after three to six weeks, performance improved to a point of asymptote where accuracy recovered. Participants were not furnished with feedback during the adaptation phase. After the molds were removed, participants were re-tested without degradation to their pre-adaptation performance. The authors believed that the lack of interference in performance in the unoccluded, post-mold adaptation condition was evidence of a process similar to acquiring a new language. In essence, the brain developed a neural network for each "set" of pinnae and applied the one most appropriate for the situation (Hofman et al., 1998). As with the results obtained in Hofman et al.'s (1998) study of human listeners, Kacelnik, Nodal, Parsons, and King (2006) found adaptive localization processes present in animal models. Their data supported that even after monaural occlusion of one ear, and thus disruption of any interaural cues, ferrets could still achieve accurate localization. The authors maintained that improvement only existed when participants were allowed to employ altered localization cues in relevant tasks. Essentially, the animals needed time and opportunity to practice with altered cues. How often this practice was delivered also effected localization improvement. Animals tested daily achieved better results faster than those tested every six days with the same number of trials. The authors surmised that the animals learned to attend to and weight cues that were not as severely disrupted by the earplug. In the case of unilateral loss, the ITDs and spectral cues were less effected. Unlike Hofman et al.'s (1998) study, an interference effect was noted after initial removal of the earplug, but this effect disappeared after one session. Kacelnik et al. (2006) cited evidence from King, Hutchings, Moore, and Blakemore's (1998) to explain this adaptation. King et al. (2006) described physiological changes in mammalian midbrain's auditory space maps that occur after behavioral adaptation to altered localization cues.

Wanrooij and Opstal (2005) demonstrated human vertical localization recovery after disrupting spectral cues using monaural and binaural pinna ear molds. The authors explained that elevation localization occurs in two phases. First, spectral information is represented as a map of the auditory space similar to an individualized perception of an HRTF. Then, cues are analyzed and compared binaurally to determine the contribution of each ear according to the listener's position relative to the sound source. Binaural cues were largely unavailable in the monaural condition. Yet, elevation accuracy improved in the monaural condition when the participants underwent an adaptation period with their earmold. The authors believed the improvement in elevation localization after adaptation in the monaurally-occluded condition was due to neural plasticity mechanisms acting at the initial stage of spectral shape matching. Of note, only seven of 12 participants gained improved within 11 days, the other five never reached performance as good as the seven participants and did not asymptote.

Auditory Training

Wanrooij and Opstal’s (2005) study addressed specific mechanisms related to localization accuracy in a vertical plane. Wright and Zhang (2009) caution that auditory learning develops as a result of several patterns depending on the dimension of sound perception. No simple rule can be employed. For example, Wright and Zhang (2009) cited the work of Wright and Fitzgerald (2001) where participants were trained to attend to and discriminate among ILDs. Performance did not generalize to frequencies not used in training. However, evidence did support, but did not definitively conclude, that ITD perception could be generalized to untrained stimuli (Wright & Fitzgerald, 2001). Wright and Zhang (2009) showed that in listeners trained in temporal interval, ITD or ILD discrimination and then tested on all three tasks, all three training groups showed improved ability on ITD tasks. In other words, participants were tested in two conditions in which they did not received training and one where they did receive prior training. Those trained on the ITD discrimination demonstrated the most improvement. Results supported that some generalization of training occurs for ITD discrimination when trained on other dimensions of spatial sound discrimination, albeit not completely.

While certain active hearing protection may interfere with sound localization, localization accuracy may improve with training while using other devices. Casali & Robinette (2015) demonstrated that azimuth localization accuracy improved using two types of TCAPS and an open ear condition. One of the devices was an electronic sound transmission earmuff (Peltor Com-Tac™ II in its “unity”, or closest to transparent setting). The other device was an electronic sound transmission earplug (EB15-LE BlastPLG® in its “Lo”, or unity gain setting). Training with the open ear was subject to practice effects. However, the steeper slopes in error reduction for the two TCAPS conditions supported that improved performance was at least partially due to auditory learning (Casali & Robinette, 2015). Prior to training, localization accuracy using either TCAPS showed performance worse than the open ear. However, post-training, results of the electronic sound transmission earplug demonstrated performance that matched that of the open ear using a "ballpark" criterion of accurate within ±15°. Thus, the greater audibility cues offered by TCAPS compared to passive hearing protection may render more cues prone to the effects of training.

Casali & Robinette (2015) first evaluated localization in a high-fidelity simulator with 12 target speakers alternating between 12 dummy speakers, drawing upon the DRILCOM test system at Virginia Tech's Auditory Systems Lab (discussed in detail later herein). Therefore, accuracy could be tested every 15°. Sound stimuli were filtered pink noise that was one octave wide and incorporated low (500-1000 Hz) and high frequency information (3000-6000 Hz), and the bands were not harmonically related to avoid tonal qualities. Two different durations were employed, 300 ms, as to preclude use of head movements, and a three second presentation to allow for use of head movement. All participants were tested in a counterbalanced method using the open ear, an in-the-ear powered earplug (EB15-LE BlastPLG®), and an over-the-ear powered circumaural earmuff (Peltor ComTac™ II). Participants then underwent 12 sessions of localization training with either the EB15-LE BlastPLG® or Peltor Comtac™ II. Each session lasted approximately one hour on consecutive days, with no more than two days between sessions. Each session consisted of three, 15-minute tasks. The three tasks were as follows:

Task 1) A sound was presented from one of the speakers. Speaker locations were identified by a unique letter label. The participant was free to swivel his or her chair to the location of the sound source, and verbally indicate the location. The participant was then provided with feedback regarding the location.

Task 2) The participant repeated Task 1, but in addition to verbally responding he or she also pointed to the speaker from which he or she believed the sound emanated.

Task 3) Speaker labels were removed. Stimuli were presented and the participant was instructed to state the letter, with a computer screen as a visual reference. The correct location was then presented through the speaker from which he or she had just selected.

After completion of the three tasks, using the same stimuli, a test was administered lasting approximately 10 minutes. No feedback was provided. In the final phase of the study, participants were tested using the TCAPS with which they trained, the other TCAPS that they did not use in training, and with the open ear. Testing conducted in the final phase of the study clearly supported that learning occurred using the assigned TCAPS. Prior to training, performance in the open ear condition was significantly better than the assigned TCAPS conditions. However, testing in the final phase showed no significant difference between the open ear and performance when using the TCAPS to which the participant conducted the training, as shown in Figure 14.

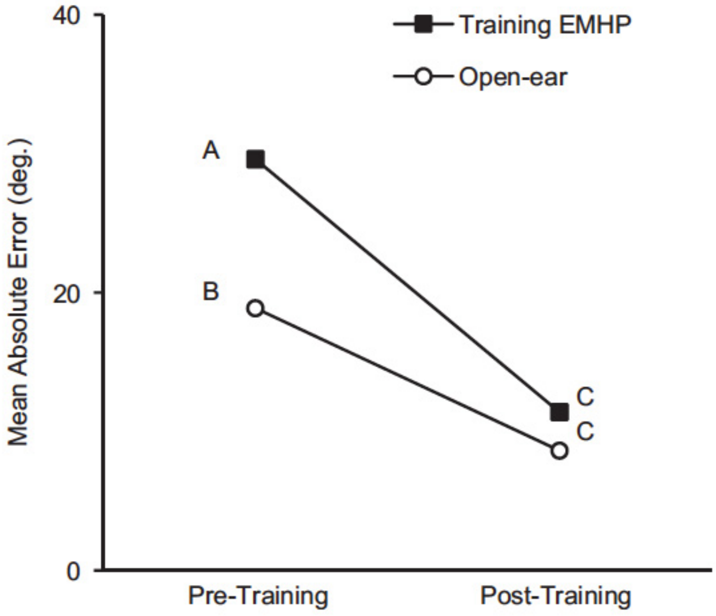


Figure . Pre-training and post training mean absolute localization error for the open ear and TCAPS with which the participant conducted training (referred to in this Figure as Training EMHP). Adapted from Casali & Robinette (2015), Figure 2, p. 5.

Of note, training with one TCAPS did not result in a transfer of learning to the TCAPS not used in training, i.e., no crossover effect occurred. However, localization accuracy with the EB15-LE BlastPLG® achieved results equivalent to that of open ear using the ballpark criterion after training occurred and improvements were significant. The same effect was also recorded using the Peltor ComTac™ II. The investigators concluded that no significant difference existed between the open-ear and the Peltor ComTac™ II prior to training due a small sample size. Similarly, after training with the device, the lack of significant difference persisted. The DRILCOM protocol has been employed in several evaluations of TCAPS, as shown in Figure 15. Most recently, using Casali & Robinette's (2015) training protocol, a separate evaluation demonstrated a similar training effect using a different TCAPS, the TEA INVISIO® X50. As of this writing, the TEA INVISIO® X50 is currently-issued by the U.S. Army for certain ground combat Soldiers. Another device that was in the prototype development stage was also tested as shown in Figure 15 as device “B”. The results in this study showed that after only five learning units, the participants using the TEA INVISIO® X50 achieved 89% accuracy compared to that of the open ear using the ±30° criterion, or ballpark accuracy (Casali & Lee, 2017). After 12 learning units, participants wearing the TEA INVISIO® X50 improved absolute accuracy, or within ±15°, from 60% to 80% (Casali & Lee, 2016b). This performance improvement rate using the TEA INVISIO® X50 was similar to that of the open ear (Casali & Lee, 2016b). The performance using the proprietary device did not approach the performance to that of open ear using both accuracy criteria levels (Casali, 2016b). After 12 learning units, localization performance using the proprietary device was approximately 30% worse than the open ear condition using the absolute criterion. The researchers attributed the poorer localization performance in the prototype device condition to a considerably different microphone design than the TEA INVISIO® X50 (Casali & Lee, 2016b). The results listed in Figure 15 demonstrate the importance of conducting psychophysical testing germane to situational awareness. Despite rigorous training, testing using the prototype device revealed poor fidelity of localization cues. The deleterious consequences of poor localization ability to ground combat Soldiers in certain situations are incontrovertible.

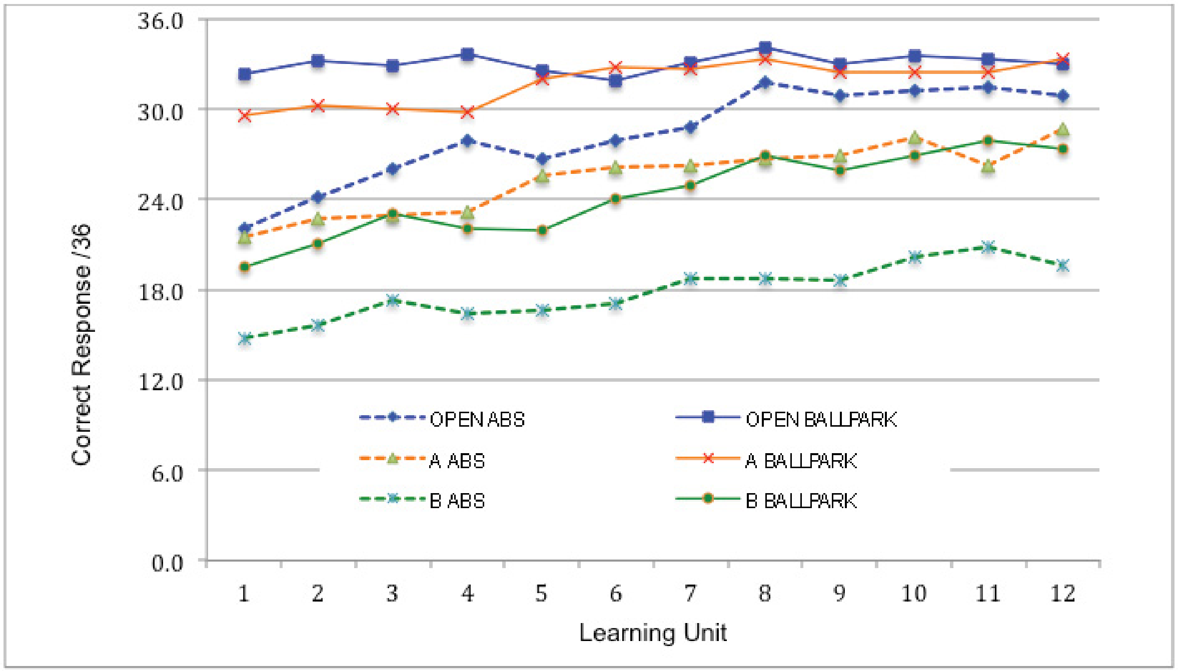


Figure . Training effect demonstrated using a training protocol delivered through the DRILCOM test battery and apparatus. Device “A” is the TEA Invisio and “B” is the prototype device. Adapted from Casali & Lee (2016), slide 47.

Auditory Training as a Clinical Rehabilitative Tool

The training used in the aforementioned Casali & Lee (2016) study employed an audiological clinical methodology known as auditory training. This method involves actively attending to auditory stimuli in order to improve auditory task performance (Henshaw & Ferguson, 2013). Clinical utility of auditory training in those with auditory impairments is often eclipsed by lack of reimbursement by health insurance providers, time-intensive delivery, and lack of evidence-based practice (Sweetow & Palmer, 2005). However, Sweetow & Palmer (2005) cite a study by Schow & Nerbonne (2002) that showed that 88% of audiologists dispensing hearing aids provide hearing aid training, but only 23% offer speech communication training. Citing the work of Kiessling et al. (2003), Sweetow & Palmer (2005) emphasized that audibility provided by hearing aids does not ensure communication. Hearing aid users also must be able to know what sounds to attend to in order to recognize an audible cue as such. Then, the listener must incorporate these sounds into an understanding of the communicated message and return communication (Sweetow & Palmer, 2005). In essence, the same concepts that apply to hearing aid rehabilitative services apply to ASA for those operating in degraded communication environments.

Systematic reviews of evidence-based efficacy of auditory training reveals a dearth of high-quality studies to support indisputable auditory training methodologies (Henshaw & Ferguson, 2013; Sweetow & Palmer, 2005). However, these same reviews did note a general improvement in the trained auditory task (Henshaw & Ferguson, 2013; Sweetow & Palmer, 2005). Plasticity of the auditory system has been confirmed by physiological means even if clinically-obtained evidence may only weakly reflect this. Mammalian plasticity in the superior colliculus of the brain is responsible for adaptation of localization, but the exact physiological underpinnings for auditory learning are unknown (Sweetow & Palmer, 2005). Sweetow & Palmer (2005) cite the work of Tremblay (2003) that referred to specific cortical changes following auditory training as support for malleable auditory mechanisms.

General Training Methodologies

Sweetow & Palmer's (2005) recommended adherence to simple learning theory principles for auditory training given the scarcity of high-quality training programs and related research. Guiding principles developed by Wolfe (1946) are still relevant in many training environments today. His principles can be recapitulated as follows: Practice does not make perfect, but perfect practice yields perfect results. Six principles comprise Wolfle's (1946) training theory:

1. *Distribution of practice*: Time compression of training can overload the learner's capabilities. Conversely, longer training sessions do not necessarily mean better performance.

2. *Active participation*: Having trainees participate versus only observing improves skill acquisition.

3. *Variation of material*: Varying material can better ensure trainees learn an underlying skill versus memorizing how to complete a specific task that cannot be generalized. Monotony can also lead to inattention; therefore, greater variation can improve motivation.

4. *Accurate records of progress:* Records enable the instructor to monitor and tailor guidance to the individual training. Keeping the trainee apprised of progress can also serve as a source of motivation and identify any areas of improvement.

5. *Knowledge of results*: Wolfle (1946) identifies this concept as one of the most important to performance improvement. Without immediate feedback, trainees can implement and reinforce the memory of incorrect training procedures. Practicing without accurate feedback can actually result in performance degradation.

6. *Systematic lesson plans:* Even the most inexperienced instructors can deliver high quality training if training content was first organized into simple logical steps. Effective instruction should initially have a clear objective and correspondingly clear methodology to achieve the objective. Instruction should first provide simple information that builds into successively more complicated material.

Additionally, current learning theory focuses on transfer-of-training. Transfer-of-training means that the trainee recognizes when to apply a certain skillset outside of the training environment and to other applications other than those learned while in training (Hays & Singer, 1989). Efficiency of training is paramount in most training environments. Training efficiency is conceptualized as the best learning achieved in the shortest amount of time, with the highest length of retention with the least amount of capital (Wickens, Hollands, Banbury, & Parasuraman, 2013). Presentation of experimental conditions does not necessarily facilitate efficient acquisition of an essential skill unless interference is possible. Interference is when stimuli have commonalities, but their responses are different or when stimuli are different but responses are similar (Wolfle, 1951). In instances when such commonalities exist, the training environment should replicate the response environment to a higher degree of fidelity (Hays & Springer, 1989). In other words, the trainee should be able to recognize when deviations in conditions and subsequent responses occur. Therefore, the goal of any training program is employing the most efficient means possible to enable the trainee when to recognize and use the skill in the right situation.

As previously noted, Wolfle (1946) and Sweetow & Palmer (2005) emphasized the importance of timely feedback on training. Figure 16 demonstrates how timely delivery of feedback can affect development of skill acquisition delivered longer than a minute after the observed behavior. In addition to providing direct verbal feedback, a method of reinforcement over feedback is called adaptive training. In general, adaptive learning systems involve monitoring learner characteristics and adjusting training content and delivery to improve learning (Shute & Zapata-Rivera, 2012). Essentially, adaptive training seeks to identify learner characteristics and uses these characteristics to tailor training content and enhance trainee learning. Use of adaptive mechanisms requires accurate diagnosis of a learner characteristic in order to prescribe optimal training content (Shute & Zapata-Rivera, 2012). McMullen & Wakefield (2017) found that sound localization accuracy improved significantly when listeners using a 3D spatialized audio headset were provided with immediate feedback. However, simple exposure to stimuli without training showed no significant improvement in azimuth localization from pre-test to post-test. However, the authors stated that performance improved within just one session of localization training (McMullen & Wakefield, 2017). In essence, how a trainee learns can assist the instructor in what content should be delivered. Wickens et al. (2013) explained that adapted training tasks should be representative and should work up to the level of difficulty required in the transferred environment, otherwise lower-level skills applied to more demanding tasks could inhibit performance (Wickens et al., 2013).

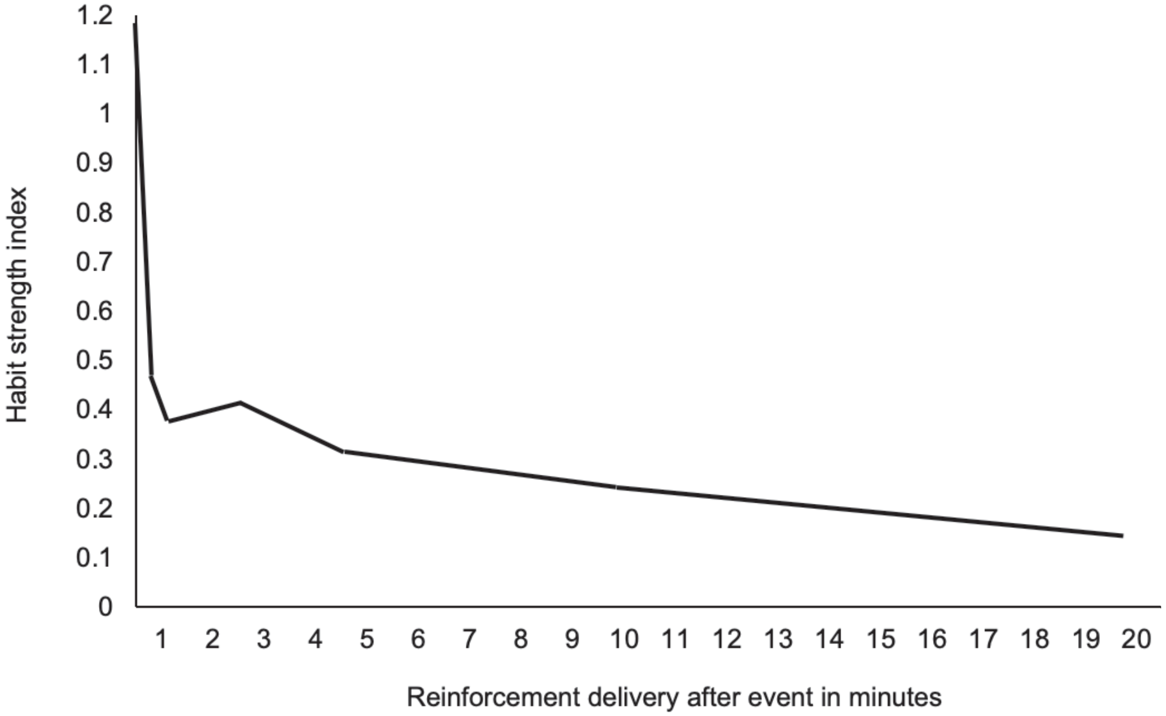


Figure . Habit strength index as a function of delay in feedback. Adapted from Wolfle (1951), p. 1268.

Integral to the role of training is the working and long-term memory capacity. Working memory reflects the capacity for an individual to temporarily hold information in his or her consciousness and performs work on the information (Wickens et al, 2013). Then, information either decays (is forgotten) or is encoded into long-term memory. The process of encoding is achieved through training. The challenge with training in sound localization is that sound and visual-spatial information is fused to result in an integrated percept. Working memory is thought to be limited by attentional control, and therefore, rehearsal of information provides opportunities for more meaningful associations between the percept and how it interacts with the environment. Training involves rehearsal of emerging skills while providing salient feedback. Thus, feedback enables the trainee to assign greater meaning to the information. Memories with stronger associated meaning increases the likelihood information is transferred and maintained in long-term memory.

Endsley (2012) states that in studies regarding attributes associated with SA, differences were identified among skill abilities. However, those with stronger SA and skill abilities possessed skills that were trainable. Automation of an acquired skill frees up working memory resources, an important component to ASA. Elements of a new situation requires combining new information with previous information and cannot be referred to externally as is possible with visually-presented information (Endsley, 2012). Therefore, auditory training shows potential in enabling a listener to free up resources in the working memory for other aspects of SA.

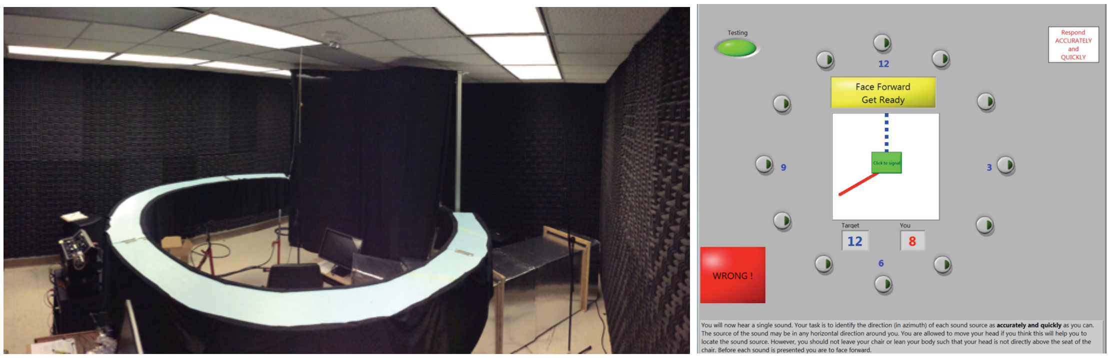
Objectives

The DRILCOM test battery and apparatus previously demonstrated its sensitivity to differences among several active and passive level-dependent hearing protectors on various aspects of auditory situation awareness (Casali & Robinette, 2015; Lee & Casali, 2019). This same instrumentation shows potential as a means to deliver training to improve localization accuracy, as demonstrated by Casali & Robinette (2015). As such, the Office of Naval Research (ONR) provided a contract to develop and validate a horizontal auditory localization training regimen. This regimen was designed for later incorporation into a portable auditory localization training system that would be sensitive to design-imposed differences among HPDs. The eventual goal of the training regimen and portable system was to improve auditory localization in U.S. Military Service Members, and others who work in jobs that require sound localization, both with and without HPDs. The focus of this study, subsequently referred to as Phase I, was the development of effective stimuli and presentation protocols for use in a training regimen for later incorporation into the portable system’s design, covered in a follow-on study. Therefore, the purpose of Phase I was to develop an improved azimuth training protocol for eventual use in a field-validated portable localization training system.

The localization protocol used in the DRILCOM auditory training regimen employed a stimulus that incorporated spectral cues needed for localization. Specifically, DRILCOM used a dissonant, non-harmonically related, tonal signal that contained 104, 295, 450, 737, 2967, 4959, 7025 and 7880 Hz (Casali & Lee, 2019). These frequencies rendered the predominant horizontal localization cues accessible using ITDs, ILDs, and pinnae spectral cues. While the dissonant signal had prior been demonstrated to be "trainable," whether this signal transfers to localization accuracy using military-relevant signals remained unknown. Therefore, to meet the objective of training optimization in the Phase I, an experiment addressed the presence of training transfer from DRILCOM's dissonant signal to untrained military-relevant stimuli.

Other objectives for this training optimization study included evaluation of the effects of decreasing training duration and using active participation, as opposed to passive observation, training techniques. The DRILCOM training strategy used by Casali & Lee (2019) incorporated 12 Learning Units (LU), with each LU lasting approximately 20 minutes for a total of about four hours. The basic experimental task involved the participant selecting on the display via mouse, "click to sound", a sound then played from one of 12 locations, and the participant would respond by selecting the speaker location on the display via mouse click. Prior to the training, participants were given a familiarization task in which they practiced initiating and responding to the experimental signal. The participants were also shown how the system would display feedback for a correct and incorrect response. Participants were instructed to respond as *accurately and quickly as possible* once they initiated signal delivery. In Casali & Lee's (2019) DRILCOM study, each LU was comprised of the following steps:

1) **Sequential** - The dissonant signal was presented around a 12-speaker “clock face” array, separated by 30° (i.e., a speaker located at each “hour”), in order, for a total of five “laps” around the 12-speaker array. The speaker array was hidden by an acoustically transparent curtain and pictured in Figure 17. The participant had prior knowledge of which loudspeaker location would play the sound, the participant would respond, and then the display would indicate the actual sound location versus participant response, as shown in Figure 17.



You will now hear a single sound. Your task is to identify the direction (in azimuth) of each sound source as **accurately and quickly** as you can. The source of the sound may be in any horizontal direction around you. You are allowed to move your head if you think this will help you to locate the sound source. However, you should not leave your chair or lean your body such that your head is not directly above the seat of the chair. Before each sound is presented you are to face forward.

Figure . DRILCOM testing facility and participant interface demonstrating the 12 response location options and system-generated feedback for an incorrect response. Adapted from Casali & Lee (2019), Figure 1, pS67 and Figure 3, pS69.

2) **Random**- The participants had no prior knowledge of which speaker would play the signal, but received immediate feedback if he/she answered correctly and the actual location of the signal. The signal was played from each speaker location five times in random order.

3) **Test**- A signal played from each location three times in random order for a total of 36 presentations. The participant would not know the signal location ahead of time and did not receive feedback regarding the response.

The secondary objective of the overarching project was to validate a new acclimation test and training system against a validated in-field localization test developed by Casali et al. (2012). Another research study, subsequently referred to as Phase II, was conducted by another graduate student (Brandon Thompson) and addressed the hardware/software configuration and equipment optimization that delivered acclimation training using TCAPS. In contrast, Phase I served to develop a training protocol to improve auditory localization for use in the follow-on equipment configuration study. A final phase, or Phase III, of the evaluation was conducted as a combined effort from both investigators to evaluate the effectiveness of the training protocol and portable equipment in a real-world environment.

Summarized Objectives

1) Identify, record, and test four military-relevant signals that have sufficient acoustical cue content that provides opportunity for auditory localization for inclusion in the optimization protocol.

2) Develop localization protocols that build upon those already implemented in DRILCOM to improve azimuth accuracy.

3) Compare new learning methodologies to determine what strategies produce the most accurate results, in an optimal number of trials, and determine when localization accuracy metrics asymptote.

4) Determine signal conditions that improve training according to:

    a) Signal

    b) Training methodology

    c) Number of signal presentations

5) Develop an improved localization training strategy, based on the DRILCOM training strategy for use in Phase II.

6) Upon completion of Phase II, evaluate the validity of the acclimation training and portable system in an in-field environment.

Hypotheses

1) Training with a dissonant, tonal complex signal transfers to localization of untrained military-relevant stimuli.

2) Shortening the number of presentations within each LU compared to previously-implemented DRILCOM protocols will result in training transfer.

3) Implementing user-engaged localization training strategies will improve localization accuracy.

4) Localization training conducted on a portable apparatus will transfer to field performance of live gun-fire stimuli.

5) Training on the portable system and testing in the field environment will be sensitive to localization performance differences with an over-the-ear and in-the-ear TCAPS, and with the open ear.

PHASE I: METHODOLOGY

Pilot Experiment 1

The first pilot sought to shorten the number of presentations within each LU. To recap, the DRILCOM protocol incorporated 12 LUs. Each LU contained two training sessions and one test session. First, the software presented five sequential "laps" around the "clock" occurring in clock-wise and counter-clockwise order. The next training subunit was comprised of five random presentations from each of the 12 speaker locations. Lastly, in the test subunit the test stimulus was played three times from each location in random order. In this pilot experiment, the number of presentations in the sequential and random subunits were reduced. Participants were randomly assigned to conditions consisting of either two, three or four-presentations. All participants underwent eight LUs with the dissonant signal. Participants were trained using the dissonant signal presented at 55 dBA with 40 dBA of pink noise in order to mask ambient sounds. The accuracy performance measure was the number of speaker locations correctly identified out of 36 presentations for each test. From this measure, the mean score according to presentation condition was computed. Specifically, this pilot was used to determine which number of presentations resulted in 90% accuracy in the least amount of time. The 90% criterion was established due to the approximately 88% absolute localization accuracy achieved in Casali & Lee (2019) that involved a more difficult task of choosing from 24 speakers versus the 12 in this pilot study. The investigators determined from the Casali & Lee (2019) study that performance reached asymptote at eight LUs. Therefore, the number of LUs were reduced from 12 to eight in the pilot study.

After completion of eight LUs with the dissonant signal, participants underwent a test using four military-relevant stimuli. The purpose of the testing with the untrained stimuli was to evaluate if training using the dissonant signal generalized to untrained stimuli. A subject matter expert panel was assembled to assist in determining which sounds were important for ground combat U.S. Service Members to localize. Seven Soldiers with at least one deployment, and with as many as six, with direct combat experience were queried individually. Ten Infantry Marines collectively responded from the Gruntworks program. Gruntworks is a U.S. Marine Corps organization tasked with research, design, acquisition, and equipping Marine rifle squads. Via email, panel members were asked to provide sounds that they felt were important for Service Members to localize. Respondents were asked to rank order their selection with 1 being the most important and 5 being less important. From the eight responses received, the most commonly ranked items in order of importance were: (1) small arms fire, namely the AK-47 rifle, (2) indirect fire (e.g. mortar and rocket-propelled grenade), (3) speech (e.g. shouted commands or native population to the conflict zone), and (4) aircraft. From the aforementioned responses, the untrained test signals were recorded in one second durations each, and consisted of the following: (1) an AK-47 three round burst, (2) a whistle tube simulator to emulate an incoming rocket-propelled grenade, (3) an utterance in Arabic, (4) an Apache helicopter. The one-second duration was maintained from the original DRILCOM experiments in order to serve as a basis of comparison (Casali & Lee, 2017). The order of presentation of the test signal was randomized. Therefore, number of presentations (two, three, and four) in each training subunit (i.e. sequential and random), served as the between-subjects factor. The untrained test stimuli (i.e. AK-47, Apache, Arabic, and whistle tube) served as the within-subjects factor. The spectral content of each stimuli is depicted in Figure 18. The dissonant signal contained the following frequencies: 104, 295, 450, 737, 2967, 4959, 7025 and 7880 Hz (Casali & Lee, 2019). In sum, Pilot 1 sought to shorten the subunits of each learning unit and evaluate transfer-of-training for untrained stimuli.

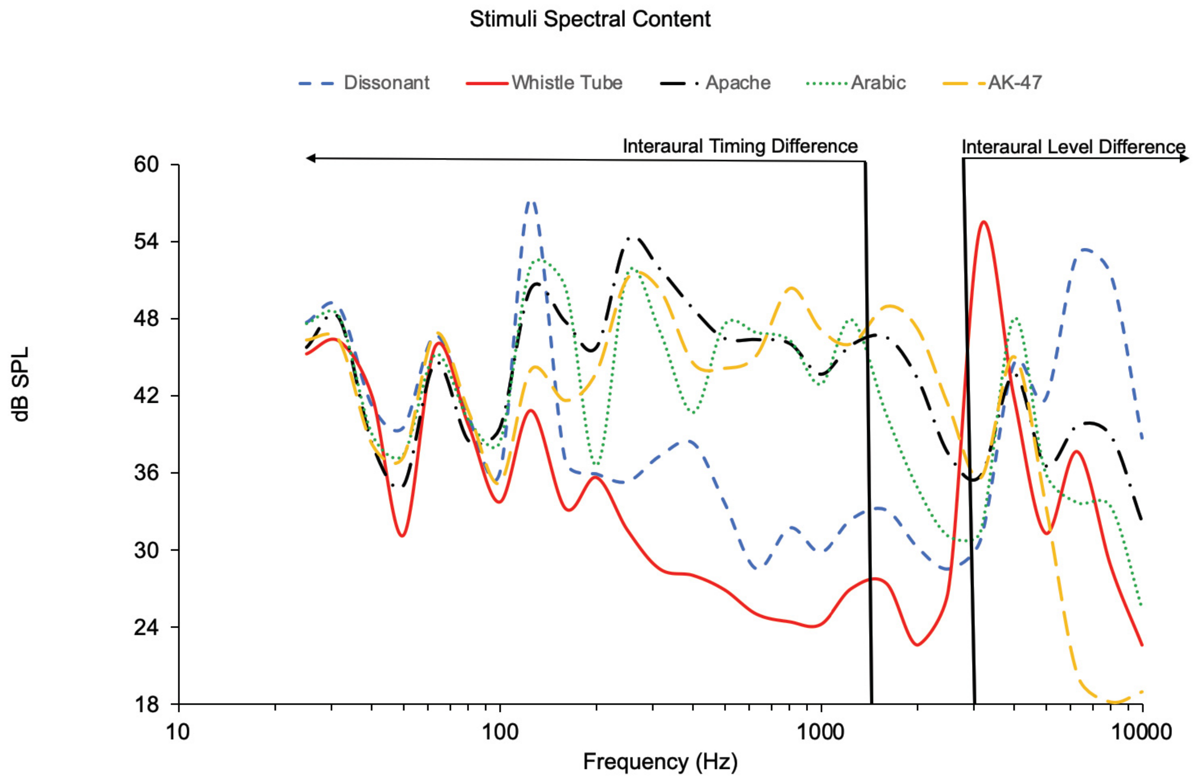


Figure . 1/3-octave band spectral content of the dissonant training signal and untrained stimuli as used in both pilot experiments and main experiment containing 104, 295, 450, 737, 2967, 4959, 7025 and 7880 Hz.

Pilot Experiment 2

In the second pilot, the purpose was to manipulate the number of presentations in the sequential portion of the testing for LUs two through eight. Participants often remarked that the sequential subunit seemed the least beneficial and the most monotonous. Therefore, the investigators evaluated the effects of maintaining four sequential presentations in LU1 only. All experimental procedures were replicated from the first pilot with the exception of the sequential subunit delivery. As determined by the results of the first pilot, four stimulus presentations from each speaker location for the random subunit were used. The objective was to compare the scaled-down sequential condition to the four-presentation condition in the first pilot. Therefore, the between-subjects variable in the second pilot was the number of presentations from each speaker location for the sequential subunit for LU2-LU8. The within-subjects factor consisted of the untrained test stimuli: AK-47, Apache, Arabic, and whistle tube. The dependent measure was consistent with the first pilot where number correct out of 36 for each test was computed. From this measure, the slopes of the simple linear regression lines across LUs were calculated and compared between the four-presentation condition in the first pilot and the results from the second pilot.

Main experiment

The main experiment evaluated the effect of training strategy on localization accuracy across LUs. Therefore, the experiment used a pretest-posttest control group, mixed-factors design. The training strategy served as the between-subjects factor. The LU served as the within-subjects factor. After the familiarization task, a pretest was administered. Then, eight LUs were administered. The test sub-unit in LU8 served as the post-test. The training strategies, or conditions, were: DRILCOM, adaptive, choose, and choose + adaptive. Table 1 provides the number of presentations from each speaker location in each LU. The results from the two pilot experiments determined the type and number of stimuli used in the main experiment.

Pilot experiment data showed that training with the dissonant signal generalized to localization performance using all of the untrained test stimuli except the whistle tube simulator. Additionally, due to the presence of a possible ceiling effect, the main experiment displayed 24 possible response locations versus 12 used in the pilot and Casali & Lee (2019) DRILCOM study. The change in response options resulted in accuracy within 15°. Four sequential presentations were maintained in the LU1. One sequential presentation was incorporated in LUs 2-8. The subunits within each experimental condition are listed in Table 3. The measurement variables were absolute percent correct and response time. An advantage to using response time is that the measure may indirectly monitor workload as localization training progresses. As practice and skill acquisition progressed, and thus listening skills became more automated, reaction time was expected to decrease (Wickens, Hollands, Banbury, & Parasuraman, 2013).

Table . Subunit composition for each training condition in the main experiment. The sequential subunit was comprised of four presentations from each location for LU1 and only one from each location for LUs 2-8.

	Training Condition (number of presentations)			
Subunit order	DRILCOM	Adaptive	Choose	Adaptive + Choose
1	Sequential	Sequential	Sequential	Sequential
2	Random (4, each speaker location)	Random (2, each speaker location)	Random (3, each speaker location)	Random (2, each speaker location)
3	Test (3, each speaker location)	Adaptive (4 for each incorrect response location from previous session)	User-selected practice (18 total presentations from location of participant’s choice)	Adaptive (2 for each incorrect response location from previous session)
4		Test (3)	Test (3)	User-selected practice (12 total presentations from location of participant’s choice)
5				Test (3, each speaker location)

The dissonant signal and pink noise were presented at 70 dBA and 55 dBA, respectively. The follow-on research on the portable localization training system that incorporated the training strategy as a result of this experiment used two active sound-transmission hearing protection devices in the form of U.S. Military Tactical Communications and Protective Systems (TCAPS) as detailed in Casali (2010b). The 70 dBA presentation level was used in order to simulate a typical operational situation with moderate level signals that need to be localized, but which are supra-threshold and easily detected, and also which, depending upon the active protector's design, may trigger the amplifier in the electronic pass-through circuitry without putting the devices into full compression.

PHASE I: EXPERIMENTAL DESIGN

Pilot 1

The experiment was a mixed-factorial, post-test design. The schematic for the experimental design of Pilot 1 is provided in Figure 19. The within-subjects factor of untrained stimuli was comprised of four levels: AK-47, Apache, Arabic, and whistle. The between-subjects factor was comprised of the number of presentations in each training subunit (i.e. sequential and random sub-units): two, three, and four presentations. The dependent measures were localization accuracy and response time. Localization accuracy was measured by percent absolute correct, where the participant reported the exact location of the speaker with 30° precision. The order of the dissonant signal's, or training stimulus, presentation location was randomized. The LabView software that facilitated the delivery of experimental protocol incorporated a pre-programmed list of randomized orders with the constraint of the number of presentations determined by the experimental condition. The stimulus presentation level was 55 dBA amidst 40 dBA of pink noise. The presentation level was chosen levels in the pilot study to maintain consistency with previous DRILCOM protocols, which used two presentation levels of 50 dBA and 85 dBA. The 85 dBA signal was not used as sounds at this level create distortion in the cochlea. The goal was to determine localization abilities at a comfortable listening level versus loud levels. In order to maintain a 15 dB S/N with follow-on studies while masking ambient noise, 55 dB was used in the pilot studies. The untrained stimuli were delivered according to a random order generator used in the Excel function.

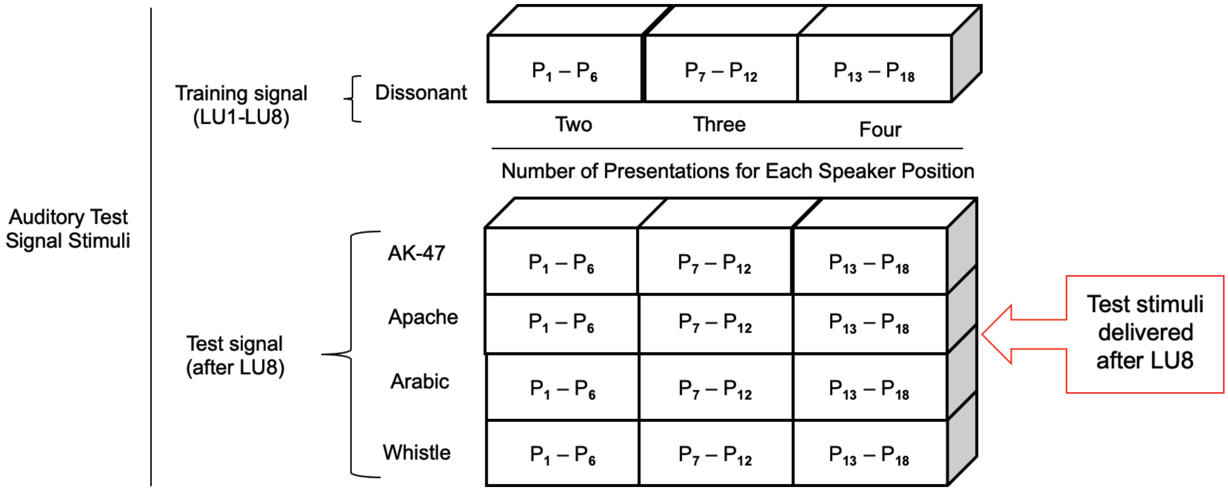


Figure . Experimental design schematic for Pilot 1.

Pilot 2

Just as in Pilot 1, the Pilot 2 study was as a post-test mixed-factorial design. Instead of manipulating the number of presentations in all the training subunits as in Pilot 1, Pilot 2 only manipulated the number of presentations in the sequential subunit. In the random subunit, four presentations were maintained. The schematic is provided in Figure 20. The schematic for Pilot 1 is provided as a reference as results from Pilot 1 were compared to Pilot 2. In other words, Pilot 2 was an ad hoc experiment added after analyzing the results of Pilot 1. Pilot 2 served as an extra condition. Therefore, the schematic of Pilot 2 is provided in the context of Pilot 1. The independent variable was the number of sequential presentations from LU2-LU8. The independent variables were the same as Pilot 1, localization accuracy measured by percent correct out of 36 trials and response time. Randomization of signal presentation was the same as Pilot 1.

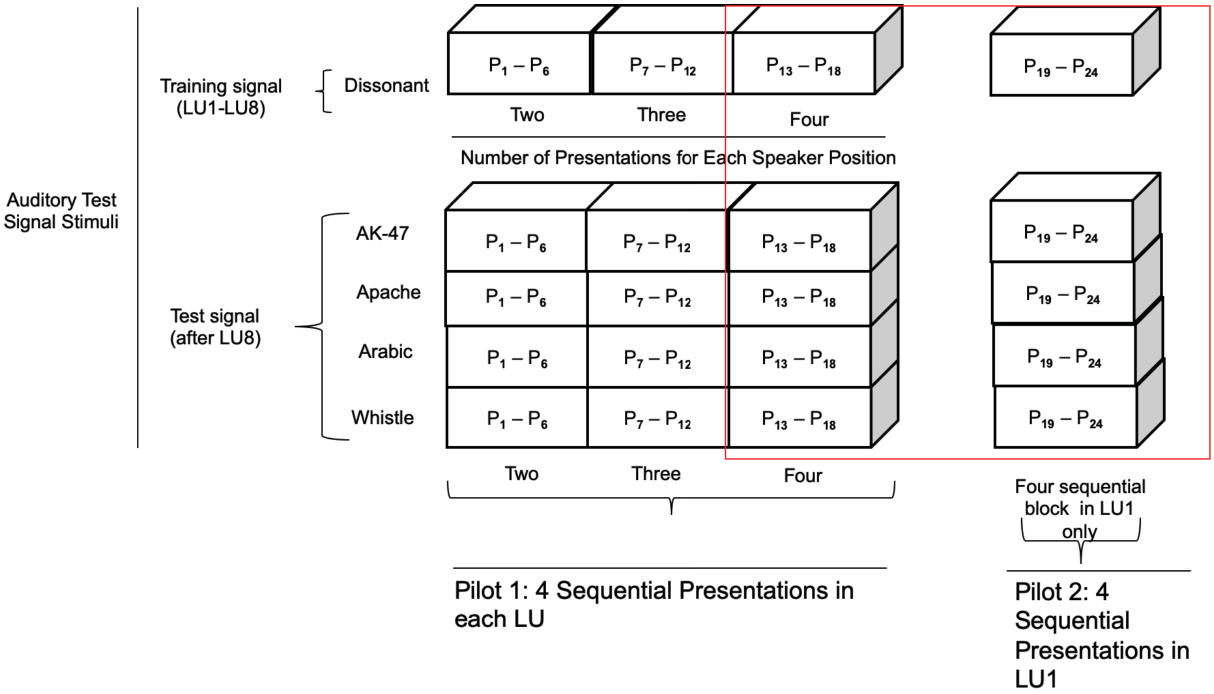
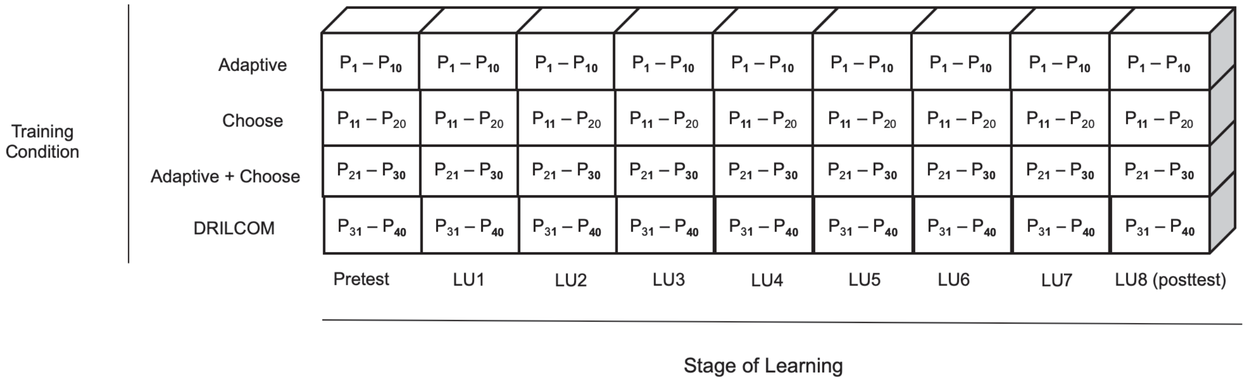


Figure . Experimental design schematic for Pilot 2.

Main experiment

The main experiment was a pretest-posttest control group, mixed factorial design. The between-subjects variable was the training strategy: adaptive, choose, adaptive + choose, and DRILCOM. The within-subjects variable was the stage of training (pretest and LU1-LU8). The dependent variables were absolute correct, ballpark correct and response time.



	Pretest	LU1	LU2	LU3	LU4	LU5	LU6	LU7	LU8 (posttest)
Adaptive	$P_1 – P_{10}$	$P_1 – P_{10}$	$P_1 – P_{10}$	$P_1 – P_{10}$	$P_1 – P_{10}$	$P_1 – P_{10}$	$P_1 – P_{10}$	$P_1 – P_{10}$	$P_1 – P_{10}$
Choose	$P_{11} – P_{20}$	$P_{11} – P_{20}$	$P_{11} – P_{20}$	$P_{11} – P_{20}$	$P_{11} – P_{20}$	$P_{11} – P_{20}$	$P_{11} – P_{20}$	$P_{11} – P_{20}$	$P_{11} – P_{20}$
Adaptive + Choose	$P_{21} – P_{30}$	$P_{21} – P_{30}$	$P_{21} – P_{30}$	$P_{21} – P_{30}$	$P_{21} – P_{30}$	$P_{21} – P_{30}$	$P_{21} – P_{30}$	$P_{21} – P_{30}$	$P_{21} – P_{30}$
DRILCOM	$P_{31} – P_{40}$	$P_{31} – P_{40}$	$P_{31} – P_{40}$	$P_{31} – P_{40}$	$P_{31} – P_{40}$	$P_{31} – P_{40}$	$P_{31} – P_{40}$	$P_{31} – P_{40}$	$P_{31} – P_{40}$

Figure . Experimental design schematic for the main experiment.

Participants

All participants were recruited from the Virginia Tech community in accordance with protocol number 19-176 approved by the university Institutional Review Board for human subjects. In the pilot studies, general trends, versus experimental power, were sought to assist in developing a training protocol. With the exception of the second pilot study, the target gender ratios of participants represented those in the U.S. Military of 75-85% male and 15-25% female (DMDC, 2019). The first pilot study included 14 males and four females (ages 19-42, mean age of 27). The second pilot included one male and five females (ages 18-28 with a mean age of 24). In the main experiment, 31 males and nine females participated (ages 19-38, with a mean age of 24). Participants had no prior experience with electronic or other level-dependent hearing protection, nor with auditory localization training from previous research participation or video games. All participants were otoscopically inspected to ensure their ear canals were generally free from obstructions such as cerumen. Participants were also pre-screened for hearing thresholds not to exceed 25 dB HL from 500-6000 Hz, with no threshold difference between each ear to exceed 15 dB HL.

Apparatus

Pilots 1 and 2 and the main experiment were conducted in the VT-ASL, as shown in Figure 17. In general, this facility incorporated a hemi-anechoic room measuring 19 feet long by 8.5 feet wide, with acoustically reflective floor tiles, acoustically-treated ceiling panels made by Celotex^{TM}, and walls treated with acoustically-treated eggshell walls made by Sonex^{TM} (Casali & Lee, 2015). The DRILCOM testing apparatus, earlier depicted as a photo in Figure 17, is also shown schematically in Figure 22. This schematic illustrates the participant orientation, the experimenter station, general orientation of speaker array, and test equipment rack. The diameter of the 12-speaker array shown in Figure 22 is slightly greater than 3 meters, with each loudspeaker mounted at 1.14 meters from the ground mounted on a steel ring (Casali & Lee, 2015). The speaker model used in DRILCOM are 30-Watt, 6.25 inch Behritone C50A powered speakers with a 5.25 inch round driver with a frequency response of 90 – 17,000 Hz, flat within ± 3 dB (Casali & Lee, 2015). For background noise generation, the system used a QSC CX1102 ^{TM} power amplifier and 4 JBL SoundPower SP215-6™ speakers (Casali & Lee, 2015). The DRILCOM system PC equipped was equipped with Windows 7, LabVIEW™ (2013 version), MatLab™ and used to record localization performance. The computer monitor and mouse in front of the participant recorded the response. An automated 15-position switch, of which 12 were used in this experiment, was routed into the patch bay that was connected to the speakers.

Daily calibration of the pink noise and target stimuli were conducted. Measurements were taken using a with a Larson-Davis Model 2900 spectrum analyzer (SN: A0280) with a ½-inch Larson-Davis 2559 microphone (SN:2575) and a Larson-Davis 9000C Preamp (SN: 0521). The microphone was calibrated at 94.0 dBA (1000 Hz tone) using a Quest QC-20 Calibrator (SN: QOA070051). The microphone was placed in the center of the array (1.5 meters) from each loudspeaker using a plumb bob. The pink noise was played using a CD recorder delivered through an Optimus 1850 CD player and delivered to the background noise speakers.

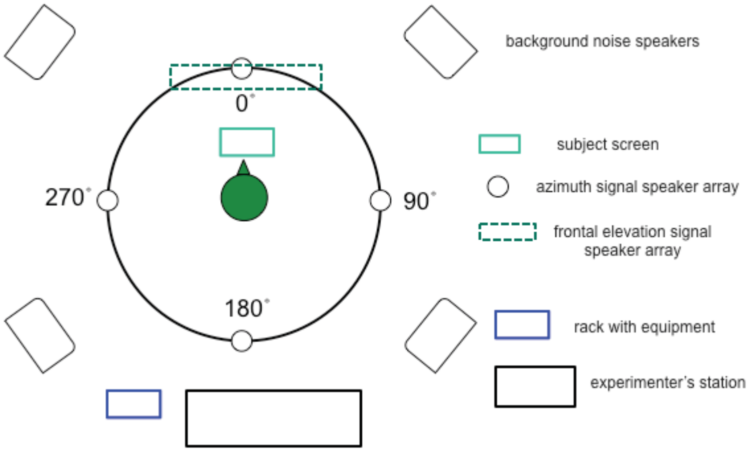


Figure . DRILCOM testing apparatus. Adapted from Casali & Lee (2015), Figure 4, p.27.

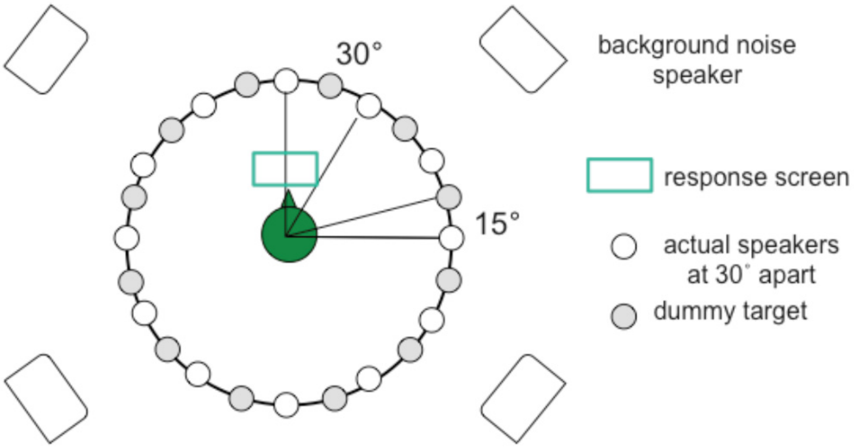


Figure . Azimuth speaker array. Adapted from Casali & Lee (2015), Figure 11, p.27.

The LabVIEW program facilitated user-initiation of the stimulus presentation, triggering recording of the response time, stimulus location, and participant response. Participants were furnished with hard-copy display examples of how to select, respond, and interpret system-generated feedback. The hard copy demonstration was immediately followed by a live demonstration on how to operate the system. The participant then selected the ‘Click to START’ prompt on the display. Upon doing so, LabVIEW signaled to the relay corresponding to a given speaker location to close, and routed the stimulus through the amplifier and relay to the target speaker. Signal presentation simultaneously triggered the response timer. Random number generators programmed into LabVIEW, with the constraint of three presentations from each speaker location, determined the presentation order for each subunit. LabVIEW recorded a response according to the participant’s selection of one of the speaker locations on the display. The participant response served as the timer offset and registered the response to a Comma Separated Values (CSV) file. Stored within the CSV file was the subject number, listening condition, test type (azimuth or elevation), loudspeaker source location, participant response location, response time, date, time, subunit, and LU. The CSV file was saved, downloaded, and scored in Excel after each participant completed the session.

PHASE I: EXPERIMENTAL PROCEDURES

Participant Recruitment

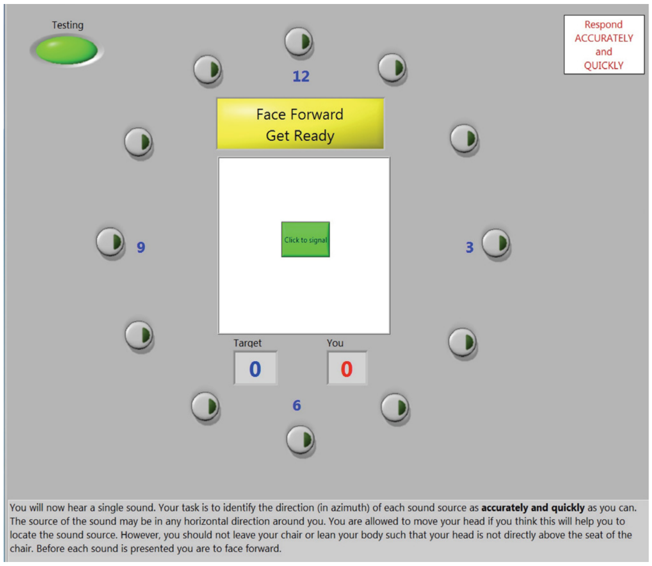
Participants were sought using flyers posted around the floor of the building where the experiment was conducted (Appendix B), the Virginia Tech graduate student listserv, emails, and word of mouth. Potential participants contacted the experiment via email to schedule the experimental session. The informed consent form is provided in Appendix C and provided to all participants at the beginning of the experimental session. The experimenter reviewed the informed consent with the participant at the test station in the center of the speaker array. The experimenter highlighted that only subtest three applied to the experiment and addressed any questions. The experimenter then returned to the experimenter station outside of the speaker array so that the participant could review the informed consent without time pressure prior to signing. Participants were otoscopically examined to ensure the ear canals were generally free from obstruction and pathology and results were recorded on the form provided in Appendix D. Also, on this form participants were screened to ensure they had no prior experience with TCAPS or localization training. All participants were audiometrically evaluated to ensure thresholds did not exceed 25 dB HL from 500-6000 Hz, with bilateral symmetry not to exceed 15 dB HL at each frequency. Participants were evaluated using a Beltone 119 audiometer (serial number: 10B0561) using a sound-treated environment in the Auditory Systems Laboratory with levels that did not exceed levels specified for normal hearing according to the DODI 6055.12. Results were recorded on a threshold screening form in Appendix E and the average thresholds for each experiment and group are provided in Table 4. All thresholds were approximately equivalent among conditions.

Table . Mean pure-tone hearing level thresholds (dBHL) by group study and group.

		Frequency (Hz)					
Main Experiment	Ear	500	1000	2000	3000	4000	6000
Adaptive	Right	4	2	-0.5	3	4.5	5.5
	Range	-5 – 10	-5 – 10	-10 – 10	-5 – 15	-10 – 10	-10 – 25
	Left	3	1.5	-2	3.5	2.5	3
	Range	0 – 15	0 – 10	-5 – 5	-10 – 15	-5 – 15	-10 – 25
Choose	Right	5.5	2	0.5	0	0.5	3
	Range	0 – 10	-10 – 10	-10 – 5	-5 – 10	-10 – 10	-5 – 15
	Left	5.5	0.5	-1.5	1.5	-3	6
	Range	0 – 10	-5 – 10	-10 – 10	-10 – 10	-10 – 15	-10 – 10
DRILCOM	Right	4.5	2.5	-0.5	-1.5	2.5	2
	Range	-5 – 15	-5 – 20	-5 - 5	-5 – 10	-5 – 5	5 – 20
	Left	5.5	3	-0.5	0	-1	5
	Range	0 – 10	-5 – 10	-5 – 10	-10 – 10	-5 – 10	-5 – 10
Adaptive + Choose	Right	6.5	6	1	-0.5	3.5	4.5
	Range	0 – 15	-5 – 15	-5 - 10	0 – 15	-5 – 10	0 – 15
	Left	7	3.5	3.5	5.5	2.5	9.5
	Range	-10 – 15	-5 – 20	-10 – 10	-10 – 15	-5 – 20	-5 – 10
Pilot 1							
Two- Presentation	Right	8	6	2	0	3	6
	Range	0 – 10	0 – 10	0 – 10	-5 – 0	-5 – 10	5 – 20
	Left	6	2	3	-3	0	12
	Range	0 – 20	0 – 15	0 – 10	-5 – 5	-5 – 10	0 – 10
Three- Presentation	Right	11.7	10.8	7.5	10	8.3	10
	Range	0 – 15	5 – 20	-10 – 20	0 – 25	0 – 25	0 – 15
	Left	9.2	12.5	7.5	11.7	13.3	10.8
	Range	5 – 15	5 – 15	0 – 15	5 – 15	0 – 20	0 – 25
Four- Presentation	Right	8.3	7.5	0.8	0.0	4.2	2.5
	Range	0 – 20	0 – 15	-10 – 5	-10 – 10	-5 – 15	-5 – 10
	Left	5.5	3	-0.5	0	-1	5
	Range	5 – 10	-5 – 15	-5 – 10	-5 – 10	0 – 5	-10 – 20
Pilot 2	Right	9.2	5.8	4.2	1.7	3.3	12.5
	Range	5—10	5 – 10	5 – 15	5 – 15	0 – 15	0 – 25
	Left	9.2	6.7	6.7	8.3	7.5	9.2
	Range	5 – 20	0 – 15	0 – 10	-5 – 10	0 – 5	10 – 15

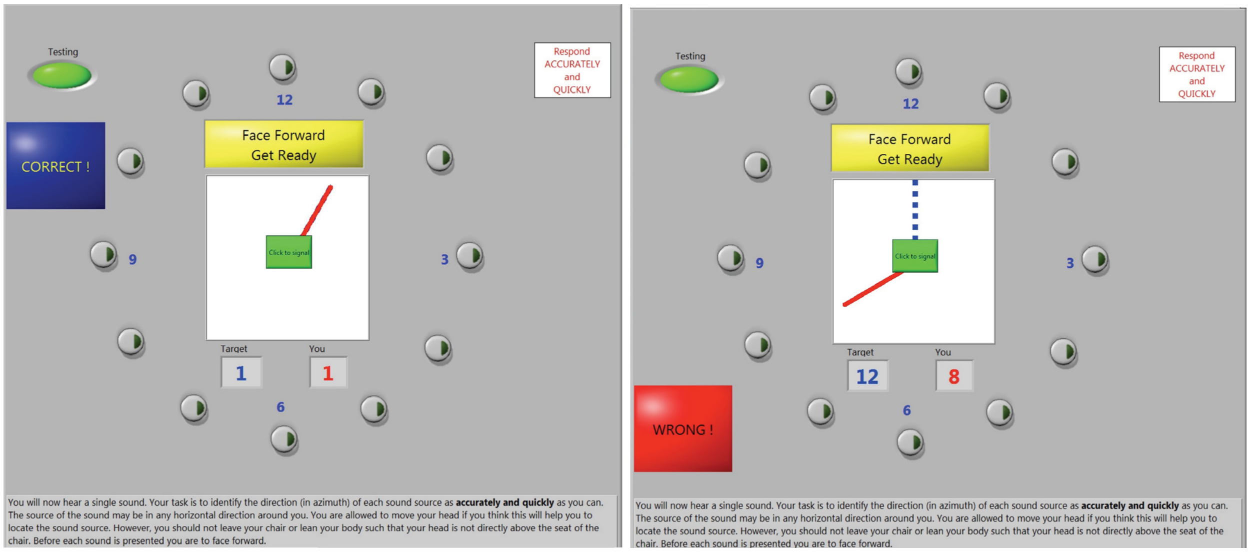
Pilots 1 and 2

Participants were initially provided with a brief orientation of the DRILCOM system. For Pilot 1, a random order generator provided by Microsoft® Excel matched the order by which the participant walked through the door to a subject number, and then random assignment to a presentation condition was stratified for every three participants. After obtaining and completing the informed consent, audiogram, and demographic questionnaire, the experimenter provided a demonstration of how to interact with the user interface, operate the software, and use the mouse. Instructions are provided in Appendix J. An “X” on the ceiling of the lab indicated participant alignment. Participants were positioned accordingly and informed that they could move their heads to better localize sound. The experimenter stood next to the participant during the orientation and demonstration. Familiarization was performed at 12, 3, 6, and 9 o’clock. The sequential display is shown in Figure 24. Feedback displayed for correct and incorrect responses are shown in Figure 25. The participant was instructed to answer the 9 o’clock position incorrectly in order to demonstrate how an incorrect response appeared on the screen. Participants were given 12 possible responses from which to choose. Participants were queried again to see if any questions about the experiment remained. Instructions for completing the sequential training were then delivered (Appendix J). An experimenter checklist, Appendix H, was used to record a hard copy of the scores in the event of hardware or software failure.



You will now hear a single sound. Your task is to identify the direction (in azimuth) of each sound source as **accurately and quickly** as you can. The source of the sound may be in any horizontal direction around you. You are allowed to move your head if you think this will help you to locate the sound source. However, you should not leave your chair or lean your body such that your head is not directly above the seat of the chair. Before each sound is presented you are to face forward.

Figure . Participant response display showing 12 sequential options. Adapted from Lee & Casali (2019), slide 12.



You will now hear a single sound. Your task is to identify the direction (in azimuth) of each sound source as **accurately and quickly** as you can. The source of the sound may be in any horizontal direction around you. You are allowed to move your head if you think this will help you to locate the sound source. However, you should not leave your chair or lean your body such that your head is not directly above the seat of the chair. Before each sound is presented you are to face forward.

You will now hear a single sound. Your task is to identify the direction (in azimuth) of each sound source as **accurately and quickly** as you can. The source of the sound may be in any horizontal direction around you. You are allowed to move your head if you think this will help you to locate the sound source. However, you should not leave your chair or lean your body such that your head is not directly above the seat of the chair. Before each sound is presented you are to face forward.

Figure . Participant response display for showing feedback for correct (left) and incorrect (right) response. Adapted from Lee & Casali (2019), slides 14 and 13, respectively.

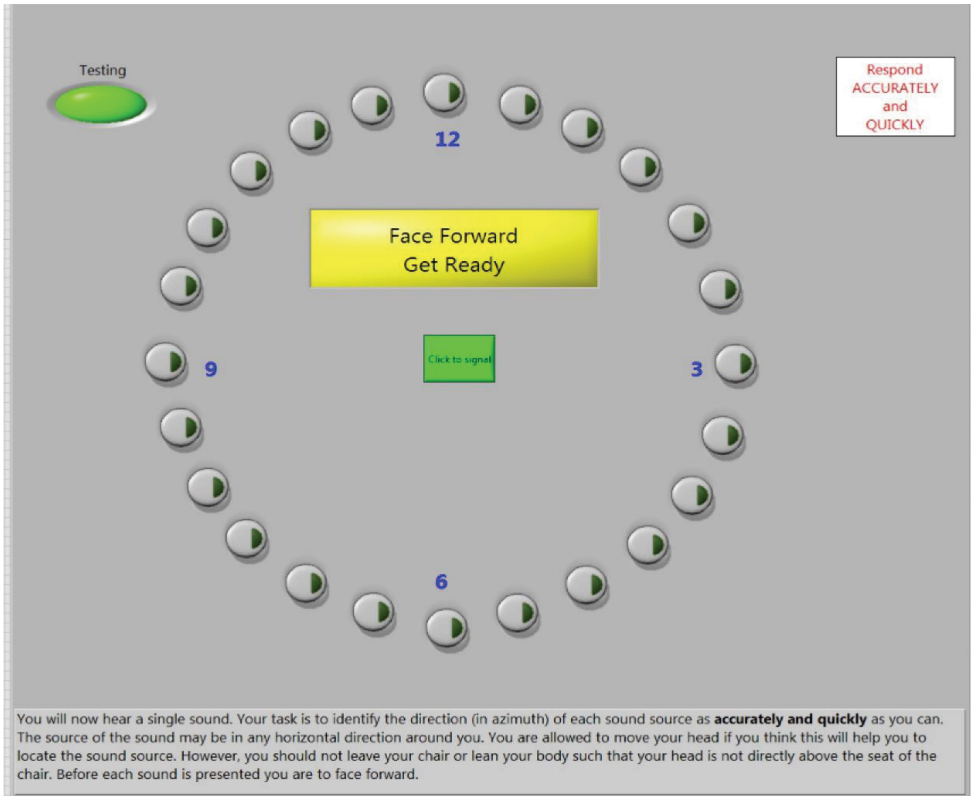
After instructing for each subunit, the experimenter moved outside of the speaker array and remained at the experimenter station until the participant completed the subunit. The participant was then instructed to initiate the sequential training by selecting the "Click to signal" icon in the center of the display, which initiated presentation of the dissonant signal in a clockwise fashion. The number of sequential presentations varied according to condition. For example, if the participant was assigned to the two-presentation condition, the signal started at 12 o'clock and presented sequentially in clockwise order for a "lap" around the "clock face", with the last presentation at 11 o'clock. Then, the order of presentation re-starts at 9 o'clock and proceed in a counterclockwise fashion. Procedures for Pilot 2 were exactly the same as the four-presentation condition for Pilot 1 with the exception of the sequential subunit exclusion from LU2-LU8.

After completion of the sequential training, the participant re-entered the array to instruct the participant on how random training is conducted. Instructions are provided in Appendix J. The participants were informed of how many presentations from each speaker they would hear according to their assigned condition. Participants were informed that they would hear the same sound, but unlike the sequential presentation, they would not know ahead of time from which location they would hear the sound. The display would then provide feedback regarding the correctness of response, just as in the sequential subunit. The participant initiated the random subunit by selecting "Click to signal." Upon completion of the random subunit, the participant proceeded to the test subunit. Participants were informed that they would hear 36 presentations of the dissonant signal. All participants were apprised of their scores after each test using the absolute and ballpark criteria. If the participant had not requested a break by the end of LU4, the participant was informed that a break would be provided. Breaks generally lasted no more than 20 minutes.

At the conclusion of LU8, participants were informed that they were about to undergo four more tests. These tests were conducted in the exact manner as the previous tests using the dissonant signal, only the tests used military signals not previously heard. The participants were informed prior to the test which signal they would hear.

Main Experiment

Participants were oriented and familiarized to experimental procedures similar to Pilots 1 and 2. However, upon examination of the data, a possible ceiling effect was observed and 24 speaker icons appeared on the participant's display for the main experiment in contrast to the 12 locations in the pilot studies. The display is provided in Figure 26. Additionally, the four tests of the untrained, military-relevant stimuli from the pilot studies were eliminated. The experimenter checklist for the main experiment is provided in Appendix H.



You will now hear a single sound. Your task is to identify the direction (in azimuth) of each sound source as **accurately and quickly** as you can. The source of the sound may be in any horizontal direction around you. You are allowed to move your head if you think this will help you to locate the sound source. However, you should not leave your chair or lean your body such that your head is not directly above the seat of the chair. Before each sound is presented you are to face forward.

Figure . Participant response display showing 24 options with only 12 active speakers. Adapted from Lee & Casali (2019), slide 15.

Four sequential presentations were maintained in the LU1 and one sequential presentation with a randomly-assigned starting position and direction (i.e. clockwise and counterclockwise) were presented for LU2-LU8. Unlike the pilot tests, a pretest was administered after familiarizing the participant. Then, eight LUs were administered. At the conclusion of LU1, a hard-copy questionnaire was administered (Appendix I). The test sub-unit in LU8 served as the post-test. Just as in the pilot test, three presentations from each speaker location comprised the test, for a total of 36 presentations. Participants were informed of their scores, both ballpark and absolute, at the conclusion of the test. The training strategies, or conditions, were: DRILCOM, adaptive, choose, and choose + adaptive. The experimenter script for each condition is provided in Appendix J.

After the posttest, the questionnaire was administered. The questionnaire differed for each condition by adding a question asking the participant to rate the perceived benefit of the training subunit(s) unique to the condition. The last item on the questionnaire also differed by condition. The item asked the participant to identify the subunit(s) where he/she noticed disengagement or disinterest. This item differed according to the composition of the LU.

In the DRILCOM condition, participants underwent the sequential training, with four presentations in LU1 and only one presentation from LU2-8. Next, the random subunit with four presentations from each location was completed. Then, a test was administered in the final subunit with three randomly-presented stimuli from each location. The experiment lasted approximately two hours and fifteen minutes.

In the adaptive condition, participants underwent sequential training as in the DRILCOM condition. Then, the random subunit was administered with two presentations from each location. Based on the incorrect responses in the random subunit, LabVIEW delivered four presentations, in a row, from the speaker location that was incorrectly answered. Therefore, the number of presentations varied according to participant performance. Participants then completed the test subunit. The testing lasted approximately two and a half hours.

For participants assigned to the choose condition, participants completed the sequential and random subunits as above. The random subunit consisted of with three, randomly-presented presentations from each speaker location. Next, the user-select subunit followed with the participant selecting the icon from which he or she wished to obtain more training, for a total of 18 presentations. Then, participants completed the test subunit. The experiment lasted approximately two and half hours.

In the adaptive + choose condition, sequential training was administered followed by the random subunit (two presentations from each location), an adaptive subunit, and the user-select subunit (12 presentations). Then, the test subunit was administered. The experiment lasted approximately three hours for each participant.

PHASE I: RESULTS

Raw data was downloaded converted from a CSV file to a Microsoft® Excel file for data reduction. Qualitative data were entered into Excel files manually. Excel files were uploaded into statistical analysis software for analysis using JMP® 14 software, IBM® SPSS® Statistics, and Excel v16.16.10.

Data Reduction Pilots 1 and 2

In Pilots 1 and 2, the mean absolute scores for each subunit test, for a total of eight LUs, were computed within each presentation condition (two, three, and four. Mean absolute scores were also computed for each test using the untrained military-relevant signals (AK-47, Apache, Arabic, and whistle). Means are shown in Figure 27. Within each presentation condition (two, three, and four), transfer-of training to untrained stimuli was evaluated by comparing the mean absolute performance at LU8 to mean test performance for the untrained stimuli. For Pilot 1, regression lines were fitted to determine the slopes from LU1 to LU8 for each presentation condition. All parametric analyses used an $\alpha = 0.05$. Regression lines are computed to predict dependent measures when an independent variable is known (Portney & Watkins, 2009). The null hypothesis for regression analyses is that the regression line slope equals zero (Portney & Watkins, 2009). Slope computation enabled assessment of the change in performance from LU1 to LU8 tests within each presentation condition. Then, regression line slopes for each presentation condition were compared using independent *t*-tests. A Bonferroni adjustment was applied to control for Type I error rate of multiple comparisons ($\alpha = \frac{0.05}{3} = 0.0167$). Slope comparisons evaluated significant differences in the change in performance among presentation conditions. An additional independent *t*-test was performed to determine if the slope of the four-presentation condition of Pilot 1 differed significantly from Pilot 2 (no sequential subunits from LU2-LU8).

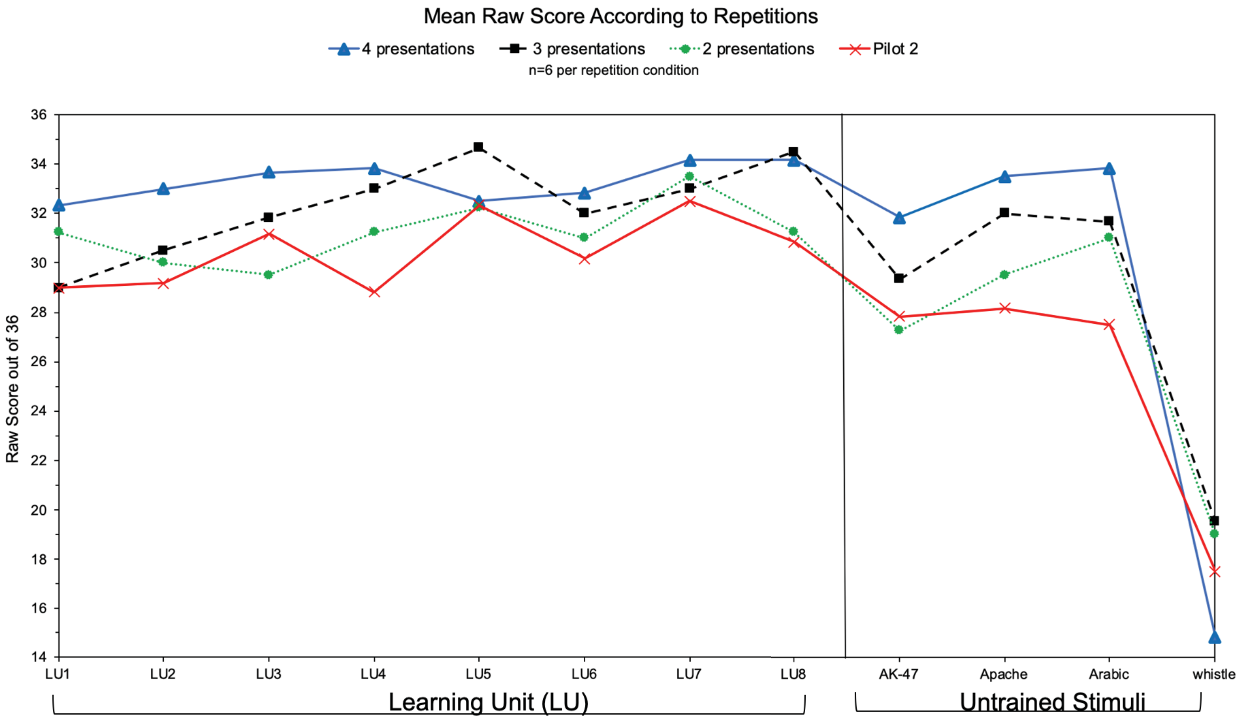


Figure . Mean score for each presentation condition at each training stage for Pilots 1 and 2.

For Pilot 1, multiple one-way analysis of variances (ANOVAs) for each untrained stimulus (AK-47, Apache, Arabic, and whistle) were conducted in order to evaluate the effect of number of presentations (two, three, and four) on the dependent variable of absolute score. In general, ANOVAs assess significant mean differences according to the level of independent variable on a dependent variable. The number of presentations in each training subunit served as the independent variable. Of note, one participant did not complete testing using the Arabic stimulus in the two-presentation condition due to her scheduling constraints. Within each presentation condition, performance was also compared using one-way, within-subjects ANOVAs among the untrained stimuli. The untrained stimuli served as the independent variable and the absolute test score for each untrained stimulus served as the dependent variable. Given that the analysis included a within-subjects factor, Mauchly's Test of Sphericity was included to evaluate the assumption of homogeneity of variance between related of group comparisons

(Portney & Watkins, 2009). When violations of the homogeneity of variance assumptions occur, the likelihood of making a Type I error can be underestimated (Portney & Watkins, 2009). If Mauchly's Test of Sphericity detects a violation, the test recommends an adjusted *p* value in the ANOVA to decrease the likelihood of making a Type I error (Portney & Watkins, 2009). A significant *p* value in the Mauchly's Test is consistent with a violation and the degrees of freedom are adjusted in the ANOVA (Portney & Watkins, 2009). In Pilot 2, where LU2-LU8 did not include sequential training, a one-way within-subjects ANOVA evaluated differences in absolute score among untrained stimuli (AK-47, Apache, Arabic, and whistle). The untrained stimuli served as the independent variable and the absolute score served as the dependent variable.

Results for Pilots 1 and 2

The mean absolute performance for each presentation condition across LUs and untrained stimuli tests, displayed in Figure 27, revealed several trends. As previously stated, in order to support the transfer-of-training from trained to untrained stimuli, the degradation in absolute percent correct was not to exceed 10% from LU8 to score obtained for the untrained stimulus. The four-presentation condition resulted in the highest mean score for the untrained stimuli at LU8 (*M*=95.8%, *SD*=5.1), with the AK-47 (*M*=88.4%, *SD*=4.8), Apache (*M*=93.1%, *SD*=4.3), and Arabic (*M*=94%, *SD*=2.6) conditions meeting the criterion for transfer-of-training. The whistle stimulus condition (*M*=41.2%, *SD*=5.3) did not meet the "trainable" criterion in the four-presentation or any other of the presentation conditions. The plotted means in Figure 27 also reveal that the two-presentation condition (at LU8 *M*=88.9%, *SD*=4.2) met the 10% transfer-of-training criterion, but the three-presentation condition did not when comparing LU8 (*M*=95.8%, *SD*=5.2) to AK-47 performance (*M*=81.5%, *SD*=11.6).

To evaluate significant differences in slopes among presentation conditions, regression lines were calculated for each presentation condition. All results were interpreted at α=0.05 using a Bonferroni adjustment for four comparisons (α=0.05/4=0.0125). The fitted regression lines for each presentation condition were as follows:

2 presentation condition absolute score=30.36 + 0.28(LU)

3 presentation condition absolute score=29.46 + 0.63(LU)

4 presentation condition absolute score=32.52 + 0.18(LU)

Slope comparison using independent *t*-tests, provided in Table 5, showed no significant difference among slopes (two versus three, *t*[94]=1.13, *p*=0.261; two versus four, *t*[94]=0.34, *p*=0.732; three versus four, *t*[94]=1.40, *p*=0.164). In order to compare the slopes of Pilot 1 and Pilot 2, the following regression line was fitted for the Pilot 2 data:

Pilot 2 absolute score=28.89 + 0.36(LU)

The independent *t*-test showed no significant difference in the slopes for Pilot 2, using α=0.0125, where no sequential training occurred than from LU2-LU8 than for the four-presentation condition (*t*[94]=0.57, *p*=0.573). The absolute performance at LU8 for Pilot 2 (*M*=31.4, *SD*=2.9) was less than the four-presentation condition (*M*=34.2, *SD*=1.8). The absolute score was considerably less in Pilot 2 than the Pilot 1 four-presentation condition.

Table . Slope comparisons for each presentation condition and Pilot 2, evaluating differences in the change in absolute score over LUs between conditions.

Source	*n*	*M*	*SD*	*t*	*p*
Two versus Three-presentation				1.13	0.261
Two-presentation	48	0.28	1.42		
Three-presentation	48	0.64	1.61		
Two versus Four-presentation				0.34	0.732
Two-presentation	48	0.28	1.42		
Four-presentation	48	0.17	1.57		
Three versus Four-presentation				1.40	0.164
Three-presentation	48	0.64	1.61		
Four-presentation	48	0.17	1.57		
Four-presentation versus Pilot 2				0.57	0.573
Four-presentation	48	0.17	1.57		
Pilot 2	48	0.36	1.55		

One-way ANOVAs for each untrained stimulus were conducted in order to evaluate the effect of number of presentations on localization accuracy (Tables 6-9). Analyses applied an $\alpha=\frac{0.05}{4}=0.0125$. No significant differences existed among presentation conditions (AK-47, *F*[2,15]=0.78, *p*=0.475; Apache, *F*[2,15]=0.83, *p*=0.457; Arabic, *F*[2,14]=0.88, *p*=0.433; whistle tube, *F*[2,15]=1.42, *p*=0.272). One-way repeated measures ANOVAs also examined the differences in performance among the untrained stimuli for each replication condition. The single subject's missing data from the Arabic stimulus was analyzed as a listwise deletion. Mauchly's test of sphericity supported that no violations occurred in the two (χ2[5]=3.10, *p*=0.702), three (χ2[5]=2.71, *p*=0.752) or four-presentation conditions (χ2[5]=9.50, *p*=0.101). Results for the two-presentation condition, provided in Table 10, showed a significant difference in absolute score according to untrained stimulus (*F*[3, 15]=20.12, *p*<0.000). Follow-up pairwise comparisons with a Bonferroni *t*-test, using a Bonferroni adjustment (Table 11), showed significant differences in performance for the Apache and Arabic stimuli compared to the whistle stimulus. For the three-presentation condition, results provided in Table 12, showed a significant difference in absolute score according to untrained stimulus ($F[3, 15]=17.75$, $p<0.000$). Follow-up pairwise comparisons using a Bonferroni adjustment (Table 13) also showed significant differences in performance for the Apache and Arabic stimuli compared to the whistle stimulus. The four-presentation condition showed a significant difference in absolute score according to the untrained stimulus ($F[3, 15]=50.60$, $p<0.000$), and results are provided in Table 14. Follow-up pairwise comparisons using a Bonferroni adjustment (Table 15) also showed significant differences in performance for the AK-47, Apache, and Arabic stimuli compared to the whistle stimulus.

Table . One-way ANOVA for the AK-47 untrained stimulus across presentation conditions.

Source	*SS*	*df*	*MS*	*F*	*p*
Presentation condition (two, three, four)	36.11	2	18.06	0.78	0.475
Error	345.67	15	23.04		
Total	381.78	17			

Table . One-way ANOVA for the Apache untrained stimulus across presentation condition.

Source	*SS*	*df*	*MS*	*F*	*p*
Presentation condition	27	2	13.5	0.83	0.457
Error	245	15	16.3		
Total	272	17			

Table . One-way ANOVA for the Arabic untrained stimulus across presentation condition.

Source	*SS*	*df*	*MS*	*F*	*p*
Presentation condition	24.89	2	12.45	0.88	0.433
Error	196.17	14	14.01		
Total	221.06	16			

Table . One-way ANOVA for the whistle untrained stimulus across presentation condition.

Source	*SS*	*df*	*MS*	*F*	*p*
Presentation condition	72.44	2	36.22	1.42	0.272
Error	381.83	15	25.46		
Total	454.28	17			

Table . One-way, within-subjects ANOVA for the two-presentation condition comparing absolute performance across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Source	*SS*	*df*	*MS*	*F*	*p*
Untrained Stimulus	**542.60**	**3**	**180.87**	**20.12**	**<0.000**
Subjects	211.30	4	52.83		
Error	107.90	12	8.99		
Total	861.80	19			

Table . Pairwise comparison by Bonferroni *t*-test for the two-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Untrained Stimulus		*M difference*	*SE*	*p*
AK-47	Apache	-2.40	1.57	1.000
AK-47	Arabic	-1.40	1.08	1.000
AK-47	Whistle	10.60	2.58	0.089
Apache	**Whistle**	13.00	2.30	0.029
Arabic	**Whistle**	12.00	2.03	0.024
Arabic	Apache	-1.00	1.38	1.000

**Note.* SE = standard error

Table . One-way, within-subjects ANOVA for the three-presentation condition comparing absolute performance across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Source	*SS*	*df*	*MS*	*F*	*p*
Untrained Stimulus	**620.46**	**3**	**206.82**	**17.75**	**<0.000**
Subjects	181.38	5	36.28		
Error	174.79	15	11.65		
Total	976.63	23			

Table . Pairwise comparison by Bonferroni *t*-test for the two-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Untrained Stimulus		*M difference*	*SE*	*p*
AK-47	Apache	-2.67	1.99	1.000
AK-47	Arabic	-2.33	1.59	1.000
AK-47	Whistle	9.83	2.65	0.083
Apache	**Whistle**	**12.50**	**1.98**	**0.009**
Arabic	**Whistle**	**12.17**	**2.09**	**0.013**
Arabic	Apache	-0.33	1.23	1.000

Table . One-way, within-subjects ANOVA for the four-presentation condition comparing absolute score across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Source	*SS*	*df*	*MS*	*F*	*p*
Untrained Stimulus	**1508.00**	**3**	**502.67**	**50.60**	**<0.000**
Subjects	233.00	5	46.60		
Error	149.00	15	9.93		
Total	1890.0	23			

Table . Pairwise comparison by Bonferroni *t*-test for the four-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Untrained Stimulus		*M difference*	*SE*	*p*
AK-47	Apache	-1.67	0.76	0.479
AK-47	Arabic	-2.00	1.39	1.000
AK-47	**Whistle**	**17.00**	**2.66**	**0.008**
Apache	**Whistle**	**18.67**	**2.39**	**0.003**
Arabic	**Whistle**	**19.00**	**1.93**	**0.001**
Arabic	Apache	0.33	0.92	1.00

In Pilot 2, to test the assumption of sphericity for the within-subjects ANOVA that evaluated differences in performance among untrained stimuli, Mauchly's Test showed no violations (χ2[5]=4.34, *p*=0.514). ANOVA testing, using α=0.05 revealed a significant effect for untrained stimulus (*F*[3, 15]=11.89, *p*<0.000). Results are provided in Table 16. Follow-up pairwise comparisons using a Bonferroni correction showed that absolute score was significantly worse for the whistle stimulus compared to the AK-47 and Arabic stimuli (Table 17).

Table . One-way, within-subjects ANOVA for Pilot 2 (no sequential training in LU2-LU8) comparing absolute score across untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Source	*SS*	*df*	*MS*	*F*	*p*
Untrained Stimulus	**481.83**	**3**	**160.61**	**11.89**	**<0.000**
Subjects	620.00	5	124.00		
Error	202.67	15	13.51		
Total	1890.0	23			

Table . Pairwise comparisons by Bonferroni *t*-test for the four-presentation condition for the untrained stimuli (bolded text indicates a significant test result at the α=0.05 significance level).

Untrained Stimulus		*M difference*	*SE*	*p*
AK-47	Apache	-0.33	1.91	1.000
AK-47	Arabic	0.33	1.12	1.000
AK-47	**Whistle**	**10.33**	**2.38**	**0.044**
Apache	Whistle	10.67	3.04	0.103
Arabic	**Whistle**	**10.00**	**2.07**	**0.028**
Arabic	Apache	-0.67	1.73	1.00

Pilot 1 and 2 Discussion

As expected, training with a signal that contains both ITD and ILD cues transferred to the untrained stimuli that also contained spectral components with these same cues. However, localization of the whistle tube simulator that contained primarily high-frequency tonal information (and little or no ITD low-frequency cues) did not benefit from the training (i.e., little or no transfer-of-training occurred). The findings are consistent with the work of Mills (1958) that showed that the minimum audible angle for signals at 3000 Hz can be as much as 7° poorer than for stimuli below 1500 Hz depending on the angle. In other words, high-frequency stimuli are more difficult to localize than low-frequency signals. Participants also described the whistle tube simulator as annoying. This perception may have increased workload and degraded motivation to respond accurately. While the results evidence that training with a broadband signal does not transfer to a high-frequency tonal one, the results do not necessarily imply that training to localize an incoming rocket-propelled grenade is infeasible. Employing an actual recording of an incoming RPG may show different results given the spectral content. In other words, different exemplars of incoming indirect fire should be evaluated as test stimuli to more accurately assess the presence of trainging transfer. Recordings of an actual incoming round, versus the simulator sound signature used in this study, may contain other spectral information that may render the signal easier to localize. For example, an actual incoming round may incorporate elevation or doppler cues, providing more localizable content. Of all the sounds identified by the SME panel, the sound of incoming indirect fire represented one of the most immediate threats to Soldiers. As such, evaluation of other exemplars of incoming rounds, localizability, and trainability deserve further investigation in future research.

The hypothesis that the slopes among presentation conditions (i.e., number of replications: two, three or four) would differ signfiicantly was unsupported. However, the data yielded sufficient information from which decisions could be rendered for the follow-on experiment. The slope comparisons among the presentation conditions revealed that the three-presentation condition resulted in the greatest slope. Unexpectedly, the three-presentation condition was the only condition that did not result in training transfer for the AK-47 condition. Unlike the Arabic and Apache conditions, the signal in the AK-47 was more intermittent. In other words, the other stimuli provided cues more continuously over time than the three-round burst that incorporated discernable pauses between each discrete gunshot. Mean results at LU1 combined with slope data show that those in the three-presentation condition initially had poorer localization ability than the other conditions.

The hypothesis that significant differences would exist in untrained stimuli performance would exist among presentation conditions (e.g. comparing performance using the AK-47 stimuli across the two-, three-, and four-presentation condition) also remained unsupported. While one-way ANOVAs demonstrated no significant difference among the presentation conditions for the untrained stimuli, the four-presentation condition revealed a clear benefit in localization accuracy for the untrained stimuli. Therefore, obvious advantages for a presentation condition were not not revealed through slope comparisons for trained stimuli or ANOVAs for each untrained stimuli. However, slopes and mean data for each test subunit suggest that the four-presentation condition better-prepared participants, and required the fewest learning units, for the tests using the military-relevant, untrained stimuli. The second pilot reinforced the findings of the first pilot by demonstrating a clear advantage of using more presentations in each LU, resulting in greater localization accuracy across LUs and with untrained stimuli. Thus, the conclusion is that the four-presentation condition yielded the best transfer-of-training to novel stimuli and should be used for follow-on experiments.

One limitation of the first pilot study was that participants were not tested with the untrained stimuli after LU4, and thus, the actual transfer after four LUs cannot be extrapolated from these data. Additionally, fatigue cannot be ruled out as a performance-limiting factor. Participants were given a break at least half-way through the study, and the effects of fatigue on performance using untrained stimuli after eight LUs was not assessed. Therefore, a tradeoff occurred when choosing the LU components for the main experiment. For the follow-on experiment, the investigator maintained four sequential presentations in the first LU, followed by one sequential presentation in the subsequent LUs. The aim of this training composition was to shorten the training duration while maintaining the benefit of sequential training. The follow-on experiment also incorporated four presentations from each loudspeaker location in random subunit for the DRILCOM condition. The component subunits of the other training strategies were designed using the DRILCOM condition as a baseline. For example, the random subunits in the choose condition were reduced to two to allow for four presentations of in the adaptive subunit in the adaptive strategy condition.

Pilot 1 and 2 Conclusions

The purpose of the pilot studies was to evaluate the effects of reducing training time. A follow-on experiment sought to evaluate an optimized training strategy, but in an efficient manner. The pilot studies supported that training with the broadband dissonant signal transfers positively to untrained military relevant stimuli, with the exception of the whistle tube simulator stimulus. Therefore, participants did not need to undergo training with a particular stimulus to glean the benefits of localization training. Moreover, the three-presentation and four-presentation conditions resulted in similar mean performance over the course of training. However, the four-presentation showed a considerable, albeit not statistically-significant, advantage when evaluating the performance of untrained stimuli. Removing the sequential presentations from LU2-LU8 did not significantly degrade performance from the four-presentation condition. However, just as in Pilot 1, the four-presentation condition resulted in higher absolute scores for the untrained stimuli and for each LU.

Main Experiment Data Reduction for Objective Performance

For the main experiment, a mixed-factor ANOVA analyzed the effect of training strategy and stage of training (pretest and LU8) on absolute score. An α level of 0.05 was adopted for all parametric tests. The training strategy served as the between-subjects factor and the stage of training served as the within-subjects factor. As justification here, a mixed-factors design enables the investigator to simultaneously evaluate the effects of an experimental treatment across all participants while examining the effects of another randomly-assigned variable of independent groups (Portney & Watkins, 2009). Analysis included three separate mixed-factors ANOVAs in order to evaluate the effects of three different time constants as the within-subjects factor. The varying time points that served as within-subjects factors and were as follows: pretest and LU8, pretest and LU4, and LU4 and LU8. Evaluating performance at different stages of training determined if performance differed significantly when training was shortened from eight LUs to four. This ad hoc analysis was included due to observation of the mean data in Figure 28 that suggested training beyond LU4 did not improve accuracy considerably. If no significant difference existed in the magnitude of improvement, then sufficient justification existed to shorten the training protocol. Analyses were also conducted using the same independent variables previously mentioned, but ballpark score served as a dependent measure to determine if a change in performance occurred over stages of training using a less stringent performance criteria.

Significant findings in the ANOVA analyses were followed up with pairwise comparisons using the Tukey's honestly significant difference (HSD). For background, the Tukey procedure applies a stringent error rate to all comparison in order to improve the probability of not making a type I error (Portney & Watkins, 2009). The Tukey procedure applies a studentized range statistic, known as q. The q test statistic represents the critical value given a desired α level, error degrees of freedom, and number of means against (Portney & Watkins, 2009). According to Portney & Watkins (2009), the minimum significant difference (MSD) is computed using the following formula:

$$q\sqrt{\frac{mean\ square\ error}{number\ of\ subjects}}$$

Then, the MSD value is compared to each value obtained in the pairwise comparison and if that value exceeds the MSD, then results are significant (Portney & Watkins, 2009).

Effect size was measured using partial eta squared (η^2_p). This measure represents the ratio of total variance in the sample that can be explained by the factor (Pituch & Stevens, 2016). Some of the effect sizes exceeded an η^2_p criterion for medium or large effect sizes of 0.06 and 0.14, respectively (Cohen, 1988). In regards to power, given that complement of the power value is the likelihood of making a Type II error, a power value of 0.8 is generally considered acceptable (Cohen, 2013).

In evaluating the effects of training strategy and training stage on response time, a similar approach of varying the time constant as a within-subjects factor was adopted. Therefore, training strategy served as the between-subjects factor. Mixed-factor ANOVAs were conducted to evaluate the effects of stage of training and training method on response time. The effects of skill acquisition, skill mastery, and lack of vigilance can have disparate effects on response time. Therefore, the objective was to describe response time and infer underlying processes, as opposed to predicting the outcome of this dependent measure.

The same time constants were applied as previously described to another set of analyses that used the difference score as the dependent variable. In doing so, the within-subjects component of the analysis was removed, but the effect of training progression was incorporated into the dependent measure. Training strategy served as the between-subjects variable and one-way, between-subjects ANOVAs were performed. Specifically, the earliest training stage score (i.e. the score at pretest or LU4) was subtracted from latest training stage (e.g. LU8 or LU4). Thus, a positive difference score reflected an improvement in absolute score over time.

Similarly, analyses were conducted using the same within-subjects variables using the response time difference score as the dependent measure. Different from the aforementioned accuracy difference score calculations, the response time differential subtracted the time at the later training stage from the earlier training stage. Therefore, the lower the difference score, the faster the response time at LU8.

In another set of analyses, multiple one-way ANOVAs evaluated the effect of strategy on the change in absolute score over time. Separate ANOVAs were conducted for different time constants (pretest to LU8, pretest to LU4, and LU4 to LU8) using the change in absolute score that occurred with training progression, or slope, as the dependent variables. Training strategy served as the independent variable. This set of analyses enabled evaluation of differences among training strategies on the change in performance with training progression.

Data were subjected to analyses to examine possible predictive relationships between the changes in absolute score and changes in response time. As the predictor variable, a difference score in adjacent LUs were calculated (later LU – earlier LU). A positive value indicated an improvement in score with training progression. The response variable was the difference in response time in adjacent units. A negative value indicated a faster response time with training progression. Regression functions were fitted to the aforementioned variables to determine if the relationship between training progression and response time differed from a slope of zero, or no relationship. Separate analyses were conducted using the following time periods: pretest to LU8, pretest to LU4, and LU4 to LU8.

Separate regression analyses were conducted for each training strategy. Specifically, analyses evaluated if between adjacent LUs, the change in absolute score between adjacent LUs predicted a change in response time. The analyses were conducted using the same time constants as above (pretest to LU8, pretest to LU4, and LU4 to LU8. The analyses totaled 12 regression functions (four training strategies for three different stages of training).

Radial plots enable observation of azimuthal localization accuracy by speaker location. The mean percent accuracy by loudspeaker locations across training strategies were plotted at LU4. Plotting this data enabled observation of trends of which stimulus were more difficult to localize than others. To evaluate participants' accuracy in choosing to practice the locations with which they had the most difficulty, the percent accuracy at each location was plotted according to the frequency count of each speaker's selection in the user-select subunit.

Correlation analysis was conducted to evaluate the relationship between accuracy in each LU and the percentage of time the loudspeaker location was chosen within that LU. Then, the same analysis was conducted at each speaker location to determine the strength of the relationship between percent accuracy and percent of the time the speaker was chosen. Sample distributions were not normal (as will be detailed later), and therefore, nonparametric correlation was used. The Spearman rank correlation coefficient used in these analyses is the non-parametric equivalent of the Pearson product-moment correlation coefficient. Lending background to the rationale, a brief explanation of the statistical approach used is in order. In general, correlation describes the relationship between two variables. Specifically, if the two analyzed variables are related, they covary, or share a portion of their variance (Portney & Watkins, 2009). In essence, the correlation coefficient reflects the extent of consistency between the two distributions (Portney & Watkins, 2009). Since the Spearman's rank correlation coefficient (r_s) does not assume underlying normal distributions, the correlation is based upon rank of ordinal data (Portney & Watkins, 2009). Both observations within each variable set are first ranked, with tie rankings assigned an average of the tied ranks. In essence, the scores within each set are converted to ranks and the difference between the two ranks are compared (Portney & Watkins, 2009). The difference score is then squared, known as d^2, and the squared values are summed ($\boldsymbol{\Sigma} d^2$), resulting in a value indicating the strength between the two variables (Portney & Watkins, 2009). Portney & Watkins (2009) state that the r_s statistic is then computed, where n=number of pairs and is as follows:

$$r_{s=1-\frac{6\Sigma d^2}{n(n^2-1)}}$$

The computed r_s value is then compared against a critical value table according to the number of pairs analyzed, and significant value is one that exceeds the tabled value (Portney & Watkins, 2009).

Main Experiment Results for Objective Performance

A mixed-factor ANOVA analyzed the effect of training strategy and stage of training (pretest and LU8) on *absolute score*. Results were interpreted as significant using α=0.05. Results are provided in Table 18. The main effect for training strategy was not significant, $F(3,36)=0.34$, $p=0.794$, $\eta_p^2=0.03$. The interaction for LU and training strategy was also not significant, $F(3,36)=2.13$, $p=0.114$, $\eta_p^2=0.15$. The main effect for LU was significant, $F(1,36)=201.26$, $p<0.001$, $\eta_p^2=0.85$. Follow-up testing for LU was not performed given that only two levels of the variable were measured and mean performance at LU8 was clearly better. A Spearman's correlation also assessed the relationship between the number total presentations and absolute score at LU8. The analyses showed no significant correlation between the number of trials and score, $r_s(40) = -0.01$, $p = 0.932$. Therefore, the difference in performance among strategies was not related to the number of total trials.

Table . Mixed-factor ANOVA results analyzing the effect of training strategy and LU on *absolute score* at **pretest and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*df*	*MS*	*F*	*p*	η_p^2
Between-Subjects					
Training Strategy	3	16.35	0.34	0.794	0.03
Error	36	47.62			
Within-Subjects					
LU	**1**	**2796.61**	**201.26**	**<0.001**	**0.85**
LU x Training	3	29.55	2.13	0.114	0.15
Error	36	13.90			
Total	79				

To determine if significant effects existed at LU4 compared to the pretest, a mixed-factor ANOVA evaluated the effect of training strategy and LU on *absolute score* at pretest and LU4. The decision level for significance was set to α=0.05. Results are shown in Table 19. The interaction for training strategy and LU was not significant, *F*(3,36)=2.60, *p*=0.070, η_p^2=0.18. The main effect for training strategy was also not significant, *F*(3,36)=0.59, *p*=0.628, η_p^2=0.05. A main effect for LU was significant, *F*(1,36)=192.09, *p*<0.001, η_p^2=0.84.

Table . Mixed factor ANOVA table analyzing the effect of training strategy and LU on *absolute score* at **pretest and LU4** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*df*	*MS*	*F*	*p*	η_p^2
Between-Subjects					
Training Strategy	3	26.72	0.59	0.628	0.05
Error	36	45.54			
Within-Subjects					
LU	**1**	**2622.05**	**192.09**	**<0.001**	**0.84**
LU x Training	3	35.52	2.60	0.070	0.18
Error	36	13.65			
Total	79				

To evaluate if significant changes in *absolute score* occurred among training strategies at LU4 and LU8, a mixed-factors ANOVA was performed. Results are shown in Table 20. Non-significant results were obtained for the following: interaction of training strategy and stage of training (F[3,36]= 0.73, p=0.542, η_p^2=0.06) training (F[3,36]=0.40, p=0.756, η_p^2=0.03), and LU (F[1,36]=0.59, p=0.448, η_p^2=0.02). The results of the mixed-factors ANOVA comparing performance at pretest, LU4, and LU8 support that no significant improvement in localization accuracy existed beyond LU4.

Table . Mixed factor ANOVA table analyzing the effect of training strategy and LU on *absolute score* at **LU4 and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*df*	*MS*	*F*	*p*	η_p^2
Between-Subjects					
Training Strategy	3	22.51	0.40	0.756	0.03
Error	36	56.75			
Within-Subjects					
LU	1	2.81	0.59	0.448	0.02
LU x Training	3	3.48	0.73	0.542	0.06
Error	36	4.79			
Total	79				

A mixed-factor ANOVA analyzed the effect of training strategy and LU on *response time* at pretest and LU8 (Table 21). Results were interpreted given α=0.05. An interaction effect for LU and training strategy was not significant, *F*(3,36)=0.36, *p*=0.780, η_p^2=0.03. The main effect for training strategy was not significant, *F*(3,36)=0.34, *p*=0.797, η_p^2=0.03. A main effect for LU was significant, *F*(1,36)=16.96, *p*<0.001, η_p^2=0.32. The results showed that response times at LU8 were significantly shorter than at pretest. The mean response time for at each LU for each training strategy is shown with the corresponding mean performance in Figures 28-32.

Table . Mixed factor ANOVA table analyzing the effect of training strategy and LU on *response time* at **pretest and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*df*	*MS*	*F*	*p*	η_p^2
Between-Subjects					
Training Strategy	3	0.15	0.34	0.797	0.03
Error	36	0.45			
Within-Subjects					
LU	**1**	**4.12**	**16.96**	**<0.001**	**0.32**
LU x Training	3	0.09	0.36	0.780	0.03
Error	36	0.24			
Total	79				

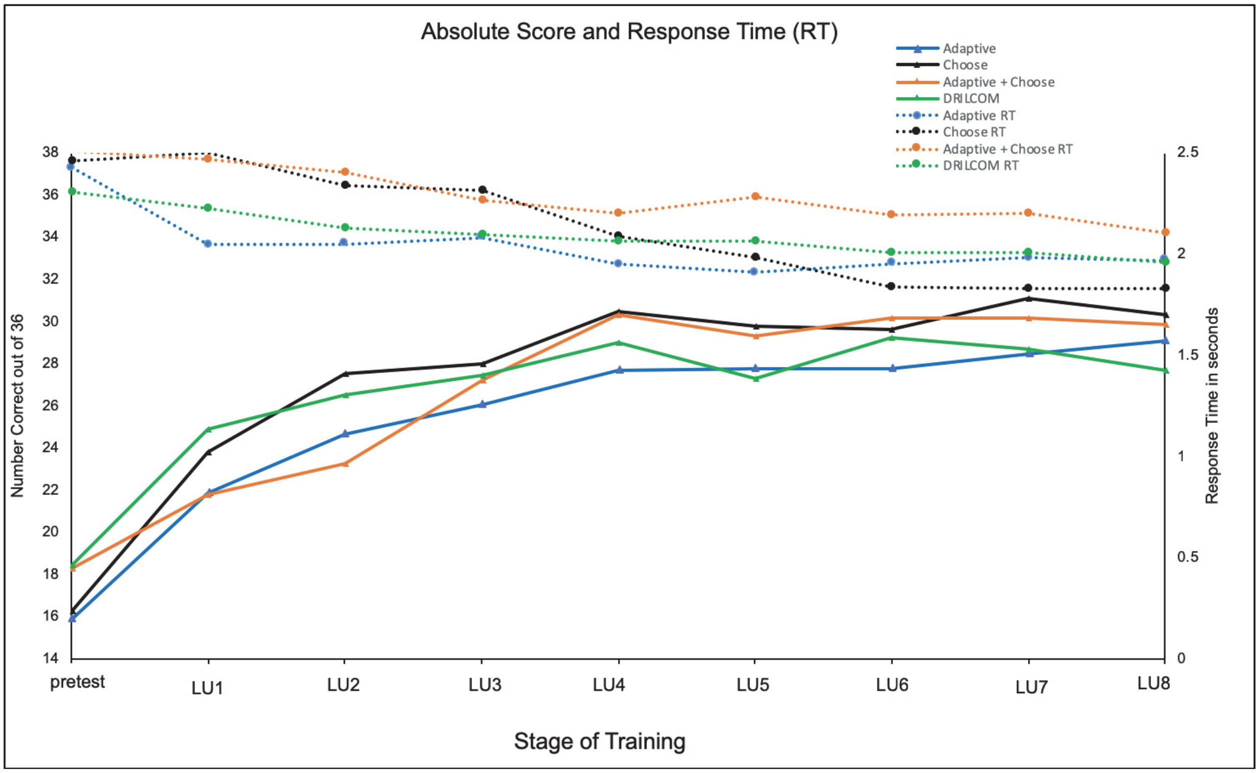


Figure . Mean *response time* and *absolute score* at each stage of training plotted according to training strategy.

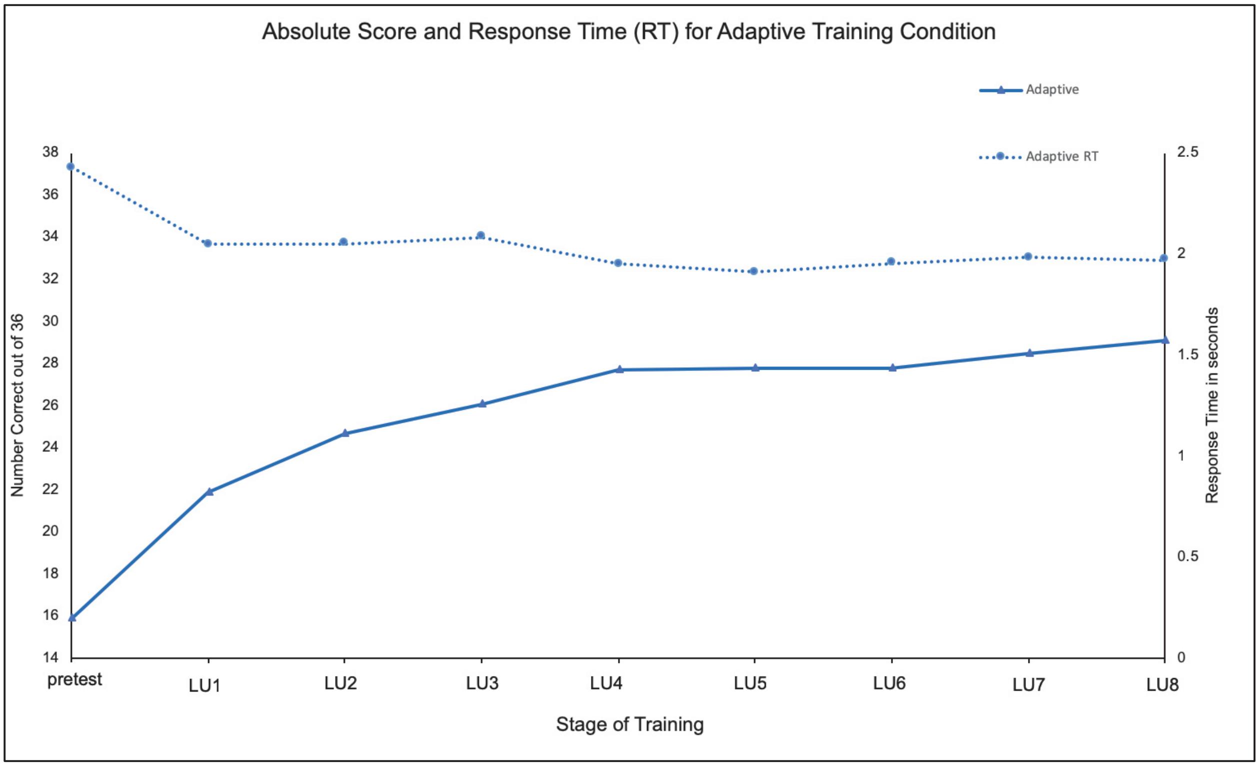


Figure . Mean *response time* and *absolute score* at each stage of training plotted according for the adaptive training strategy.

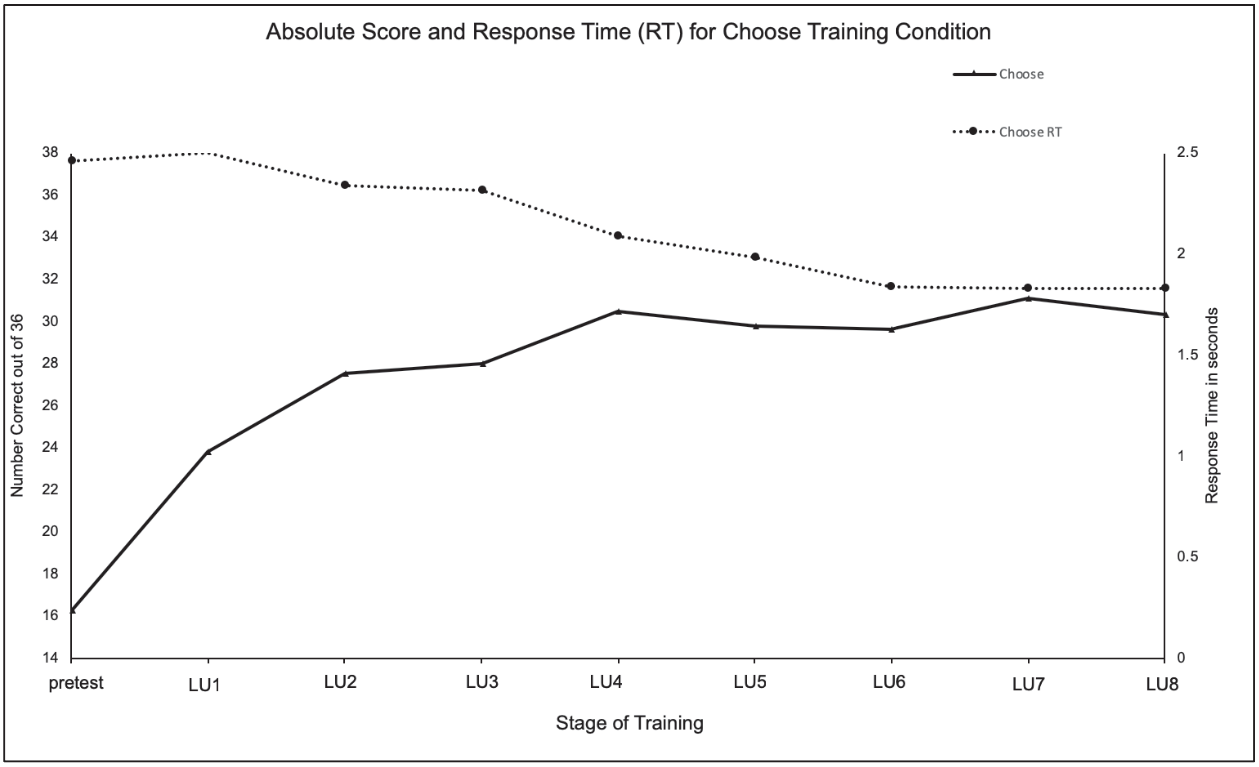


Figure . Mean <u>*response time*</u> and <u>*absolute score*</u> at each stage of training plotted according for the choose training strategy.

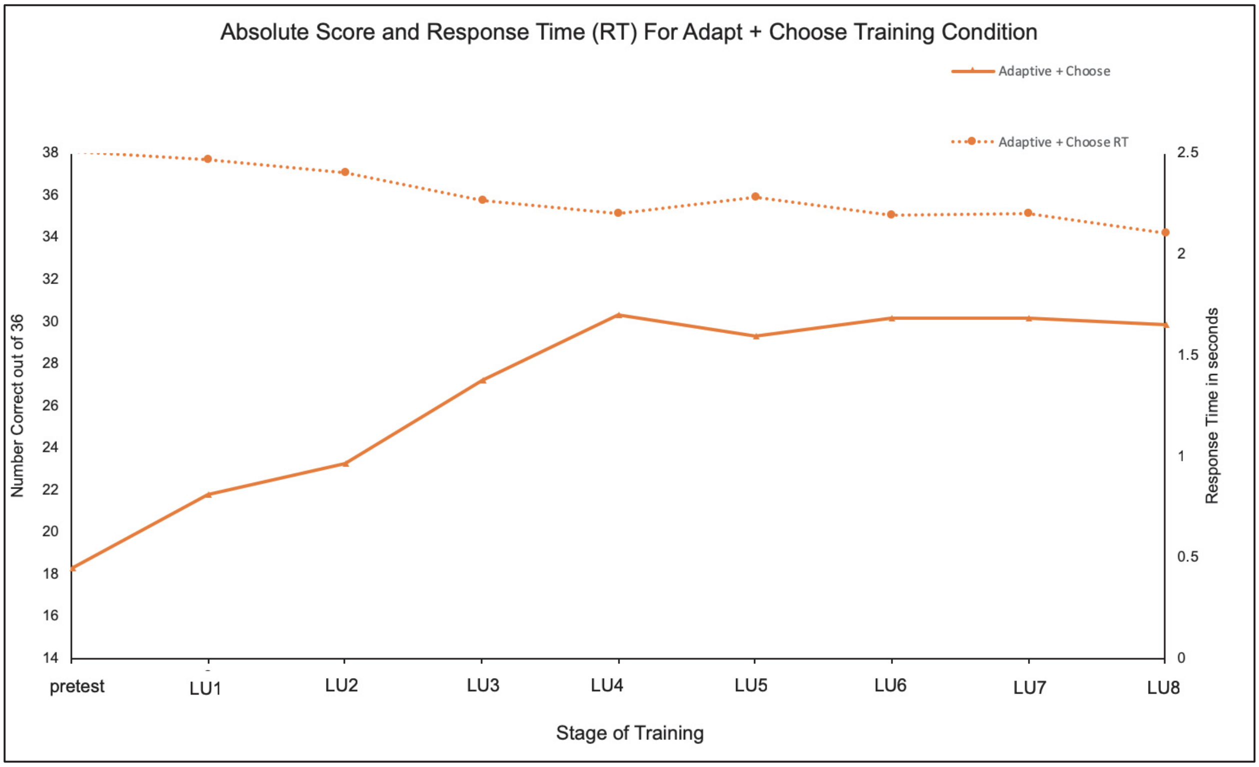


Figure . Mean *response time* and *absolute score* at each stage of training plotted according for the adapt + choose training strategy.

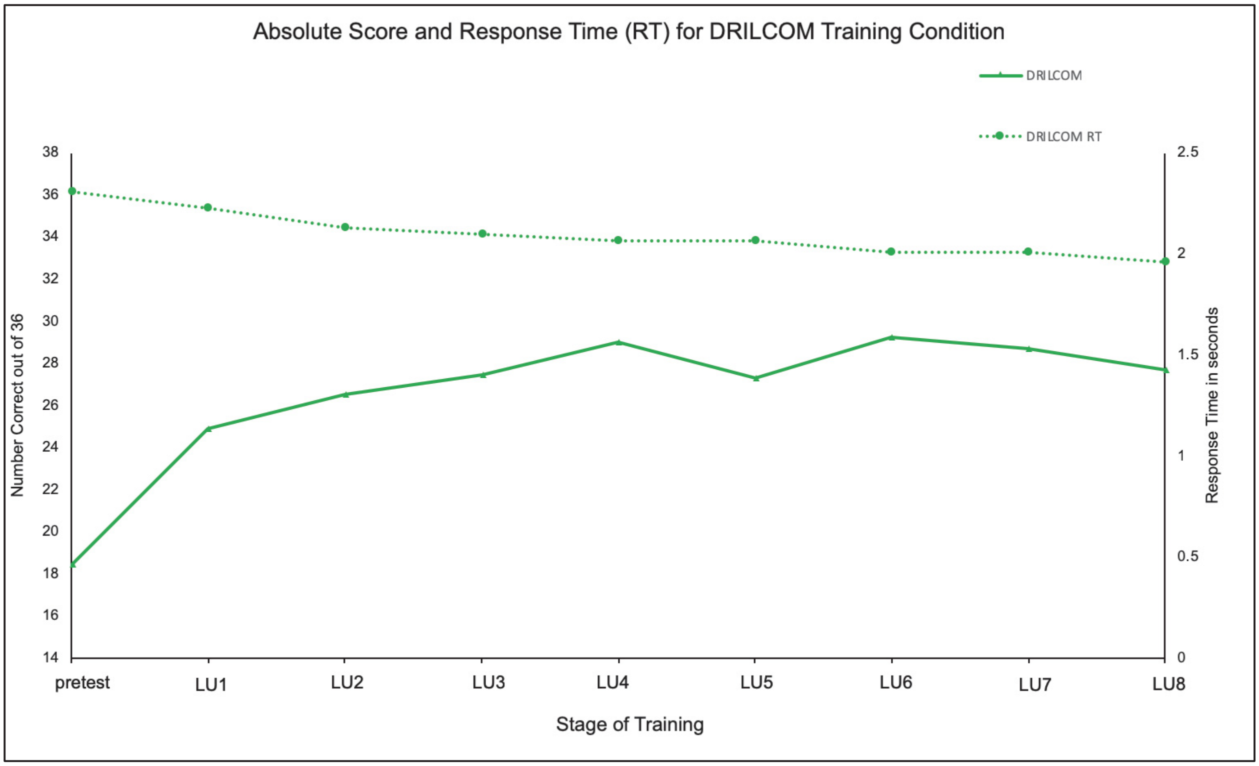


Figure . Mean *response time* and *absolute score* at each stage of training plotted according for the DRILCOM training strategy.

A mixed-factor ANOVA analyzed the effect of training strategy and LU on *response time* at pretest to LU4 (Table 22). Results were interpreted as significant at α=0.05. An interaction effect for LU and training strategy was not significant, $F(3,36)=0.46$, $p=0.710$, $\eta_p^2=0.04$. The main effect for training strategy was not significant, $F(3,36)=0.26$, $p=0.857$, $\eta_p^2=0.02$. A main effect for LU was significant, $F(1,36)=12.41$, $p=0.001$, $\eta_p^2=0.26$. Another mixed-factor ANOVA was performed to determine the change in response time comparing LU4 to LU8 (Table 23). An interaction effect for LU and training strategy was not significant, $F(3,36)=1.02$, $p=0.397$, $\eta_p^2=0.08$. The main effect for training strategy was not significant, $F(3,36)=0.35$, $p=0.790$, $\eta_p^2=0.03$. A main effect for LU was significant, $F(1,36)=4.57$, $p=0.039$, $\eta_p^2=0.11$. Results supported that response time was significantly faster at LU8 than LU4.

Table . Mixed factor ANOVA table analyzing the effect of training strategy and LU on *response time* at **pretest and LU4** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*df*	*MS*	*F*	*p*	η_p^2
Between-Subjects					
Training Strategy	3	0.14	0.26	0.857	0.02
Error	36	0.55			
Within-Subjects					
LU	**1**	**2.18**	**12.41**	**0.001**	**0.26**
LU x Training	3	0.08	0.46	0.710	0.04
Error	36	0.18			
Total	79				

Table . Mixed factor ANOVA table analyzing the effect of training strategy and LU on *response time* at **LU4 and LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*df*	*MS*	*F*	*p*	η_p^2
Between-Subjects					
Training Strategy	3	0.17	0.35	0.790	0.03
Error	36	0.48			
Within-Subjects					
LU	**1**	**0.31**	**4.57**	**0.039**	**0.11**
LU x Training	3	0.07	1.02	0.397	0.08
Error	36	0.07			
Total	79				

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *absolute difference score* at LU8 compared to pretest (LU8 $_{\text{absolute}}$ – Pretest $_{\text{absolute}}$). Results were interpreted at α=0.05. Results, shown in Table 24, showed no significant effect of training strategy on difference score, *F*(3,36)=2.13, *p*=0.114 (Table 24). Descriptive statistics for this analysis are provided in Table 25. These results corroborate with the mixed-factors ANOVA that showed no effect on score according to training strategy.

Table . One-way ANOVA results for measuring effect of training strategy on the *absolute difference score* from **LU8 compared to pretest**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	177.28	3	59.09	2.13	0.114
Error (Within-Subjects)	1000.50	36	27.79		
Total	1177.78	39			

Table . Descriptive statistics for the *absolute difference score* **from LU8 compared to pretest.**

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	13.20	1.74	9.26-17.14
Choose	10	14.00	2.06	9.34-18.66
Choose +Adapt	10	11.60	1.65	7.87-15.33
DRILCOM	10	8.50	1.06	6.11-10.89

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *absolute difference score* at LU4 compared to pretest (LU4 $_{\text{absolute}}$ – Pretest $_{\text{absolute}}$). All results were interpreted using α=0.05. Results showed no significant effect of training strategy on difference score, *F*(3, 36)=2.60 *p*=0.067 (Table 26). Descriptive statistics for this analysis are provided in Table 27.

Table . One-way ANOVA results for measuring effect of training strategy on the *absolute difference score* from **LU4 compared to pretest**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	213.10	3	71.03	2.60	0.067
Error (Within-Subjects)	982.80	36	27.30		
Total	1195.90	39			

Table . Descriptive statistics for using the *absolute difference score* from **LU4 compared to pretest**.

Condition	*n*	*M*	*SE*	*95% Confidence Interval (CI)*
Adaptive	10	11.8	1.65	8.45-15.15
Choose	10	14.2	1.65	10.85-17.55
Choose +Adapt	10	12.0	1.65	8.65-15.35
DRILCOM	10	7.8	1.65	4.45-11.15

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *absolute difference score* at LU8 compared to LU4 ($LU4_{absolute} – LU8_{absolute}$). Analyses were conducted using α=0.05. Results showed no significant effect of training strategy on difference score, *F*(3, 36)=0.73, *p*=0.543 (Table 28). Descriptive statistics for this analysis are provided in Table 29.

Table . One-way ANOVA results for measuring effect of training strategy on the *absolute difference score* from **LU8 compared to LU4**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	20.88	3	6.96	0.73	0.543
Error	344.50	36	9.57		
Total	365.38	39			

Table . Descriptive statistics for the *absolute difference score* from **LU4 compared to pretest.**

Condition	*n*	*M*	*SE*	*95% Confidence Interval (CI)*
Adaptive	10	1.40	0.98	-0.58-3.38
Choose	10	-0.20	0.98	-2.18-1.78
Choose +Adapt	10	-0.40	0.98	-2.38-1.58
DRILCOM	10	0.70	0.98	-1.28-2.68

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *ballpark difference score* at LU8 compared to pretest ($LU8_{ballpark} – Pretest_{ballpark}$). Results were interpreted using α=0.05. Results showed no significant effect of training strategy on difference score, $F(3, 36)=0.73$, $p=0.542$ (Table 30). Descriptive statistics for this analysis are provided in Table 31.

Table . One-way ANOVA results for measuring effect of training strategy on the <u>*ballpark difference score*</u> from **LU8 compared to pretest**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	44.68	3	14.89	0.73	0.542
Error	735.70	36	20.44		
Total	780.38	39			

Table . Descriptive statistics for the <u>*ballpark difference score*</u> from **LU8 compared to pretest.**

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	5.00	5.00	1.15-2.39
Choose	10	3.80	3.80	1.79- -0.24
Choose +Adapt	10	4.50	4.50	1.60-0.88
DRILCOM	10	2.20	2.20	1.04- -0.16

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the <u>*ballpark difference score*</u> at LU4 compared to pretest ($\text{LU4}_{\text{ballpark}} - \text{Pretest}_{\text{ballpark}}$). Results, provided in Table 32, showed no significant effect of training strategy on difference score, $F(3, 36)=1.45$, $p=0.245$ (Table 32). Descriptive statistics for this analysis are provided in Table 33.

Table . One-way ANOVA results for measuring effect of training strategy on the <u>*ballpark difference score*</u> for **LU4 compared to pretest.**

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	71.88	3	23.96	1.45	0.245
Error	595.10	36	16.53		
Total	666.98	39			

Table . Descriptive statistics for the <u>*ballpark difference score*</u> from **LU4 compared to pretest.**

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	4.90	1.29	2.29-7.51
Choose	10	3.90	1.29	1.29-6.51
Choose +Adapt	10	5.40	1.29	2.79-8.01
DRILCOM	10	1.90	1.29	-0.71-4.51

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the <u>*ballpark difference score*</u> at LU8 compared to LU4 ($LU8_{\text{ballpark}} – LU4_{\text{ballpark}}$). Results were interpreted at α=0.05. Results showed no significant effect of training strategy on difference score, *F*(3, 36)=0.35, *p*=0.790 (Table 34). Descriptive statistics for this analysis are provided in Table 35.

Table . ANOVA results for measuring effect of training strategy on ballpark score using the <u>*ballpark difference score*</u> from **LU8 compared to LU4**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	8.30	3	2.77	0.35	0.790
Error	284.80	36	7.91		
Total	293.10	39			

Table . Descriptive statistics for the <u>*ballpark difference score*</u> from LU8 compared to LU4.

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	3.03	0.96	-2.07-2.27
Choose	10	3.25	1.03	-2.42-2.22
Choose +Adapt	10	2.42	0.77	-2.63-0.83
DRILCOM	10	2.45	0.78	-1.45-2.05

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the <u>*response time difference score*</u> from pretest to LU8 ($\text{Response time}_{\text{pretest}} – \text{Response time}_{\text{LU8}}$). Results were interpreted for significance at α=0.05. A positive value indicates an improvement, or faster, response time. Results were non-significant, *F*(3, 36)=0.36, *p*=0.780, and shown in Table 36. Descriptive statistics are provided in Table 37.

Table . One-way ANOVA results for measuring effect of training strategy *response time difference score* from **pretest to LU8**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	.53	3	0.18	0.36	0.780
Error	17.5	36	0.49		
Total	18.03	39			

Table . Descriptive statistics for the *response time difference score* from **pretest to LU8**.

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	0.46	0.22	0.01-0.90
Choose	10	0.64	0.22	0.19-1.08
Choose +Adapt	10	0.40	0.22	-0.05-0.85
DRILCOM	10	0.32	0.22	-0.13-0.77

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *response time difference score* from pretest to LU4 (Pretest $_{\text{response time}}$ – LU4 $_{\text{response time}}$). Results were interpreted at α=0.05. Results were non-significant, *F*(3, 36)=0.46 *p*=0.714, and shown in Table 38. Descriptive statistics are provided in Table 39.

Table . One-way ANOVA results for measuring effect of training strategy *response time difference score* from **pretest to LU4.**

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	0.48	3	0.16	0.46	0.714
Error (Within-Subjects)	12.62	36	0.35		
Total	13.11	39			

Table . Descriptive statistics for the *response time difference score* from **pretest to LU4.**

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	0.47	0.19	0.10-0.85
Choose	10	0.37	0.19	-0.01-0.75
Choose +Adapt	10	0.30	0.19	-0.08-0.68
DRILCOM	10	0.17	0.19	-0.21-0.55

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *response time difference score* from LU4 to LU8 ($\text{LU4}_{\text{response time}} - \text{LU8}_{\text{response time}}$). Results were interpreted for significance at α=0.05. Results were non-significant, *F*(3, 36)= 1.02, *p*=0.397, and shown in Table 39. Descriptive statistics are provided in Table 40.

Table . One-way ANOVA results for measuring effect of training strategy *response time difference score* from **LU4 to LU8.**

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	0.41	3	0.14	1.02	0.397
Error (Within-Subjects)	4.85	36	0.14		
Total	5.26	39			

Table . Descriptive statistics for the *response time difference score* from LU4 to LU8.

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	0.44	0.14	-0.33-0.29
Choose	10	0.28	0.09	0.06-0.47
Choose +Adapt	10	0.39	0.12	-0.18-0.38
DRILCOM	10	0.34	0.11	-0.09-0.39

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *slope of the absolute score* from pretest to LU8. Results were interpreted at α=0.05. Results were not significant, *F*(3, 36)=2.25, *p*=0.099, and shown in Table 42. Descriptive statistics are provided in Table 43.

Table . One-way ANOVA results for measuring effect of training strategy on mean *slope* from **pretest to LU8**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	3.19	3	1.06	2.25	0.099
Error (Within-Subjects)	16.98	36	0.47		
Total	20.17	39			

Table . Descriptive statistics for *slope* measured from **pretest to LU8** for each training strategy.

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	1.34	0.22	0.91-1.78
Choose	10	1.40	0.22	0.96-1.84
Choose +Adapt	10	1.46	0.22	1.02-1.90
DRILCOM	10	0.76	0.22	0.31-1.20

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the *slope of the absolute score* from pretest to LU4. Results were interpreted at α=0.05. Results were significant, *F*(3, 36)=3.04, *p*=0.041, and provided in Table 44. Descriptive statistics are provided in Table 45. Results supported that the degree of change in the absolute score (as indicated by line slope) from pretest to LU4 differed significantly according to training strategy. Tukey HSD pairwise comparisons (Table 46) showed that the slope in the choose condition was significantly greater than the DRILCOM condition.

Table . One-way ANOVA results for measuring effect of training strategy on *slope* from **pretest to LU4** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	**14.33**	**3**	**4.77**	**3.04**	**0.041**
Error (Within-Subjects)	56.45	36	1.57		
Total	70.77	39			

Table . Descriptive statistics for *slope* measured from pretest to **LU4 for each training strategy** (bolded text in the table indicates a significant test result at the α=0.05 significance level**).**

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	2.78	0.40	1.98-3.58
Choose	10	3.26	0.40	2.46-4.06
Choose +Adapt	10	2.94	0.40	2.14-3.74
DRILCOM	10	1.67	0.40	0.87-2.47

Table . Tukey HSD pairwise comparisons for each training strategy using slope from **pretest to LU4** as the dependent measure (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Listening Condition		*M*	*SE Difference*	*p*
Choose	**DRILCOM**	**1.59**	**0.56**	**0.036**
Choose + Adaptive	DRILCOM	1.27	0.56	0.125
Adaptive	DRILCOM	1.11	0.56	0.214
Choose	Adaptive	0.48	0.56	0.827
Choose	Choose+ Adaptive	0.32	0.56	0.940
Choose + Adaptive	Adaptive	0.16	0.56	0.992

A one-way, between-subjects ANOVA analyzed the effect of training strategy on the slope of the absolute score from LU4 to LU8. Results were interpreted at α=0.05. Results were not significant, *F*(3, 36)=1.24, *p*=0.311, and shown in Table 47. Descriptive statistics are provided in Table 48.

Table . One-way ANOVA table results for measuring effect of training strategy on slope from **LU4 to LU8**.

Source	*SS*	*df*	*MS*	*F*	*p*
Training Strategy	0.03	3	0.01	1.24	0.311
Error (Within-Subjects)	0.31	36	0.01		
Total	0.34	39			

Table . Descriptive statistics for _slope_ measured **from LU4 to LU8** for each training strategy.

Condition	*n*	*M*	*SE*	*95% Confidence Interval* (CI)
Adaptive	10	0.01	0.03	-0.05- 0.07
Choose	10	-0.07	0.03	-0.13- -0.01
Choose +Adapt	10	-0.03	0.03	-0.09- 0.03
DRILCOM	10	-0.04	0.03	-0.10- 0.02

Regression analyses evaluated the predictive value of _changes in absolute score_ on _changes in response time between adjacent sessions_ from pretest to LU8. Results were interpreted for significance at α=0.05. The change in difference score between adjacent LUs served as a significant predictor of change in response time for adjacent LUs (*F*[1, 318]=8.76, *p*=0.003, R^2=0.03). The scatter plot is provided in Figure 33. The ANOVA summary table for pretest to LU8 is provided Table 49. The linear regression equation was as follows:

adjacent difference score response time= -0.03 – 0.01(adjacent difference score)

Table . ANOVA table and parameter estimates for linear regression for _change in response time given a change in absolute score_ measured from **pretest to LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*SS*	*df*	*MS*	*F*	R^2	*p*
Model	**0.95**	**1**	**0.95**	**8.76**	**0.03**	**0.003**
Error	34.66	318	0.10			
Total	35.61	319				

Variable	Parameter Estimate	*SE*	*t*	*p*
Intercept	-0.04	0.02	-1.85	0.065
Adjacent absolute difference score	**-0.01**	**0.00**	**-2.96**	**0.003**

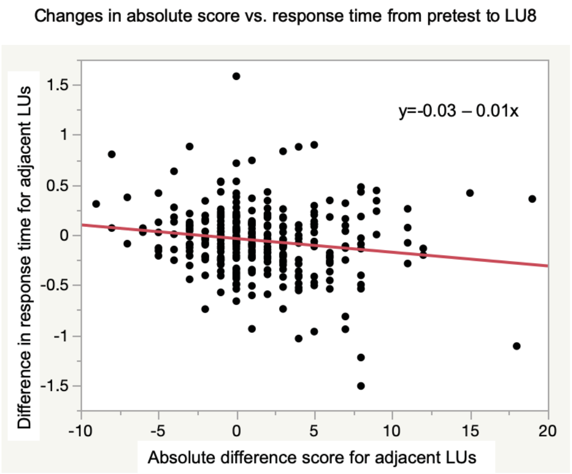


Figure . *Changes in absolute score* versus *response time* from pretest to LU8 with the regression line and equation plotted.

As above, regression analysis was repeated to evaluate the predictive value of *changes in absolute score* on *changes in response time* between adjacent sessions from pretest to LU4. Results were interpreted at α=0.05. In this training duration, the difference score between adjacent LUs did not significantly predict a change in response time for adjacent LUs. The ANOVA summary table for pretest to LU8 is provided Table 50 and the scatterplot is provided in Figure 34. The linear regression equation was as follows:

adjacent difference score response time = -0.05 - 0.01(adjacent difference score)

Table . ANOVA table and parameter estimates for linear regression for *change in response time* given a *change in absolute score* measured from **pretest to LU4**.

Source	*SS*	*df*	*MS*	*F*	*R-square*	*p*
Model	0.44	1	0.44	3.05	0.02	0.083
Error	22.53	158	0.14			
Total	22.96	159				

Variable	Parameter Estimate	*SE*	*t*	*p*
Intercept	-0.05	0.04	-1.30	0.195
Adjacent absolute difference score	-0.01	0.01	-1.75	0.082

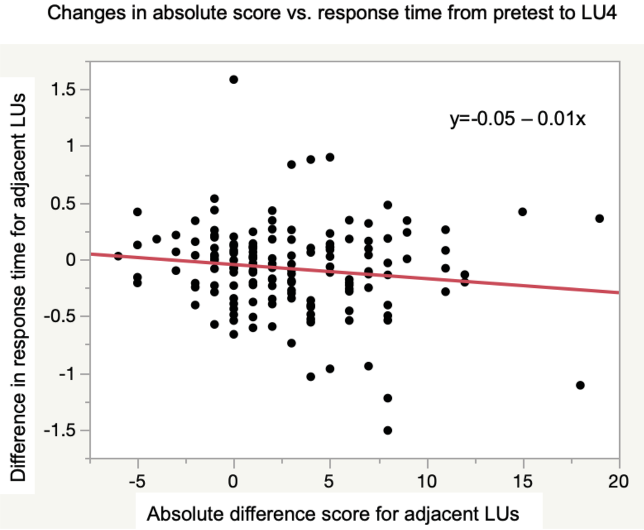


Figure . *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted.

Regression analyses evaluated the predictive value of *changes in absolute score* on *changes in response time* between adjacent sessions from LU4 to LU8. Results were interpreted at α=0.05. The change in difference score between adjacent LUs served as a significant predictor of change in response time for adjacent LUs (*F*[1, 158]=4.28, *p*=0.040, R^2=0.02). The ANOVA summary table for LU4 to LU8 is provided Table 51, and the scatterplot is provided in Figure 35. The linear regression equation was as follows:

adjacent difference score response time = -0.03 - 0.01(adjacent difference score)

Table . ANOVA table and parameter estimates for linear regression for *changes in absolute score* versus *response time* measured from **LU4 to LU8** (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Source	*SS*	*df*	*MS*	*F*	R^2	*p*
Model	**0.33**	**1**	**0.33**	**4.28**	**0.02**	**0.040**
Error	12.11	158	0.08			
Total	12.44	159				

Variable	Parameter Estimate	*SE*	*t*	*p*
Intercept	-0.03	0.02	-1.36	0.177
Adjacent absolute difference score	**-0.01**	**0.01**	**-2.07**	**0.040**

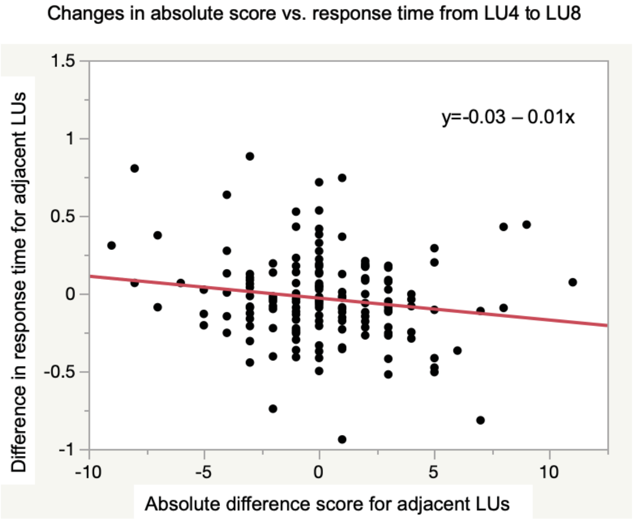


Figure . *Changes in absolute score* versus *response time* from **LU4 to LU8** with the regression line and equation plotted.

Separate regression analyses were conducted for each training strategy with the *change absolute score* from the adjacent LU serving as the predictor variable and the *change in response time* from the adjacent LU serving as the response variable. Regression analyses were conducted from pretest to LU8 and results are provided in Table 52. Scatterplots are proved in Figures 36-39. Only the adaptive training strategy resulted in a significant predictive relationship between the improvement in absolute score and a decrease in response time (*F*[1, 78]=11.81, *p*=0.001, R^2=0.13). Linear regression equations for each condition are as follows:

Adaptive: adjacent difference score response time = -0.01 - 0.03(adjacent difference score)

Choose: adjacent difference score response time = -0.07 - 0.00(adjacent difference score)

Adaptive + choose: adjacent difference score response time = -0.02 - 0.02(adjacent difference score)

DRILCOM: adjacent difference score response time = -0.04 + 0.01(adjacent difference score)

Table . ANOVA results and linear regression parameter estimates for each training strategy analyzing the <u>*changes in absolute score*</u> versus <u>*response time*</u> from **pretest to LU8** (bolded text in the table indicates a significant test result at the the α=0.05 significance level).

Training Strategy (Source)		*SS*	*df*	*MS*	*F*	*R-square*	*p*
Adaptive		**1.31**	**1**	**1.31**	**11.81**	**0.13**	**0.001**
	Error	8.63	78	0.11			
	Total	9.94	79				
Choose		0.02	1	0.02	0.20	0.00	0.655
	Error	8.23	78	0.11			
	Total	8.25	79				
Adaptive + Choose		0.46	1	0.46	2.92	0.04	0.092
	Error	12.23	78	0.16			
	Total	12.69	79				
DRILCOM		0.03	1	0.03	0.49	0.01	0.484
	Error	4.64	78	0.06			
	Total	4.67	79				

Variable by Training Strategy		Parameter Estimate	*SE*	*t*	*p*
Adaptive					
	Intercept	-0.03	0.02	-1.36	0.177
	Adjacent absolute difference score	**-0.01**	**0.01**	**-2.07**	**0.040**
Choose					
	Intercept	-0.07	0.04	-1.81	-0.075
	Adjacent absolute difference score	0.00	0.01	-0.45	0.656
Adaptive + Choose					
	Intercept	-0.02	0.04	-0.51	0.613
	Adjacent absolute difference score	-0.02	0.01	-1.71	0.092
DRILCOM					
	Intercept	-0.05	0.03	-1.62	0.110
	Adjacent absolute difference score	0.01	0.01	0.70	0.484

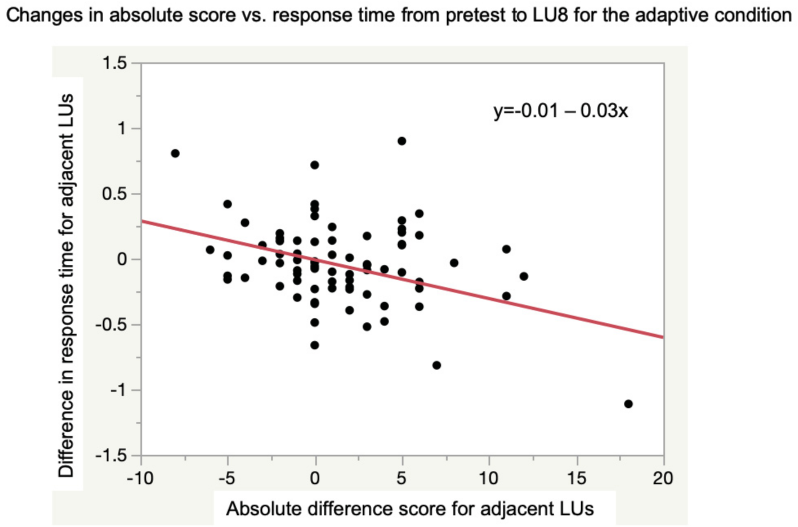


Figure . *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the adaptive condition.

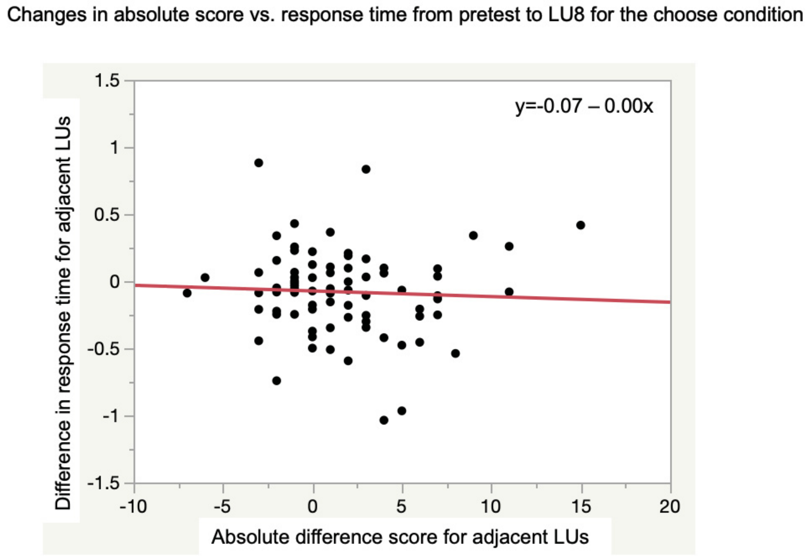


Figure . *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the choose condition.

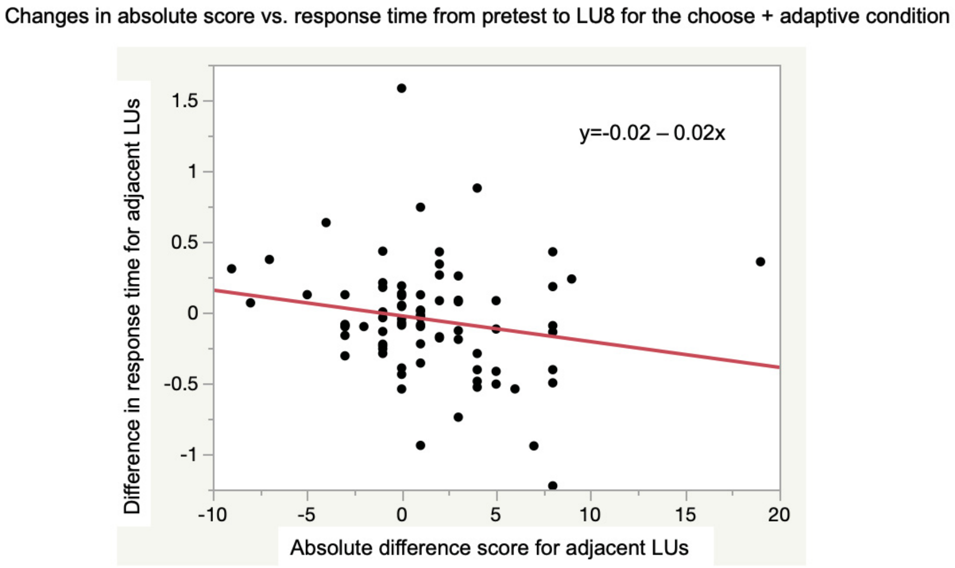


Figure . *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the adaptive + choose condition.

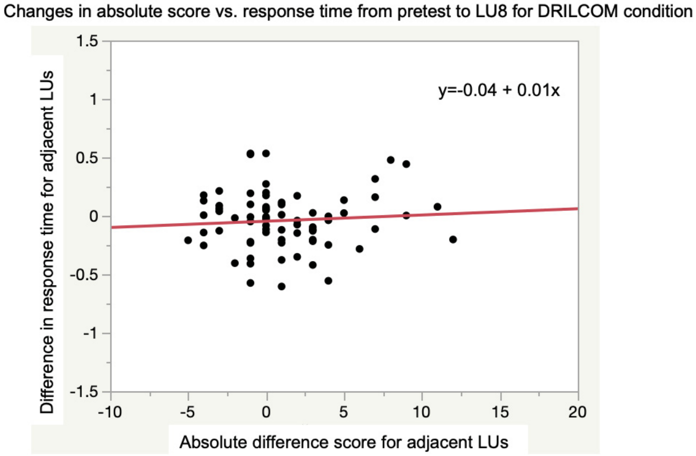


Figure . *Changes in absolute score* versus *response time* from **pretest to LU8** with the regression line and equation plotted for the DRILCOM condition.

As above, separate regression analyses were conducted for each training strategy. Results were interpreted at α=0.05. The *change in absolute score* from the adjacent LU served as the predictor variable and the *change in response time* from the adjacent LU served as the response variable. However, results were analyzed from pretest to LU4 and provided in Table

53. Scatterplots are provided in Figures 40-43. Only the adaptive training strategy resulted in a significant predictive relationship between the improvement in absolute score and a decrease in response time (*F*[1, 38]=4.15, *p*=0.049, R^2=0.10). Linear regression equations for each condition are as follows:

Adaptive: adjacent difference score response time = -0.04 - 0.03 (adjacent difference score)

Choose: adjacent difference score response time = -0.08 - 0.00 (adjacent difference score)

Adaptive + choose: adjacent difference score response time = -0.02 - 0.02 (adjacent difference score)

DRILCOM: adjacent difference score response time = -0.06 + 0.01(adjacent difference score)

Table . ANOVA results and linear regression parameter estimates for each training strategy analyzing the *change in response time* given a *change in absolute score* measured from **pretest to LU4** (bolded text in the table indicates a significant test result at the the α=0.05 significance level).

Training Strategy (Source)		*SS*	*df*	*MS*	*F*	*R-square*	*p*
Adaptive		**0.61**	**1**	**0.61**	**4.15**	**0.10**	**0.049**
	Error	5.55	38	0.15			
	Total	9.94	39				
Choose		0.01	1	0.01	0.05	0.00	0.829
	Error	4.83	38	0.13			
	Total	4.83	39				
Adaptive + Choose		0.24	1	0.24	1.00	0.03	0.324
	Error	9.01	38	0.24			
	Total	9.24	39				
DRILCOM							
	Error	0.02	1	0.02	0.30	0.01	0.590
	Total	2.60	38	0.07			

Variable by Training Strategy		Parameter Estimate	*SE*	*t*	*p*
Adaptive					
	Intercept	-0.04	0.07	-0.50	0.620
	Adjacent absolute difference score	**-0.03**	**0.01**	**-2.04**	**0.049**
Choose					
	Intercept	-0.08	0.07	-1.13	0.266
	Adjacent absolute difference score	-0.00	0.01	-0.22	0.829
Adaptive + Choose					
	Intercept	-0.02	0.09	-0.21	0.836
	Adjacent absolute difference score	-0.02	0.02	-1.00	0.324
DRILCOM					
	Intercept	-0.06	0.05	-1.18	0.244
	Adjacent absolute difference score	0.01	0.01	0.55	0.588

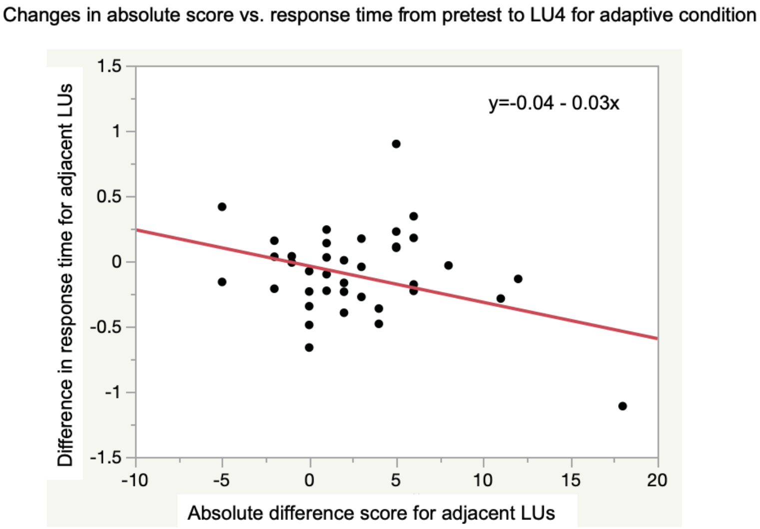


Figure . *<u>Changes in absolute score </u>*versus *<u>response time</u>* from **pretest to LU4** with the regression line and equation plotted for the adaptive condition.

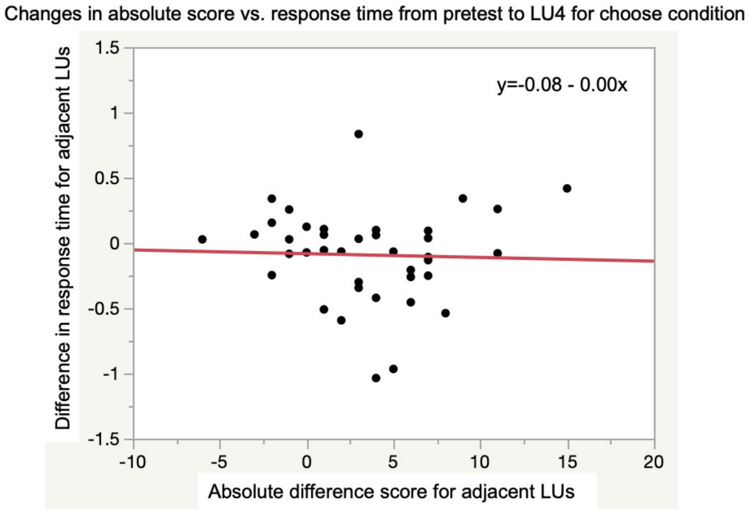


Figure . *<u>Changes in absolute score </u>*versus *<u>response time </u>*from **pretest to LU4** with the regression line and equation plotted for the choose condition.

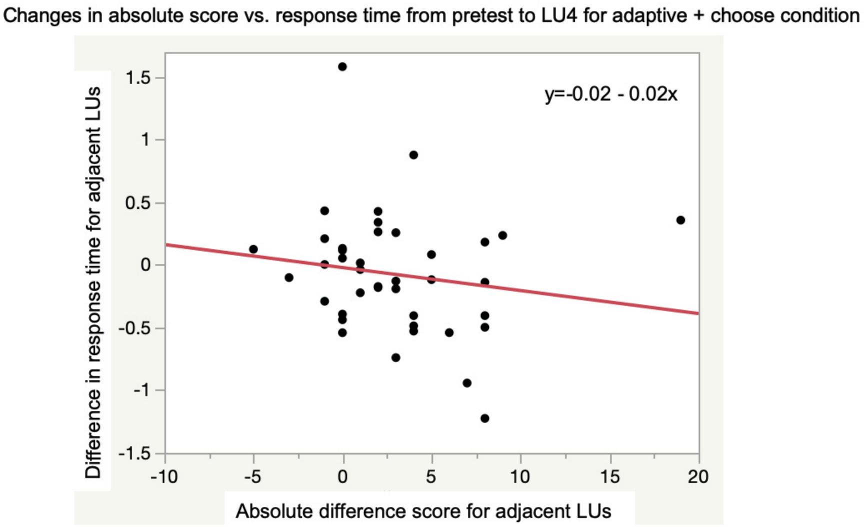


Figure . *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted for the adaptive + choose condition.

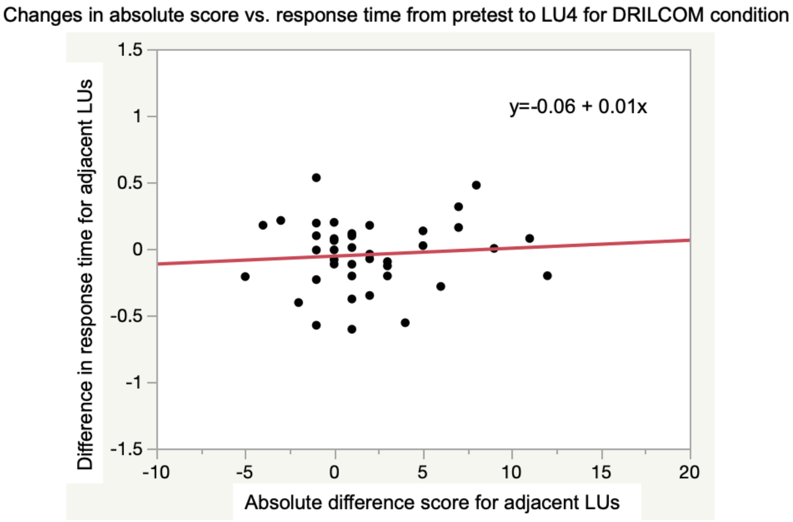


Figure . *Changes in absolute score* versus *response time* from **pretest to LU4** with the regression line and equation plotted for the DRILCOM condition.

Separate regression analyses were conducted for each training strategy from LU4 to LU8. Results were interpreted at α=0.05. The change in absolute score from the adjacent LU served as the predictor variable and the change in response time from the adjacent LU served as the response variable. Results are provided in Table 54, and the scatterplots are provided in Figures

44-47. Only the adaptive training strategy resulted in a significant predictive relationship between the improvement in absolute score and a decrease in response time (*F*[1, 38]=5.55, p=0.024, R^2=0.13). Linear regression equations for each condition are as follows:

Adaptive: adjacent difference score response time = -0.01 - 0.03 (adjacent difference score)

Choose: adjacent difference score response time = -0.07 - 0.01 (adjacent difference score)

Adaptive + choose: adjacent difference score response time = -0.03 - 0.02 (adjacent difference score)

DRILCOM: adjacent difference score response time = -0.04 + 0.01(adjacent difference score)

Table . ANOVA results and linear regression parameter estimates for each training strategy analyzing the change in response time given a change in absolute score measured from LU4 to LU8 (bolded text in the table indicates a significant test result at the α=0.05 significance level).

Training Strategy (Source)		*SS*	*df*	*MS*	*F*	*R-square*	*p*
Adaptive		**0.44**	**1**	**0.44**	**5.55**	**0.13**	**0.024**
	Error	3.04	38	0.08			
	Total	3.48	39				
Choose		0.01	1	0.01	0.06	0.00	0.814
	Error	3.40	38	0.09			
	Total	3.41	39				
Adaptive + Choose		0.17	1	0.17	2.04	0.05	0.162
	Error	3.22	38	0.08			
	Total	3.40	39				
DRILCOM		0.01	1	0.01	0.25	0.01	0.619
	Error	2.04	38	0.05			
	Total	2.05	39				

Variable by Training Strategy		Parameter Estimate	*SE*	*t*	*p*
Adaptive					
	Intercept	0.14	0.04	0.32	0.750
	Adjacent absolute difference score	**-0.03**	**0.01**	**-2.36**	**0.024**
Choose					
	Intercept	-0.07	0.05	-1.40	0.169
	Adjacent absolute difference score	-0.01	0.02	-0.24	0.814
Adaptive + Choose					
	Intercept	-0.03	0.05	-0.59	0.557
	Adjacent absolute difference score	-0.02	0.01	-1.43	0.162
DRILCOM					
	Intercept	-0.04	0.04	-1.04	0.304
	Adjacent absolute difference score	0.00	0.01	0.50	0.619

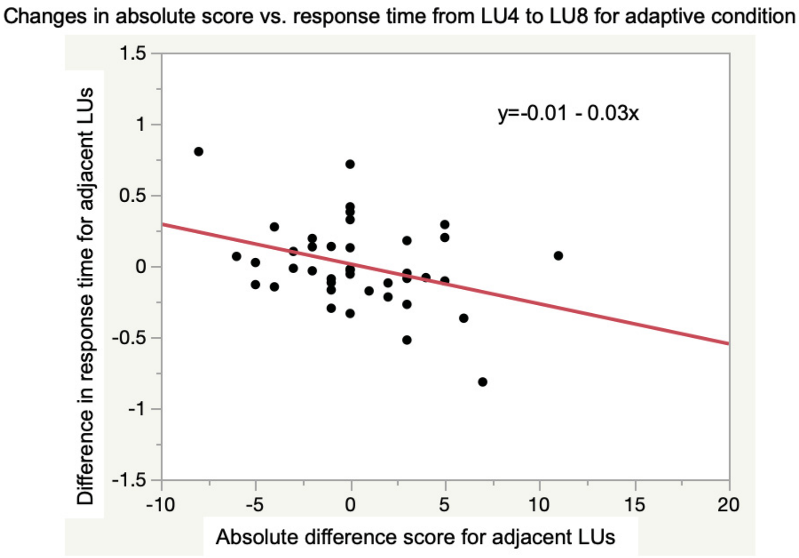


Figure . <u>*Changes in absolute score*</u> versus <u>*response time*</u> from **LU4 to LU8** with the regression line and equation plotted for the adaptive condition.

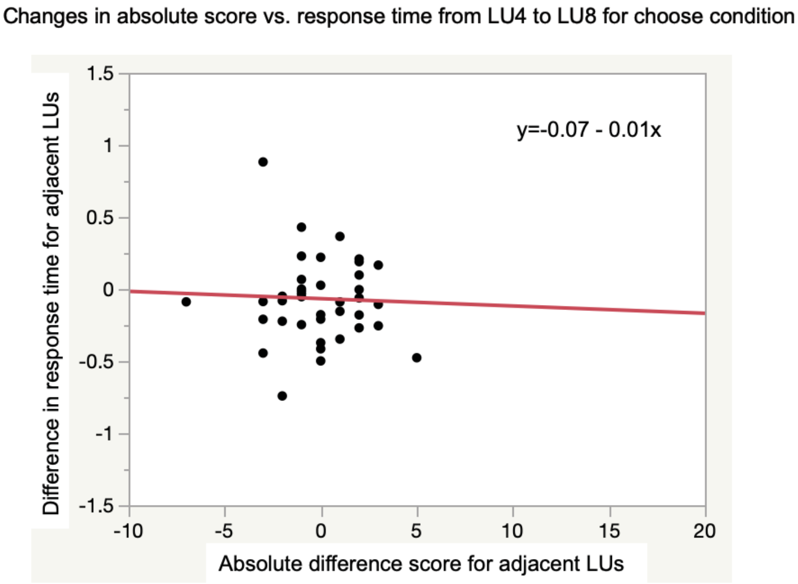


Figure . <u>*Changes in absolute score*</u> versus <u>*response time*</u> from **LU4 to LU8** with the regression line and equation plotted for the choose condition.

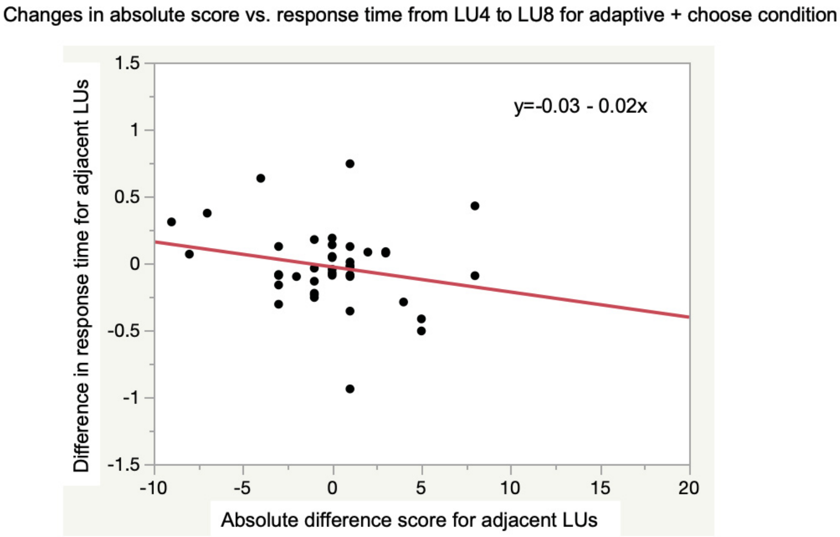


Figure . *<u>Changes in absolute score</u>* versus *<u>response time</u>* from **LU4 to LU8** with the regression line and equation plotted for the adaptive + choose condition.

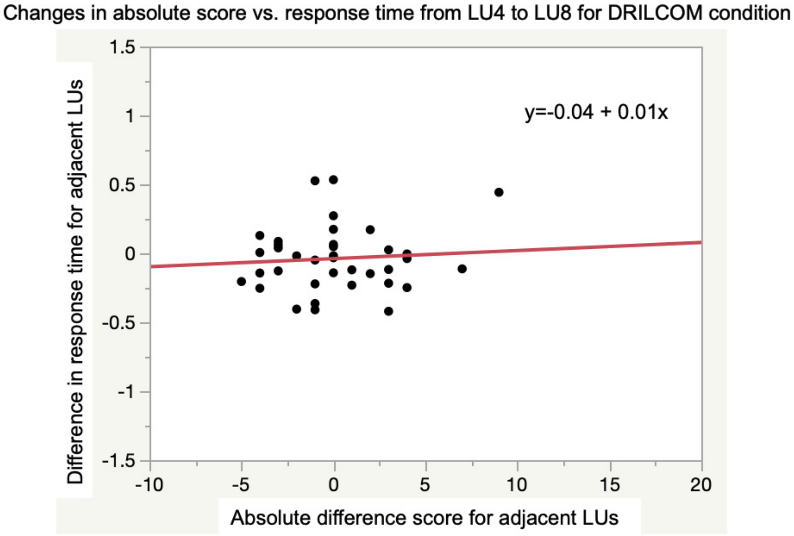
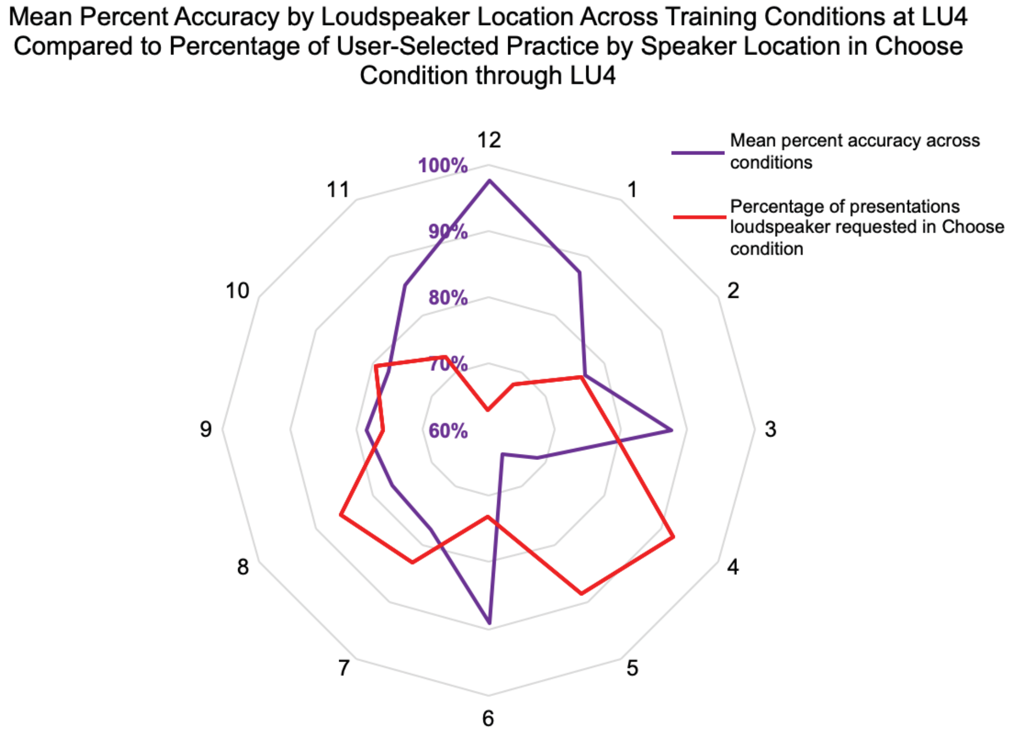


Figure . *<u>Changes in absolute score</u>* versus *<u>response time</u>* from **LU4 to LU8** with the regression line and equation plotted for the DRILCOM condition.

Main Experiment Data Reduction by Location

The author evaluated the participants' ability to identify problematic locations (for more practice) that lead to reduced accuracy. To do so, correlations were calculated between absolute percent accuracy and percent chosen within the same LU for the choose condition. Given the difference scales of comparison, scores were converted to *Z* scores. Spearman correlation showed that a significant negative relationship existed between percent accuracy at the speaker and percentage of the time the speaker was chosen, (r_s[480]=-0.14, p<0.002). Polar plots showed the relationship between mean percent accuracy at each speaker location and percentage of the time the loudspeaker was chosen for practice at LU4 (Figures 48 and 49). In order to observe general trends of localization accuracy at LU4, the percent accuracy according to loudspeaker location was plotted and shown in Figure 50. Polar plots results revealed poorer accuracy at two, four, and five o'clock compared to the left-sided equivalent.



Mean Percent Accuracy by Loudspeaker Location Across Training Conditions at LU4 Compared to Percentage of User-Selected Practice by Speaker Location in Choose Condition through LU4

Figure . Mean percent accuracy at LU4 for each speaker location for the choose subunit compared to the user-selected practice subunit.

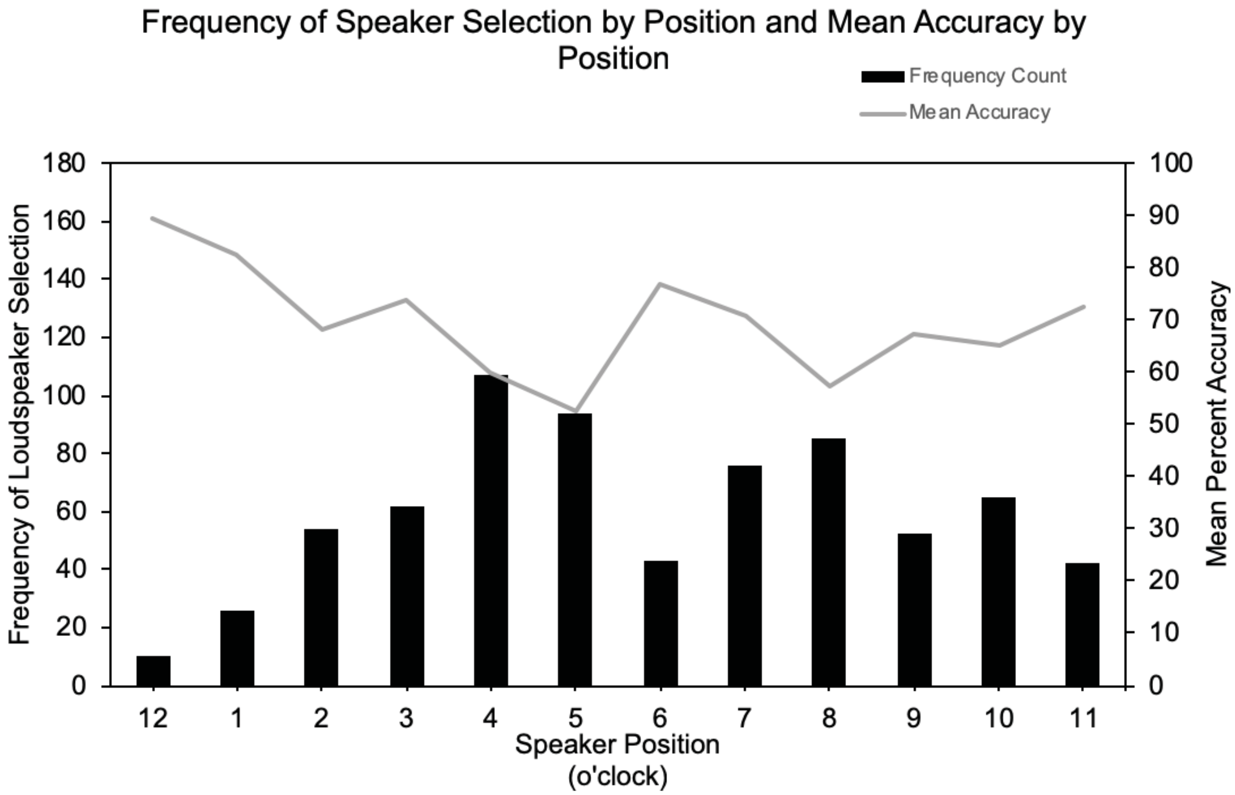


Figure . Frequency of loudspeaker selection compared to mean percent accuracy at each speaker location for the user-selected practice subunit.

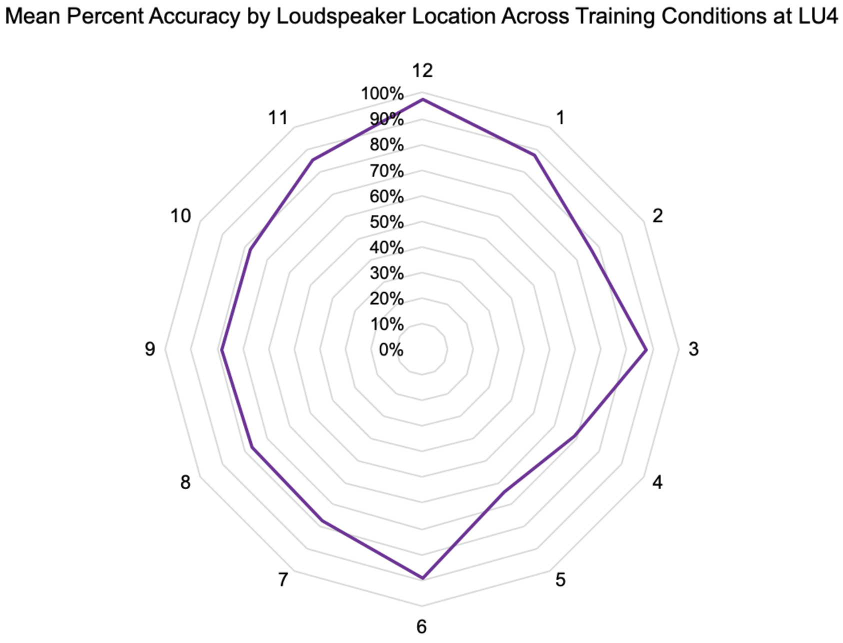


Figure . Mean percent accuracy at LU4 for each speaker location.

Spearman correlations were computed for each speaker location to evaluate the relationship between percentage of time chosen and percent accuracy for the choose condition. No significant relationship was revealed between participant choice of speaker in the user-select subunit and percent accuracy at that particular speaker location. Results as well as descriptive statistics are provided in Table 55.

Table . Spearman correlations using Z scores for percentage of trials a speaker location was chosen and the percent correct performance at that location.

Location	Measure	Percent Chosen (r_s)	*M*	*SD*	*p*
12 o'clock	Percent Accuracy	-0.18	0.48	0.60	0.258
	Percent Chosen		-0.99	0.56	
1	Percent Accuracy	0.19	0.43	0.89	0.250
	Percent Chosen		-0.67	0.58	
2	Percent Accuracy	0.01	0.15	0.89	0.936
	Percent Chosen		-0.11	0.81	
3	Percent Accuracy	0.04	0.15	0.89	0.788
	Percent Chosen		0.05	0.78	
4	Percent Accuracy	0.04	-0.28	1.09	0.790
	Percent Chosen		0.94	0.93	
5	Percent Accuracy	-0.27	-0.58	1.18	0.087
	Percent Chosen		0.68	0.99	
6	Percent Accuracy	0.05	0.13	0.92	0.761
	Percent Chosen		-0.34	0.96	
7	Percent Accuracy	0.09	0.13	0.92	0.584
	Percent Chosen		-0.34	0.96	
8	Percent Accuracy	-0.19	-0.31	1.08	0.240
	Percent Chosen		0.51	1.08	
9	Percent Accuracy	0.09	-0.20	1.10	0.600
	Percent Chosen		-0.15	0.79	
10	Percent Accuracy	0.25	-0.12	1.10	0.114
	Percent Chosen		0.10	1.06	
11	Percent Accuracy	-0.16	0.23	0.82	0.334
	Percent Chosen		-0.35	0.71	

PHASE I: MAIN EXPERIMENT OBJECTIVE MEASURES DISCUSSION AND CONCLUSIONS

Objective Measures Discussion

The hypothesis that absolute localization performance would differ significantly with training progression according to training strategy was not supported. However, the results of the mixed-factors ANOVAs demonstrated that localization performance improved significantly from pretest to LU8 and pretest to LU4. The results demonstrate that regardless of training strategy, performance improved significantly when compared to pretest performance. The large effect sizes of η_p^2=0.8 associated with the LU independent variable support that a large portion of the variance was attributed to training progression. Unfortunately, neither using a difference score, thereby removing LU as a within-subjects variable, nor relaxing the scoring criteria to ballpark yielded a greater and significant effect for training strategy. The non-significant findings from LU4 to LU8, especially for the stage of training independent variable, supported halving the training time, i.e., using only LU0 (pretest) to LU4. With an hour-long training target, this analysis supported this objective for Soldier training.

With regards to the independent variable of training strategy, while statistical significance was not achieved in the ANOVA analysis, the “choose” condition demonstrated an apparent advantage in magnitude. Additionally, slope analyses from pretest to LU4 demonstrated a significant difference in the change absolute score between the highest scoring condition, choose, and the lowest one, DRILCOM. The significant negative correlation between accuracy and percentage of time the speaker was chosen for additional practice supports that participants were aware of their problem locations. Correlations for each location using the same variables were not significant, however, trends in Figures 48 and 49 revealed that participants practiced more frequently with locations where they had difficulty. Of note, participants had the greatest difficulty discerning between the four and five o’clock positions. The radial plot in Figure 50 suggests that the left-sided response bias may be attributable to room effects as the room was not perfectly symmetrical (for reasons stated earlier). In Phase III, which is discussed in a subsequent chapter, an opposite response bias was recorded in a different room for a similar localization task. The choose strategy implemented a strategy aligned with Wolfle’s (1946) training theory of active participation. Specifically, the choose condition showed an advantage on the dependent measures of absolute score, ballpark score, and response time from pretest to LU4. Given the impact of active participation, albeit non-significant, future investigations should focus on this component as an incentive to shortening training. For example, perhaps the participant could initiate a user-select subunit prior to a random subunit, and then followed by another user-selected number of presentations for the choose-to-practice training strategy. For the purpose of experimental control, the presentation number in the choose trial was controlled. Future investigations of improving training strategies should quantify the effect of using a participant-driven number of user-select sessions versus an experimenter-driven one.

In addition to demonstrating significant improvements with training progression, the results also demonstrated an inverse relationship between accuracy improvement and response time. The ANOVA analyses supported that the response time improved significantly from pretest to LU8 and LU4 to LU8. The regression analyses supported that the change in score from pretest to LU8 and LU4 to LU8 predicted the change in response time. Surprisingly, the response time improved, or shortened, significantly from LU4 to LU8. The investigator postulated that the participants would become bored, and performance would decrease with response time. However, the response time results supported that skill acquisition continued to occur until LU8. This finding was evidenced by the significant decrease in response time and medium to large effect sizes in the analyses. Of note, the practical implications of improving response time in a localization task by tenths of a second remains unknown. For the purposes of this experiment, the measure of response time served to indirectly measure underlying cognitive processes in the context of performance. A shorter response time for a localization task in a combat scenario is an obvious desirable outcome. However, how much of a change in response time results in a functional impact remains unknown, especially since response time incorporates the motor response. In other words, change in response time may have been partially or completely due to an improved motor response time without improving localization skill acquisition.

A limitation to this study is that localization learning decay was not assessed, because that requires a longitudinal assessment after training has been completed. While the long-term effects of the training have yet to be evaluated, the results from this study can be employed to converge on a solution to prevent attrition of auditory localization skills. Another limitation, which was addressed in the follow-on study of Phase III, is the external validity of training in a laboratory environment. Service Members often have more annual training requirements than are technically feasible in a year. To better ensure that the benefit of training exceeds the time and money invested in training, the impact on actual job performance should be quantified. Implications for a localization training protocol include: a more comprehensive ASA fitness-for-duty evaluation, a means to evaluate hearing protection effects on localization, and remediation of localization inaccuracies with and without hearing protection. Exactly how much localization accuracy is needed for Service Members to maintain auditory situation awareness remains unknown, and it also will vary among different U.S. Military occupational specialties (MOS’s) and for different tactical missions. However, measuring the impact of localization loss by using the DRILCOM testing protocol and apparatus and associating lab performance with actual duty-related tasks could be used to better predict acuity requirements. Subsequent experimental investigations could focus on assisting those with localization cue loss, either from hearing loss or hearing protection, to remediate localization errors and improve functioning.

The ad hoc analysis of performance between pretest and LU4 and between LU4 and LU8 yielded useful data for the follow-on study. Specifically, this finding resulted in decreasing the training time by about half in the follow-on studies of Phases II and III whereby a portable system was developed and field tested, respectively. However, given that subsequent studies incorporated advanced hearing protection devices and TCAPS, it was left open that four learning units may have insufficiently captured the range of training effects.

Objective Measures Conclusions

The findings of the main experiment supported that across training strategies, absolute and ballpark scores improved significantly from pretest to LU4 and pretest to LU8. However, results did not support that a significant change in performance existed from LU4 to LU8. Therefore, reducing the number of LUs for the follow-on studies in Phase II and III of the overarching experiment was not expected to adversely impinge upon training ability. The limitation to shortening the design is the unknown degree of training transfer to different types of untrained stimuli. Given the evidence in the pilot studies, training on the dissonant signal transferred to three of the four untrained stimuli, but that effect was only measured after LU8, as opposed to the four LUs used in Phase II and III. Given that previous DRILCOM studies demonstrated diminishing returns after eight LUs, the finding that the diminishing returns existed after only four LUs was unexpected. Therefore, the decision to deliver untrained stimuli tests was rendered a priori. Additionally, the response time grew shorter with improvements in absolute score, but the improvement may have been at least partially attributable to an improved motor response. This result may be consistent with training benefits beyond LU4, but the time investment of an additional hour for a non-significant improvement in absolute score does not support the use of additional LUs.

The choose strategy resulted in the highest mean absolute, ballpark, and absolute difference score (LU4- pretest) localization accuracy performance. However, results did not show significant differences among training strategies. Mean differences in performance among training strategies, especially between the adaptive + choose condition, were not significant. Nevertheless, the number of trials, and therefore the time to complete training, for the choose condition were fewer, thus rendering the training more efficient with that strategy.

Main Experiment Subjective Performance Data Reduction

At the conclusion of the pretest subunit in LU1 and LU8, participants completed a questionnaire. The questionnaires differed according to training strategy condition and are provided in Appendix I. One additional question was included in LU8 stating, “Please rate your reaction time from before to after all the training that you have received so far.” For each training condition (adaptive, choose, adaptive + choose, and DRILCOM), Wilcoxon signed-rank tests compared differences in response ratings from LU1 to LU8 for questions one through five and eight. Questions one through five and eight asked the participant to rate the following items on a seven-point Likert scale: 1) confidence, 2) difficulty, 3) ability, 4) usefulness, 5) confidence in abilities, and 8) appropriateness of time allotted.

For background on the analysis technique applied, the Wilcoxon signed-ranks test serves as the non-parametric equivalent of the dependent samples *t*-tests. Specifically, the test evaluates the direction and extent of differences of paired scores (Portney & Watkins, 2009). First, difference scores between the pair are computed. Next, signs are assigned to the ranks. In the case of a tie, a mean rank is computed and assigned. A significant Wilcoxon signed-rank test results when the number of positive and negative ranks is significantly different (Portney & Watkins, 2009). In this study, a significant finding reflected a significant difference between ratings for LU1 compared to LU8. An α level of 0.05 was adopted for all non-parametric tests.

To examine the effects of training condition on ratings at LU1 and LU8, Kruskal-Wallis analyses were conducted at each LU. The Kruskal-Wallis analysis of variance by ranks was selected because it is the non-parametric equivalent of a one-way ANOVA, and the rating data were treated as ordinal, even though the scales (Appendix I) were of equal-appearing intervals. The test statistic, *H*, is computed by first aggregating all of the data from the groups and assigning ranks in ascending order (Portney & Watkins, 2009). Then, ranks are separated according to group and summed (Portney & Watkins, 2009). The *H* statistic is computed using the total number of cases, *N*, the number of cases in each sample, *n,* and the sum of ranks for each sample, *R* (Portney & Watkins, 2009). The formula is as follows:

$$H = \frac{12}{N(N+1)} \Sigma \frac{R^2}{n} - 3(N+1)$$

The *H* statistic is then compared to a critical value based on a chi-square distribution and *k*, or number of groups, minus 1. Mean ratings were also plotted for each training strategy for each question to observe rating response trends. Kruskal-Wallis analyses were also conducted for questions one through five and eight using the difference score between LUs (LU8 – LU1) to evaluate significantly different changes in ratings among training strategies. Significant results were followed up with Mann-Whitney *U* tests, the non-parametric equivalent of independent *t*-tests. The procedure for Mann-Whitney *U* is similar to the Kruskal-Wallis in that scores are initially ranked in ascending order (Portney & Watkins, 2009). The ranks within each group are then summed. Sufficiently large differences between the groups supports rejecting the null hypothesis that the ratings between groups are equal.

Main Experiment Subjective Performance Results

Question 1. Perceived Confidence

Participants were asked to respond to the following: Rate how **confident you are** in your ability to locate sounds from 1 (no confidence) to 7 (extremely confident). Wilcoxon signed-ranks test were performed to conduct a within-subjects comparison of ratings at LU1 and LU8 for each training condition (e.g. for the choose condition, comparing scores at LU1 versus LU8). Wilcoxon results showed that only the choose condition (Z=-2.32, p=0.020) resulted in significantly higher ratings of confidence from LU1 to LU8 (Table 56). To compare differences in ratings for each training strategy, a between-subjects analysis, Kruskal-Wallis analysis of the variance by ranks, were conducted at LU1 and LU8 (e.g., at LU1 comparing ratings for adaptive, choose, DRILCOM, and adaptive + choose, and the same for LU8). No significant differences were found among training strategies for analysis conducted at LU1 or LU8 (Table 57). Mean ratings for each training strategy for LU1 and LU8 are shown in Figure 51. Means are provided in lieu of the median scores typically furnished in non-parametric analyses as changes in mean scores better capture how the whole sample changed versus how the middle scores changed. No significant differences existed among training strategies, but mean ratings showed an improvement in confidence. To examine the effect of training strategy on a change in confidence ratings, rating difference scores were calculated (LU8–LU1). Kruskal-Wallis analysis was conducted to compare the effect of training strategy on the dependent measure of a difference score in ratings from LU1 to LU8 (LU8-LU1). Results were not significant (H[3]=1.52, p=0.667) and means are provided in Table 58 and Figure 52. Spearman's correlations were calculated to determine if a relationship existed between confidence ratings and absolute score. Spearman's correlations were significant for LU1, r_s=0.52, p=0.001 and LU8, r_s=0.49, p=0.001. Results supported that a positive significant relationship existed between confidence and absolute score.

Table . Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy for Question 1, Perceived Confidence (bolded text in the table indicates a significant test result at $p<0.05$.)

Listening Condition			LU1 (*n*=10)		LU8 (*n*=10)	
	Z	*p*	*M*	*SD*	*M*	*SD*
Adaptive	-1.73	0.084	5.40	1.03	6.10	0.88
Choose	**-2.32**	**0.020**	**4.80**	**0.92**	**5.90**	**0.74**
Adaptive + choose	-2.11	0.035	5.20	0.63	5.90	0.99
DRILCOM	-2.11	0.035	5.20	1.03	5.90	0.74

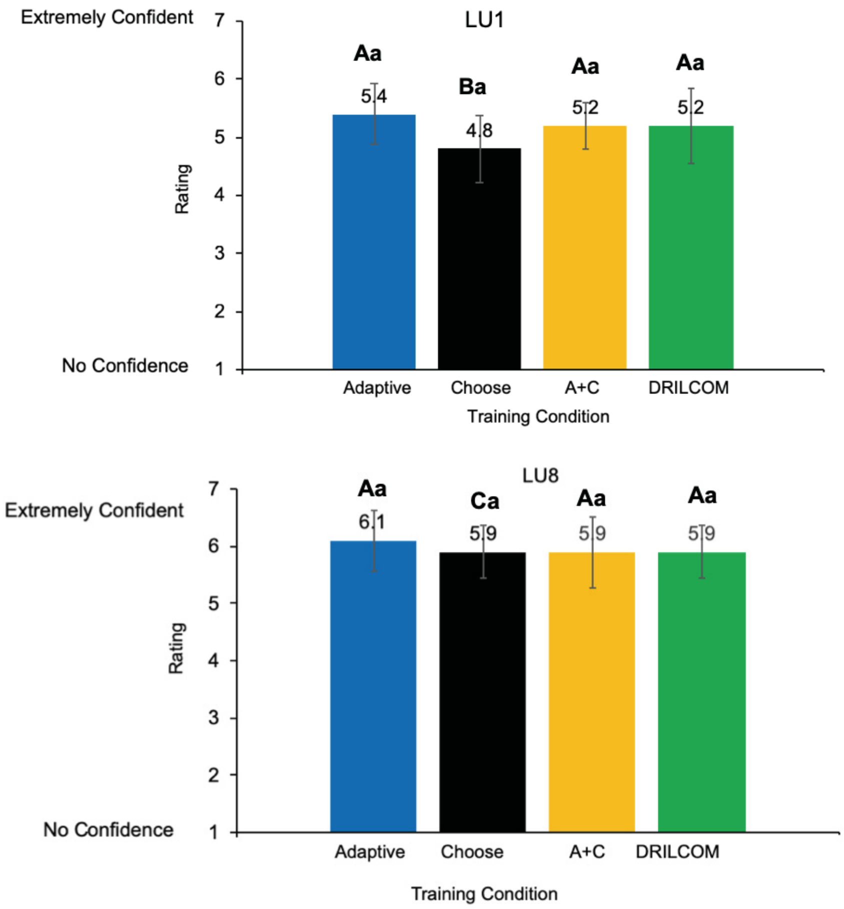


Figure . Mean ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 1, Perceived Confidence. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Table . Kruskal-Wallis results and mean ranks comparing training strategies at LU1 and LU8 for Question 1, Perceived Confidence (bolded text in the table indicates a significant test result at *p*<0.05).

	H	n	*df*	*p*
LU1	2.44	40	3	0.487
LU8	0.83	40	3	0.842

Mean ranks at LU1 and LU8 across all conditions for Question 1, Perceived Confidence (bolded text in the table indicates a significant test result at *p*<0.05).

Condition	Rank	
	LU1	LU8
Adaptive	24.10	23.10
Adaptive + choose	20.80	19.30
DRILCOM	20.65	20.30
Choose	16.45	19.30

Table . Kruskal-Wallis mean ranks results comparing ratings among training strategies using the difference score (LU8-LU1) for Question 1, Perceived Confidence. (bolded text in the table indicates a significant test result at *p*<0.05).

Condition	Rank
Choose	24.05
Adaptive + choose	19.75
DRILCOM	19.75
Adaptive	18.45

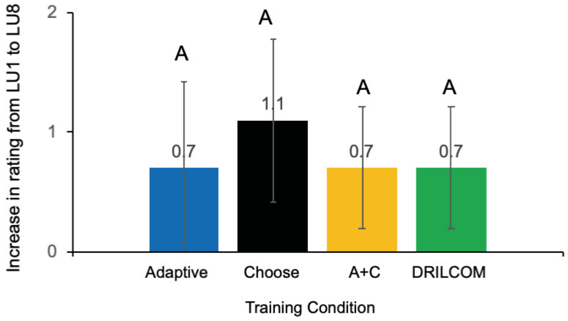


Figure . Mean changes in ratings from LU1 and LU8 (LU8-LU1) with confidence intervals for each training strategy for Question 1, Perceived Confidence. Mean ratings with a different upper-case letter between Training Condition are significantly different at *p*<0.05 using a Kruskal-Wallis test.

Question 2. Perceived Difficulty

Participants were asked to respond to the following: Rate how **difficult** it was to judge the location of the sounds in this session from 1 (extremely difficult) to 7 (extremely easy). Wilcoxon results showed no significant differences between LU1 and LU8 for each training condition (Table 59). Mean difficulty ratings for each LU are provided in Figure 42. Kruskal-Wallis analyses conducted for LU1 and LU8, comparing all training conditions were not significant at each training stage (Table 60). Kruskal-Wallis analysis was conducted using the dependent measure of a difference score in ratings from LU1 to LU8. Results were not significant ($H[3]=2.98$, $p = 0.394$) and means are provided in Table 61 and Figure 43, respectively.

Table . Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy for Question 2, Perceived Difficulty (bolded text in the table indicates a significant test result at $p<0.05$.)

Listening Condition			LU1 (*n*=10)		LU8 (*n*=10)	
	Z	*p*	*M*	*SD*	*M*	*SD*
Adaptive	-1.59	0.112	4.40	0.97	5.10	1.29
Choose	-0.35	0.725	3.90	0.99	4.10	1.69
Adaptive + choose	-0.72	0.472	4.10	1.29	3.80	1.48
DRILCOM	-0.85	0.393	4.30	1.34	4.80	1.14

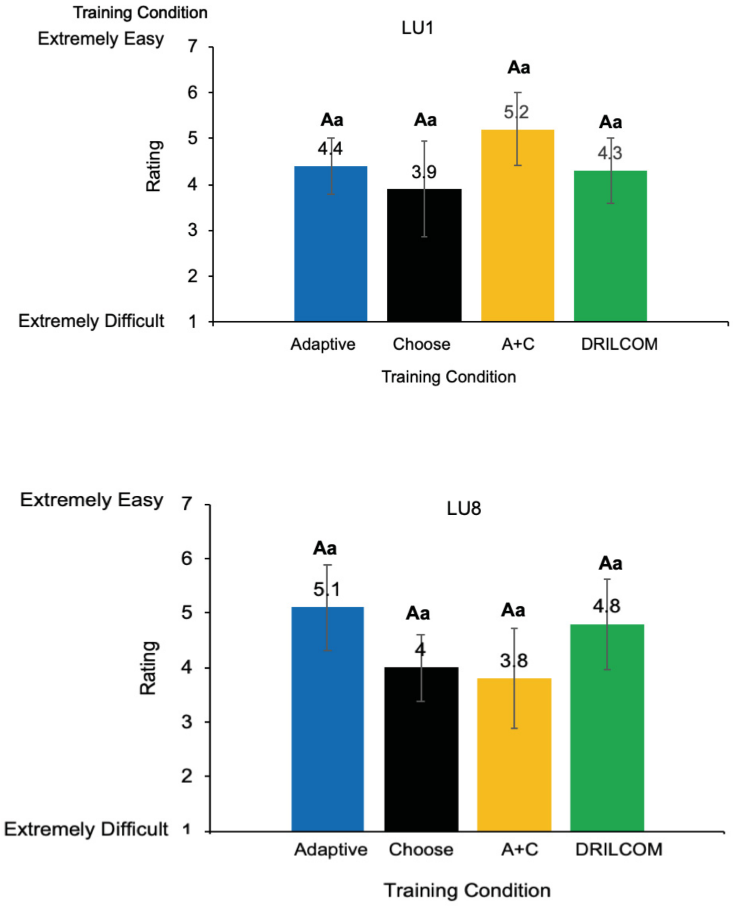


Figure . Mean ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 2, Perceived Difficulty. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Table . Kruskal-Wallis results comparing all training strategies at LU1 and LU8 for Question 2, Perceived Difficulty (bolded text in the table indicates a significant test result at $p<0.05$).

	H	n	*df*	*p*
LU1	1.24	40	3	0.745
LU8	5.52	40	3	0.138

Mean ranks at LU1 and LU8 across all conditions (bolded text in the table indicates a significant test result at $p<0.05$).

Condition	Rank	
	LU1	LU8
Adaptive	22.75	25.80
Adaptive + choose	21.90	23.40
DRILCOM	19.70	17.20
Choose	17.65	15.60

Table . Kruskal-Wallis mean ranks results comparing ratings for training strategies using the difference score in ratings (LU8-LU1) for Question 2, Perceived Difficulty (bolded text in the table indicates a significant test result at $p<0.05$).

Condition	Rank
Choose	24.70
Adaptive + choose	21.95
DRILCOM	18.95
Adaptive	16.40

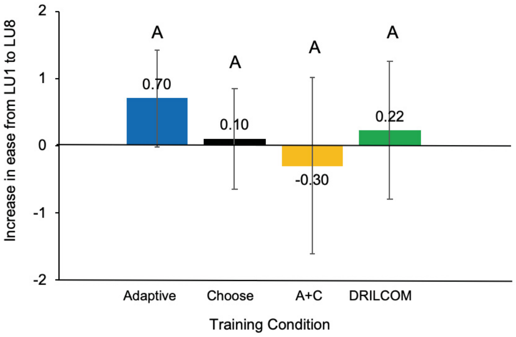


Figure . Mean changes in ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 2, Perceived Difficulty. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Question 3. Perceived Ability

Participants were asked to respond to the following: Rate how much you feel your **ability** to determine sound location improved as a result of training in **this session** from 1 (no improvement) to 7 (maximum improvement). Wilcoxon results showed significant differences between LU1 and LU8 for the adaptive (Z=-2.11, p=0.035) and the choose (Z=-2.26, p=0.024) conditions (Table 62). Mean ratings of perceived ability for each training strategy for LU1 and LU8 are shown in Figure 44. Kruskal-Wallis analysis conducted for LU1 and LU8, comparing all training conditions was not significant (Table 63). Kruskal-Wallis analysis was conducted using the dependent measure of a difference score in ratings from LU1 to LU8. Results were not significant (H[3]= 0.11, p = 0.990) and means are provided in Table 64 and Figure 45, respectively.

Table . Wilcoxon results comparing ability ratings for LU1 versus LU8 for each training strategy group for Question 3, Perceived Ability (bolded text in the table indicates a significant test result at p<0.05.)

Listening Condition			LU1 (*n*=10)		LU8 (*n*=10)	
	Z	*p*	*M*	*SD*	*M*	*SD*
Adaptive	**-2.11**	**0.035**	**5.30**	**0.82**	**6.00**	**0.82**
Choose	**-2.26**	**0.024**	**5.10**	**1.10**	**6.00**	**0.67**
Adaptive + choose	-2.12	0.034	5.30	1.41	6.30	0.48
DRILCOM	-1.99	0.046	4.80	1.13	5.60	1.07

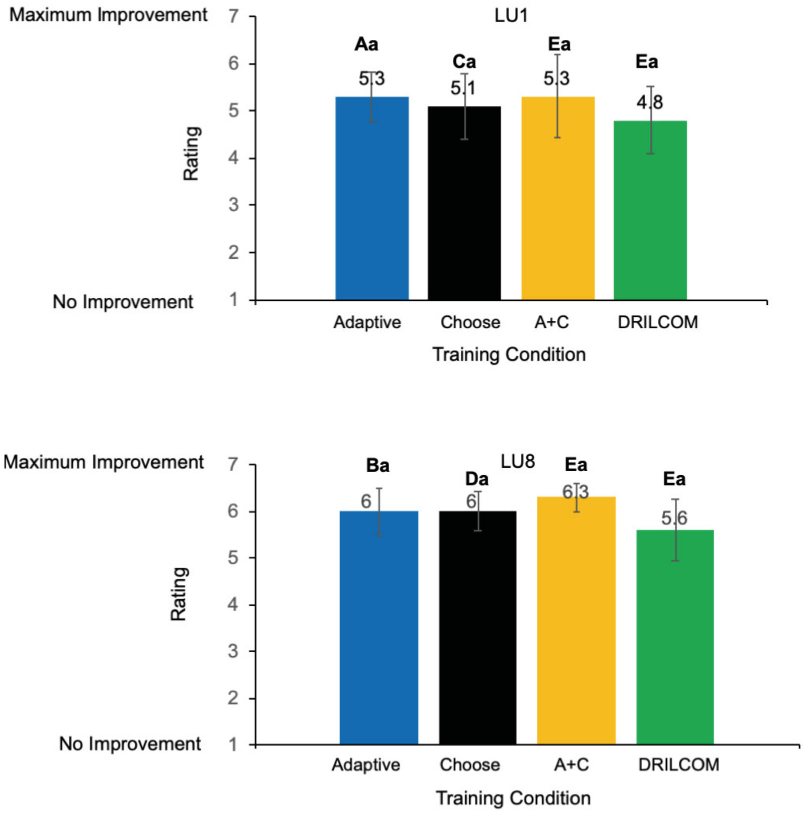


Figure . Mean response ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 3, Perceived Ability. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Table . Kruskal-Wallis results comparing ratings of ability for all training strategies at LU1 and LU8 for Question 3, Perceived Ability (bolded text in the table indicates a significant test result at $p<0.05$).

	H	n	*df*	*p*
LU1	2.24	40	3	0.524
LU8	3.55	40	3	0.314

Table . Kruskal-Wallis mean ranks results comparing training strategies using the difference score in ratings (LU8-LU1) for Question 3, Perceived Ability (bolded text in the table indicates a significant test result at $p<0.05$).

Condition	Rank
Choose	21.30
Adaptive + choose	20.60
DRILCOM	20.45
Adaptive	19.65

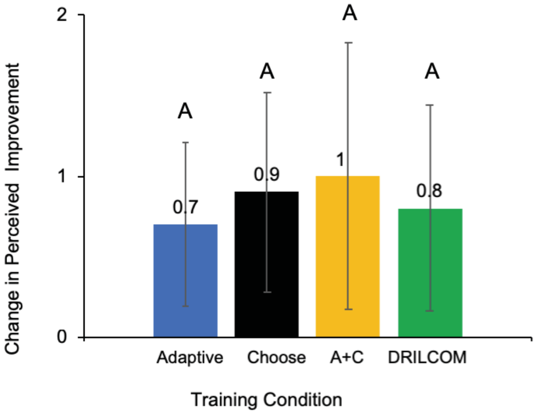


Figure . Mean changes in difficulty ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 3, Perceived Ability. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Question 4. Training Usefulness

Participants were asked to respond to the following: If you had a job that would require you to determine sound direction, how **useful** would you find this training from 1 (not useful) to 7 (extremely useful). Wilcoxon results showed no significant difference on ratings of usefulness for each condition between LU1 and LU8, (Table 65). Mean ratings of perceived usefulness for each training strategy for LU1 and LU8 are shown in Figure 57. Kruskal-Wallis analysis conducted for LU1 and LU8, comparing all training conditions was not significant (Table 65). Kruskal-Wallis analysis was conducted using the dependent measure of a difference score in ratings from LU1 to LU8. Results were not significant ($H[3]=1.01$, $p=0.800$) and means are provided in Table 67 and Figure 58, respectively.

Table . Wilcoxon results comparing confidence ratings for LU1 versus LU8 for each training strategy group for Question 4, Perceived Usefulness (bolded text in the table indicates a significant test result at $p<0.05$).

Listening Condition			LU1 (*n*=10)		LU8 (*n*=10)	
	Z	*p*	*M*	*SD*	*M*	*SD*
Adaptive	-1.63	0.102	6.20	0.63	6.00	0.82
Choose	-1.89	0.059	5.70	1.25	6.20	0.79
Adaptive + choose	-1.34	0.180	6.00	1.05	6.30	0.82
DRILCOM	-2.06	0.039	5.20	1.62	6.00	0.82

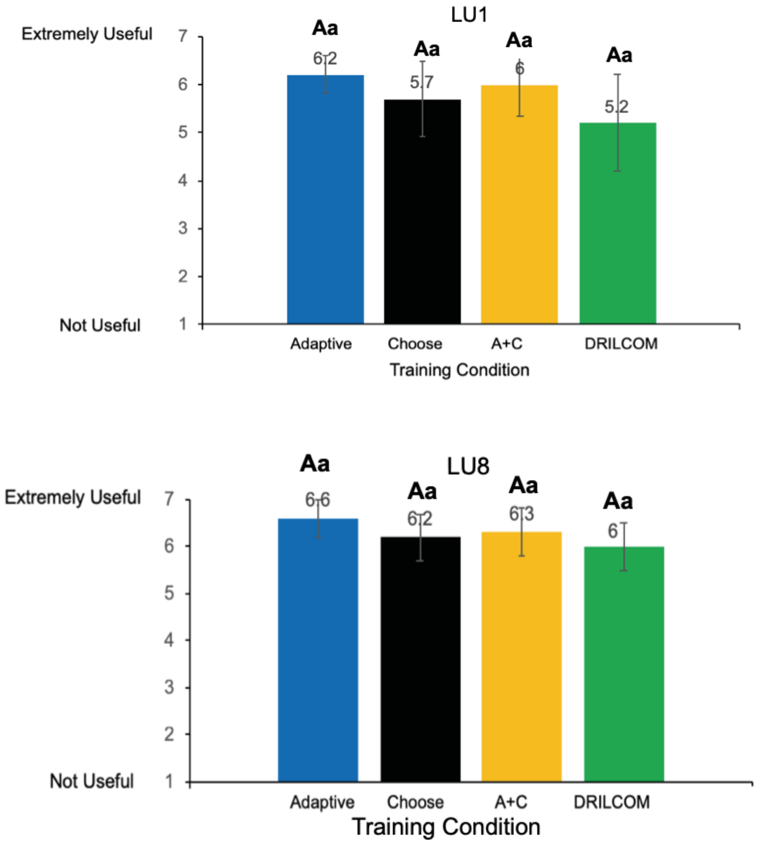


Figure . Mean response ratings of usefulness at LU1 and LU8 with confidence intervals for each training strategy for Question 4, Perceived Usefulness. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Table . Kruskal-Wallis results comparing ratings of usefulness for all training strategies at LU1 and LU8 for Question 4, Perceived Usefulness (bolded text in the table indicates a significant test result at $p<0.05$).

	H	*n*	*df*	*p*
LU1	2.40	40	3	0.494
LU8	3.27	40	3	0.352

Table . Kruskal-Wallis results comparing ratings of usefulness for all training strategies at LU1 and LU8 for Question 4, Perceived Usefulness (bolded text in the table indicates a significant test result at $p<0.05$).

Condition	Rank
Choose	23.10
Adaptive + choose	20.35
DRILCOM	20.15
Adaptive	18.40

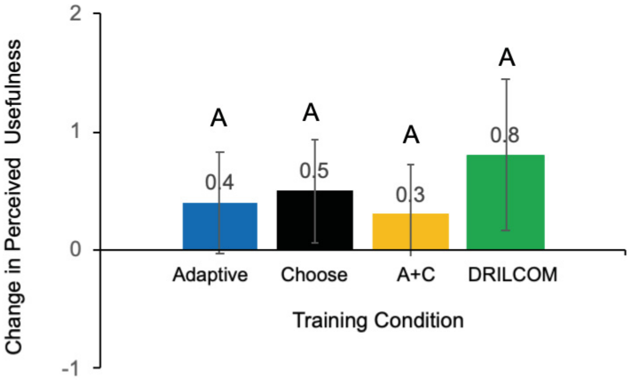


Figure . Mean changes in usefulness ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 4, Perceived Usefulness. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Question 5. Confidence in abilities

Participants were asked to respond to the following: Please rate how **confident you were in your ability** to locate sounds at the end of this session from 1 (no confidence) to 7 (extremely confident). Wilcoxon results showed a significant difference on ratings of confidence in abilities for only the choose condition ($Z=-2.06$, $p=0.039$) between LU1 and LU8 (Table 68). Mean ratings of confidence in abilities for each training strategy for LU1 and LU8 are shown in Figure

59. Kruskal-Wallis analysis conducted for LU1 and LU8, comparing all training conditions were not significant (Table 69). Kruskal-Wallis analysis was conducted using the dependent measure of a difference score in ratings from LU1 to LU8 on ratings of confidence in abilities. Results were not significant ($H[3]= 0.64$, $p=0.800$) and means are provided in Table 70 and Figure 49, respectively.

Table . Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy group for Question 5, Confidence in Abilities (bolded text in the table indicates a significant test result at $p<0.05$).

Listening Condition			LU1 (n=10)		LU8 (n=10)	
	Z	p	M	SD	M	SD
Adaptive	-1.28	0.202	5.40	0.84	6.00	1.15
Choose	**-2.06**	**0.039**	**5.20**	**1.03**	**6.00**	**0.67**
Adaptive + choose	-2.00	0.046	5.85	0.75	6.25	0.42
DRILCOM	-2.24	0.025	5.50	0.85	6.00	0.67

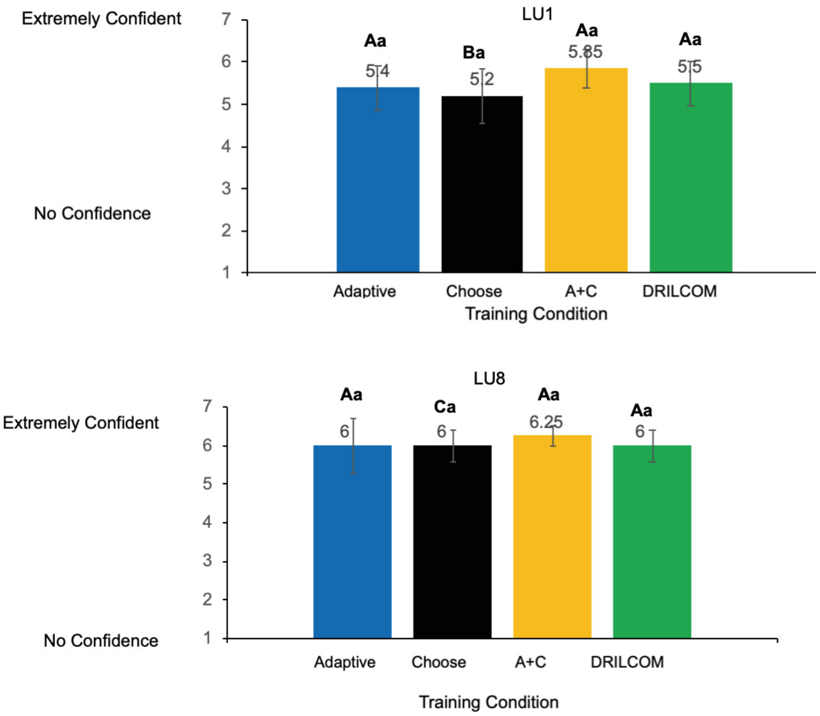


Figure . Mean response ratings at LU1 and LU8 with confidence intervals for each training strategy, Question 5, Confidence in Abilities. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at *p*<0.05 using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at *p*<0.05 using a Kruskal-Wallis test.

Table . Kruskal-Wallis results comparing ratings for all training strategies at LU1 and LU8 for Question 5, Confidence in Abilities (bolded text in the table indicates a significant test result at *p*<0.05).

	H	*n*	*df*	*p*
LU1	1.82	40	3	0.611
LU8	1.35	40	3	0.718

Table . Kruskal-Wallis mean ranks results comparing ratings among training strategies using the difference score in ratings (LU8-LU1) for Question 5, Confidence in Abilities (bolded text in the table indicates a significant test result at $p<0.05$).

Condition	Rank
Adaptive + choose	24.25
DRILCOM	20.05
Adaptive	19.80
Choose	17.90

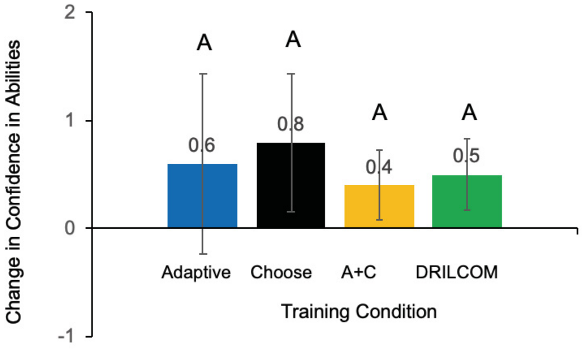


Figure . Mean changes in ratings from LU1 to LU8 with confidence intervals for each training strategy for Question 5, Confidence in Abilities. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Question 6. Ability to Localize from Before to After Training

Participants were asked to respond to the following: Please rate your **ability** to locate the sound's direction from **before to after all the training that you have received so far** from 1 (performance was worse) to 7 (performance improved). Kruskal-Wallis analysis comparing all training conditions at LU8 showed no significant difference among training strategies ($H[3]=4.89$, $p= 0.180$). Mean ratings of confidence are shown in Figure 61. Mean ranks are provided in Table 71. Spearman's correlation, one-tailed, was conducted to evaluate a relationship between improvement in absolute score and higher ratings on changes in abilities. Results were significant, $r_s=0.44$, $p=0.002$.

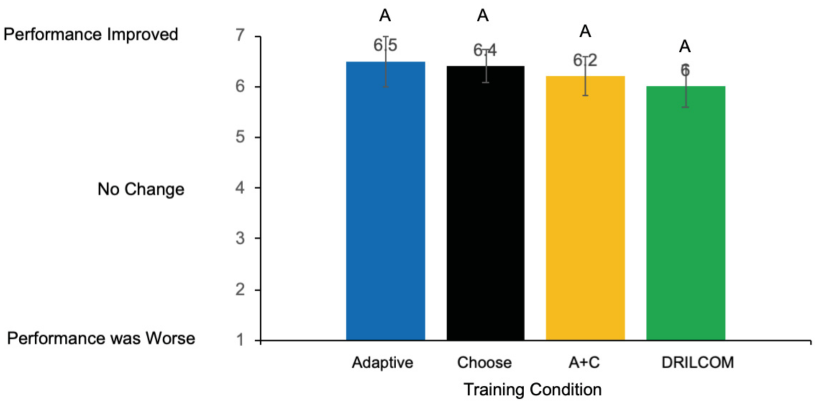


Figure . Mean response ratings for each training strategy for Question 6, Ability to Localize from Before to After Training. Mean Training Condition ratings with a different upper-case letter within the LU1 and within the LU8 graphs are significantly different at *p*<0.05 using a Kruskal-Wallis test.

Table . Table 70. Mean ranks for perceived ability to localize from before to after the training for Question 6, Ability to Localize from Before to After Training (bolded text in the table indicates a significant test result at *p*<0.05).

Condition	Rank
Adaptive	25.75
Choose	21.70
Adaptive + choose	18.75
DRILCOM	15.80

Question 7. Response Time from Before to After Training

Participants were asked to respond to the following: Please rate your **reaction time** in determining sound direction from **before to after all the training that** you have received **so far** from 1 (slower) to 7 (faster). Kruskal-Wallis analysis comparing all training conditions at LU8 showed a significant difference among training strategies (*H*[3]=8.82, *p*=0.032). Mean ratings of change in response time are shown in Figure 51. Mann-Whitney *U* testing for pairwise comparisons applied a Bonferroni-adjusted alpha level (α=0.5/6=0.008). Pairwise comparisons of each training strategy showed non-significant findings and are provided in Table 72.

Spearman's correlation evaluated the relationship between the perceived change in response time rating and objective improvement in response time. Results showed a weak positive relationship between the perceived and actual response time, r_s=0.28, p=0.41.

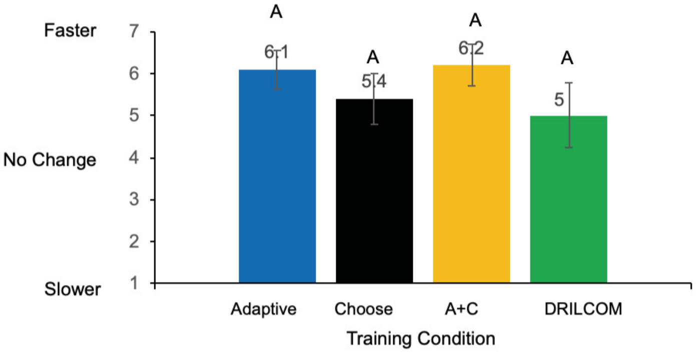


Figure . Mean response ratings with confidence intervals for each training strategy for Question 7, Response Time from Before to After Training. Mean Training Condition ratings with a different upper-case letter within the LU1 and within the LU8 graphs are significantly different at p<0.008 using Mann-Whitney *U* tests following a Kruskal-Wallis test.

Table . Mann-Whitney U pairwise comparisons for perceived response time ratings from after LU1 to after the training for Question 7, Response Time from Before to After Training (bolded text in the table indicates a significant test result at p<0.008).

Condition	*U*	*Median*	*Z*	*p*
Adaptive versus Choose	29.5	6.0 vs 5.0	-1.65	0.100
Choose versus Adaptive + choose	27.0	5.5 vs 6.0	-1.83	0.068
Adaptive versus Adaptive + choose	46.0	6.0 vs 6.0	-0.33	0.744
Adaptive versus DRILCOM	22.0	5.0 vs 6.0	-2.26	0.024
Choose versus DRILCOM	43.0	5.0 vs 6.0	-0.50	0.575
Adaptive+ choose versus DRILCOM	20.0	6.0 vs 6.0	-2.39	0.017

Question 8. Appropriateness of time allotted

Participants were asked to respond to the following: Please rate the extent that the amount of time allotted for training was appropriate for the amount of improvement using the anchors of 1 (needed more time), 4 (training took the right amount of time), and 7(training took too much time). Wilcoxon results showed a significant difference in ratings of appropriateness of time allotted for the adaptive (*Z*=-2.45, *p*=0.014) and choose conditions (*Z*=-2.32, *p*=0.026) between LU1 and LU8, (Table 73)**.** Mean ratings of appropriateness of time allotted for each training strategy for LU1 and LU8 are shown in Figure 63. Kruskal-Wallis analyses conducted for LU1 and LU8, comparing all training conditions were not significant (Table 74). Kruskal-Wallis analysis was conducted using the dependent measure of a difference score in ratings from LU1 to LU8 on ratings of appropriateness of time allotted. Results were not significant (*H*[3]=0.68, *p*=0.878) and means are provided in Table 75 and Figure 53, respectively.

Table . Wilcoxon results comparing ratings for LU1 versus LU8 for each training strategy group for Question 8, Appropriateness of Time Allotted (bolded text in the table indicates a significant test result at *p*<0.05).

Listening Condition			LU1 (*n*=10)		LU8 (*n*=10)	
	Z	*p*	*M*	*SD*	*M*	*SD*
Adaptive	**-2.45**	**0.014**	**3.7**	**0.48**	**4.20**	**0.63**
Choose	**-2.32**	**0.026**	**3.5**	**0.85**	**4.50**	**0.71**
Adaptive + choose	-2.53	0.011	3.80	0.42	4.60	0.70
DRILCOM	-2.26	0.024	3.70	0.48	4.60	0.97

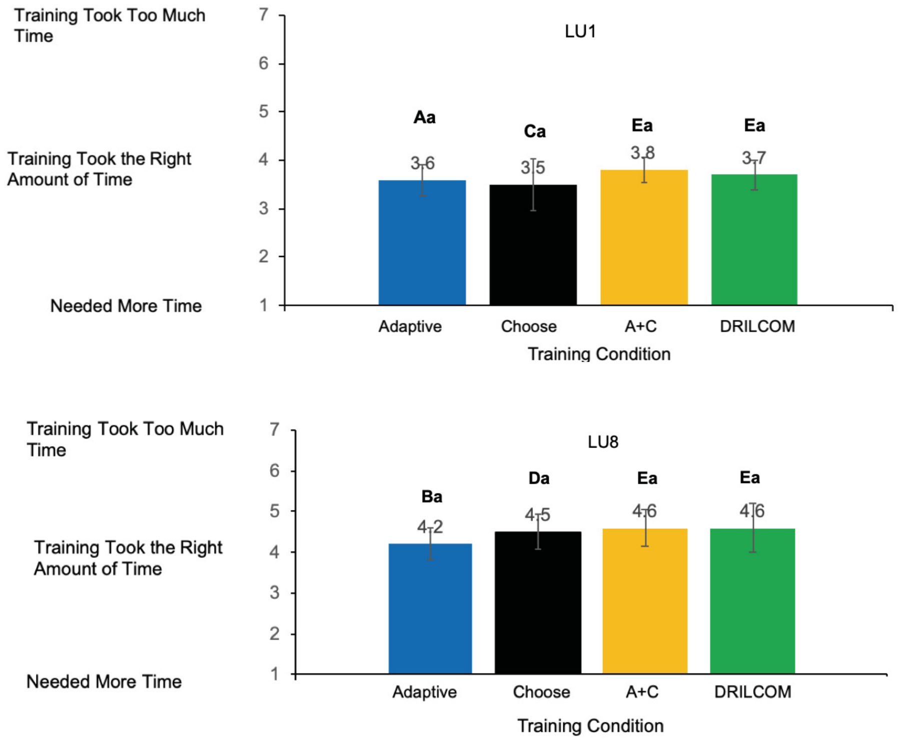


Figure . Mean response ratings at LU1 and LU8 with confidence intervals for each training strategy for Question 8, Appropriateness of Time Allotted. Mean ratings with a different upper-case letter between the LU1 and LU8 graphs, for a given Training Condition, are significantly different at $p<0.05$ using a Wilcoxon test. Mean Training Condition ratings with a different lower-case letter within the LU1 and within the LU8 graphs are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Table .Kruskal-Wallis results comparing ratings of usefulness for all training strategies at LU1 and LU8 for Question 8, Appropriateness of Time Allotted (bolded text in the table indicates a significant test result at $p<0.05$).

	H	*n*	*df*	*p*
LU1	0.95	40	3	0.814
LU8	1.56	40	3	0.669

Table . Kruskal-Wallis mean ranks results comparing ratings among training strategies using the difference score in ratings (LU8-LU1) for Question 8, Appropriateness of Time Allotted (bolded text in the table indicates a significant test result at $p<0.05$).

Condition	Rank
Choose	21.95
Adaptive + choose	21.05
DRILCOM	20.80
Adaptive	18.20

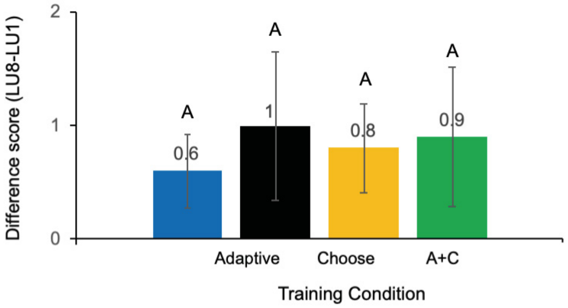


Figure . Mean changes in ratings from LU1 to LU8 with confidence intervals for each training strategy Question 8, Appropriateness of Time Allotted. Mean ratings with a different upper-case letter between Training Condition are significantly different at $p<0.05$ using a Kruskal-Wallis test.

Question 9-13. Subunit (Learning Unit) benefit

Participants were asked to respond to the following: On each of the following aspects of your training, rate how beneficial you felt the particular part of training was in learning where sounds were located using the anchors of 1 (no benefit) and 7 (maximum benefit). The question differed by condition as the subunit components varied accordingly (i.e., sequential, random, adaptive, user-selected practice, and test). As such, the objective was to analyze general trends to guide system design versus analytics with goal of harnessing statistical power. Trends are presented in Figures 65 through 69.

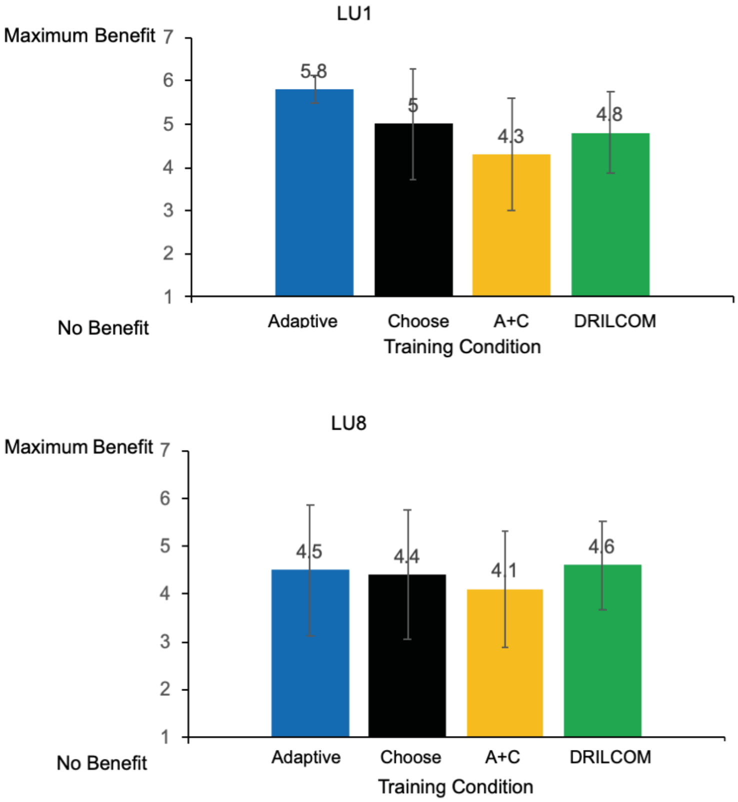


Figure . Mean ratings and confidence intervals for the **sequentia**l subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C.

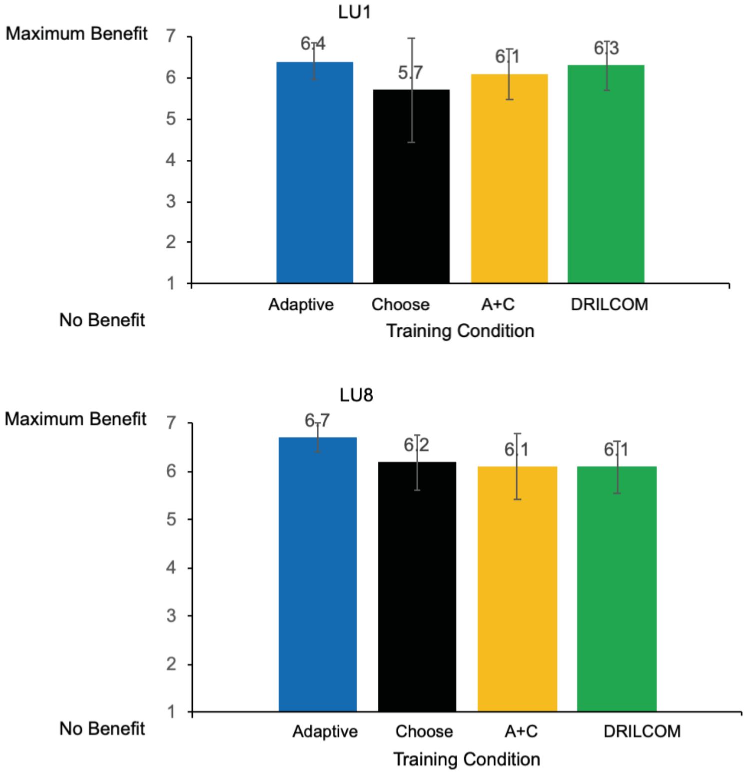


Figure . Mean ratings and confidence intervals for the **random** subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C.

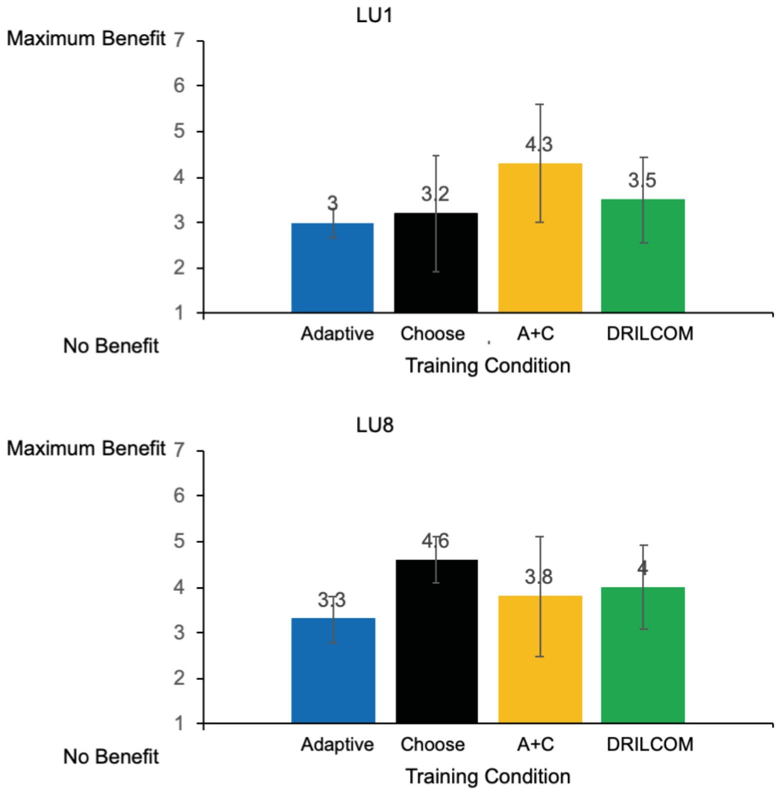


Figure . Mean ratings and confidence intervals for the **test** subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C.

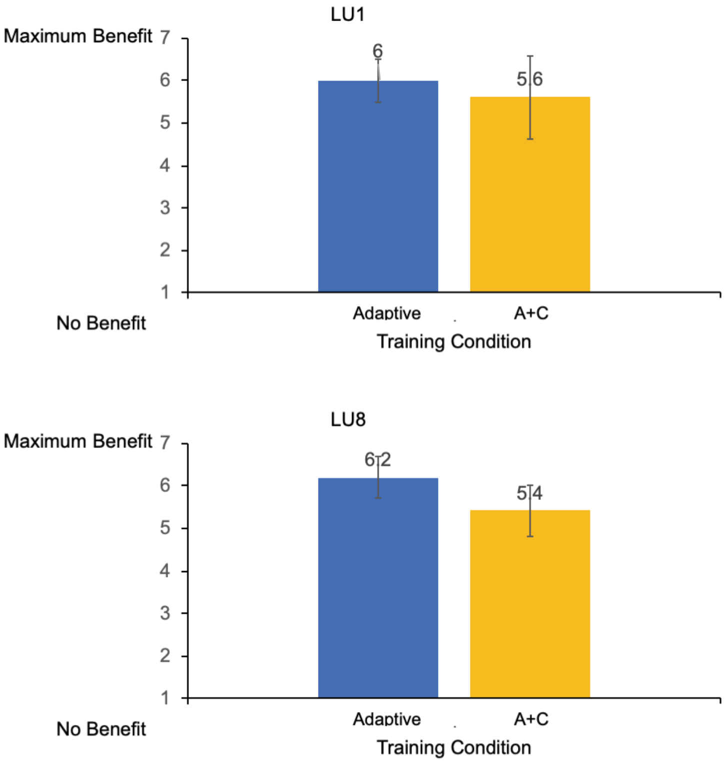


Figure . Mean ratings and confidence intervals for the **adaptive** subunit separated by training condition and training stage. The adaptive + choose condition is represented as A + C.

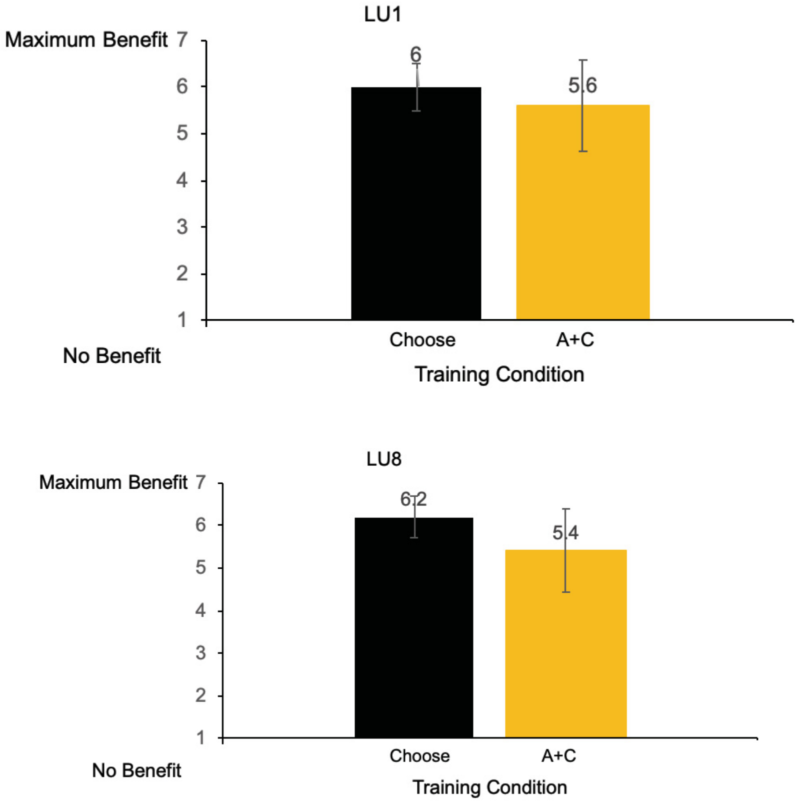


Figure . Mean ratings and confidence intervals for the **choose** subunit separated by training condition and training stage.

Question 13. *Subunits (Learning Units) participants identified where they felt disinterested/disengaged.*

Participants were asked to respond to the following: At any point in the training did you feel yourself becoming disinterested or disengaged? You can circle one option below or describe this time in the space provided. Depending on the training condition, different subunits were provided (i.e., sequential, random, adaptive, user-selected practice, and test). Therefore, not all participants had an opportunity to respond to each type of subunit. The percentage of participants reporting that they grew disinterested or disengaged are provided in Figures 70 and 71 for LU1 and LU8, respectively. The numerator was the number of times the subunit block was selected divided by the number of participants in each group (10).

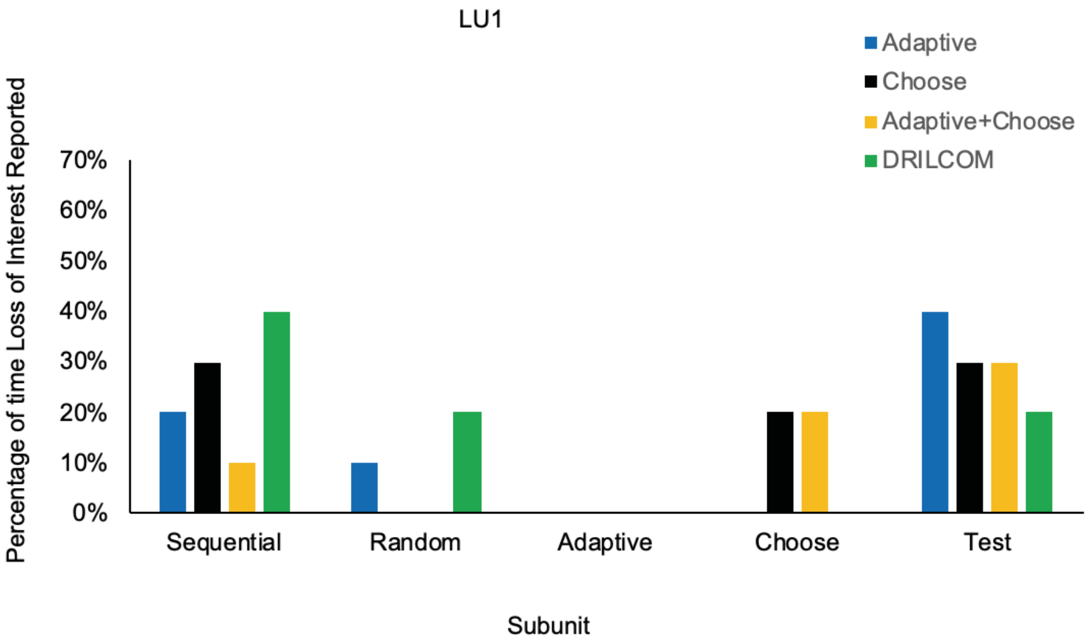


Figure . Percentage of loss of interest reported by subunit for LU1.

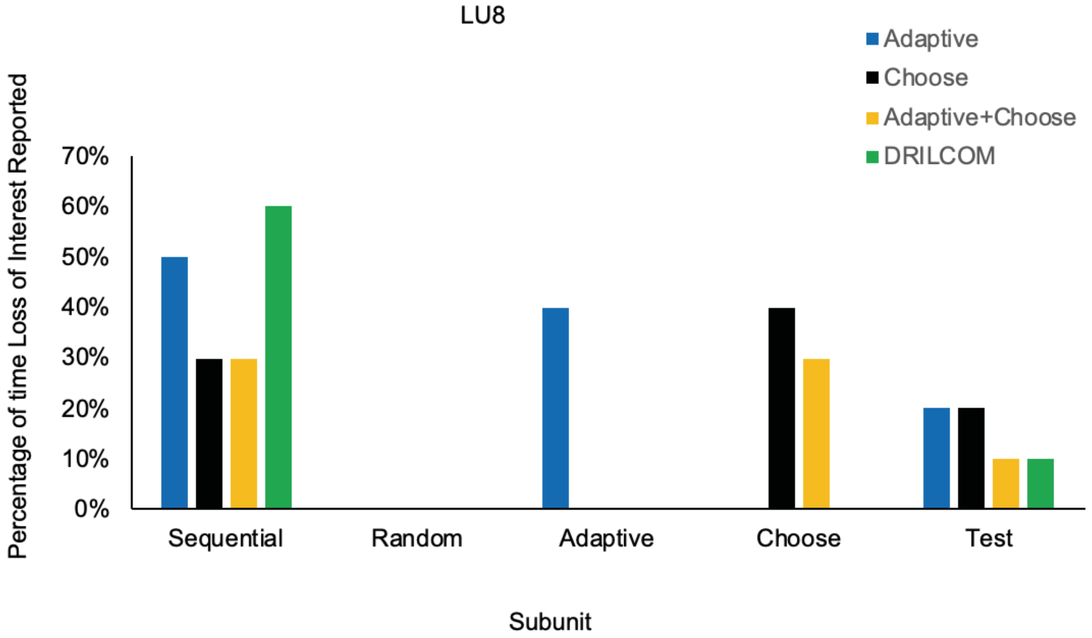


Figure . Percentage of loss of interest reported by subunit for LU8.

Main Experiment Subjective Measures Discussion

On ratings of confidence (question 1), comparison of LU1 to LU8 confidence ratings for each training strategy showed that only the choose condition resulted in a statistically-significant improvement. However, all strategies demonstrated an improvement in confidence ratings. The positive correlation between confidence ratings in absolute score was suggestive of participants' perception of performance aligning with the actual performance.

As ratings of confidence increased, ratings of difficulty (question 2) decreased for all strategies with the exception of the adaptive + choose condition. The ratings reflected that the training was rated as easier for the adaptive, choose, and DRILCOM conditions at LU8 versus LU1, with the choose condition demonstrating the greatest reduction in difficulty. The choose condition showed the highest ratings of confidence and reduction in difficulty, and perhaps giving the participant control over a portion of the training instilled this perception.

Similarly, ratings of perceived abilities, Question 3, improved in all conditions, but only significantly so for the choose and the adaptive conditions from LU1 to LU8. The expectation would be that an increase in confidence and a decrease in difficulty would align with ratings of ability, but the choose condition did not show significant changes in difficulty. Likewise, the adaptive + choose condition did not show significant changes in confidence or difficulty. However, for the question asking participants to rate confidence in abilities (Question 5), the choose condition was the only training strategy that showed a significant change from LU1 to LU8. This result aligned with the questions querying perceived confidence (Question 1) and perceived ability (Question 3) where the choose condition also resulted in significant improvement from LU1 to LU8. Objective measures showed that the change in localization accuracy from LU1 to LU4 showed significant improvement for the choose condition from LU1 to LU4. Subjective measures employed in this study have not been subjected to validation testing whereby shared variance among the responses could be reasonably inferred. However, these measures do enable the measurement of the change in perception over the course of the experiment. For the question that asked participants to rate their change in abilities from before to after training (Question 6), all analyses were not significant. Therefore, the item that pointedly asked about perceived change in ability did not align with objective performance or the significant change found for Question 3.

The questionnaire item regarding usefulness (Question 4) of the training showed neither significant change in ratings over the course of training nor significant differences among training strategies. Unfortunately, a limitation to this study is that participants were purposefully selected not to have any prior military or other experience with tactical localization. Therefore, this question was posed to those without any occupational or other experiential appreciation for the importance of localization.

For the question regarding response time (Question 7), despite non-significant differences among training strategies, all reported an improvement in reaction time. Of note, the questionnaire did not differentiate response time versus reaction time. Participants may have felt their motor response was faster or that they responded to the task faster. Therefore, the ratings may not have reflected the participants' perceptions that localization cue processing speed improved.

For the question regarding appropriateness of time allotted (Question 8), the adaptive and choose conditions resulted in significantly different ratings from LU1 to LU8. All training conditions resulted in mean ratings less than four, meaning that participants felt they needed more time, after LU1. However, all conditions resulted in mean ratings over four but less than five, consistent with training taking too much time. Surprisingly, the longest in duration, the adaptive + choose condition, did not result in significantly different perceptions of appropriateness of time allotted. In general, the ratings supported that the training time was insufficient for LU1 and too long by LU8. Unfortunately, ratings were not administered at LU4 to determine if appropriateness of time allotted would be rated more closely aligned with training took the right amount of time.

When examining the mean ratings of benefit for each subunit (Questions 9-13), two general trends emerge. At LU1, the mean ratings for the sequential and test subunits received the lowest ratings of perceived benefits, with the subunit receiving the lowest ratings. This trend carried over to LU8. Secondly, the sequential ratings of benefit decreased and the test ratings increased. The random subunit was the most highly rated in LU1. The adaptive and choose ratings were similar at LU1, but the adaptive subunit was rated higher than the choose condition by LU8, despite the choose training group out-performing the adaptive training group by LU8.

Generally, perceived benefit by subunit aligned with participants' ratings of where they felt themselves become disinterested or disengaged most often in the sequential and test subunit during LU1. In LU8, the random subunit was not selected, meaning participants did not find themselves growing disinterested. By LU8, the test subunit dropped in percentage of time participants grew disinterested or bored, but the sequential and test subunits increased. Therefore, by the time LU8 occurred, the random phase appeared to maintain participants' attention. The choose ratings demonstrated increased boredom despite better performance in training conditions that incorporated this subunit. Perhaps the task of the choose subunit would have been rated more highly if participants were allowed to choose the number of trials.

Main Experiment Subjective Measures Conclusions

Overall, subjective ratings showed that regardless of training strategy, confidence improved and that is of great importance for military applications. Confidence improved significantly for the choose condition, possibly because of the higher locus of control associated with this strategy. Ratings of confidence correlated with increases in objective measure of absolute score. However, as difficulty ratings increased, confidence ratings decreased. Perceived abilities also improved with training, as did perceived reaction time. As with measures of confidence, regardless of training strategy, all participants reported improvements in response time. By LU8, mean ratings showed that participants in all training conditions rated the training as taking just over the middle anchor of "took the right amount of time." By LU8 the sequential subunit was most frequently reported as the subunit where participants lost interest. The random subunit was least frequently reported as the subunit where participants lost interest.

Implications from Phase I

In order to discuss the combined implications of Phases I and III on military implementation, the findings from Phase I are recapitulated below. From Pilot 1 the main objectives and findings were as follows:

- **Pilot 1**
    - Objective 1: determine via a panel of SMEs which sounds were most important for ground combat Service Members to localize.
        - Feedback from seven soldiers showed the following results, with the corresponding number of respondents noted in parentheses: incoming mortar (7), speech (5), small arms or AK-47 specifically (4), and helicopter (3).

    - A panel of 10 Marines from the Gruntworks program responded collectively, but after Pilot 1 was conducted, with the following in rank order for combat environments: gunshot, explosion, vehicles, rotary wing [aircraft], and crack-thump [from a round].
    - Given the Army respondents' feedback, the following one-second test stimuli were used in Pilot 1: AK-47 three-round burst, Apache helicopter, spoken Arabic, and whistle from a ground burst simulator.
- Objective 2: determine the number of presentations in each training subunit that resulted in the highest localization accuracy in the shortest duration.
    - The slopes of the two-, three-, and four-presentation conditions were not significantly different from each other from LU1 to LU8.
    - The highest mean performance was obtained at LU5 in the three-presentations condition.
    - At LU8, the highest mean performances were very close for the three- (*M*=34.5) and four-presentation (*M*= 34.2) conditions.
- Objective 3: determine if training using a broadband stimulus (DRILCOM dissonant tonal complex) transferred to untrained, military relevant stimuli.
    - The four-presentation condition resulted in the highest mean performance for the AK-47, Apache, and Arabic stimuli. The four-presentation condition yielded the worst performance compared to the other presentation conditions.
    - Performance using the whistle tube simulator was significantly worse within every presentation condition (two, three, and four) compared to

every other test stimulus, thus no transfer-of-training from the broadband signal was found for the whistle tube signal (and only for this signal).

Pilot 2

Given the findings of Pilot 1, evidence supported that transfer-of-training occurred from the training broadband stimulus to the untrained, military-relevant stimuli of the AK-47, Apache, and Arabic stimuli. The training stimulus did not transfer to the whistle test stimulus regardless of the number of presentations. Given the highest training transfer results in Pilot 1, the four-presentation condition was incorporated into the next experiment, and the broadband stimulus was used for training. However, participants often remarked that the sequential portion of the training subunits seemed monotonous. In an effort to shorten the training sessions, Pilot 2 was conducted where participants only underwent four sequential presentations in LU1. All experimental conditions were held constant from Pilot 1 with the exception of the sequential subunit elimination from LU2-LU8. Pilot 2 results were compared to the four-presentation condition. Results showed a significantly steeper slope for Pilot 2 versus Pilot 1 from LU1-LU8. For the untrained stimuli, only the Arabic stimulus yielded significantly higher performance in Pilot 1 than Pilot 2.

Main experiment

The main experiment incorporated the four-presentation protocol from Pilot 1, but only one sequential presentation from LU2-LU8 was included. In this experiment, the training strategy was manipulated. The four training strategies employed were DRILCOM, adaptive, choose-to-practice, and adaptive + choose-to-practice. Twenty-four locations versus the 12 given in the pilot studies were included in order to avoid suspected ceiling effects in the pilot studies. The objectives were as follows:

- Objective 1: Determine the training strategy that yields the greatest localization accuracy in the fewest LUs.
    - The choose condition resulted in the highest mean difference score from pretest to LU8 and pretest to LU4.
    - Using the measure of difference score from pretest to LU8 and pretest to LU4, no significant differences existed among training strategies.
    - Comparing the slopes for each training condition from pretest to LU4 showed a significant difference in training strategies with the choose condition demonstrating the greatest slope (i.e., most rapid learning) and DRILCOM demonstrating the smallest slope.
    - Although not significant from pretest to LU4, the response time decreased as training progressed for all conditions except the adaptive training strategy
    - The choose, adaptive + choose, and DRILCOM strategies showed improvements in confidence from LU1 to LU8. These improvements did not exist when asked on a later question to rate the extent that they believed their localization ability to improve from before to after training.
    - Adaptive was the only condition that resulted in decreased difficulty ratings (i.e. the task became easier) from LU1 to LU8.
    - Perceived ability ratings improved significantly from LU1 to LU8 for the adaptive, choose, and adaptive + choose training strategies.
    - No significant differences existed among training strategies on the measure of perceived usefulness at LU8.

  - Confidence in abilities improved from LU1 to LU8 for the choose, adaptive + choose, and DRILCOM conditions.
  - Adaptive and adaptive + choose resulted in a significantly faster perceived response time at LU8 than the DRILCOM condition.
  - All participants rated the training as taking too much time at LU8.
  - The random subunit received the highest mean ratings, in regards to benefit of the subunit, for all training conditions.
- Objective 2: Determine how many LUs are needed to achieve the highest localization accuracy.
  - No significant differences in performance existed between LU4 and LU8 using the difference score and slope as measures.

Given the highest mean performance difference score and slope, the choose training strategy was implemented in Phases II and III of the follow-on experiments. To recapitulate, the objective of Phase II was to evaluate the efficacy of localization training conducted in laboratory-grade training apparatus compared to training conducted in a portable apparatus. The objective of Phase III was to evaluate the presence of transfer-of-training from localization training conducted on a portable system to a field environment. To improve the training for the follow-on phases, the training was shortened to five LUs to better ensure the range of training transfer was appropriately captured. Twenty-four locations were incorporated into the follow-on studies to avoid ceiling effects and given that follow-on portable localization training system designs will incorporate a 24-speaker design in azimuth, providing signals from 15° increments of arc.

PHASE III: IN-FIELD INVESTIGATION OF TRANSFER-OF-TRAINING

Phase III: Objectives

The primary objective of the Phase III in-field experiment was to evaluate the transfer-of-training effects of conducting azimuthal localization training in-lab, using the PALAT system, on in-field localization performance. The experiment also investigated the sensitivity of the in-lab training and in-field testing to differences among listening conditions. As such, the evaluation incorporated three listening conditions: open ear (unoccluded), in-the-ear TCAPS device (TEP-100), and over-the-ear TCAPS device (ComTac™ III). A secondary objective was to evaluate the validity of using the in-lab PALAT system results to predict localization performance under all three listening conditions in a military operational setting. To meet these objectives, localization performance was compared between trained and untrained participants using an in-lab pretest and an in-field posttest using live (blank) gunshots.

Phase III: Methodology

This investigation aimed to measure the transfer-of-training effect instilled by the PALAT system using the localization training protocol developed in Phase II and Phase I, respectively. A series of auditory localization studies conducted at the Virginia Tech Auditory Systems Laboratory previously measured localization ability in terms of accuracy, response time, and subjective rankings of participant-perceived localization ability in a single environment, either the laboratory (Casali & Robinette, 2014; Casali & Lee, 2016a; Casali & Lee, 2016b) or a field environment (Talcott et al., 2012). Unique to the Phase III experiment, participants trained and tested on a localization task in an office setting using a dissonant tonal signal, but tested in a field using a military-relevant signal. Previous VT-ASL experiments did not incorporate lab and field and environments within the same study nor did participants train on signals different from the test stimuli.

The participant first signed a consent form and then was audiometrically and demographically screened. Next, the participant underwent pretesting in the PALAT system under all three listening conditions (i.e., open ear, TEP-100, and ComTac™ III) using the dissonant training signal. Figure 72 displays a participant's progression through the experiment. The PALAT system was located in an academic building office space at Virginia Tech. The pretest order was counterbalanced and the participant's order was maintained throughout the stages of the study. After the pretest, half of the participants were randomly selected to undergo localization training on the PALAT system. In the trained group, or experimental group, participants underwent the localization training protocol developed in Phase I of the overarching research effort (Cave, Thompson, Lee, & Casali, in press; Cave, 2019). The training sessions consisted of three, one-hour sessions of localization training using the PALAT system. Training occurred under one listening condition during each one-hour session. Each session included all five learning units of localization training and testing. The three training sessions were completed within three days of the pretest date with a maximum of two training sessions per day and at least two hours separation between any training session. Within one to three days after training completion under all listening conditions, the trained and untrained groups underwent field testing. The field site for the posttest experiment was designed to simulate a scenario where a U.S. Military service member listens for enemy threats in a lightly-wooded field. The same field site previously used in Casali et al. (2012) was used in this experiment. Results from the Subject Matter Expert survey conducted by Cave et al. (2019) showed the most prominent enemy threat facing U.S. Armed Forces ground combat personnel was gunshots (Cave, 2019). As a result, gunshots from Fiocchi .22 caliber long rifle blanks were used as the posttest stimuli.

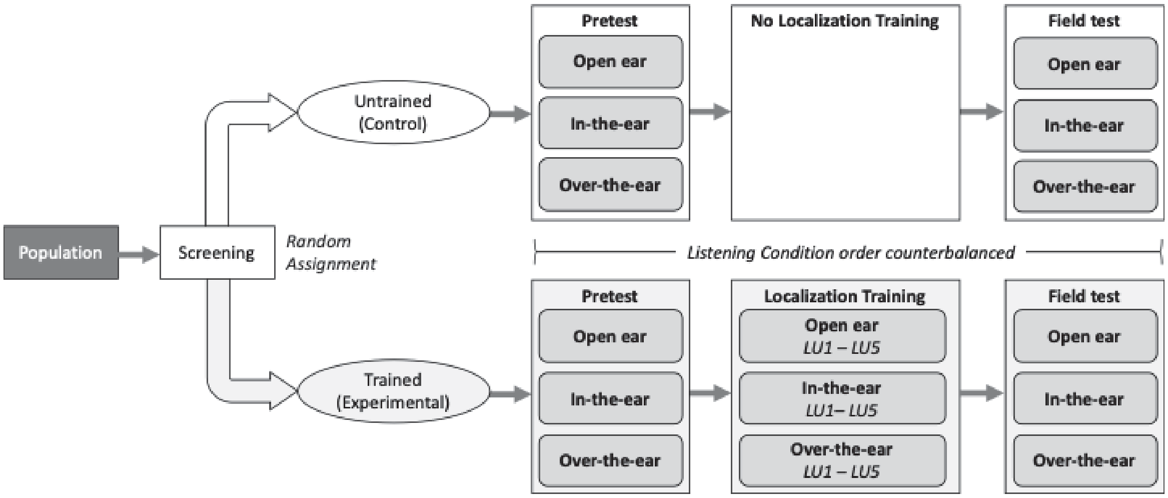
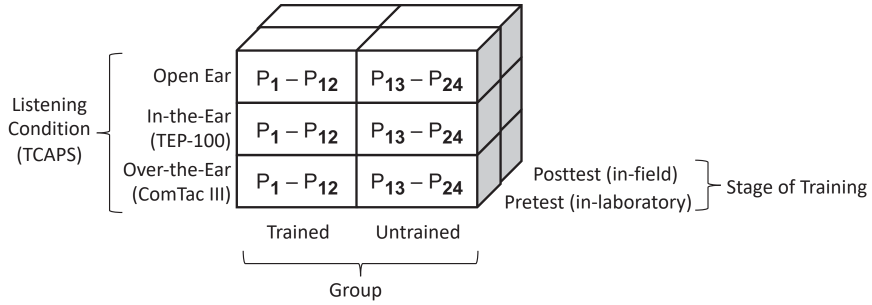


Figure . Phase III experimental design order.

Phase III: Experimental Design

Phase III consisted of a pretest-posttest, control group design experiment involving 24 normal-hearing participants (Figure 73) who had neither prior experience in localization testing nor TCAPS use. The 2 x 3 x 2 mixed factor design involved three independent variables: a between-subjects factor of group with two levels (trained and untrained) and two within-subjects factors: listening condition with three levels (Open Ear, TEP-100, and ComTac™ III) and stage of training with two levels (pretest and posttest). Results were measured using three groups of dependent measures: localization accuracy, response time, and participant subjective responses.



Experimental Design

1. Mixed-factor, pretest posttest control group design
   Between-subjects
   - 2x Group: Trained (Experimental) and Untrained (Control)
   Within-subjects
   - 3x Listening conditions: Open ear, In-the-ear, Over-the-ear
   - 2x Stage of training: Pretest (In-laboratory) and Posttest (In-field)
2. Order of:
   - Group assignment – Random
   - Listening condition – Latin square, counterbalanced order
   - Stimulus presentation azimuth – Random during testing
3. Participants
   - 18-45 years of age
   - 75-85% male to generalize military population
   - Normal Hearing
   - No experience with PALAT system, localization training, or TCAPS

Dependent Variables

1. Localization accuracy
   - Percent absolute accurate (response matches signal azimuth)
   - Percent ballpark accurate (response ±15° signal azimuth)
   - Front-back confusion percent
2. Response time (seconds, with msec precision)
3. Perceived localization performance rating
   - Confidence in accuracy for each listening condition
   - Perceived effect of training on localization accuracy
   - Perceived effect on response time for each listening condition

Figure . Experimental design for Phase III, with independent variables, experimental order, participant assignment, and dependent measures listed.

A Microsoft® Excel random number generator was used to assign 24 participant numbers to an arrival order. Participants who were assigned numbers 1 to 12 were assigned to the trained group, and participants 13 to 24 were assigned to the untrained group. Two participants were replaced during the experiment, for reasons explained later. The replacement participants were assigned the participant number of the participant who they replaced. In order to generalize to the U.S. Military population of 84% male and 16% female, participants were limited to 18 males and six females with nine males and three females randomly assigned to both the trained and untrained group (U.S. Department of Defense, 2018). Four sets of an identical 3 x 6 Latin square were repeated to counterbalance the listening condition order for each participant. The participant listening condition order was maintained throughout the study. Table 76 displays the participant order for the Phase III experiment by sex, group assignment, and listening condition order.

Table . Participant study order by sex, group assignment (random assignment based on arrival order), and listening condition (counterbalanced using a repeating 3 x 6 Latin square).

Arrival order	Participant Number	Group Assignment	Listening Condition order 1	2	3
M1	P20	Untrained	ITE	OTE	Open
M2	P13	Untrained	Open	ITE	OTE
M3	P21	Untrained	OTE	Open	ITE
M4	P2	Trained	ITE	OTE	Open
M5	P4	Trained	OTE	ITE	Open
M6	P18	Untrained	ITE	Open	OTE
M7	P12	Trained	ITE	Open	OTE
M8	P7	Trained	Open	ITE	OTE
M9	P19	Untrained	Open	ITE	OTE
M10	P15	Untrained	OTE	Open	ITE
M11	P1	Trained	Open	ITE	OTE
M12	P3	Trained	OTE	Open	ITE
M13	P8	Trained	ITE	OTE	Open
M14	P10	Trained	OTE	ITE	Open
M15	P24	Untrained	ITE	Open	OTE
M16	P16	Untrained	OTE	ITE	Open
M17	P23	Untrained	Open	OTE	ITE
M18	P11	Trained	Open	OTE	ITE
F1	P6	Trained	ITE	Open	OTE
F2	P14	Untrained	ITE	OTE	Open
F3	P22	Untrained	OTE	ITE	Open
F4	P17	Untrained	Open	OTE	ITE
F5	P5	Trained	Open	OTE	ITE
F6	P9	Trained	OTE	Open	ITE

Independent Variables (IVs)

Independent Variable – Group

Per Figure 73, two between-subjects group levels were used in this investigation: trained (experimental) and untrained (control). Participant age range, mean and median were very similar between the two levels (Table 77). Participants were informed which group they were assigned to prior to the pretest in accordance with the VT IRB requirements.

Table . Participant demographics by group.

	Trained (n=12)	Untrained (n=12)
Age (years)		
Range	19 - 30	19 - 34
Median	26	27
Mean	25.6	26.9
SD	2.9	4.6

Open ear

The open ear listening condition was included in this investigation for several reasons. First, testing the open ear condition established a baseline performance, enabling a within-subjects comparison of training effect for each TCAPS device. Secondly, the open ear condition is the most commonly-encountered listening condition for U.S. Service Members in training and combat environments where hazardous noise exposure is not imminent or expected, but threat or hazard localization remains paramount. Lastly, several studies have identified barriers to HPDs and TCAPS compliance. Abel (2008) and Bevis et al. (2014) specifically described discomfort and a perceived loss of auditory situation awareness as reasons for non-compliance by U.S. Service Members. In Bevis et al. (2014), all 16 focus groups mentioned that auditory localization was negatively affected by hearing protection devices. One British Army Soldier stated, “If you can’t locate that position then you’re redundant” (Bevis et al., 2019, p131). Therefore, by examining localization performance in the open ear, the influence of device-imposed changes to environmental cues and comfort could be eliminated. Furthermore, the open ear condition addressed the secondary objective of assessing the validity of using PALAT system-obtained results to generalize to auditory localization in the field.

In-the-ear TCAPS

The earplug-style 3M™ PELTOR™ TEP-100 Tactical Earplug is an active, or powered electronic sound transmission, in-the-ear hearing protection device, shown in Figure 74. The TEP-100 Tactical Earplugs are issued as a set of two identical, rechargeable electronic earplugs with a recharging case. For testing purposes, the investigator designated a right and left ear device in each set according to serial numbers. The right and left device designations were maintained throughout the study to reduce confounding effects of differences between earplugs. The 3M™ PELTOR™ level-dependent technology is advertised to “provide hearing protection, and helps improve situational awareness and communication” (3M, 2016a, p1). As a passive earplug, the TEP-100 is advertised to provide a mean attenuation of 23 NRR according to the EPA-required labeling on the device (3M, 2016a). The TEP-100 is compatible with several styles of eartips including the 3M™ PELTOR™ Ultrafit eartips shown in Figure 74 which are the standard issue version for the U.S. Military. As a result, each participant in this experiment were fitted with one of the three sizes of Ultrafit eartips with the TEP-100. A professional U.S. military audiologist, K. Cave, conducted a visual inspection of each participant’s ear canal and ensured the participant was fitted with the proper Ultrafit eartip size.

Figure . 3M™ PELTOR™ TEP-100 electronic earplug-style TCAPS device.

The TEP-100 tactical earplug is equipped with two volume settings, “normal” and “high,” that is operated by a single button. The investigator tested the TEP-100 volume settings to identify the unity gain setting. Unity gain was previously defined by Casali & Lee (2016a) as the state where the electronic gain control is set to overcome or offset the passive attenuation of the earplug and provide as close to natural hearing as possible. Four TEP-100 devices loaned to the Virginia Tech Auditory Systems Laboratory, two devices from U.S. Army PEO Soldier and two devices from 3M^{TM}, were tested in a reverberation chamber to identify the unity gain setting during Phase II of the overarching experiment. One of the TEP-100 devices was found to have significant differences in sound pressure level measurements and was not used during this investigation. The remaining three TEP-100 devices were evenly assigned between the trained and untrained groups so that participants from each group used all three devices.

The following steps were performed to identify the unity gain setting for the TEP-100. A ½ inch Larson-Davis 2575 measurement microphone (SN: 2559) and Larson-Davis 9000C Preamp (SN: 0521) were placed in the center of the reverberation chamber and connected to a Larson-Davis 2900 Model Spectrum Analyzer (SN: A0280) at an investigator table located outside of the chamber. The microphone was calibrated at 94.0 dBA (1000 Hz tone) using a

Quest QC-20 Calibrator (SN: QOA070051). A pink noise signal was generated via a MATLAB® program and measured at 70 dBA, 10 second Leq, fast time constant. Next, an acoustical test manikin, known as KEMAR (Knowles Electronics Manikin for Acoustic Research by GRAS), was positioned in the center of the reverberation chamber and the measurement microphone was fitted inside the left ear canal of the KEMAR and the right ear canal was occluded with putty. The pink noise signal was measured in the open ear listening condition at 77.6 dBA which served as the reference level for unity gain. Each TEP-100 earplug was then fitted in the left ear of the KEMAR and the sound pressure level of the pink noise signal was measured three times at each volume setting: off, or passive, setting, normal volume, and high volume.

The “normal” volume setting provided the closest unity gain for the TEP-100 and thus was the setting used for this experiment (Table 78). Figure 75 displays the sound pressure level measurements of the 70 dBA pink noise signal at each 1/3 octave-band frequency for the open ear and TEP-100 at normal volume setting on the KEMAR manikin. The sound pressure levels measured under the TEP-100 are noticeably lower from 100 Hz to 315 Hz than the open ear levels. The TEP-100 also did not transmit the pink noise at the 10,000 Hz 1/3 octave-band frequency.

Table . Sound pressure level (SPL) of 70 dBA pink noise measured using KEMAR manikin for three sets of TEP-100 devices and mean (by left and right ear designation). Levels compared with the open ear 77.6 dBA reference level displayed as the delta (Δ), open ear – TEP-100.

		Device 1		Device 2		Device 3		Mean	
Listening Condition	**Gain level**	**SPL (dBA)**	**Δ**	**SPL (dBA)**	**Δ**	**SPL (dBA)**	**Δ**	**SPL (dBA)**	**Δ**
Open Ear (Reference Level)		77.6		77.6		77.6		77.6	
TEP-100		Left (SN: 64174)		Left (SN: 61389)		Left (SN: 36576)		Left ear	
	Off (passive)	31.3	(-46.3)	31.8	(-45.8)	29.1	(-48.5)	30.7	(-46.9)
	Normal	**77.4**	**(-0.2)**	**76.9**	**(-0.7)**	**77.2**	**(-0.4)**	**77.2**	**(-0.4)**
	High	88.1	(10.5)	87.4	(9.8)	88.0	(10.4)	87.8	(10.2)
		Right (SN: 64517)		Right (SN: 64343)		Right (SN: 36173)		Right ear	
	Off (passive)	31.6	(-46)	32.0	(-45.6)	29.8	(-47.8)	31.1	(-46.5)
	Normal	**79.7**	**(2.1)**	**78.2**	**(0.6)**	**75.2**	**(-2.4)**	**77.7**	**(0.1)**
	High	90.3	(12.7)	89.1	(11.5)	86.0	(8.4)	88.5	(10.9)

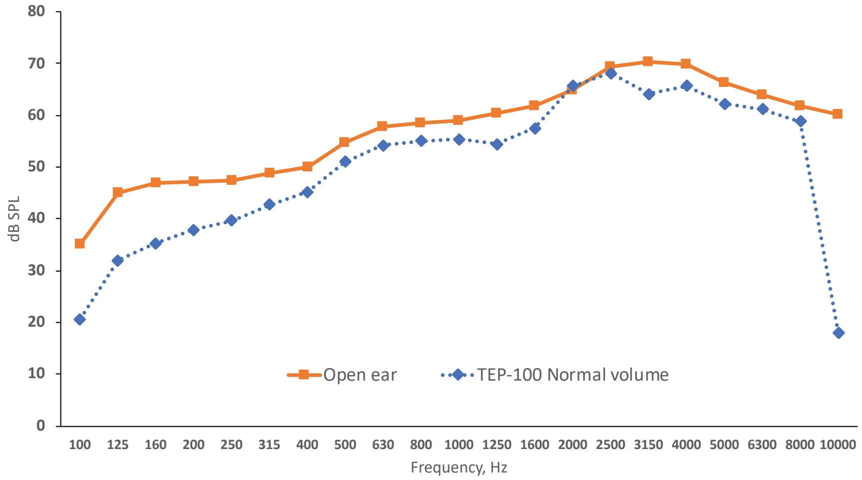


Figure . Mean frequency response of 70 dBA pink noise measured using KEMAR under open ear and TEP-100 devices on normal volume setting by 1/3 octave-band frequencies.

Over-the-Ear TCAPS

The earmuff-style 3M™ PELTOR™ ComTac™ III headset is an active, or electronic sound transmission, over-the-ear hearing protection device, shown in Figure 76. This battery-powered TCAPS is equipped with four volume settings and an additional boost mode to amplify low level external sounds to audible, but not hazardous levels, and pass them through the muff. According to the manufacturer’s literature, the 3M™ PELTOR™ ComTac™ III utilizes a proprietary digital audio circuit to compress hazardous noise to a permissible safe exposure level of less than 82 dBA (3M, 2016b). As a passive headset, the 3M™ PELTOR™ ComTac™ III is advertised to provide a NRR of 23 (3M, 2016b).

Figure . 3M™ PELTOR™ ComTac™ III electronic earmuff-style TCAPS device.

Three ComTac™ III headsets were loaned to the Virginia Tech Auditory Systems Laboratory, one headset from U.S. Army PEO Soldier and two headsets from 3M^{TM}, for the study. All three headsets were tested to identify the unity gain setting using the same procedure described above. The highest volume setting, or fourth increase from default, provides the closest unity gain for the ComTac™ III and was used for this experiment (Table 79). Figure 77 displays the sound pressure level measurements of the 70 dBA pink noise signal at each 1/3 octave-band frequency for the open ear and ComTac™ III at the high volume setting on the KEMAR manikin. The sound pressure levels measured under the TEP-100 are noticeably lower from 4,000 Hz to 10,000 Hz than the open ear levels.

Table . Sound pressure level (SPL) of 70 dBA pink noise measured using KEMAR manikin for three sets of ComTac™ III devices and mean. Levels compared with the open ear 77.6 dBA reference level displayed as the delta (Δ), open ear – ComTac™ III.

Listening Condition	**Gain level**	**Device 1 (SN: 7500)**		**Device 2 (SN: 7607)**		**Device 3 (SN: 1099)**		**Mean**	
		SPL (dBA)	**Δ**	**SPL (dBA)**	**Δ**	**SPL (dBA)**	**Δ**	**SPL (dBA)**	**Δ**
Open Ear (Reference Level)		77.6		77.6		77.6		77.6	
ComTac™ III	Off (passive)	38.6	(-39.0)	38.0	(-39.6)	40.1	(-37.5)	38.9	(-38.7)
	1 (Low)	56.4	(-21.2)	57.5	(-20.1)	57.4	(-20.2)	57.1	(-20.5)
	2	62.2	(-15.4)	63.4	(-14.2)	63.3	(-14.3)	63.0	(-14.6)
	3	68.2	(-9.4)	69.4	(-8.2)	69.3	(-8.3)	69.0	(-8.6)
	4 (High)	**74.2**	**(-3.4)**	**75.4**	**(-2.2)**	**75.2**	**(-2.4)**	**74.9**	**(-2.7)**

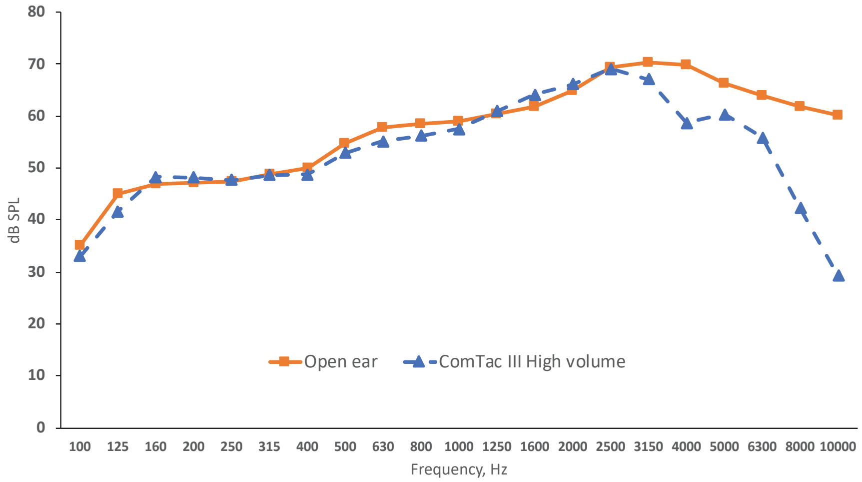


Figure . Mean frequency response of 70 dBA pink noise measured using KEMAR under open ear and ComTac™ III devices at high volume setting by 1/3 octave-band frequency.

Independent Variable – Stage of Training

Two within-subjects stage of training levels were used in this experiment: pretest and posttest (Figure 73). Three sets of pretests and posttests were administered to each participant, one for each listening condition, following the prescribed counterbalanced order for each participant. Each pretest and posttest for a listening condition consisted of a total of 36-presentations, three presentations from each of the 12 loudspeaker or 12 remote firing device locations. For all participants, the pretests were completed in one, approximately 30-minute testing session and the posttests were completed in one, approximately one-hour testing session.

As noted earlier, posttests were administered within one to three days of the pretest for the untrained group (mean=1.8 days, *SD*=0.8 days) and within one to three days of the final training session for the trained group (mean=1.7 days, *SD*=0.9 days).

Pretests were conducted in-office using the PALAT system. The pretest employed a dissonant, non-harmonically related, tonal signal comprised of 104, 295, 450, 737, 2967, 4959, 7025, and 7880 Hz (Casali & Lee, 2019). A series of auditory localization studies successfully demonstrated that the dissonant signal provides binaural and monaural cues necessary to test and train localization (Casali & Robinette, 2014; Casali & Lee, 2016a; Casali & Lee, 2016b; Cave et al., 2019). The dissonant tone frequencies were selected to provide the predominant localization cues accessible using the following mechanisms: interaural timing differences (ITDs), interaural level differences (ILDs), and pinnae spectral cues. ITD cues dominate localization of sounds below 1500 Hz as the wavelength must be able to "bend around" the diameter of the head to render timing cues (Moore, 1997). ILDs occur when the near ear receives a more intense signal than the far ear (Emanuel, Maroonroge, & Letowski, 2009). ILD cues are dominated by higher frequencies with frequencies above 2000-3000 Hz providing the most information (Moore, 1997). Lastly, the highly contoured surface of the pinnae and successive funneling into the ear canal resonates higher frequencies (Emanuel et al., 2009). This contouring creates spectral changes in the signal even with small changes in sound location, particularly in the 3000-8000 Hz region (Pickles, 1988). Figure 78 displays the spectral content of the 70 dBA dissonant signal presented by the PALAT system.

Posttests were conducted in-field on a rural, lightly wooded farm using a military relevant auditory stimulus in the form of a .22 caliber blank gunshot. Both the experimental site and stimulus were previously used in a sound localization study using active HPDs (Talcott et al., 2012). The investigator measured the spectral content of the .22 blank gunshot in situ and verified that the stimulus contains frequency content necessary to provide interaural timing differences (ITDs) and interaural level differences (ILDs) (Figure 78).

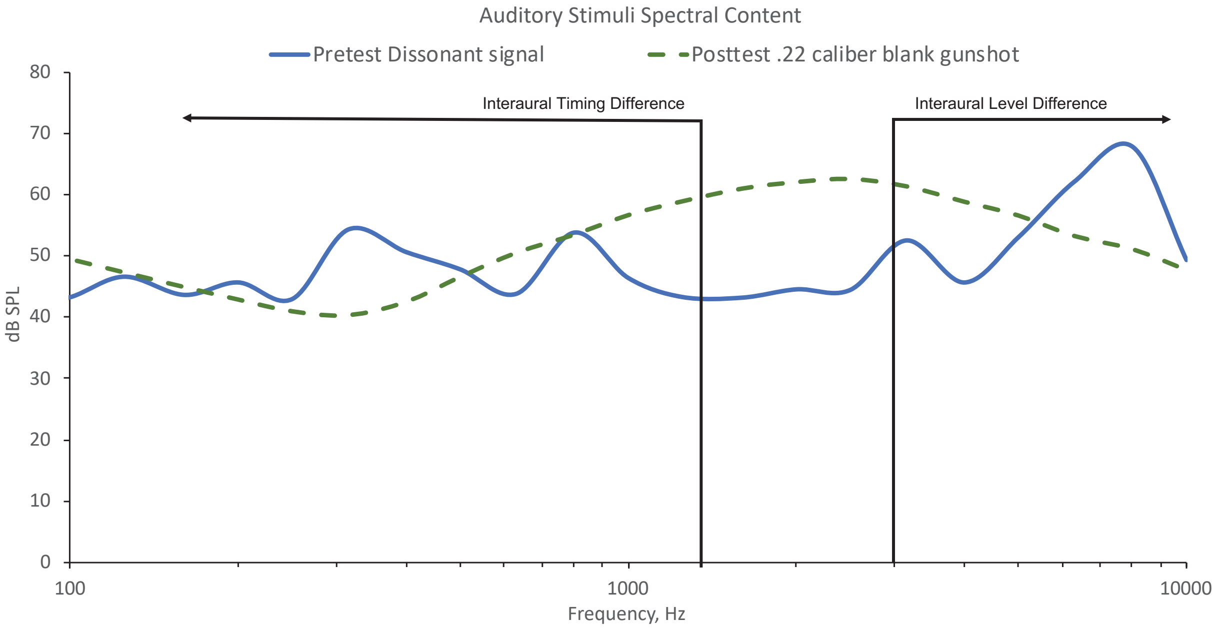


Figure . 1/3 octave-band spectral content of pretest Dissonant training signal (played on PALAT system) and posttest .22 caliber blank gunshot by 1/3 octave-band frequency. Recorded at the participant’s ear in office environment (for PALAT) or at outdoor field site (for transfer-of-training test). Overall sound pressure level of 70 dBA for both signals.

Dependent Measures

Three classes of dependent measures were used to test localization performance: 1) localization accuracy, 2) response time, and 3) subjective ratings (listed in Figure 72). The following sections describe each dependent measure in detail.

Localization accuracy

Three measures of localization accuracy were recorded and analyzed: 1) absolute correct responses, 2) ballpark correct responses, and 3) number of front-back errors. Each test in this investigation presented three signals (dissonant tone or gunshot) from 12 locations in random order for a possible maximum score of 36 correct on each test. The 12 signal locations were separated azimuthally by 30° resembling the 12-hour positions on an analog clock face. U.S. Military Service Members are trained to identify and communicate threat direction or points-of-interest using the 12 clock face number positions with 12 o'clock serving as the frontal midline reference (Department of the Army, 2017). For example, if a military unit were on a patrol walking through the woods in a northerly direction and heard gunshots from an enemy located directly to the east, the members of the unit would yell, "contact, enemy three o'clock." Thus, the investigator decided to present signals from the 12 clock face azimuthal locations. However, 24 response locations were provided to allow the participant to select a direction between two adjacent clock face positions if they were unsure of the exact signal location. Figure 79 shows the test screen from the Surface Pro computer tablet that the participant used during the pretest and posttest, displaying 24 response options (black circles) and 12 signal locations (black circles marked with yellow numbers).

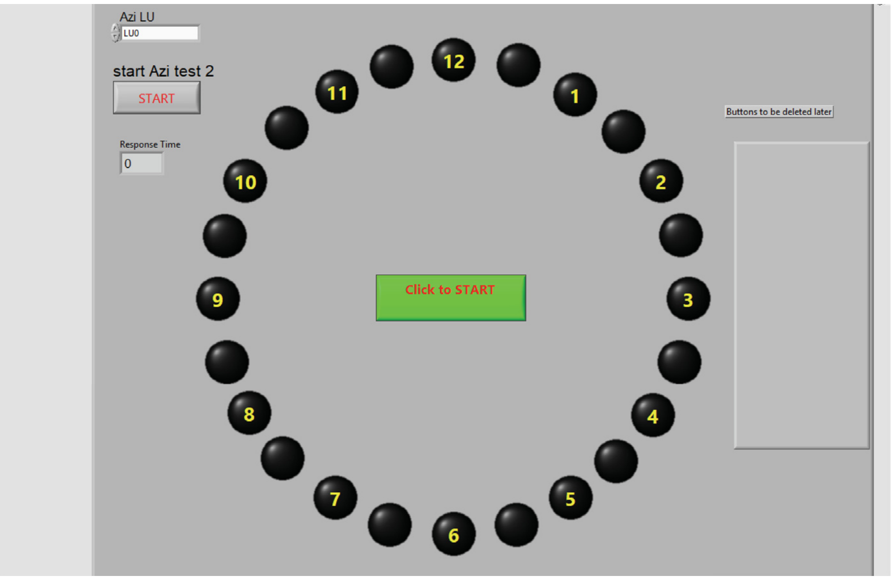


Figure . Participant pretest and posttest screen displaying 24 response options (black circles) and 12-signal locations (black circles with yellow numbers).

1. *Absolute correct responses:* the total number of occurrences in which the participant responded with the exact azimuthal location of the signal location. Figure 80 displays an example of an absolute correct response indicated by the arrow if the signal originated from the one o'clock position.

2. *Ballpark correct responses:* the total number of occurrences in which the participant responded with an azimuthal location within ±15° of the location of the presented signal. A ballpark score is achieved if the participant response matches the exact location of the signal location (absolute correct) or if the response identifies a speaker location directly adjacent, left or right, to the presented signal location. Figure 80 displays an example of the range for a ballpark correct response indicated by the grey shaded region of a signal originating from one o'clock.

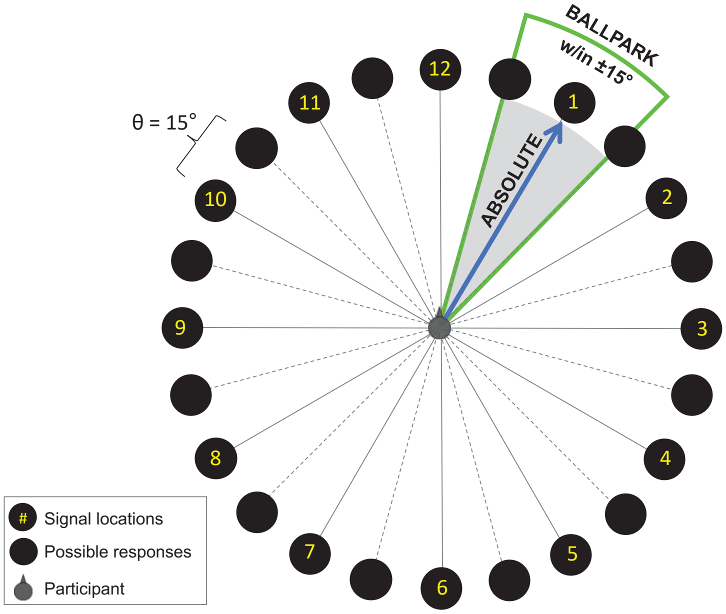


Figure . Absolute correct response (arrow) and ballpark correct response region (grey shaded region) when the signal emanates from the one o'clock position.

3. *Front-back reversal errors:* the total number of occurrences in which the participant responded with an azimuthal location in the back (to the rear of participant) from four o'clock to eight o'clock (120-degees to 240-degrees) when the signal was presented in the front from ten

o’clock to two o’clock (300-degrees to 60-degrees) and vice-versa. This window for front and back reversals is consistent with the new ANSI 3.71 standard window from 290-degrees to 70-degrees in front of the participant and 110-degrees to 250-degrees behind the participant (American National Standards Institute (ANSI), 2019). However, this experiment’s operational definition of front-back reversal differs from the ANSI standard by allowing front-back reversals to occur if the difference between the source and response crosses the median plane. For example, a front-back reversal occurs in this experiment if a sound originates from the seven o’clock position and the participant responds with the one o’clock position. The investigator felt this offered a more realistic operational definition of front-back reversals for auditory situation awareness in U.S. Military operations. If a U.S. service member perceived a gunshot from the one o’clock position (in front of them) that actually originated from seven o’clock (behind them), then the service member would have made a front-back reversal that could be detrimental to survivability. Figure 81 displays the front and back regions where either the signal originated and the response was selected to constitute a front-back reversal error if the signal and response were in opposite regions.

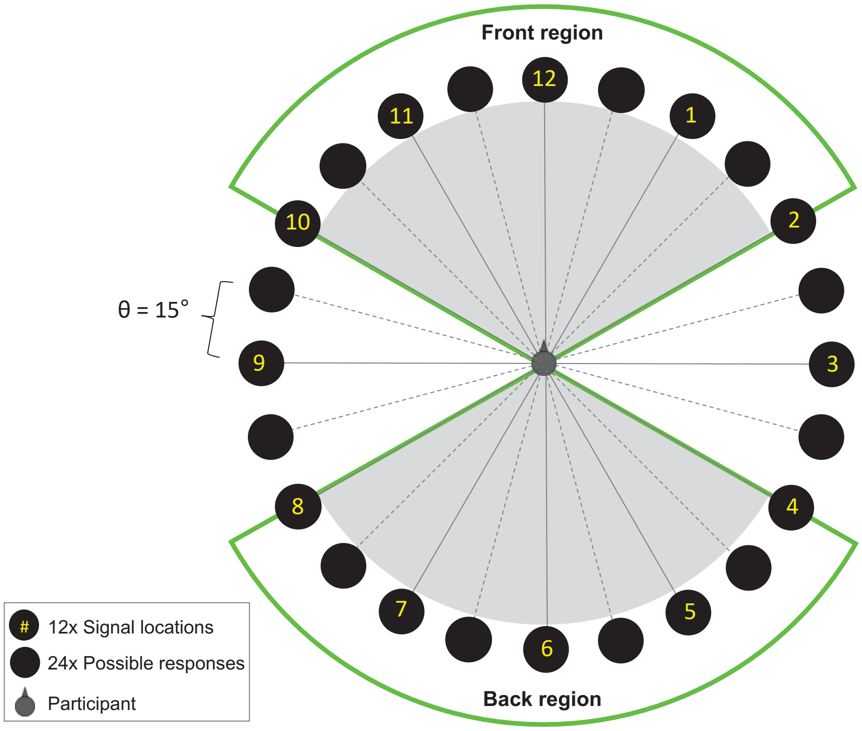


Figure . Front and back regions (shaded regions) depicting the range of signal locations and response locations for possible front-back reversal errors.

Response time

Response time was measured as the duration of time occurring from signal onset, dissonant tone (PALAT) or gunshot (field test), to the participant response selection on the computer tablet. Response time was automatically calculated via the LabView computer program on the Surface Pro computer tablet. The response time clock onset was triggered by the participant selecting the green ‘Click to START’ icon located in the center of the test screen (Figure 79). Selecting the ‘Click to START’ icon simultaneously presented the dissonant tone or the gunshot for the tests conducted in the PALAT or field, respectively. The response time clock offset occurred when the participant selected a speaker icon on the response display. A window located on the left side of the test screen displayed the running clock. After response selection, the display showed the most recent response time, allowing the participant to view their response time. Response times were recorded in 100 millisecond resolution. The maximum allowable response time was set at 10 seconds. Mean response times were calculated for each pretest and posttest and used as the dependent measure score.

Subjective ratings

Participants completed a questionnaire at the conclusion of every pretest and posttest for every listening condition (Appendix K). Every questionnaire included the same six questions so that comparisons could be made between tests. Participants in the trained group were asked an additional question at the conclusion of the posttest related to the effect of training on localization performance. All questions used a semantic differential, bipolar rating scale with seven discrete choices (example shown in Figure 82). Question 7 was used only for trained group after each listening condition during the posttest.

Every questionnaire included the same six questions so that comparisons could be made between tests. Participants in the trained group were asked an additional question at the conclusion of the posttest related to the effect of training on localization performance. All questions used a semantic differential, bipolar rating scale with seven discrete choices (example shown in Figure 82). Question 7 was used only for trained group after each listening condition during the posttest.

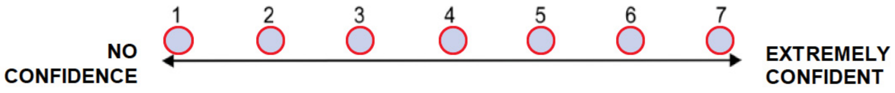


Figure . Example of semantic differential rating scale.

1. *Confidence in ability to localize*:

> “Rate how **confident you were** in your ability to locate sounds under this listening condition” from 1 (no confidence) to 7 (extremely confident).

2. *Perceived localization accuracy*:
   "Rate your perceived **accuracy** to determine sound location under this listening condition" from 1 (highly inaccurate) to 7 (highly accurate).
3. *Difficulty in judging location of sound*:
   "Rate how **difficult** it was to judge the **location** of the sounds under this listening condition" from 1 (extremely difficult) to 7 (extremely easy).
4. *Perceived reaction time*:
   "Rate your perceived **reaction time** in determining the sound location under this listening condition" from 1 (extremely slow) to 7 (extremely fast).
5. *Comfort of TCAPS device or open ear*:
   "Please rate how **comfortable this hearing protection device condition** (or open ear) was while wearing it during the experiment" from 1 (extremely uncomfortable) to 7 (extremely comfortable).
6. *Likelihood to use TCAPS device*:
   "How likely would you be to wear this hearing protection device during a task similar to this experiment that required sound localization if you had access to this hearing protection device" from 1 (extremely unlikely) to 7 (extremely likely).
7. *Preparedness from training*:
   "Rate the degree of preparedness you felt as a result of the training on the localization system (ring of loudspeakers) compared to the task of localizing .22 blank gunshot sounds" from 1 (extremely unprepared) to 7 (extremely prepared).

Participants

This human-subjects experiment was approved by the Western Institutional Review Board (WIRB protocol #20190789, VT-IRB #19-176) which acted as the assigned review board for Virginia Tech as of the date of this research. Participants were required to be between the ages of 18 to 45 years with up to 25% females in order to generalize to the U.S. Military population (Defense Manpower Data Center (DMDC), 2017). The study sample consisted of 24 participants: 18 males and 6 females, age 19 to 34 years with a mean age of 26.3 years (*SD*=3.8).

Two additional male participants were involved in the study but were replaced due to one failing to complete the study (illness) and one participant's performance resulting in a statistical outlier on two performance measures (discussed later in section 3.4 Results). Participants were recruited from Virginia Tech and the surrounding communities. Participants were compensated $10 per hour and received a $30 bonus upon completion of the study. Transportation was provided to the field site for all but one participant who was reimbursed for mileage at the Virginia Tech-assigned rate of $0.58 per mile.

Participants were required to have normal hearing and no previous experience with localization studies or training. All participants were screened for hearing thresholds not to exceed 25 dB HL at 250, 500, 1000, 2000, 3000, 4000, 6000, and 8000 Hz, with no threshold difference between each ear to exceed 15 dB (bilaterally-symmetrical). Following the participant's informed consent, they were otoscopically inspected to check for ear canal obstructions, irritation, or infections that could affect localization performance. One of the investigators, an Active Duty Army Audiologist, performed the otoscopic inspections and administered the hearing tests. If the participant passed otoscopic inspection, a manual pure-tone audiogram using a standard Hughson-Westlake procedure was conducted using a Beltone Electronics Corporation Model 119 Audiometer (SN: 10B0561, calibrated 26 December 2019). The test was performed in the VT-ASL portable test booth located in the same room as the PALAT system (Figure 83). Table 80 displays the mean pure-tone hearing level thresholds (dBHL) for all participants and by group. Following the audiogram, participants were screened to ensure no prior experience with localization training or TCAPS devices (Appendix L).

Table . Mean pure-tone hearing level thresholds (dBHL) for all participants and by group.

		Frequency (Hz)							
	Ear	250	500	1000	2000	3000	4000	6000	8000
All participants	Right	10.2	6.3	4.6	0.0	4.2	2.3	5.4	9.2
	Left	8.5	6.3	5.2	-0.2	2.1	4.6	5.0	8.5
Trained	Right	10.0	5.8	2.5	-0.8	3.3	1.7	6.7	7.5
	Left	9.2	7.9	5.8	1.3	3.8	5.0	6.3	12.9
Untrained	Right	10.4	6.7	6.7	0.8	5.0	2.9	4.2	10.8
	Left	7.9	4.6	4.6	-1.7	0.4	4.2	3.8	4.2

Figure . Portable audiometric booth (left) co-located with the PALAT system in the office environment used for training and pretesting.

Apparatus

The Phase III investigation was conducted in two locations: an office environment on the campus of Virginia Tech where training and pretesting occurred and an outdoor field-conducted posttest on a rural farm in Pulaski County, Virginia.

In-Office: PALAT System

The pretest and localization training were conducted using the PALAT system apparatus. The PALAT system was located in a small office room on the fifth floor of Whittemore Hall at Virginia Tech. The PALAT system apparatus was operated in the same office during Phase II and was found to provide a similar localization testing and training experience and performance results as the full-scale, laboratory grade DRILCOM system used in previous localization experiments (Casali & Robinette, 2014; Casali & Lee, 2016a; Casali & Lee, 2016b). The

PALAT room was approximately 13.5 feet by 12.5 feet and contained typical office furniture including a desk, chairs, wooden bookshelf, metal storage cabinets, dry-erase board, metal window blinds, carpeted floor, and dropped panel ceiling. In addition, a metal portable audiometric booth was located in the corner of the room (Figure 83). The small office space was selected due to its semi-reverberant environment that represents a typical setting where the military (or civilian industry) would employ the PALAT system. Likewise, the investigator left the acoustically reflective furniture inside the office assuming that users of the PALAT system would have the constraint of using the system in rooms designated for other purposes; thus, there was no attempt to "optimize" the office for uniformity of reflections or other acoustic considerations. In other words, the office environment used to evaluate the PALAT system was believed to be as realistic as possible, representative of that which would be typically encountered in actual military training practice. The PALAT system was positioned in the room so that no speaker was within two feet of any reflective surface but the system was not centered in the room. Centering the PALAT system in the room would be preferred in order produce a more uniform reflective surface. Hartmann (1983) found that early reflections from side walls created the largest decremental effect on localization due to spectral information that conflicted with the direct sound wave. The investigator decided against centering the portable system assuming that future users of the PALAT system may have similar limitations due to varying room sizes and shapes.

The small semi-reverberant office used for PALAT testing and the hemi-anechoic laboratory room housing the DRILCOM system (for comparison) were tested to find the ambient noise floor and reverberation time (RT60). Measurements were made with a Larson-Davis Model 831 Sound Level Meter (SN: 0002486) with a ½-inch Larson-Davis 2575 measurement microphone (SN:LW131180) and Larson-Davis PRM831 Preamp (SN: 017153). The microphone was calibrated at 94.0 dBA (1000 Hz tone) using a Quest QC-20 Calibrator (SN: QOA070051). The investigator performed five measurements, in the center of the room and approximately one-meter from each room corner. RT60 measurements were taken using an impulse noise at approximately 120 dBA produced by hitting together two-wooden 2x4 blocks. The RT60 measurements were calculated using a 30 dB decrease in level to avoid limitations posed by the noise floor. The calculation to extrapolate the RT30 values to RT60 values was performed automatically by the sound level meter. Noise floor and RT60 values are shown in Table 80 for both the DRILCOM and PALAT rooms.

Table . Mean noise floor and reverberation time (RT60) measurements of the DRILCOM and PALAT rooms, as measured in octave bands.

		Frequency (Hz)					
		250	500	1000	2000	4000	8000
DRILCOM Room	Noise Floor (dB SPL)	34.0	30.5	31.3	33.9	36.6	40.0
	RT60 (ms)	408	272	182	144	119	110
PALAT Room	Noise Floor (dB SPL)	42.1	33.1	31.5	33.5	36.6	40.0
	RT60 (ms)	407	402	348	339	410	396

The PALAT system is a 2-meter diameter circular, horizontal and vertical (front) localization apparatus consisting of 32 loudspeakers with 24 loudspeakers (horizontal) and eight additional vertical loudspeakers housed in a semi-reverberant room (Figure 84). Two of the horizontal loudspeakers are used during elevation testing to provide 10 vertical loudspeakers. Only azimuth speakers were used in this study. All loudspeakers are separated by an angle of 15° from the center of the apparatus, or center head position of the participant. The horizontal loudspeakers are located one meter from the participant. The loudspeakers are mounted on a portable, collapsible frame consisting of 12 telescopic poles. The telescopic poles allow for horizontal loudspeaker heights of 43.5, 45.5, and 47.5 inches above the floor, to accommodate seating heights of different individuals. The speaker heights were set to 45.5 inches above the floor for the duration of the in-office experiment. The PALAT system is controlled by the user seated in the middle of the loudspeaker array via a Microsoft® Surface Pro running a LabView software program. The system uses Cambridge Audio Minx Min 12 loudspeakers with a 2.25-inch single cone driver, a Stewart Audio AV30MX-2 two channel stereo mixer amplifier, and a Numato 32 channel USB relay module. Under the participant chair, a pink noise generator mounted inside the equipment case emits 55 dBA of pink noise. To generate the pink noise signal, a Mystic Marvels LLC. PNG-400 Pink Noise Generator routed through a Stewart Audio AV30MX-2 amplifier and Minx Min 12 loudspeaker is used (Figure 84). The 55 dBA pink noise masks extraneous sounds during the experiment while maintaining a +15 dB signal to noise ratio given the 70 dBA dissonant signal. Figure 85 displays the spectral content of the dissonant training signal and the pink noise as measured in the PALAT system. The microphone was placed in the center of the PALAT system at the approximate height of the participant's head. The two poles housing the elevation speakers were removed during all in-office testing and training in order to present a more uniform apparatus for azimuthal testing (all 24 speakers aligned on one horizontal plane).

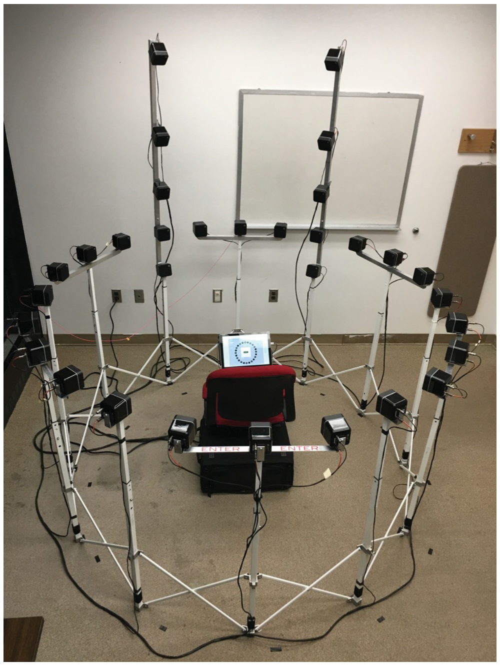


Figure . PALAT system apparatus located in a semi-reverberant office room at Virginia Tech.

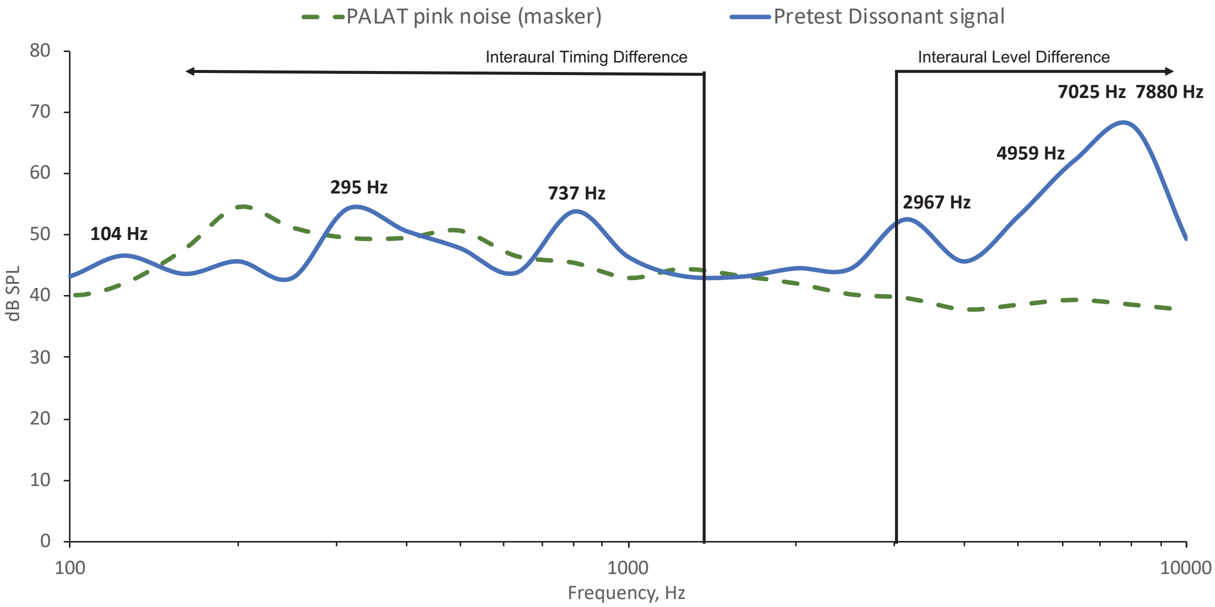
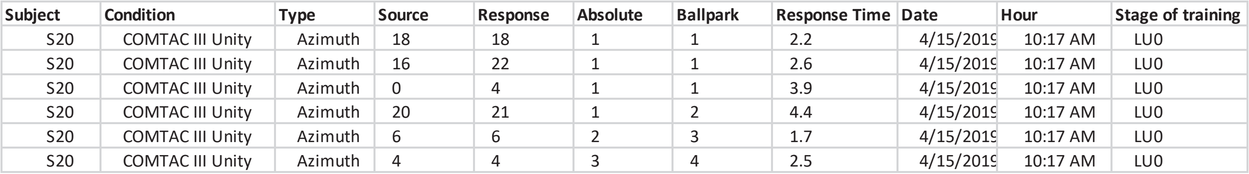


Figure . 1/3-octave band spectral content of PALAT system pink noise (green dashed line) and dissonant signal (blue solid line) in 1/3 octave-band frequency. Eight pure tones comprising the dissonant tone are labeled above the respective frequency.

The participant-controlled computer tablet was used to initiate the dissonant signal presentation and record the user response, azimuth and time. The participant used a stylus pen to operate the controls on the computer tablet. Prior to testing, the participant was given instructions and received a demonstration on how to control the computer tablet and software program. When ready, the participant selected the ‘Click to START’ icon on the computer tablet. The LABVIEW program then sent a signal to the USB-controlled relay switch to close the relay corresponding to the presentation speaker. Simultaneously, LABVIEW transmitted the dissonant audio signal through a 3.5 mm audio cable through the amplifier and relay switch to the presentation speaker. Signal transmission from LABVIEW also triggered the onset of the response timer. To randomize the speaker location, the software program used a random number generator to select a loudspeaker position with the constraint of requiring three presentations from each loudspeaker location during each test. The participant registered a response by selecting one of the 24 loudspeaker locations by touching the stylus to one of the black circles corresponding to the loudspeaker locations, as accurately and quickly as possible. Upon response, the LABVIEW software stopped the response timer and recorded the response to a Comma Separated Values (CSV) file. The software program recorded the subject number, listening condition, test type (azimuth or elevation), loudspeaker source location, participant response location, response time, date, time, and stage of training for each trial. The software program also calculated and displayed a running total absolute score and running total ballpark score for the test (example shown in Figure 86). The CSV file was saved to a shared folder so participant scores could be accessed after each test.



Subject	Condition	Type	Source	Response	Absolute	Ballpark	Response Time	Date	Hour	Stage of training
S20	COMTAC III Unity	Azimuth	18	18	1	1	2.2	4/15/2019	10:17 AM	LU0
S20	COMTAC III Unity	Azimuth	16	22	1	1	2.6	4/15/2019	10:17 AM	LU0
S20	COMTAC III Unity	Azimuth	0	4	1	1	3.9	4/15/2019	10:17 AM	LU0
S20	COMTAC III Unity	Azimuth	20	21	1	2	4.4	4/15/2019	10:17 AM	LU0
S20	COMTAC III Unity	Azimuth	6	6	2	3	1.7	4/15/2019	10:17 AM	LU0
S20	COMTAC III Unity	Azimuth	4	4	3	4	2.5	4/15/2019	10:17 AM	LU0

Figure . Example CSV file output of PALAT system.

In-Field Site

The field-conducted posttest was conducted outdoors on a rural farm located in Pulaski County, Virginia. The participant stood in the center of 50-foot circular clearing, surrounded by a lightly wooded forest of relatively mature trees in which twelve hard-wired, but remotely operated blank-firing devices (Figure 87). The site was previously used for a sound localization study where eight firing positions surrounded the participant (Talcott et al., 2012). Due to the extended time period between the 2012 and current experiments, and the addition of four firing positions, the field site was re-cleared of obstructions and reoriented to align the twelve firing positions so that no large trees were in the direct line of sight (or direct sound ray) of each remote firing device. The clearing in the woods was situated at the top of a small rolling hill. The hill rolled off in a nonuniform pattern with the steepest roll-off in the northeast (12 o’clock), fairly flat terrain to the southeast (three o’clock) and northwest (nine o’clock), and a gradual roll-off to the southwest (six o’clock). Based on the terrain features, the investigator established the 12 o’clock remote firing position to the northeast at a magnetic azimuth of 32°. The investigator used an optical level and U.S. Military tritium lensatic compass to mark the 12 firing position azimuths separated by 30° angles measured from the center point. Table 82 displays the magnetic azimuth of 12 firing device positions measured from the center point where the participant stood.

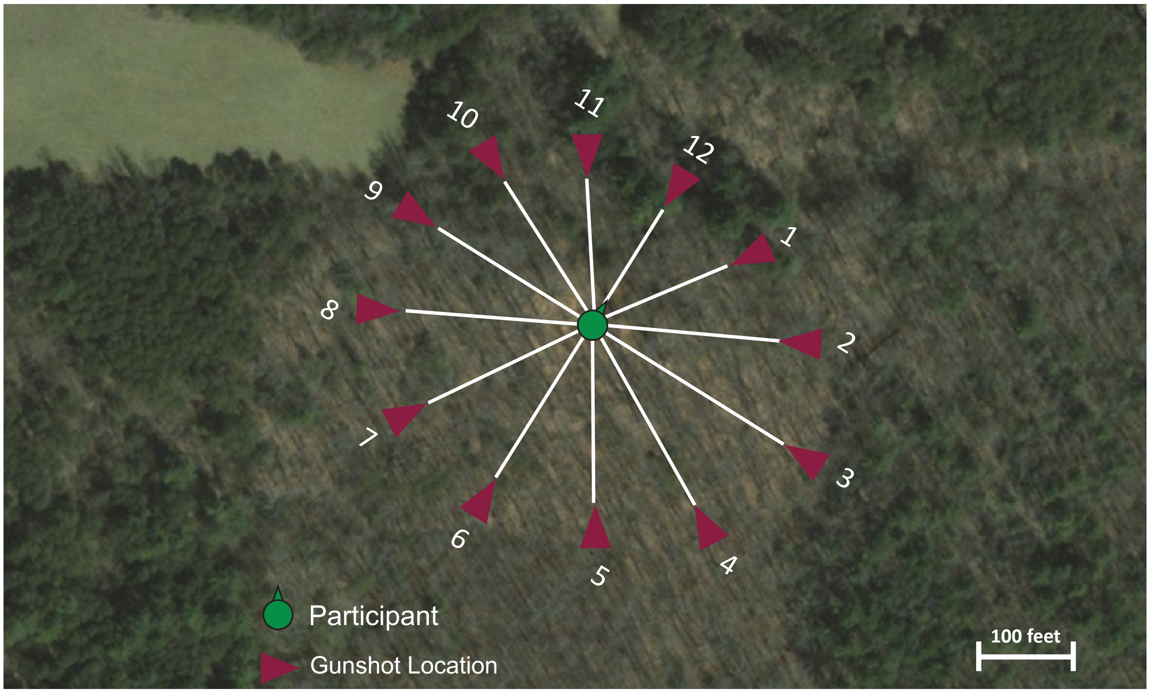


Figure . Aerial view of field site layout with 12-remote firing device positions, located around the participant.

Table . Magnetic azimuth (degrees), radial distance (feet), and sound pressure level (dBA max) at center position of each remote firing device location.

Remote device location	12	1	2	3	4	5	6	7	8	9	10	11
Magnetic azimuth (degrees)	32	62	92	122	152	182	212	242	272	302	332	2
Radial distance (feet)	150	160	200	250	215	210	220	200	200	192	175	160
SPL (dBA max)	69.4	71.7	70.2	72.1	67.7	70.5	70.8	70.9	72.5	70.9	71.6	72.0

The investigator matched the blank gunshot signal in the posttest with the 70 dBA-max signal used in the pretest. Sound levels were controlled by adjusting the distance of the remote firing devices from the center to achieve the target level. The 12 distances to the firing devices were adjusted until the mean sound pressure level from three .22 caliber blank gunshots resulted in approximately 70 dBA-max at each location. The ambient noise floor was measured to ensure that the 70 dBA-max gunshot signal was easily detectable and provided interaural timing and interaural level cues used in sound localization (Figure 88). Table 82 displays the radial distance and mean sound pressure level measured for each firing position. Of note, the investigator discovered slight inconsistencies in sound pressure levels produced by blank rounds fired from the same remote firing device. These variations were possibly due to variations in the manufacturing process or an effect of how the gunpowder was ignited in each blank round, but were obviously not under experimental control and thus constituted a small random variance source. These fluctuations were randomly distributed throughout all experiments effecting both groups and each listening condition.

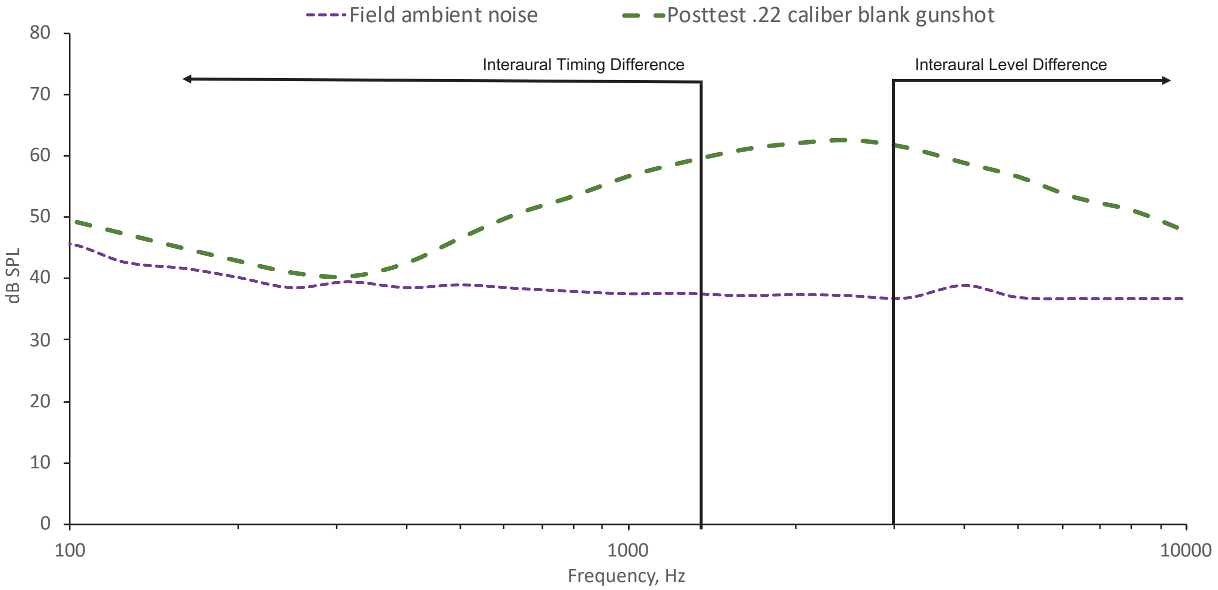


Figure . Mean 1/3-octave band spectral content of .22 caliber blank gunshot from all 12 firing device positions (dashed green line) versus ambient noise floor (dotted purple line) measured in 1/3 octave-band frequency at the participant center head position.

Figure 89 shows a panoramic photograph of the field site from the participant's perspective marked with each firing device location by clock face number. A sign displaying the number 12 was placed approximately 50 feet directly in front of the participant along the 12 o'clock azimuth (and only at that position) to help orient the participant and give them a visual reference point to focus on prior to firing each blank gunshot (Figure 90).

Figure . Field site panoramic picture (from left at 7 o'clock to right at 6 o'clock), with clock face positions identified. Only the 12 o'clock position actually included a sign during the conduct of the field experiment.

Figure . Participant view of computer tablet used to operate the field posttest and 12 o'clock reference sign to orient participant.

Outdoor Azimuthal Gunshot Presentation System

In Talcott et al.'s (2012) study at the same field site, .22 caliber blanks were fired from revolver pistols by the investigators. Three investigators walked to each of the eight firing positions located at a radial distance of 150 feet in a predetermined order. Each participant spent approximately 4 hours at the field site conducting sound localization testing in the Talcott et al. (2012) protocol. In the present study, the number of firing locations increased to 12, and the radial distance increased by 100 to 150 feet at most firing locations. Due to the labor and time-intensive requirements of the field experiment, the investigator automated the .22 caliber blank gunshot delivery. One of the investigators, Brandon Thompson, designed a remote firing device that consisted of three electric magnetic locks, three spring loaded firing pins, and a bar to hold three .22 caliber blanks (Figure 91). The electric magnetic lock contained a solenoid attached to a small lever with a hook that released a U-shaped lock when supplied with approximately 12 Volts. A key ring attached to the top of the spring-loaded firing pin inserted into the U-shaped lock. The firing pin was held under tension above the .22 caliber blank rounds. The three firing pins were aligned directly over a ¾ by 1-inch aluminum bar that held three .22 caliber blanks. When the firing pin released, the pin struck the rim of the rim-fired .22 caliber blank.

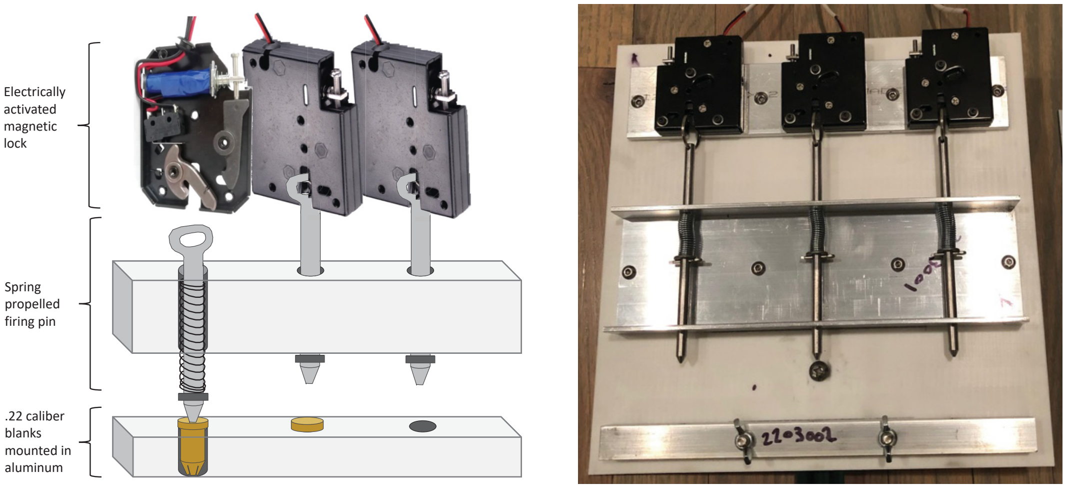


Figure . Remote firing device design concept sketch and final product.

The field localization test employed 12 remote firing devices each containing three separately controlled firing mechanisms (electric magnetic lock, spring loaded firing pin, and .22 caliber blank). An additional firing mechanism was built on the reverse side of the four firing devices that were placed at the 12 o'clock, 3 o'clock, 6 o'clock, and 9 o'clock positions. These additional devices served as a single familiarization round prior to each posttest. The remote firing devices were mounted on a steel u-post at a height of approximately 4 feet above ground (Figure 92). The 12 o'clock firing device was adjusted to a height of 6 feet above the ground to compensate for the steep roll-off directly in front of the participant. Sound measurements were recorded and analyzed to determine the optimal orientation of the remote firing device to present a relatively consistent sound signature from each firing position. The investigator found that orienting the firing device perpendicular to the participant reduced the visual signature and resulted in the most consistent sound levels (Figure 92).

Figure . One remote firing device containing three remote firing mechanisms mounted on a steel u-post located in the wooded forest at the field localization site (Left: Front view of remote firing device, Right: Profile (side) view as seen from direction of the participant), which reduced the visual signature.

The remote devices were hard-wired to a control box containing 12, four-position rotary switches, an LED power indicator light, a safety toggle switch, 8-ampere fuse, and Numato® USB 16-channel relay switch (Figure 93). The remote firing system was initiated by a participant-controlled Microsoft® Surface Pro computer tablet running a localization testing LabView code, almost identical to the interface used to control the PALAT system. The remote firing devices were located between 150 to 300 feet away from the participant and connected to the control box by 18-gauge wires. Due to the voltage drop across the long distance of electrical wire, the investigator used two, 12-Volt gel car batteries connected in series to supply 24 Volts from the control box. The resulting voltage at each firing device measured between 10 to 15 Volts, depending upon the radial distance of each remote firing position. One of the car batteries powered the 12-Volt relay switch. The control box was operated by an investigator seated approximately 10 feet behind the participant. The investigator ensured the correct firing device and firing mechanism were selected on the control box prior to the participant initiating the blank round by clicking on the 'Click to START' button on the computer tablet. Figure 94 displays the components and general wiring of the remote firing system (Appendix M).

Figure . Remote firing device control box.

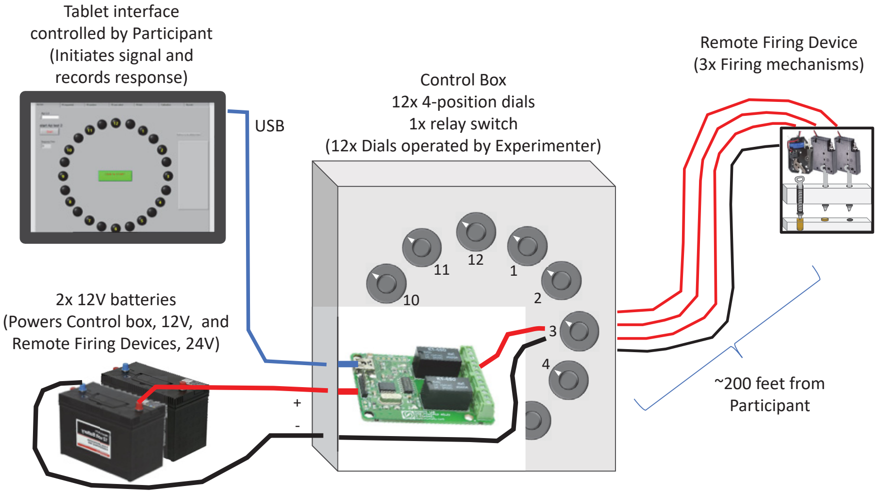


Figure . Remote firing device block diagram with all major components.

Outdoor Auditory Localization Data Capture System

The same participant-controlled computer tablet used in the in-office pretest was used to initiate the .22 blank gunshot and record the user response, response azimuth and time. The computer tablet was placed on a music stand located directly in front of the standing participant, but well below the participant's head to prevent interference with gunshot's direct sound wave (Figure 94). The participant was given instructions and received a demonstration on how to control the computer tablet and software program prior to testing. Prior to each trial, or blank gunshot, one of the investigators seated behind the participant set the rotary dial to the proper firing position and turned on the control box switch. The investigator then informed the participant that the system was armed by saying "READY." When ready, the participant selected the 'Click to START' button on the computer tablet. The LabView program then signaled to the USB-controlled relay switch to close the corresponding switch located inside the control box. Closing the switch allowed 24 Volts, supplied by the two car batteries, to be routed through relay switch, to the investigator-activated rotary dial, and to the desired firing mechanism. Six randomly-generated firing sequences were preprogramed into the LabView software. One sequence was randomly assigned to each posttest with the constraint that a unique sequence was used for all three listening conditions. The investigator used a sequence order checklist to ensure the correct rotary dial selection. The hard copy sequence checklist was synchronized with the order in LabView to enable automated scoring according to absolute and ballpark criteria. Just as in the in-lab study, after signal presentation the participant selected their response on the 24-icon display via stylus. The participant was then prompted via display to speak their response. A second investigator seated behind the participant recorded the verbalized response as a backup data source. The participant response on the computer tablet triggered response timer offset and recorded the time to a Comma CSV file. The software program recorded the subject number, listening condition, test type (azimuth or elevation), loudspeaker source location, participant response location, response time, date, time, and stage of training for each trial. The software program also calculated and displayed a running total using the absolute and ballpark criteria for the test for the investigator (example shown in Figure 86). The CSV file was saved to a Dropbox folder and shared through a mobile hotspot so that participant scores could be accessed after each test by the investigator.

Figure . Participant-controlled computer tablet placed on music stand at the center of the field experiment site.

Phase III: Experimental Procedures

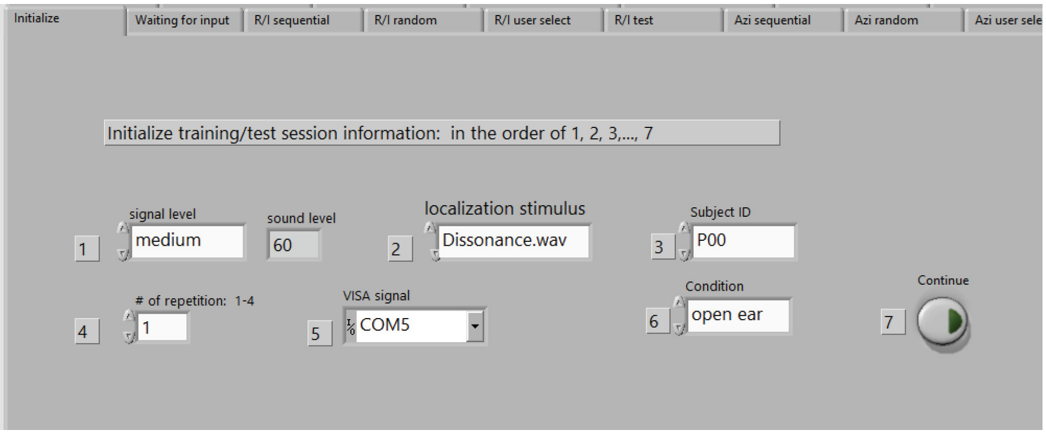
The experimental procedure for this investigation involved four distinct phases: 1) recruitment and screening, 2) pretest, 3) training (experimental condition only), and 4) posttest. The following sections detail the experimental procedures for each of these phases.

Recruitment and screening

Participants were recruited from the Virginia Tech community and surrounding areas via posted flyers (Appendix N), emails, and word of mouth. Potential participants contacted one of the investigators through email and a screening date was scheduled. A copy of the Phase III informed consent (Appendix O) was sent by email to the interested participant prior to their scheduled screening date for review. The potential participants were notified that a hard copy of the informed consent would also be provided at the time of their screening session. At the time of the screening, one of the investigators read aloud through the informed consent with the participant and answered all of the participant's questions. The participant was then given as much time as needed to read through the informed consent before signing. After signing the informed consent, an otoscopic inspection and an audiogram were administered (discussed in section 3.3.1). Following the audiogram, the participant was screened to ensure adherence with demographic requirements (Appendix D). The participant was then scheduled for the remainder of their tests and training sessions.

In-office pretest

On each day and prior to participant arrival, the investigator conducted a calibration of the PALAT system to ensure signal delivery of approximately 70 dBA. The investigator then populated the LabView fields with the participant number, listening condition, and auditory stimulus (Figure 96).



Initialize training/test session information: in the order of 1, 2, 3,..., 7

1 signal level: medium
sound level: 60
2 localization stimulus: Dissonance.wav
3 Subject ID: P00
4 # of repetition: 1-4: 1
5 VISA signal: COM5
6 Condition: open ear
7 Continue

Figure . PALAT system initialization screen.

Upon arrival, the participant was given an overview of the PALAT system. Participant instructions appear in Appendix P. The participant was informed of the purpose of the study and given a demonstration of how to operate the computer tablet user interface and localization software program. The investigator ensured the participant was seated in the center of the loudspeaker array. For TCAPS listening conditions, the investigator ensured the TCAPS were turned on and set to the unity gain prior to fitting the participant. Then, to ensure consistency of proper fit, the participant was fitted with their assigned TCAPS device by the investigator. The investigator ensured the TCAPS devices were comfortable and informed the participant to notify the investigator if they experienced any discomfort or acoustic feedback from the TCAPS. Figure 97 shows how a participant was seated and operated the PALAT system for each pretest and training session. The investigator was present in the room during every session. The investigator sat outside of the loudspeaker array behind or to the side of the participant and guided the participant through the testing and training.

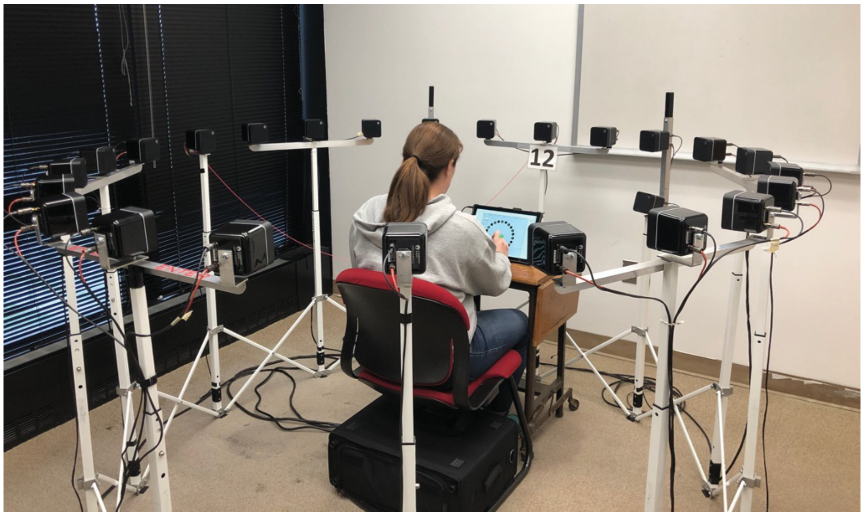


Figure . Participant operating the PALAT system.

Prior to each pretest, the participant received a familiarization unit consisting of a total of four presentations of the dissonant signal from the 12, 3, 6, and 9 o'clock loudspeaker positions. The familiarization unit was conducted to orient the participant to the PALAT system, familiarize the participant with the dissonant tonal signal, and allow the participant to practice operating the computer tablet. To perform the familiarization, the investigator selected the 'H sequential training' button from the main menu screen (Figure 98). The investigator selected Learning Unit 0 (LU0) from the dropdown menu on the sequential training screen and then selected the 'Start' button (Figure 99). The participant was then handed the stylus pen and instructed to wait for the investigator to position themselves outside of the loudspeaker array. Once in position, the investigator instructed the participant to begin the familiarization whenever they were ready by clicking on the 'Click to Sound' button located in the center of the sequential training screen and to respond by touching one of the 24 response buttons represented by the black circles on the sequential training screen. Prior to starting the familiarization unit, the participant was informed that the sequence of signal presentation would be emitted from 12 o’clock, then 3 o’clock, then 6 o’clock, and lastly 9 o’clock. However, the participant was allowed to respond on the computer tablet with any of the 24 response options. The participant was instructed to face forward and look at the white sign with the black number ‘12’ prior to clicking on the ‘Click to Sound’ button. The participant was informed that they were free to move their head and rotate their body at the onset of the dissonant signal to aid in localization and target identification. Head movement can be used to overcome a lack of localization cues by creating momentary changes in the sound spectrum at each ear. As a result, localization errors are reduced when the listener is allowed to move their head (Muller & Bovet, 1999; Thurlow & Mergener, 1970). Head movements were allowed in this experiment in order to more closely replicate a U.S. service member operational situation. The participant was reminded to “respond as accurately and as quickly as possible.” During the familiarization unit, the participant was allowed to ask the investigator questions. The investigator assisted the participant if they were having trouble operating the system. The PALAT system automatically returned to the main menu screen at the completion of the familiarization unit.

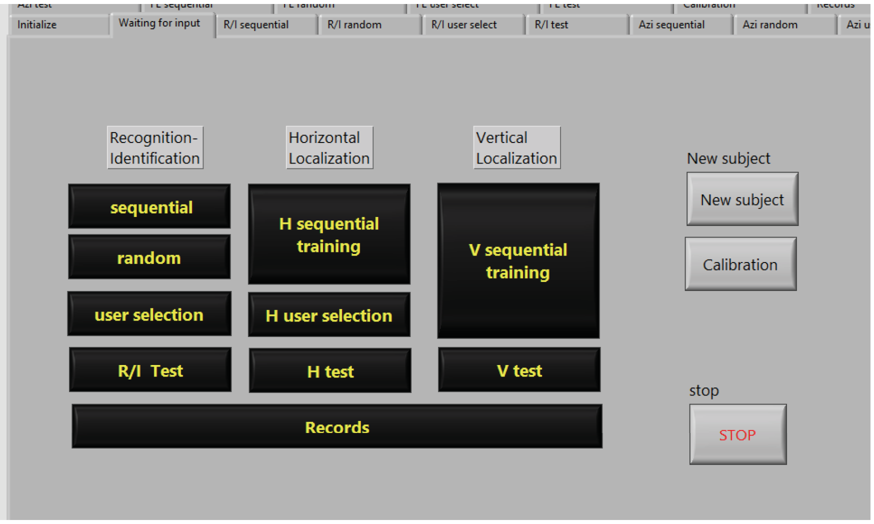
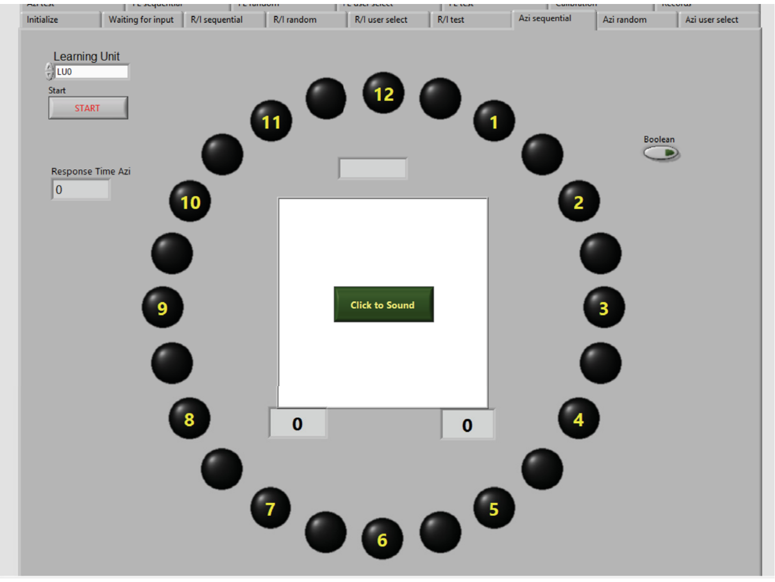


Figure . PALAT system main menu screen on computer tablet.

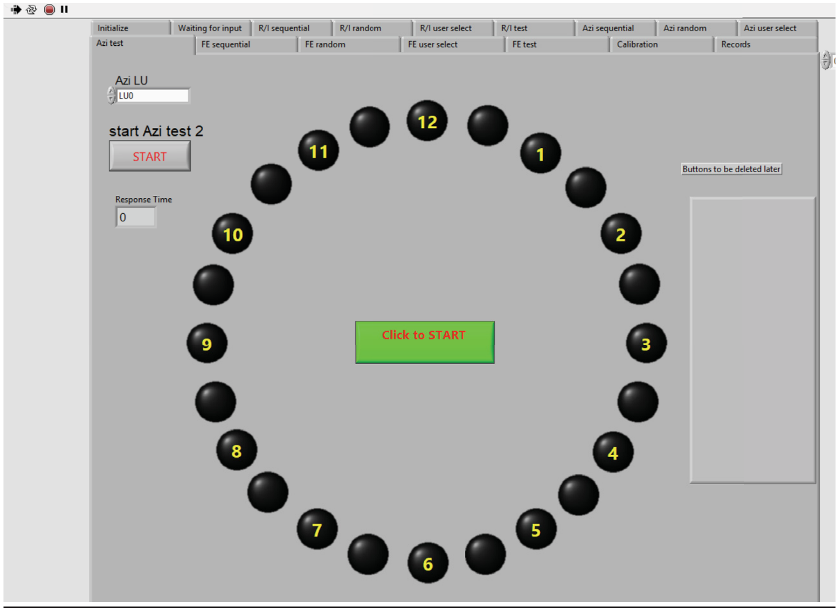


Learning Unit LU0

Response Time Azi 0

Figure . PALAT system sequential training screen on computer tablet.

Following the familiarization, the participant was asked if they had any questions about the task or how to operate the computer tablet. The participant was not allowed to retake the familiarization but the investigator would answer any questions and re-demonstrate how to use the computer tablet if the participant was confused. Once the participant was ready, the investigator would select the ‘H test’ button from main menu to navigate to the PALAT system testing screen (Figure 100). The testing screen was built to look very similar to the training screen to reduce operating errors. The main difference between the training and testing screens was the removal of the white box in the middle of the training screen that provided feedback to the participant. The investigator then selected Learning Unit 0 (LU0), or pretest, from the dropdown menu and clicked on the ‘Start’ button to initialize the system (Figure 100). The participant was then handed the stylus pen and instructed to wait for the investigator to get setup outside of the loudspeaker array. The participant was reminded that the pretest consisted of 36 random presentations, or trials, with each of the 12 numbered loudspeakers playing three times during the test. The participant was instructed to select on the touchscreen where they perceived the dissonant tone originated by clicking on one of the 24 black circles representing the 12 loudspeaker locations and 12 positions between each loudspeaker. The participant was instructed to face forward and look at the white sign with the black number ‘12’ prior to clicking on the ‘Click to START’ button. The participant was informed that they were free to move their head and rotate their body at the onset of the dissonant signal to aid in localization. The participant was then instructed to “respond as accurately and as quickly as possible.”


Azi LU: LU0

Response Time: 0

Figure . PALAT system test screen on computer tablet.

Once in position, the investigator instructed the participant to begin the pretest whenever they were ready by clicking on the ‘Click to START’ button located in the center of the testing screen and to respond by touching one of the 24 response buttons represented by the black circles on the sequential training screen. Each time the participant selected the ‘Click to START’ button, a dissonant signal was emitted from one of the 12 loudspeakers located at the 12 clock face positions while triggering the response timer onset. Upon selecting one of the 24 response buttons, the response timer stopped and a new row of data was stored in the CSV file. At the completion of the 36 presentations and participant responses, the PALAT system informed the participant that the test was completed by a pop-up window stating, 'This completes the test.' The system returned to the main screen after the participant clicked the 'Ok' button.

Following the pretest under a TCAPS listening condition, the investigator removed and turned off the TCAPS device. The investigator then accessed the pretest score from the Dropbox file and informed the participant of their absolute and ballpark score. The participant was then asked to complete a questionnaire on the computer tablet. The investigator was available to answer any questions concerning the questionnaire. After completing the questionnaire, the investigator changed the PALAT system listening condition in the computer tablet and prepared the participant for the next pretest. The familiarization unit and pretest were repeated for the remaining two listening conditions. Following the completion of all three pretests, the investigator confirmed to the participant the schedule for the next training session or field posttest.

Training Session (Experimental group only)

The auditory localization training employed in this study was originally designed by Lee & Casali (2017) and modified during Phase I of the overarching experiment. Participants randomly assigned to the experimental group conducted three, one-hour localization training sessions consisting of five learning units (LUs). The participant underwent training with one listening condition per session. Each LU consisted of the following subunits:

1) Sequential- For LU 1, the dissonant signal played in sequential order around the 12-loudspeaker array for four "laps," for a total of 48 presentations. The progression of the sequential presentations was as follows:

a) starting at 12 o'clock and moving clockwise through 11 o'clock,
b) starting at 9 o'clock and moving counterclockwise through 10 o'clock

c) starting at 3 o’clock and moving clockwise through 2 o’clock, and

d) starting at 6 o’clock and moving counterclockwise through 5 o’clock

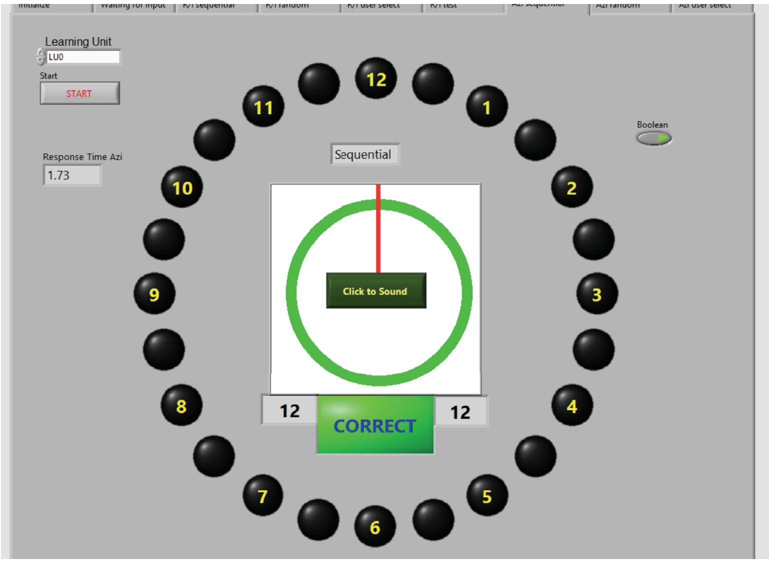
For LU 2-5, the system delivered only one “lap” around the clock face, totaling 12 presentations, with a randomly-assigned starting position and direction of progression. As in the familiarization, the participant had prior knowledge of the signal location via computer tablet display. Figure 101 displays the feedback for an absolute correct response. Figure 102 displays the feedback for an incorrect response.

2) Random- The participant did not have prior knowledge of which loudspeaker would present the signal, but knew immediately if he/she answered correctly. The signal was played from each signal location three times in random order, for a total of 36 presentations.

3) User-select- The participant had 18 trials to choose loudspeaker locations from which they wished to hear additional presentations (Figure 103).

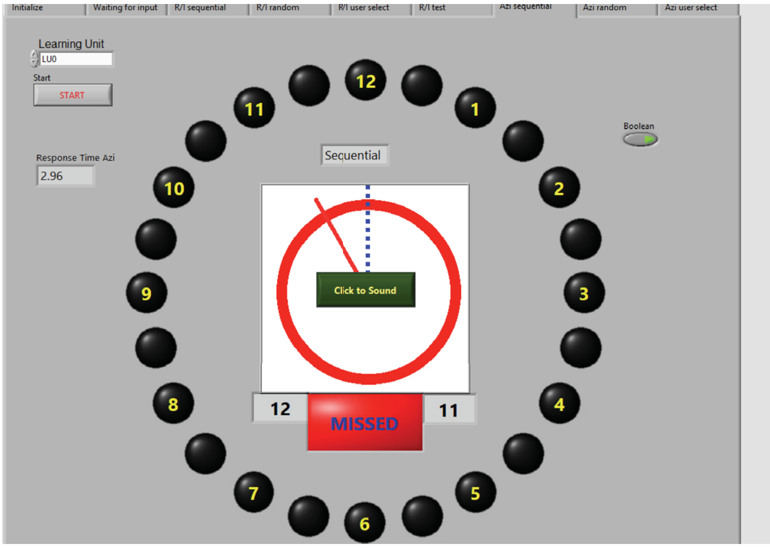
4) Test- A signal played from each of the 12 loudspeaker locations three times in random order, for a total of 36 presentations. The participant did not have prior knowledge of the sound location nor did they receive immediate feedback regarding their response.

At the end of each session, the participant completed the same questionnaire administered after the pretest.


Learning Unit LU0

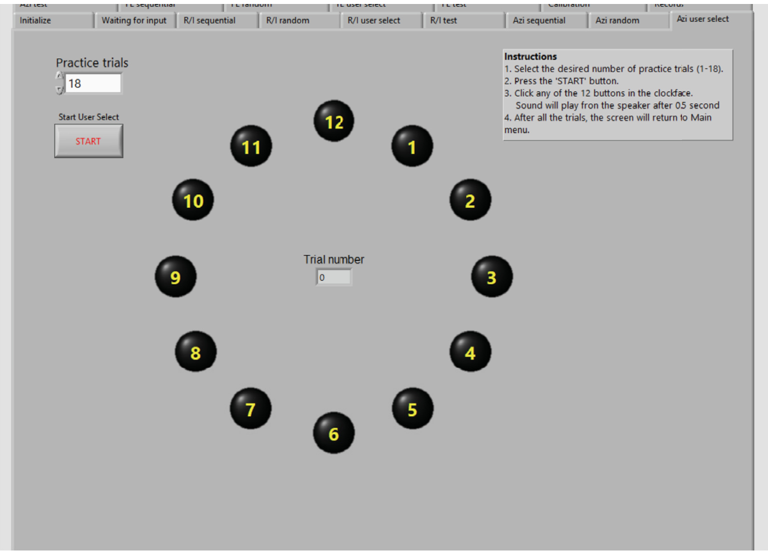
Response Time Azi 1.73

Figure . Participant interface on the computer tablet demonstrating the 24 response location options and system-generated feedback for an absolute correct response.


Learning Unit LU0

Response Time Azi 2.96

Figure . Participant interface on the computer tablet demonstrating feedback provided for an incorrect response.



Instructions
1. Select the desired number of practice trials (1-18).
2. Press the 'START' button.
3. Click any of the 12 buttons in the clockface.
Sound will play fron the speaker after 0.5 second
4. After all the trials, the screen will return to Main menu.

Figure . Example of the user-select display interface on the computer tablet.

In-field posttest

The participant met the investigator at Whittemore Hall and were driven about 45 minutes to the field site. Upon arrival at the field site, the participant was offered bottled water, insect repellant, and sunscreen. The participant was then escorted on foot approximately 150 feet uphill to the 9 o'clock firing position. The investigator showed the participant the blank rounds and component parts of the remote firing device, for full disclosure purposes. The participant then walked approximately 200 feet up the hill to the center of the remote-firing device array. A sign with the number '12' was placed 50 feet from the center point to orient the participant and provide a visual reference point. The participant stood facing the 12 o'clock target, but was allowed to move his/her head to aid in localization. An investigator oriented the participant to the field site layout, the posttest procedure, and the response procedures. The participant the signal and registered their response using the same computer tablet as used during the pretest and localization training. After responding on the computer tablet, the participant was instructed to verbalize their response location so that the investigator could record the response as a backup data source. Following instructions for TCAPS posttests, the investigator fitted the participant with the TCAPS device. The two investigators were seated behind the participant between the 6 o'clock and 7 o'clock positions (Figure 104). One investigator observed the participant and recorded the response locations. The other investigator operated the control box used to route the electrical signal to the firing mechanism containing the blank .22 caliber rounds.

Figure . Field posttest site layout with participant standing in center of remote firing devices facing 12 o'clock target position and two investigators seated behind participant between the 6 o'clock and 7 o'clock positions.

Prior to each posttest, the participant received a familiarization unit similar to the pretest. The participant was instructed prior to starting the familiarization unit that a blank gunshot would be initiated from 12 o'clock, 3 o'clock, 6 o'clock, and lastly 9 o'clock. The participant was instructed to respond to the signal as accurately and quickly as possible. As in the training and pretest, the participant had the option of selecting one of 24 possible blank gunshot locations, 12 active gunshot locations and 12 inactive. Following the familiarization unit, the participant was administered a posttest. As in the pretest and LU-generated tests, the participant heard three presentations, or gunshots, from each of the 12 locations in a randomized order for a total of 36 presentations. The participant conducted one posttest under each of the three listening conditions in the same counterbalanced order as their pretest.

After each listening condition and at the conclusion of the posttest, the participant completed the same questionnaire used during the pretest and localization training. The trained group had one additional question that queried their perceived degree of preparedness as a result of lab-conducted training.

The wind speed was measured and recorded at the start of every posttest. Mean, minimum and maximum wind speeds for all posttests are shown in Table 83. No posttests were conducted if wind gusts measured above 8 miles per hour (mph). In addition, the posttests were suspended in inclement weather more than a very light rain mist. The temperature and humidity were measured and recorded at the start of each posttest. During the posttest, the investigator monitored the wind speeds and weather conditions to ensure gusts did not exceed 8 mph. In order to mitigate weather delays during testing, the investigator monitored the weather forecast prior to scheduling the posttest. Testing ceased only for changes in wind versus those in temperature or humidity as wind caused masking effects. In other words, wind creates noise unrelated to the original sound. Whereas variation in humidity and temperature contributed to the external validity of the experiment without drastic change in the overall sound source.

Table . Weather conditions during the field test.

	Mean	Minimum	Maximum
Wind Speed	0.7 mph	0 mph	2 mph
Temperature	69°	50°	81°
Humidity	62%	38%	87%

Phase III: Results

Data reduction and calculations were performed using Microsoft® Excel. Statistical analysis was performed with JMP® 14 software, IBM® SPSS® Statistics, and Excel v16.16.10.

Outlier Analysis

After running 24 participants through the full experiment, a Dixon *Q*-test was performed on all dependent measures to identify outliers. The outlier analysis was performed separately each of the three quantitative dependent measures for each listening condition for pretest and posttest. The resulting sample size for each Dixon *Q*-test was n=24, 12 untrained and 12 trained participants. To perform the Dixon *Q*-test, the subset of data for each test was arranged sequentially from lowest to highest value. A Dixon *Q*-test was then performed manually using one of the following formulae:

$$(1) \qquad Q = \frac{|x_n - x_{n-1}|}{|x_n - x_1|} \quad \text{or} \quad Q = \frac{|x_2 - x_1|}{|x_n - x_1|}$$

where n is the sample size and the x represents the ordered values from $x_1 < x_2 < \ldots < x_n$ (Dixon, 1951). The numerator in equation (1) represents the gap between the two lowest values or the gap between the two highest values. The denominator in equation (1) represents the range of the data. The equation that resulted in the largest gap was used to identify the existence of a single outlier for each data subset. The calculated *Q*-value for each data subset was compared to Dixon's r_{10} table for n=24 using a 95% confidence interval (Dixon, 1951). If $Q \geq 0.34$, then an outlier was deemed present.

Two significant outlier data points were found using the Dixon *Q*-test. A significant outlier was present for the absolute correct score on the pretest under the TEP-100 listening condition (Q=0.37). The outlier data point of an absolute correct score=36 was associated with participant 23 from the untrained group. Figure 105 displays the absolute scores on the pretest under the TEP-100 listening condition by group.

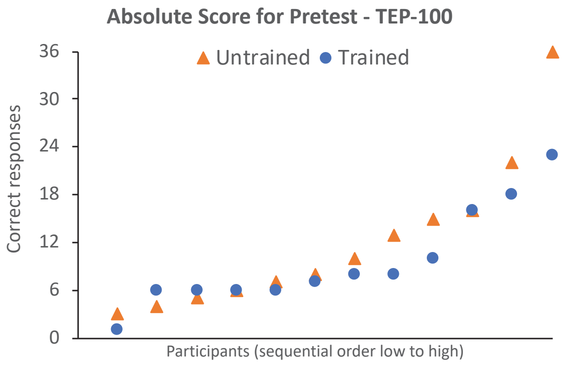


Figure . Absolute correct score on pretest under TEP-100 listening condition for all participants. Values ordered and plotted from lowest to highest score by group.

A significant outlier was present for response time on the posttest under the open ear listening condition (*Q*=0.37). The outlier data point of a response time=4.86 seconds was associated with participant 23 from the untrained group. Figure 106 displays the response times on the posttest under the open ear listening condition by group.

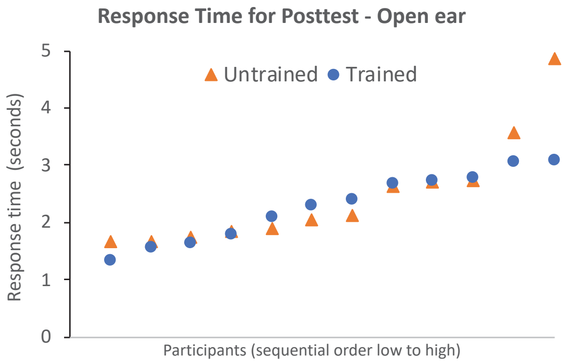


Figure . Response time on posttest under open ear listening condition for all participants. Values ordered and plotted from lowest to highest response time by group.

Both outlier data points were associated with participant 23 who was assigned to the untrained group. In both instances, participant 23's absolute correct score and response time would have biased the mean data in favor of *better* performance for the trained group. Additionally, the outlier would have increased the pretest absolute correct score for the untrained group, and thus, exaggerating the mean difference between pretest and posttest. Inclusion of participant 23's response time would have resulted in an exaggerated increase in response time from the pretest to the posttest. As a result, the investigator decided to replace outlier participant 23 with a new participant. This replacement was assigned to the same group (untrained) and counterbalancing scheme for listening condition (open ear, TEP-100, and ComTac™ III) as the outlier. All data points and analyses were then performed using the new participant 23's results. With the new data set inclusive of the outlier replacement, another Dixon *Q*-test was performed on all dependent measure data and no significant outliers were detected.

Objective performance

Analysis technique overview

Auditory localization performance on a set of measures (absolute correct, ballpark correct, and response time) according to training condition (untrained versus trained), listening condition (open ear, TEP-100, and ComTac™ III), and stage of training (pretest versus posttest) was first examined using a multivariate analysis of variance (MANOVA). The MANOVA evaluated the presence of significant mean differences according to experimental manipulation on a composite set of dependent measures. The test statistic employed in the MANOVA was Wilk's Lambda (λ), which represents the percentage of variance unexplained by the manipulation of the independent variables (Pituch & Stevens, 2016). Generally, the analysis stayed within the significance level for the mixed factor MANOVA using $\alpha=0.05$. However, given the lack of experimental control in the field study, the α level was set to a less stringent α=0.10 for follow-up univariate tests in order to further analyze results.

Another statistic which was applied was partial eta squared (η^2_p), which represents the ratio of total variance in the sample that can be explained by the factor (Pituch & Stevens, 2016). Some of the effect sizes exceeded an η^2_p criterion for medium or large effect sizes of .06 and .14, respectively (Cohen, 1988), implying a meaningful effect. In regards to power, given that complement of the power value is the likelihood of making a Type II error, a power value of 0.8 is generally considered acceptable (Cohen, 2013). However, given the nature of the field study, 0.7 was considered as an acceptable power value.

The degree of auditory localization needed to perform ground combat-related tasks has yet to be validated or quantified. As such, the ballpark correct localization criterion (i.e., a response within ±15° of the location of the actual presented signal) was included in the analysis to describe performance using a less stringent standard should this requirement become known. The intent of applying different criteria to results of this study was to enable interpretation and relevancy according to differing standards. However, the investigator maintains that the 30° (±15°) accuracy criterion enables vision to better overcome localization blur. Azimuthal separation factor has implications for point of sound origin separation distance at effective range of military relevant signals. Auditory localization orients the listener to the direction of the sound and cues the visual modality, effectively reducing the response time in target identification (Wickens et al., 2013). A wider azimuthal separation angle at the listener translates to a broader field of view search as the distance increases from the listener. Larger separation angles become problematic in ground combat scenarios where military threats originate from long distances. Table 84 shows a comparison of the resulting visual field search distances associated with 30°

and 45° azimuthal separation for military weapons originating from their effective range distances.

Table . Visual field of view search distances associated with 30° and 45° azimuthal separation for military threats originating from their effective range distances (USMC, 2017)

Military Threat	Effective Range (m)	Field of view search distance at effective range (m)		
		30°	45°	Difference (Δ)
AK-47 gunshot	300	155	229	74
Rocket Propelled Grenade (RPG)	500	258	383	125
AK-74 Sniper rifle	800	414	612	198
PKM machine gun	1000	517	756	239
82mm mortar launch	3000	1553	2296	743
107mm rocket launch	> 5000	2588	3826	1238

The analysis included a within-subjects factor with more than two levels. Therefore, Mauchly's Test of Sphericity evaluated the assumption of homogeneity of variance between related of group comparisons (Portney & Watkins, 2009). This test is not applied to the between-subjects variable or a variable with only two levels, such as stage of training in this study. *F*-tests on a source of variance that is associated with violations of the homogeneity of variance assumption can underestimate the likelihood of making a Type I error (Portney & Watkins, 2009). Mauchly's Test of Sphericity evaluates the need for adjusting the *p* value when violations are detected (Portney & Watkins, 2009). A significant *p*-value in the Mauchly's Test indicates a violation occurred and an adjustment to reduce the degrees of freedom is made (Portney & Watkins, 2009). As a result, the critical value needed to achieve a significant finding is greater, correcting for the greater likelihood of making a Type I error (Portney & Watkins, 2009). Two estimators, the Greenhouse-Geisser and Huynh-Feldt, provide measures of Epsilon (ϵ) that describes the deviation of the covariance matrix from sphericity, and both were applied with the results in Table 85. A value of 1 indicates no deviation, and thus adherence to the assumption of sphericity (Pituch & Stevens, 2016). The Greenhouse-Geisser and Huynh-Feldt can underestimate and overestimate ϵ , respectively (Pituch & Stevens, 2016). As such the Greenhouse-Geisser is usually the first estimate applied if the correction results in a significant finding (Portney & Watkins, 2009). If results are not significant, the Huynh-Feldt is typically applied (Portney & Watkins, 2009). Mauchly's Test of Sphericity (Table 85) did not result in any significant values; therefore, the assumption of sphericity was met for the MANOVA and no corrections were required.

Following the MANOVA, univariate ANOVAs were conducted. Main effects were followed up with pairwise comparisons using a Bonferroni adjustment. Simple-effects *F*-tests followed up any interactions in order to partition the data according to the sources of most interest. Given that each factor analyzed in the simple-effects *F*-tests procedure only had two levels, follow-up pairwise comparisons were not conducted.

Finally, in graphing the data, for most graphs that follow, arithmetic mean values are plotted in bar graph form, with 95% confidence limits shown about the mean, and means with different letters indicative of statistical significance. In addition, in all tables, statistically-significant effects are denoted by boldface font.

Results: Evaluation of transfer-of-training effects from the in-lab to in-field localization performance

The mixed-factors MANOVA (Table 85) for the effect of training group, listening condition, and stage of training on the set of dependent measures showed a significant effect for listening condition, Wilk's λ (6, 17=0.04, F=1.79, p< 0.000). Follow-up univariate ANOVAs were conducted for each dependent measure (Table 86). For the between-subjects variable of group, a significant difference was found using the measure of ballpark score, F(1, 22)=3.05,

p=0.095 , η_p^2=0.12. For the within-subjects variable of listening condition, significant differences existed using the measures of absolute correct score, $F(2, 44)$=120.44, p<0.000, η_p^2=0.85, ballpark correct score, $F(2,44)$=143.94, p<0.000, η_p^2=0.87, and response time $F(2, 44)$=11.35, p<0.000, η_p^2=0.34.

Table . Mauchly's Test of Sphericity and MANOVA results evaluating the effects of training group, listening condition, and stage of training on absolute score, ballpark score, and response time (bolded text indicates a significant test result at the α=0.05 significance level).

Mauchly's Test of Sphericity on Within-Subjects Variables						Epsilon (ϵ)	
Variables	Measure	Mauchly's Criterion	Chi-Square	df	p	Greenhouse-Geisser	Huynh-Feldt
Listening Condition	Absolute	0.91	1.92	2	0.384	0.92	1
	Ballpark	0.77	5.47	2	0.065	0.81	0.91
	RT	0.97	0.58	2	0.750	0.97	1
Listening Condition x Stage of Training	Absolute	0.99	0.25	2	0.882	0.99	1
	Ballpark	0.99	0.17	2	0.920	0.99	1
	RT	0.92	1.74	2	0.419	0.93	1

Source	Wilk's λ	df	F value	p	η_p^2
Between-Subjects					
Training Group (Trained; Untrained)	0.88	(3, 20)	0.92	0.447	0.12
Within-Subjects					
Listening Condition (Open; TEP-100; ComTac™III)	**0.04**	**(6, 17)**	**1.79**	**< 0.000**	**0.96**
Stage of Training (Pretest; Posttest)	0.73	(3, 20)	2.51	0.088	0.27
Stage of Training x Training Group	0.83	(3, 20)	1.41	0.270	0.17
Listening condition x Training Group	0.77	(6, 17)	0.83	0.566	0.23
Stage of Training x Listening condition	0.82	(6, 17)	0.61	0.720	0.12
Listening condition x Stage of Training x Training Group	0.61	(6, 17)	1.79	0.161	0.39

Table . Univariate ANOVA results for each dependent measure for each independent variable (bold text indicates significant values at the α=0.10 significance level).

Source	df	Mean Square	*F value*	*p*	η_p^2
Between Subjects					
Group (G)					
Absolute	1	100.00	1.39	0.252	0.06
Ballpark	**1**	**142.01**	**3.05**	**0.095**	**0.12**
Response Time	1	0.60	0.26	0.618	0.01
Error (S/G)					
Absolute	22	72.10			
Ballpark	22	46.62			
Response Time	22	2.34			
Within Subjects					
Listening Condition (C)					
Absolute	**2**	**3295.26**	**120.44**	**<0.000**	**0.85**
Ballpark	**2**	**3516.90**	**143.94**	**<0.000**	**0.87**
Response Time	**2**	**3.19**	**11.35**	**<0.000**	**0.34**
C x G					
Absolute	2	10.65	0.39	0.680	0.02
Ballpark	2	0.76	0.03	0.970	0.00
Response Time	2	0.33	1.19	0.310	0.05
Error (*C x S/G*)					
Absolute	44	27.36			
Ballpark	44	24.43			
Response Time	44	0.28			
Stage of training (T)					
Absolute	1	0.03	0.00	0.975	0.00
Ballpark	1	15.34	0.77	0.390	0.03
Response Time	**1**	**1.96**	**6.10**	**0.022**	**0.22**
T x G					
Absolute	**1**	**110.25**	**4.05**	**0.057**	**0.16**
Ballpark	1	47.84	2.39	0.136	0.10
Response Time	1	0.30	0.94	0.344	0.04
Error (*T x S/G*)					
Absolute	22	27.25			
Ballpark	22	19.98			
Response Time	22	0.32			
C x T					
Absolute	2	28.38	1.55	0.223	0.07
Ballpark	2	28.38	1.46	0.244	0.06
Response Time	2	0.06	0.60	0.553	0.03
C x T x G					
Absolute	**2**	**52.94**	**2.90**	**0.066**	**0.12**
Ballpark	2	21.05	1.08	0.348	0.05
Response Time	2	0.03	0.28	0.758	0.01
Error (C x T x S/G)					
Absolute	44	18.27			
Ballpark	44	19.47			
Response Time	44	0.09			
Total	429	7634.77			

Group Main Effect: Post hoc test for Ballpark Correct Score

Pairwise comparisons for the effect of group was not conducted given that only two levels of the independent variable, trained and untrained, were used in this experiment. Instead, mean ballpark scores were examined and showed that the trained group (*M*=20.39, *SD*=8.96) scored higher than the untrained group (*M*=18.40*, SD*=8.30). The means for each group and 95% confidence intervals are plotted below in Figure 107.

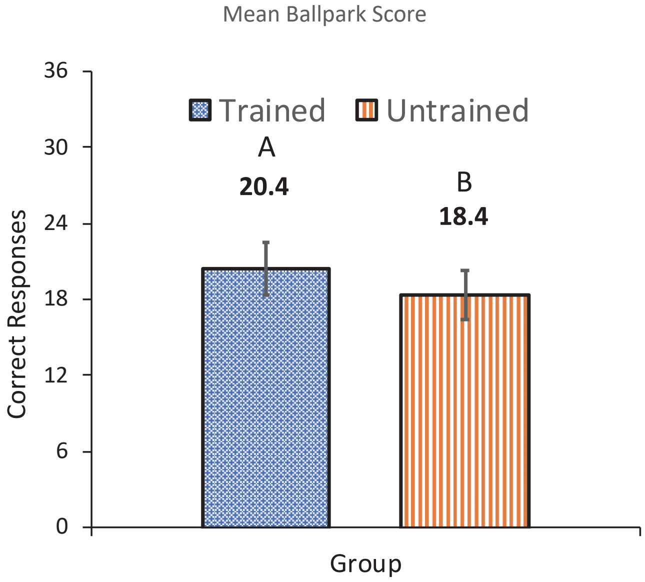


Figure . Mean ballpark correct scores with 95% confidence intervals plotted. Different letters indicate significant differences at $p \leq 0.10$.

Listening Condition Main Effect: Post hoc test for Absolute Correct Score

Pairwise comparisons were conducted for each listening condition (within the main effect of listening condition) using the measure of absolute correct score Table 87. All pairwise comparisons used a Bonferroni adjustment, which results in α=0.167 given that three comparisons were made (α=0.05/3). The mean absolute correct score for the open ear condition

(*M*=26) differed significantly from mean scores obtained in the TEP-100 (*M*=10.9) and ComTac™ III (*M*=12.6) conditions (Figure 108).

Table . Pairwise comparisons for listening condition using the absolute correct score.

Listening Condition		*M*	*SE*	*p*
Open ear	**TEP-100**	**15.15**	**1.01**	**<0.000**
Open ear	**ComTac™ III**	**13.40**	**0.91**	**<0.000**
ComTac™ III	TEP-100	-1.75	1.19	0.462

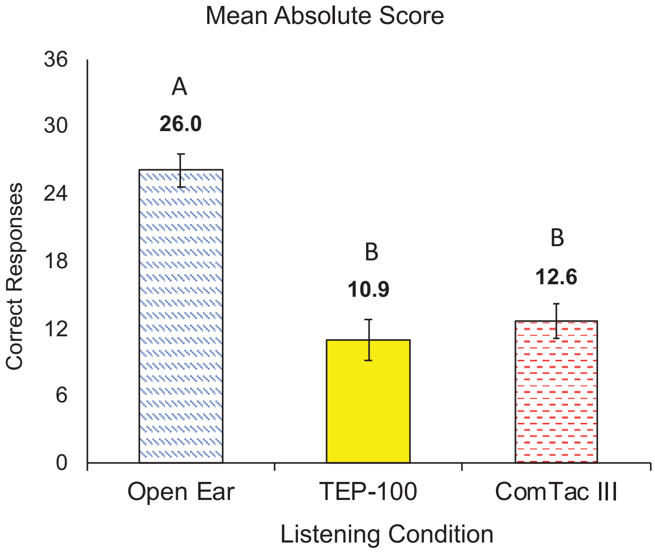


Figure . Mean absolute correct scores for each listening condition with 95% confidence intervals plotted. Different letters indicate significant differences at $p \leq 0.10$.

Listening Condition Main Effect: Post hoc test for Ballpark Correct Score

Pairwise comparisons conducted using the ballpark correct measure, shown in Table 88, also resulted in significant differences between the open (*M*=29.2) and TEP-100 (*M*=13.3) conditions and between the open and ComTac™ III (*M*=15.7) conditions. Mean ballpark scores within the main effect of listening condition are shown in Figure 109.

Table . Listening condition significant pairwise comparisons using the **ballpark** correct score.

Listening Condition		*M*	*SE*	*p*
Open ear	**TEP-100**	**15.85**	**1.04**	**<0.000**
Open ear	**ComTac™ III**	**13.52**	**0.76**	**<0.000**
ComTac™ III	TEP-100	2.33	1.19	0.186

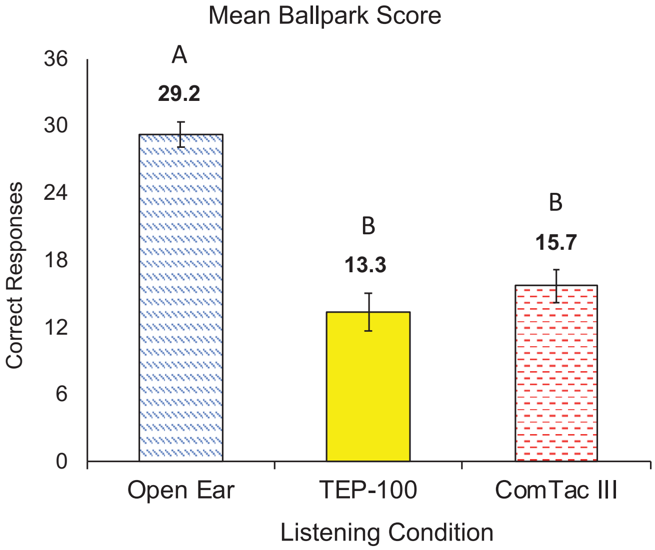


Figure . Mean ballpark correct scores for each listening condition with 95% confidence intervals plotted. Different letters indicate significant differences at $p \leq 0.10$.

Listening Condition Main Effect: Post hoc test for Response Time

Pairwise comparisons on the measure of response time in seconds, shown in Table 88, showed significant differences between the mean response time for open ear (*M*=2.2) and TEP-100 (*M*=2.6) and between the open ear and the ComTac™ III (*M*=2.6). Mean response times within the main effect of listening condition are shown in Figure 110.

Table . Listening condition significant pairwise comparisons using the **response time** score.

Listening Condition		*M*	*SE*	*p*
Open ear	**TEP-100**	**-0.48**	**0.12**	**<0.000**
Open ear	**ComTac™ III**	**-0.41**	**0.11**	**0.001**
ComTac™ III	TEP-100	-0.07	0.10	1.000

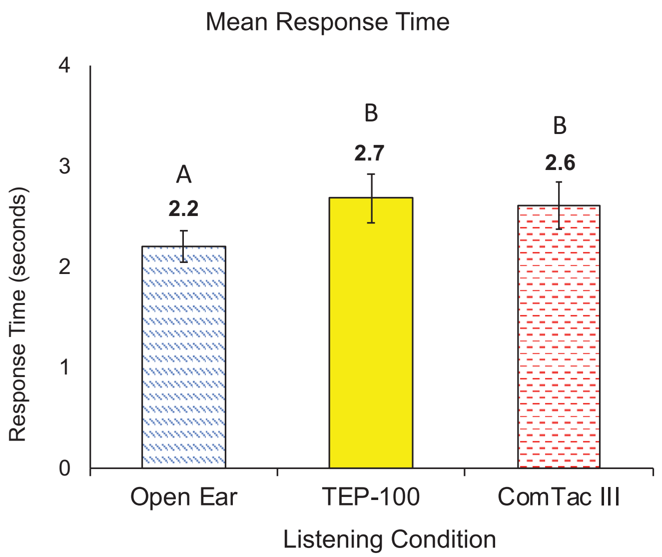


Figure . Mean response times for each listening condition with 95% confidence intervals plotted. Different letters indicate significant differences at $p \leq 0.10$.

Stage of Training Main Effect: Post hoc test for Response Time

The repeated measures ANOVA conducted for the stage of training main effect on the measure of response time was significant, $F(1,22)=6.10$, $p=0.022$, $\eta_p^2=0.2$. The mean response time at pretest ($M=2.4$) was significantly lower than the posttest ($M=2.6$) (Figure 111). Given only two levels of the independent variable, pairwise comparisons were not conducted.

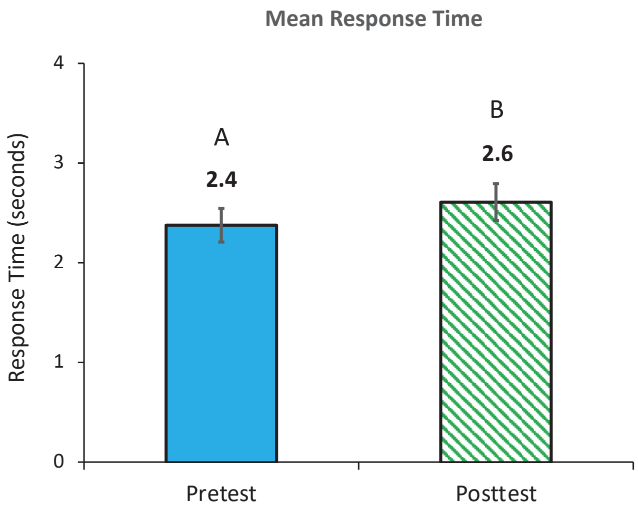


Figure . Mean response times from pretest to posttest. Different letters indicate significant differences at $p \leq 0.10$.

Stage of Training x Group Interaction: Post hoc test for Absolute Correct Score

Simple-effects *F*-tests further analyzed the significant interaction, at $\alpha = 0.10$, for stage of training (pretest versus posttest) by group (trained versus untrained) using the absolute correct localization measure. Specifically, to determine if the groups significantly differed at pretest and posttest, two between-subjects ANOVAs were conducted at each training stage. To evaluate the assumption of homogeneity of variances, Levene's tests were conducted. The Levene's test calculates the deviation scores of the participants within each group from the group mean and then coverts the scores to absolute values (Pituch & Stevens, 2016). An ANOVA is then conducted comparing the mean absolute deviation scores between groups (Pituch & Stevens, 2016). A result of $p < 0.05$ for the Levene's test indicates a violation to the assumption of homogeneity of the variance. In this analysis, the Levene's tests supported equality of the variances for the ANOVA conducted at the pretest stage examining group differences, $F(1,70)=1.06$, $p=0.297$, and at the posttest stage, $F(1,70)=1.45$, $p=0.233$; therefore, no corrections for heterogeneity were necessary. Simple-effects *F*-tests, shown in Table 90, ensued, yielding no significant differences between groups at pretest, *F*(1,22)=0.00, *p*=0.946, but exhibiting significance at posttest, *F*(1,22)=7.17, *p*=0.011. Mean absolute correct scores for each group at pretest and posttest are displayed in Figure 112. Results supported group equivalence at pretest collapsed across listening conditions, while showing significantly higher scores for the trained group over the untrained group at posttest.

Table . Simple-effect *F*-tests for trained group versus untrained group at each stage of training using the absolute correct score.

Source	*SS*	*df*	*MS*	*F*	*p*
G at Pretest	0.13	1	0.13	0.00	0.946
G at Posttest	**210.13**	**1**	**210.13**	**7.71**	**0.011**
Error (T x S/G)	599.40	22	27.24		
Total	809.65	24			

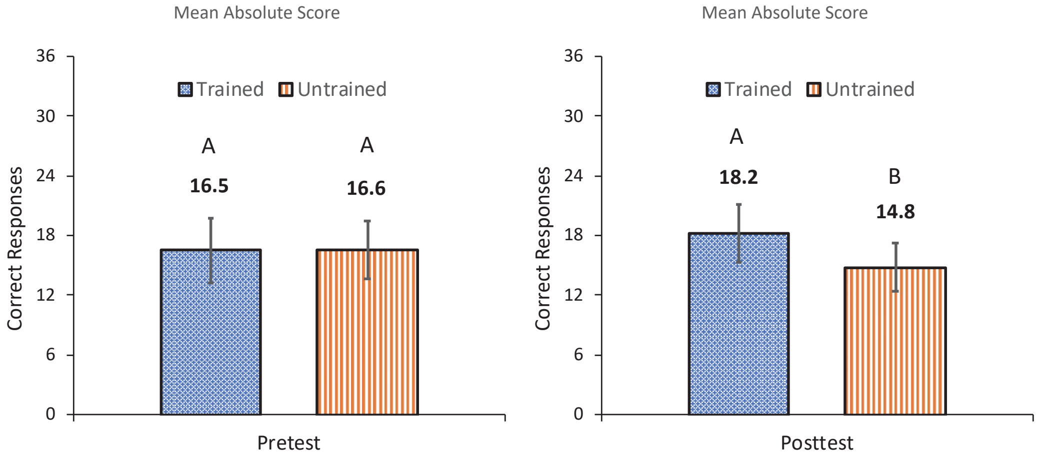


Figure . Mean absolute correct scores for each group at pretest and posttest. Different letters indicate significant differences at $p \leq 0.10$.

To further analyze the significant interaction of stage of training and group on the measure of absolute correct, simple-effects *F*-tests ANOVA were conducted. Separate ANOVAs were run for each group comparing performance between pretest and posttest. Analyzing only two levels of the repeated measures precluded sphericity testing. No significant differences were found between pretest and posttest absolute scores in the untrained group, $F(1, 22)=2.09$, $p=0.163$, or the trained group, $F(1, 22)=1.96$, $p=0.175$. ANOVA results are listed in Table 91 and means are displayed in Figure 113.

Table . Simple-effects *F*-tests for each group examining pretest versus posttest performance collapsed across listening conditions using the absolute correct score.

Source	*df*	*SS*	*MS*	*F*	*p*
Untrained	1	56.89	56.89	2.09	0.163
Trained	1	53.39	53.39	1.96	0.175
Error (T x S/G)	22	599.40	27.24		

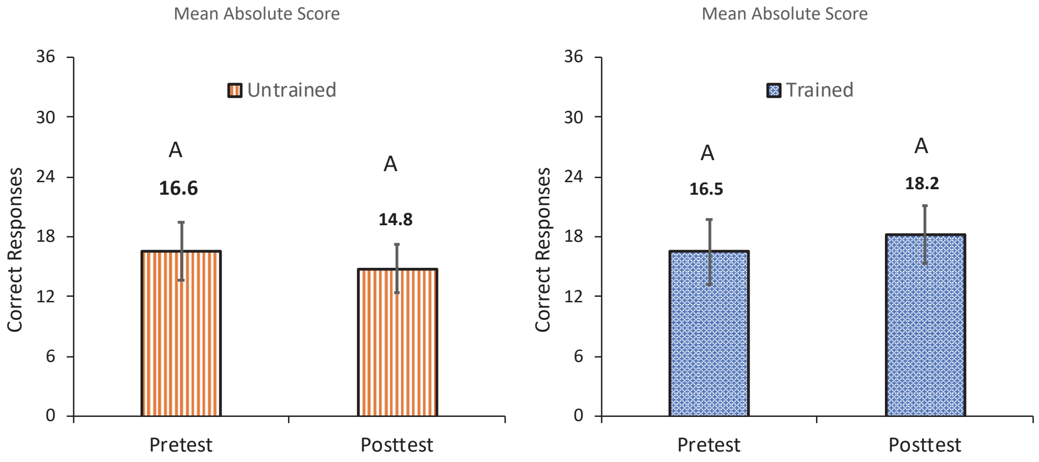


Figure . Mean absolute correct scores for each training group comparing pretest and posttest performance. Different letters indicate significant differences at $p \leq 0.10$.

Listening Condition x Stage of Training x Group Interaction: Post hoc test for Absolute Correct Score

To analyze the significant interaction of listening condition, stage of training, and group on the absolute correct measure, simple-effects *F*-tests were conducted. Separate ANOVAs conducted for each listening condition (open ear, TEP-100, and ComTac™ III) evaluated differences in scores for the trained versus untrained group at each stage of training. The Levene test supported equality of variances at pretest for the open ear, *F*(1,22)=0.13, *p*=0.726 and TEP-100, *F*(1,22)=0.09, *p*=0.771, but not for the ComTac™ III *F*(1,22)=6.10, *p*=0.022. However, ANOVA *F*-tests tend to be robust to violations of variance equality given equal group sizes, as was the case in this study. Simple-effects results, listed in Table 92 and Figure 114, show no significant differences between trained and untrained participants at pretest for the open ear, *F*(1, 44)=0.23, *p*=0.634, the TEP-100, *F*(1, 44)=0.33, *p*=0.569, and the ComTac™ III, *F*(1, 44)=0.82, *p*=0.370. Thus, these results indicated group equivalence at pretest in each listening condition. The simple-effects test above was then repeated, comparing the trained versus untrained group for each listening condition, but for the posttest stage of training. Levene's statistic supported equality of variances for the open ear, *F*(1,22)=0.003, *p*=0.959, TEP-100, *F*(1,22)=0.427, *p*=0.520, and ComTac™ III, *F*(1,22)=2.46, *p*=0.131 (Table pp). The *F*-tests showed significant differences between the trained and untrained groups for the open ear, *F*(1, 44)=13.18, *p*=0.001, and the TEP-100 conditions, *F*(1, 44)=3.83, *p*=0.057. In all listening conditions, the mean absolute correct score was higher for the trained group when tested in the field, as shown in Table 93 and Figure 115.

Table . Simple-effect *F*-tests examining absolute correct score differences at **pretest** in trained versus untrained participants for each listening condition (open ear, TEP-100, and ComTac™ III)

Source	*df*	*SS*	*MS*	*F*	*p*
Open ear	1	4.17	4.17	0.23	0.634
TEP-100	1	6.00	6.00	0.33	0.569
ComTac™ III	1	15.04	15.04	0.82	0.370
Error (C x T x S/G)	44	817.08	18.27		
Total	47	842.29	43.48		

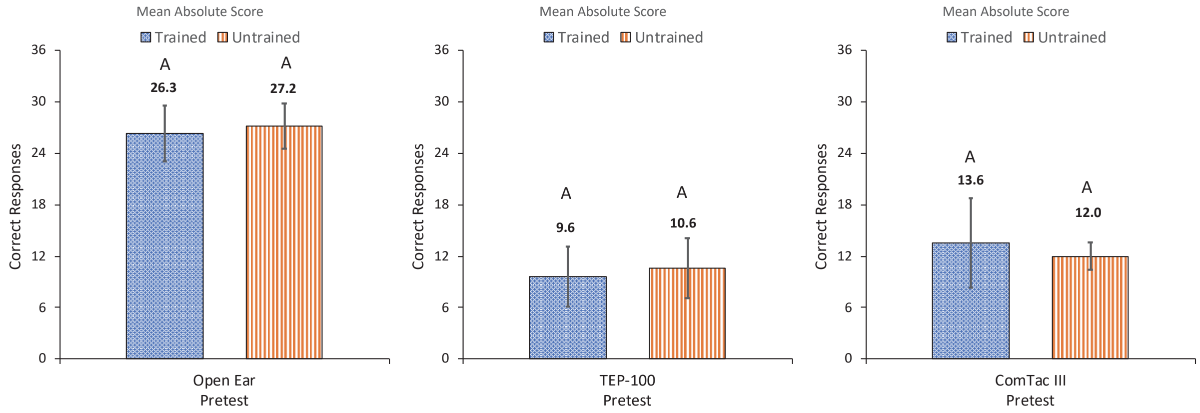


Figure . Mean absolute correct scores for the trained versus untrained groups at pretest for each listening condition. Different letters indicate significant differences at $p \leq 0.10$.

Table . Simple-effect *F*-tests examining absolute correct score differences at **posttest** in trained versus untrained participants for each listening condition (open ear, TEP-100, and ComTac™ III)

Source	*df*	*SS*	*MS*	*F*	*p*
Open ear	**1**	**240.67**	**240.67**	**13.18**	**0.001**
TEP-100	**1**	**70.04**	**70.04**	**3.83**	**0.057**
ComTac™ III	1	1.50	1.50	0.08	0.227
Error (Listening Condition x Stage of training/Group*)*	44	817.08	18.27		
Total	47	1129.29	330.48		

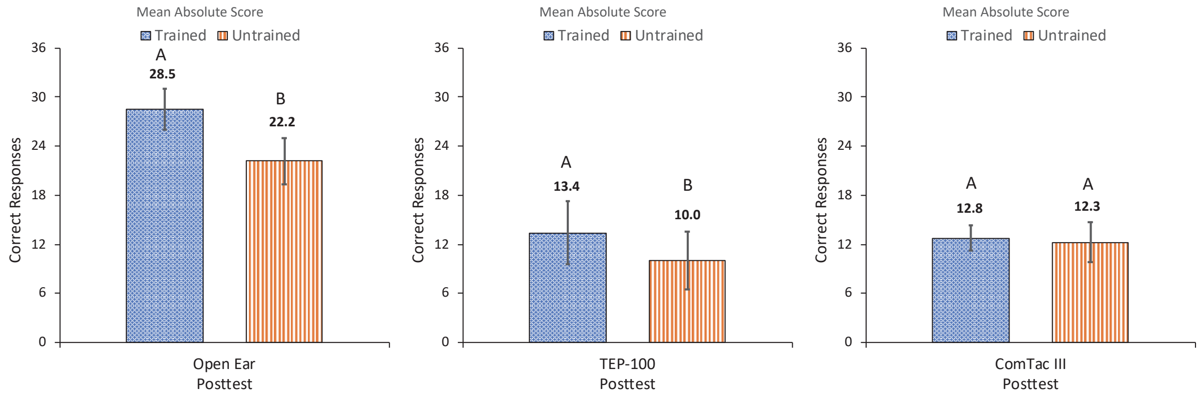


Figure . Mean absolute correct scores at posttest for each listening condition comparing the trained and untrained groups. Different letters indicate significant differences at $p \leq 0.10$.

Additional simple-effects testing was conducted on the 3-way interaction of listening condition, stage of training, and group. Simple-effects *F*-tests examined pretest versus posttest performance for each group and for each listening condition. The untrained group demonstrated significantly lower mean performance (Figure 116) in the open condition in the posttest versus pretest, *F*(1, 44)=8.21, *p*=0.006 (Table 94). Significant differences between pretest and posttest were not found for the TEP-100, *F*(1, 44)=0.021, *p*=0.885, or ComTac™ III conditions, *F*(1, 44)=0.021, *p*=0.885. For the trained group, the TEP-100 condition resulted in significantly higher scores in the posttest versus the pretest, *F*(1, 44)=4.83, *p*=0.033. No significant differences were found in the open ear, *F*(1, 44)=1.54, *p*=0.221, or ComTac™ III conditions, *F*(1, 44)=0.23, *p*=0.634. Results of the trained group simple-effects *F*-tests and means are provided in Table 95 and Figure 117, respectively.

Table . Simple-effect *F*-tests examining absolute correct score differences at pretest versus posttest for the **untrained** group. Different letters indicate significant differences at $p \leq 0.10$.

Source	*df*	*SS*	*MS*	*F*	*p*
Open ear	**1**	**150.00**	**150.00**	**8.21**	**0.006**
TEP-100	1	2.04	2.04	0.11	0.742
ComTac™ III	1	0.38	0.38	0.02	0.885
Error (C x T x S/G)	44	803.88	18.27		
Total	47	956.3	170.69		

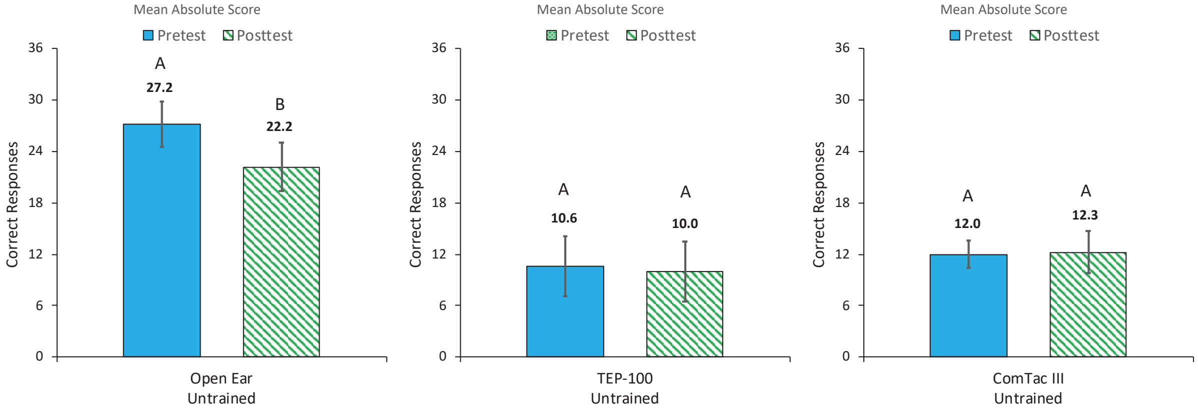


Figure . Mean absolute correct scores for each listening condition for the **untrained** group **comparing pretest and posttest** performance. Different letters indicate significant differences at $p \leq 0.10$.

Table .Simple-effect *F*-tests examining absolute correct score differences at pretest versus posttest for the trained group.

Source	*df*	*SS*	*MS*	*F*	*p*
Open ear	1	28.17	28.17	1.54	0.221
TEP-100	**1**	**88.16**	**88.16**	**4.83**	**0.033**
ComTac™ III	1	4.17	4.17	0.23	0.634
Error (C x T x S/G)	44	817.08	18.27		
Total	47	937.58	138.77		

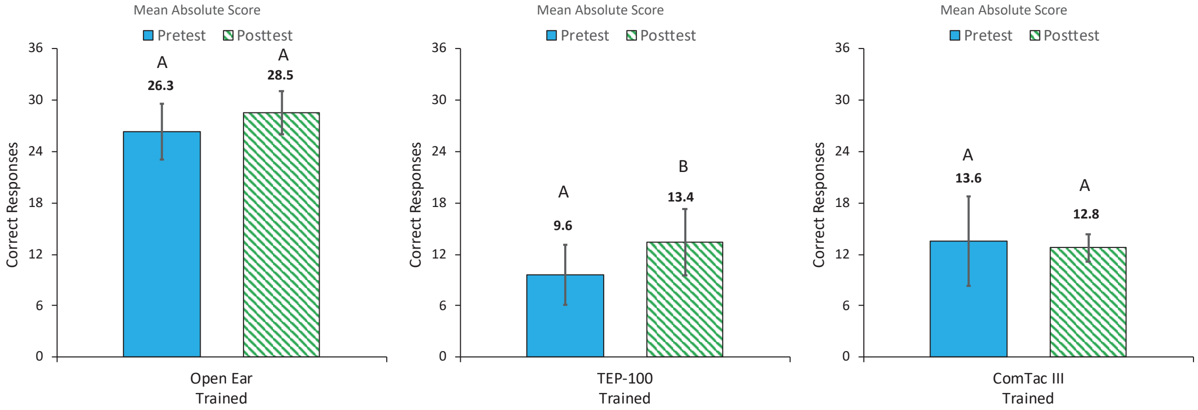


Figure . Mean absolute correct scores for each listening condition for the **trained** group **comparing pretest and posttest** performance. Different letters indicate significant differences at $p \leq 0.10$.

Front-back Reversal Errors

In addition to analyses using number of absolute and ballpark correct responses and response time measures, analyses were conducted using the dependent variable of the number of front-back reversal errors out of 36 trials. A front-back reversal occurred when the participant responded that a sound originating from 4 o'clock through 8 o'clock positions was located in the 10 o'clock through 2 o'clock positions, and vice-versa. As such, this type of error is known as a 120° arc front-back reversal. A mixed-factors ANOVA was conducted in order to examine the effect of group (between-subjects), listening condition (within-subjects), and stage of training (within-subjects) on the mean number of 120° arc front-back reversals. Mauchly's test of sphericity showed that the within-subjects factors met the assumption of equality of the variances (Table 96). The post hoc testing was not performed on stage of training given that only two levels were used in the analysis. Results from the mixed-factors ANOVA, using $\alpha=0.10$, showed only a main effect for listening condition was significant, $F(2, 44)=78.9$, $p<0.000$. All ANOVA results are provided in Table 97. Plotted means for each listening condition (Figure 118) showed that the highest number of front-back reversal errors occurred in the ComTac™ III condition ($M=8.4$), followed by the TEP-100 ($M=7.9$), and then the open-ear (1.1). Follow-up pairwise comparisons for the main effect for listening condition using a Bonferroni correction showed that the mean errors for open ear condition differed significantly from mean errors obtained in the TEP-100 and ComTac™ III conditions (Table 98).

Table . Mauchly's test of sphericity for mixed ANOVA for the effect of group, listening condition, and stage of training on **front-back errors** using the **120-degree arc criterion**.

Mauchly's Test of Sphericity					Epsilon (ϵ)	
Variables	Mauchly's Criterion	Chi-Square	df	p	Greenhouse-Geisser	Huynh-Feldt
Listening Condition	0.98	0.41	2	0.82	0.98	1
Listening Condition x Stage of training	0.81	4.40	2	0.11	0.84	1

Table . Mixed-factor ANOVA table evaluating differences in front-back reversal errors using the 120-degree arc criterion according to group, listening condition, and stage of training.

Source	df	Mean Square	*F value*	p	η_p^2
Between Subjects					
Group (G)	1	8.51	0.49	0.49	0.02
Error(S/G)	22	17.27			
Within Subjects					
Listening Condition (C)	**2**	**784.92**	**78.90**	**<0.000**	**0.78**
C x G	2	0.34	0.03	0.97	0.002
Error (C x S/G)	44	9.95			
Stage of Training (T)	1	31.17	2.65	0.12	0.11
T x G	1	2.01	0.17	0.68	0.008
Error (T x S/G)	22	11.76			
C x T	2	3.34	0.64	0.53	0.03
C x T x G	2	0.34	0.07	0.94	0.003
Error (C x T x S/G)	44	5.23			
Total	143	874.84			

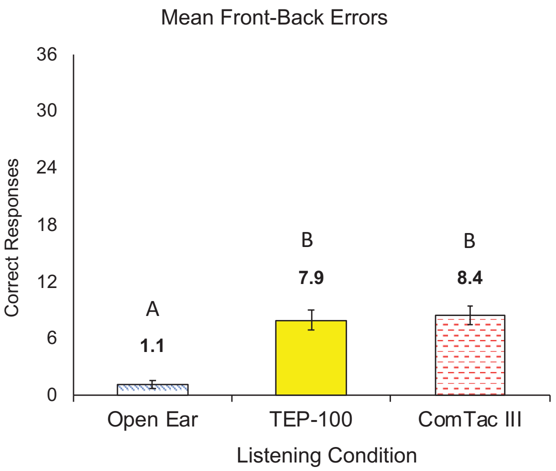


Figure . Mean front-back reversal errors using the 120-degree arc criterion for each listening condition. Different letters indicate significant differences at $p \leq 0.10$.

Table . Significant pairwise comparisons between listening conditions for front-back reversal errors using the 120° arc criterion with a Bonferroni adjustment

Listening Condition		*M*	*SE*	*p*
Open ear	**TEP-100**	**-6.73**	**0.67**	**<0.000**
Open ear	**ComTac™ III**	**-7.25**	**0.60**	**<0.000**
TEP-100	ComTac™ III	-0.52	0.66	1.000

Regression Analysis

In order to determine if performance on the portable system predicted in-field localization accuracy, especially in those who conducted training, regressions analysis was conducted. Specifically, the absolute score correct score was used in-office to predict in-field performance to assess the validity of using the in-office environment localization as a means to improve in-field performance. Therefore, post hoc regression was calculated to predict in-field performance based on in-lab results using the absolute correct score. In general, regression describes the magnitude of the relationship between the independent and dependent variables (Portney & Watkins, 2009). The regression line can be used to predict values of the dependent variable given a value of independent variable (Portney & Watkins, 2009). The null hypothesis for linear regression analyses is that the slope of the regression line is equal to zero (Portney & Watkins, 2009). In other words, a change in the independent variable results in no change in the dependent variable. As part of regression analyses, the *r*, or correlation coefficient is calculated. The *r* value reflects how closely the data matches the predicted values of the regression line, or goodness of fit (Portney & Watkins, 2009). Squaring the correlation coefficient, known as r^2, reflects the percentage of variance of the dependent variable accounted for by the independent variable. An α=0.10 value was used as the criterion for a significant linear regression in these analyses. A significant finding would indicate that given an absolute pretest score obtained in-office given a certain group membership (trained versus untrained) and listening condition (open ear, ComTac™ III, and TEP-100), the change in posttest score obtained in-field could be predicted.

Therefore, linear regression was conducted to examine the predictive value of pretest score for each combination of group and listening condition on posttest score. Regression analyses did not result in significant values for the following conditions: open ear condition for the trained group (*F*[1, 10]=2.51, *p*=0.144), open ear condition for the untrained group (*F*[1, 10]=0.41, *p*=0.536), TEP-100 condition for the untrained group (*F*[1, 10]=2.12, *p*=0.168), ComTac™ III condition for the trained group (*F*[1, 10]=2.12, *p*=0.176), and ComTac™ III condition for the untrained group (*F*[1, 10]=0.556, *p*=0.473).

A significant regression was found for the trained group using the TEP-100, *F*(1, 10)=3.71, *p*=0.083. Trained participants' posttest scores on the TEP-100 increased by 0.57 in the field for each correct answer on the pretest score (Figure 119). Table 99 shows the results of the ANOVA table. The resulting prediction equation is as follows:

Trained, TEP-100 Posttest Absolute Correct Score=7.06 + 0.57(Trained, TEP-100 Pretest Absolute Correct Score)

Table . ANOVA table for pretest prediction of posttest absolute correct scores in the trained group, TEP-100 condition.

Source	*SS*	*df*	*MS*	*F*	R^2	*p*
Model	**138.90**	**1**	**138.90**	**3.71**	**0.27**	**0.083**
Error	374.02	10	37.40			

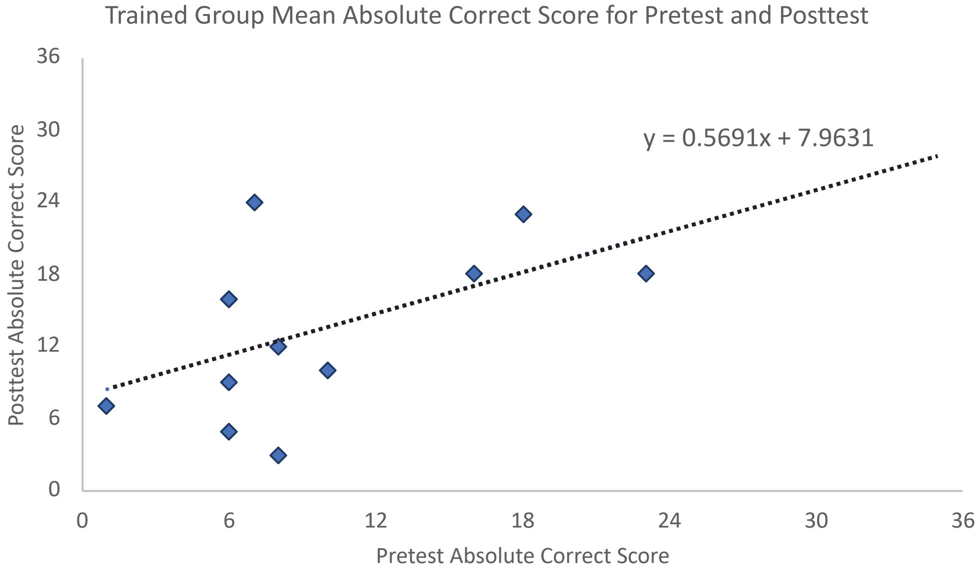


Figure . Mean absolute correct score on pretest and posttest for trained group. Regression line and equation plotted.

Independent-samples *t*-tests were then conducted to assess if the slope of the regression lines differed significantly between the trained and untrained groups from pretest to posttest for *each* listening condition, using the absolute correct scores. The α level was set to 0.10 and divided by the number of planned comparisons (3), resulting in a significance criterion value of 0.033. Levene's tests showed no violations to equality of the variance assumptions occurred for the open ear, *F*(1, 22)=0.29, *p*=0.595, TEP-100, *F*(1, 22)=0.37, *p*=0.551 and ComTac™ III, *F*(1,22) =2.36, *p*=0.139. *T*-test results, provided in Table 100, showed that significant group differences existed between the trained and untrained groups in the open ear condition, *t*(22)=3.20, *p*=0.004 (Figure 120), and the TEP-100 condition, *t*(22)=2.49, *p*=0.021 (Figure 121).

No significant difference existed between the trained and untrained group in the ComTac™ III condition. Examining the means for each group at for the open ear and TEP-100 conditions showed that training improved the participants' absolute correct scores from pretest to posttest, but performance declined in the posttest without training.

Table . Descriptive statistics and independent-samples *t*-tests, using the 0.033 corrected alpha level comparing group differences within each listening condition measured by the slope of the regression line from pretest to posttest for absolute correct score.

Variable	*n*	*M*	*SD*	*df*	*t*	*p*
Open ear				**22**	**3.20**	**0.004**
Trained	12	0.35	0.76			
Untrained	12	-0.83	1.03			
TEP-100						
Trained	12	1.00	1.07	**22**	**2.49**	**0.021**
Untrained	12	-0.10	1.10			
ComTac™ III						
Trained	12	0.38	1.36	22	0.75	0.139
Untrained	12	0.04	0.76			

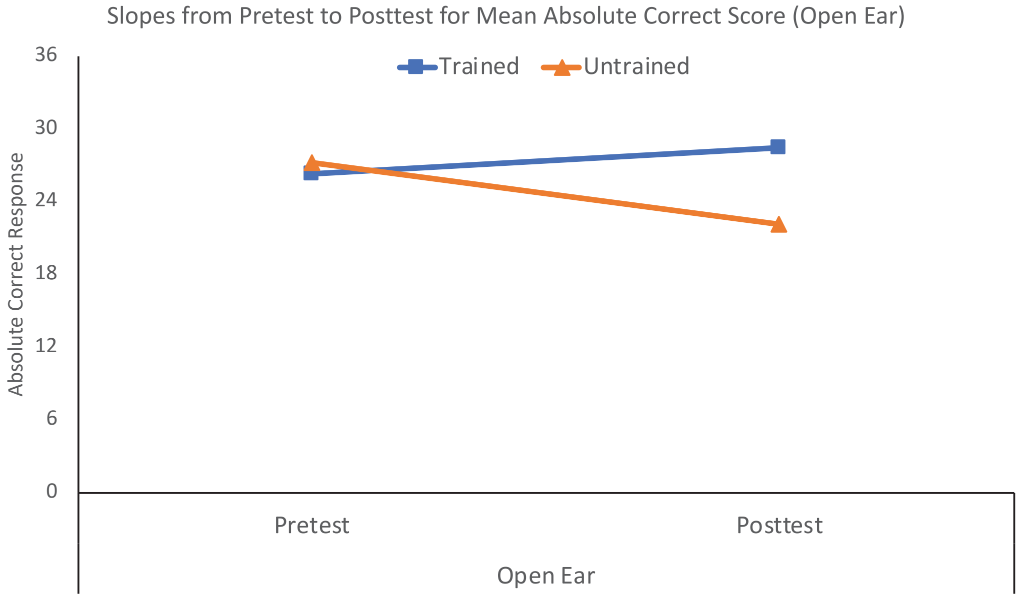


Figure . Slopes from pretest to posttest for mean absolute correct score for open ear by group.

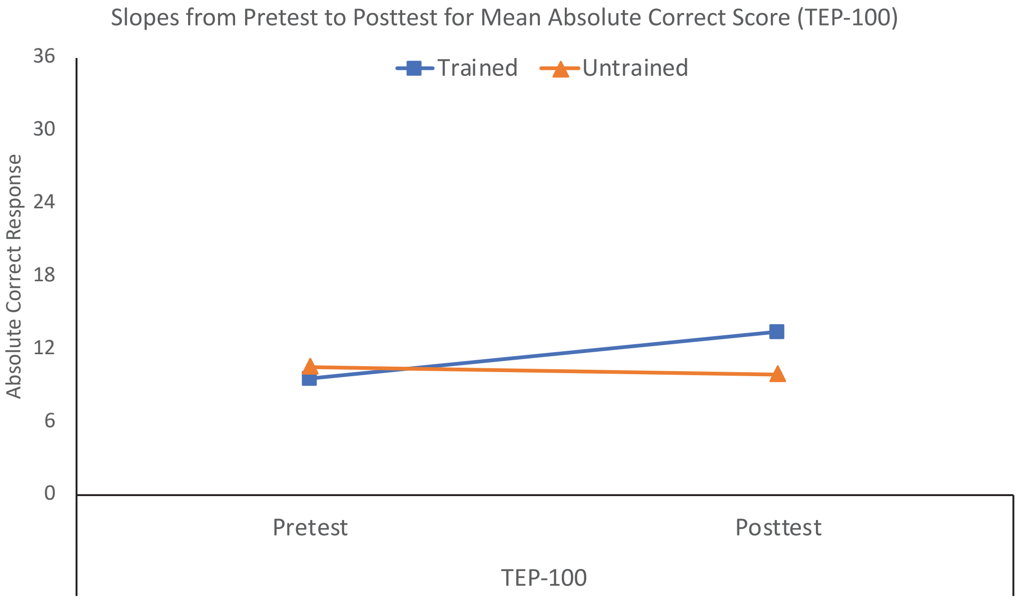


Figure . Slopes from pretest to posttest for mean absolute correct score for open ear by group.

Subjective ratings

Following the pretest and posttest, all participants answered a questionnaire which consisted of various bipolar rating scales. Participants in the trained condition also completed a questionnaire after the test in the LU5 subunit for each listening condition. The questionnaires are provided in Appendix K. To assess paired differences from pretest to posttest for each group and each training condition, Wilcoxon signed-rank tests were performed. This test procedure is the non-parametric equivalent to dependent samples *t*-tests. The Wilcoxon signed-ranks test assesses the direction and magnitude of differences of paired scores (Portney & Watkins, 2009). In this procedure, difference scores are ranked, disregarding the +/-sign and eliminating pairs with difference scores equal to zero (Portney & Watkins, 2009). Then, respective signs are assigned to the ranks (Portney & Watkins, 2009). If a participant's difference scores result in a tie, a mean rank is assigned (Portney & Watkins, 2009). Rejecting the null hypothesis for a Wilcoxon signed-ranks test means an unequal number of positive and negative ranks existed

(Portney & Watkins, 2009). Conversely, supporting the null hypothesis indicates that an equal number of positive and negative ranks existed (Portney & Watkins, 2009). In this study, a significant finding indicated ratings were significantly different at $\alpha=0.05$, given a condition, from pretest to posttest.

In order to assess group differences for each listening condition given a certain stage of training, Mann-Whitney *U* tests were performed. This test is the non-parametric counterpart to an independent samples *t*-test. The testing procedure involves ranking all of the scores, regardless of group membership, in ascending order (Portney & Watkins, 2009). The ranks for each group are then summed, with equal sums for groups supporting the null hypothesis (Portney & Watkins, 2009). Adequately large differences between the sums for each group results in rejecting the null hypothesis (Portney & Watkins, 2009). Given that non-parametric tests do not have a parallel procedure for mixed-factors ANOVAs, between-subjects and within-subjects testing was evaluated separately. An α level of 0.05 was adopted for all non-parametric tests.

Question 1. Perceived Confidence

Participants were asked to respond to the following: Rate how **confident you were** in your ability to locate sounds under this listening condition from 1 (no confidence) to 7 (extremely confident). Wilcoxon signed-ranks tests showed no significant differences from pretest compared to posttest for each group and listening condition combination (Table 101). Mean ratings for each listening condition, group, and stage of training are plotted in Figure 122.

Table . Wilcoxon results comparing ratings for pretest versus posttest for each listening condition and group for Question 1, Perceived Confidence.

Listening Condition	Group	Z	*p*
Open ear	Trained	-0.58	0.564
Open ear	Untrained	-0.71	0.480
TEP-100	Trained	-0.50	0.620
TEP-100	Untrained	-1.31	0.190
ComTac™ III	Trained	-0.92	0.357
ComTac™ III	Untrained	-0.14	0.887

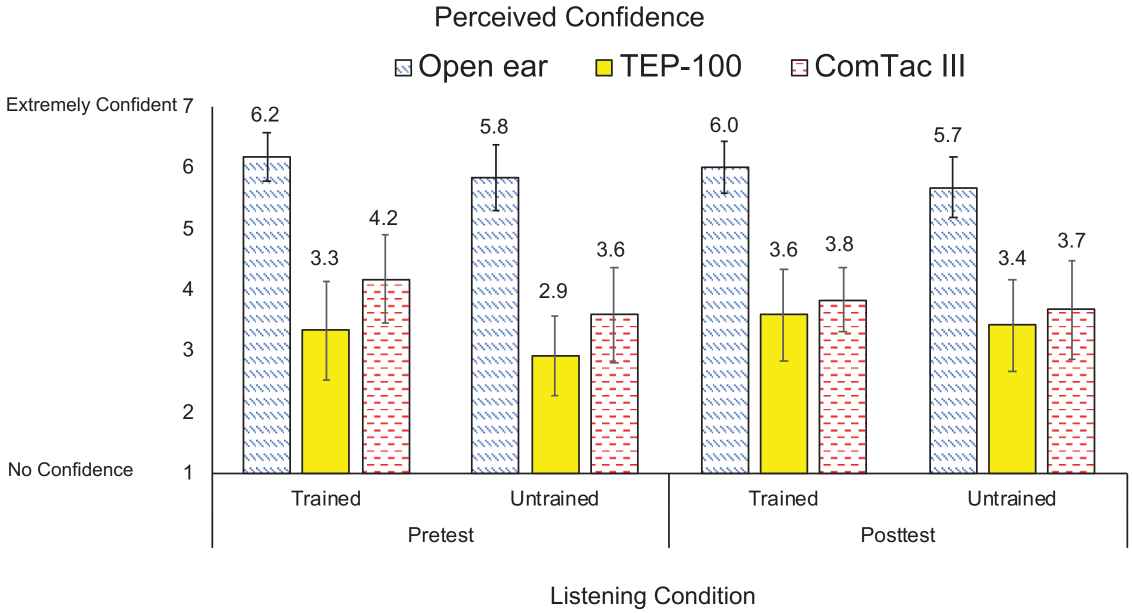


Figure . Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 1, Perceived Confidence.

To evaluate differences in ratings of confidence in the trained versus untrained groups, Mann-Whitney *U* tests were conducted at pretest (Table 102) and posttest (Table 103) collapsed across all listening conditions. Results were not statistically significant for group differences at pretest or posttest, across listening conditions.

Table . Results of the Mann-Whitney *U* test assessing group differences at pretest collapsed across all listening conditions on ratings of confidence for Question 1, Perceived Confidence.

Training Condition	N	Mean Rank	*U*	*p*
Trained	36	39.21	550.5	0.265
Untrained	36	33.79		

Table . Results of the Mann-Whitney *U* test assessing group differences at posttest collapsed across all listening conditions for Question 1, Perceived Confidence.

Training Condition	N	Mean Rank	*U*	*p*
Trained	36	37.86	599	0.574
Untrained	36	35.14		

Mann-Whitney *U* tests were conducted to evaluate group differences at pretest (Table 104) and then at posttest (Table 105) for each listening condition on ratings of confidence. Results showed no significant differences in confidence ratings at pretest between the trained and untrained group for each device. Figures displaying non-significant findings are provided in Appendix Q given that the added volume and complexity of such figures would not add to the main body of the document. As such, results of the analyses are graphed in Figure 146, Appendix Q. For the posttest ratings, Mann-Whitney U tests conducted for each listening condition comparing groups, shown in Table 105 and Figure 147, showed no significant differences in ratings of confidence between trained and untrained groups for each listening condition.

Table . Results of Mann-Whitney U tests comparing confidence ratings between training groups for each listening condition at pretest for Question 1, Perceived Confidence.

Training Condition	N	Mean Rank	*U*	*p*
Open				
Trained	12	13.67	58	0.387
Untrained	12	11.33		
TEP-100				
Trained	12	13.42	61	0.51
Untrained	12	11.58		
ComTac™ III				
Trained	12	13.92	55	0.312
Untrained	12	11.08		

Table . Results of Mann-Whitney U tests comparing confidence ratings between training groups for each listening condition at posttest for Question 1, Perceived Confidence.

Training Condition	N	Mean Rank	*U*	*p*
Open				
Trained	12	13.75	57	0.354
Untrained	12	11.25		
TEP-100				
Trained	12	13.33	62	0.552
Untrained	12	11.67		
ComTac™ III				
Trained	12	12.88	67.5	0.788
Untrained	12	12.13		

Within-subjects non-parametric analyses were conducted to compare confidence ratings among devices, i.e., listening conditions, for each group at pretest and then for each group at posttest. In order to compare ratings of the perceived confidence across listening conditions for the trained group in the field, a Friedman two-way analysis of variance by ranks was performed. The Friedman test is the non-parametric counterpart to the repeated measures ANOVA. In the test procedure, subjects are treated as an independent variable (Portney & Watkins, 2009). Data are then organized with subjects arranged in rows and levels of conditions in columns (Portney & Watkins, 2009). Ranks are then assigned for each participant, ranking results across the row (e.g., three treatments would result in three rankings for each participant) (Portney & Watkins, 2009). Ties are handled by assigning an average value for the row. Ranks are then generated for each column, or treatment level (Portney & Watkins, 2009). The null hypothesis supports that the ranks for the columns are equal (Portney & Watkins, 2009). The alternative hypothesis is that at least one pair of treatment levels are different (Portney & Watkins, 2009). At the pretest, Friedman tests showed significant differences in ratings of confidence among the listening conditions for the trained ($\chi^2[2]=19.86$, $p< 0.00$) and untrained ($\chi^2[2]=18.67$, $p< 0.00$) groups. Results are provided in Table 106 and Figure 123. Follow-up pairwise comparisons used a criterion α level of 0.016, given that the overall α level was set to 0.05, but three comparisons (0.05/3=0.016) were conducted for each Friedman's test. For the trained condition, follow-up pairwise comparisons using the Wilcoxon tests, Table 107, at pretest showed significant differences in the open, M=6.2, versus TEP-100 condition, M=3.3, Z=2.96, p=0.003 , open M=6.2, versus ComTac™ III, M=4.2, Z=2.96, p=0.003, and TEP, M=3.3, versus ComTac™ III condition, M=4.2, Z=2.49, p=0.013. For the untrained condition, Wilcoxon signed-rank tests, Table 108, showed significant differences at pretest in ratings of confidence between the open ear, M=5.8 and TEP-100, M=2.9, Z=3.08, p=0.002, and between the open ear and ComTac™ III, M=3.6, Z=3.08, p=0.002. Significant differences in confidence ratings were also found at posttest among listening conditions for trained ($\chi^2[2]=18.47$, $p< 0.002$) and untrained ($\chi^2[2]=16.33$, $p< 0.000$) groups (Table 109 and Figure 124). For the trained group at posttest, Wilcoxon signed-rank tests, Table 110, showed significant differences in ratings of confidence between the open ear, M=6.0, and TEP-100, M=3.6, Z=2.95, p=0.003, and between the open ear and ComTac™ III, M=3.8, Z=3.09, p =0.002. For the untrained group at posttest, Wilcoxon signed-rank tests, Table 111 showed significant differences at posttest in ratings of confidence between the open ear, *M*=5.7, and TEP-100, *M*=3.8, *Z*=2.96, *p*=0.003 and between the open ear and ComTac™ III, *M*=3.7, *Z*=2.82, *p*=0.005.

Table . Friedman test results demonstrating significant differences among listening conditions for the trained and untrained groups at pretest for Question 1, Perceived Confidence.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**19.86**	**12**	**2**	**<0.000**
Untrained	**18.67**	**12**	**2**	**<0.000**

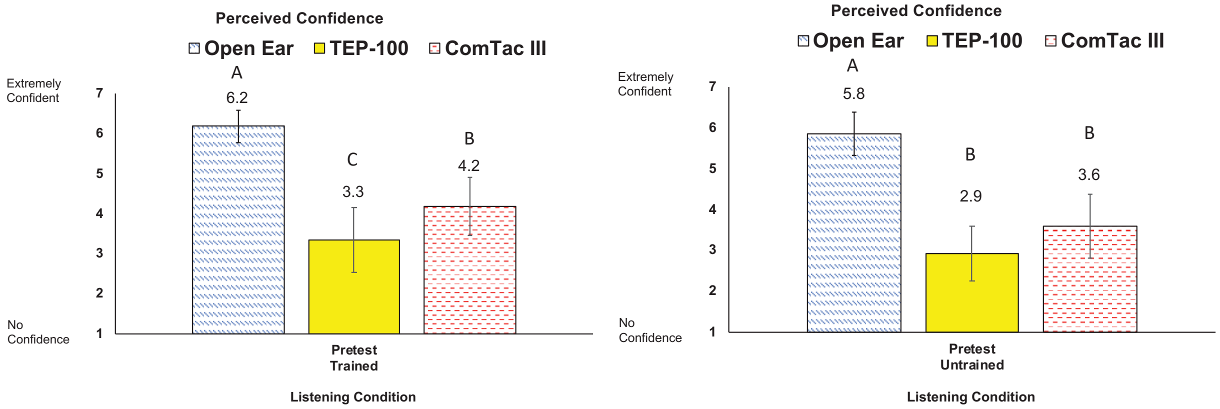


Figure . Mean ratings for each group at pretest for Question 1, Perceived Confidence.

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the trained group at pretest for Question 1, Perceived Confidence.

Listening Condition	Z	*p*
Open - TEP 100	**2.96**	**0.003**
Open - ComTac™ III	**2.96**	**0.003**
TEP-100 – ComTac™ III	**2.49**	**0.013**

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the untrained group at pretest for Question 1, Perceived Confidence.

Listening Condition	Z	*p*
Open - TEP 100	**3.08**	**0.002**
Open - ComTac™ III	**3.08**	**0.002**
TEP-100 – ComTac™ III	1.12	0.263

Table . Friedman test results demonstrating significant differences among listening conditions for the trained and untrained groups at posttest for Question 1, Perceived Confidence.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**18.47**	**12**	**2**	**<0.002**
Untrained	**16.33**	**12**	**2**	**<0.000**

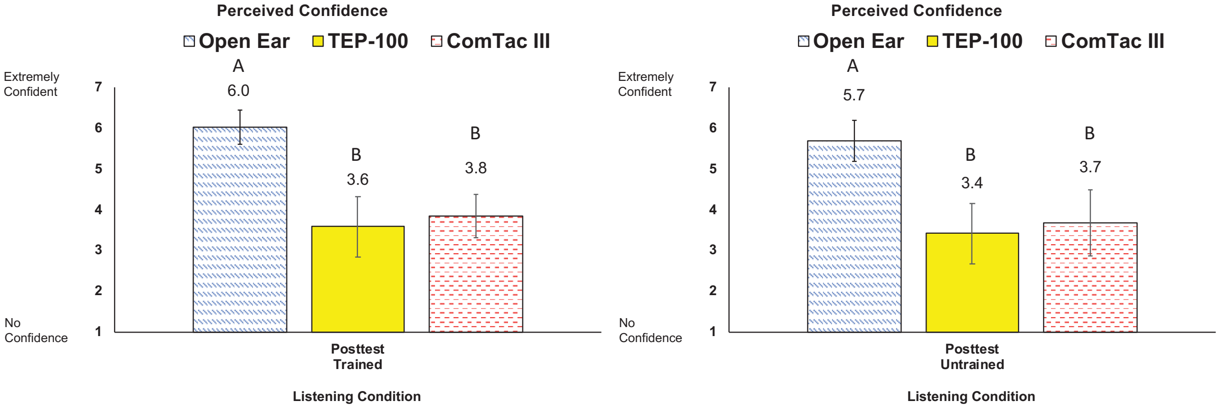


Figure . Mean ratings for each group at posttest for Question 1, Perceived Confidence.

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the trained group at posttest for Question 1, Perceived Confidence.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.95**	**0.003**
Open - ComTac™ III	**3.09**	**0.002**
TEP-100 – ComTac™ III	0.30	0.763

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the untrained group at posttest for Question 1, Perceived Confidence.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.96**	**0.003**
Open - ComTac™ III	**2.82**	**0.005**
TEP-100 – ComTac™ III	0.37	0.714

Question 2. Perceived Accuracy

Participants were asked the following: Rate your perceived **accuracy** to determine sound location under this listening condition from 1 (highly inaccurate) to 7 (highly accurate). On ratings of perceived accuracy, Wilcoxon signed-ranks tests showed no significant differences from pretest compared to posttest that were evaluated for each group and listening condition combination (Table 112). Mean ratings of confidence for each listening condition, group, and stage of training are provided in Figure 125. To evaluate differences in perceived accuracy in the experimental versus control groups, separate Mann-Whitney *U* tests were conducted at pretest (Table 113) and posttest (Table 114) collapsed across all listening conditions. Results were not significant for group differences in ratings of perceived accuracy at pretest or posttest, collapsed across listening conditions for perceived accuracy.

Table . Wilcoxon signed-ranks results comparing confidence ratings for pretest versus posttest for each listening condition and group for Question 2, Perceived Accuracy.

Listening Condition	Group	*Z*	*p*
Open ear	Trained	-0.63	0.527
Open ear	Untrained	-0.58	0.564
TEP-100	Trained	0	1
TEP-100	Untrained	0	1
ComTac™ III	Trained	-1.61	0.107
ComTac™ III	Untrained	-0.75	0.454

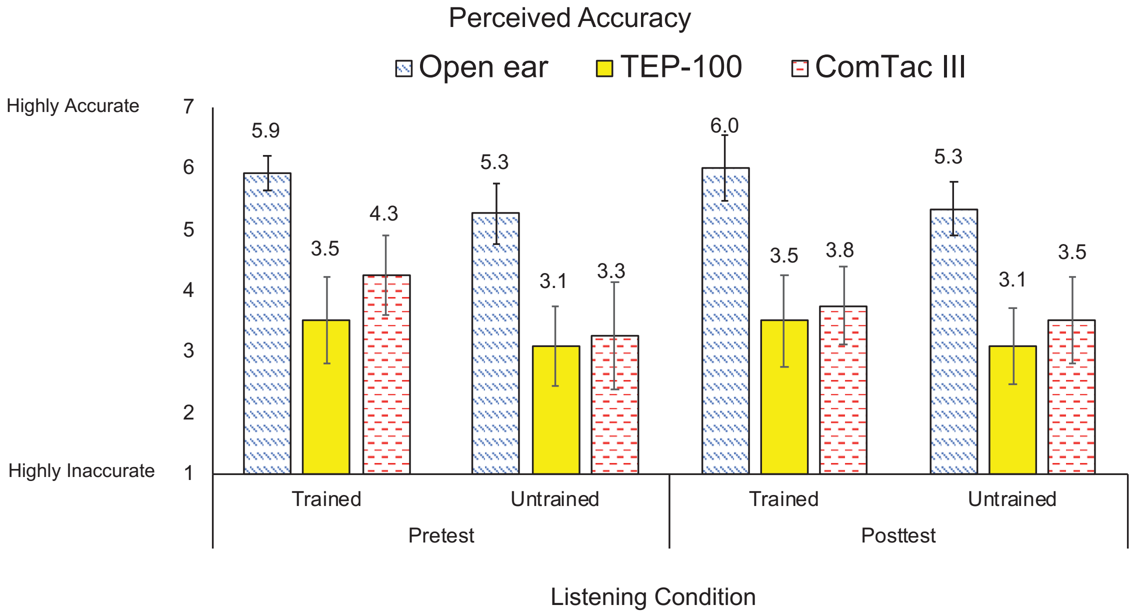


Figure . Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 2, Perceived Accuracy.

Table . Results of the Mann-Whitney *U* test assessing group differences at pretest collapsed across all listening conditions for Question 2, Perceived Accuracy.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	41.08	483	0.058
Untrained	36	31.92		

Table . Results of the Mann-Whitney *U* test assessing group differences at posttest collapsed across all listening conditions for Question 2, Perceived Accuracy.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	39.15	552.5	0.274
Untrained	36	33.85		

Mann-Whitney *U* tests were conducted to evaluate group differences in ratings of perceived accuracy at pretest (Table 115) and posttest (Table aaf and Figure 126) for each listening condition. At pretest, the trained (*M*=5.9) versus untrained group (*M*=5.3) showed a significant difference in the open ear condition, *U*=37, *p*=0.027. Figures of non-significant findings are displayed in Appendix Q, Figure 148. At posttest, groups showed no significant differences in ratings of perceived accuracy according to listening condition (Table 116).

Table . Results of Mann-Whitney U tests comparing confidence ratings between training groups for each listening condition at pretest for Question 2, Perceived Accuracy.

Training Condition	N	Mean Rank	*U*	*p*
Open				
Trained	**12**	**15.42**	**37.0**	**0.027**
Untrained	**12**	**9.58**		
TEP-100				
Trained		13.63	58.5	0.413
Untrained		11.38		
ComTac™ III				
Trained	12	14.92	43.0	0.088
Untrained	12	10.08		

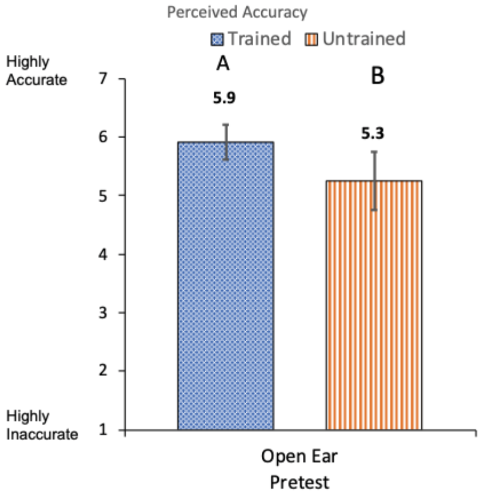


Figure . Mann-Whitney *U* results comparing trained and untrained ratings at pretest in the open ear condition for Question 2, Perceived Accuracy.

Table . Results of Mann-Whitney U tests comparing confidence ratings between training groups for each listening condition at posttest for Question 2, Perceived Accuracy.

Training Condition	*n*	Mean Rank	*U*	*p*
Open				
Trained	12	15.08	41.0	0.060
Untrained	12	9.92		
TEP-100				
Trained	12	13.67	58.0	0.400
Untrained	12	11.33		
ComTac™ III				
Trained	12	13.04	65.5	0.699
Untrained	12	11.96		

Within-subjects non-parametric analyses were conducted to compare perceived accuracy ratings among devices for each group at pretest and then for each group at posttest. At pretest, Friedman tests showed significant differences in ratings of confidence among the listening conditions for the trained ($\chi^2[2]=18.73$, $p<0.000$) and untrained ($\chi^2[2]=17.30$, $p<0.000$) groups. Results are provided in Table 117 and Figure 127. Follow-up pairwise comparisons used a criterion significance of α=0.016, for reasons discussed earlier. For the trained condition, follow-up pairwise comparisons using Wilcoxon tests, Table 118, at pretest showed significant differences in the open, *M*=5.9, versus TEP-100 condition, *M*=3.5, *Z*=3.10, *p*=0.002, and between open versus ComTac™ III, M= 4.3, Z=2.84, *p*=0.005. For the untrained condition, Wilcoxon signed-rank tests, Table 119, showed significant differences at pretest in ratings of confidence between the open ear, *M*=5.3 and TEP-100, *M*=3.1, *Z*=3.10, *p*=0.003, and between the open ear and ComTac™ III, *M*=3.3, *Z*=2.96, *p*=0.003. Significant differences in confidence ratings were also found at posttest among listening conditions for trained (χ^2[2]=14.68, *p*<0.001) and untrained (χ^2[2]=18.53, *p*<0.000) groups (Table 120 and Figure 128). For the trained group at posttest, Wilcoxon signed-rank tests, Table 121, showed significant differences in ratings of perceived accuracy between the open ear, *M*=6.0, and TEP-100, *M*=3.5, *Z*=2.87, *p*=0.004, and between the open ear and ComTac™ III, *M*=3.8, *Z*=2.95, *p*=0.003. For the untrained group at posttest, Wilcoxon signed-rank tests, Table 122, showed significant differences in ratings of perceived accuracy between the open ear, *M*=5.3, and TEP-100, *M*=3.1, *Z*=3.11, *p*=0.002 and between the open ear and ComTac™ III, *M*=3.5, *Z*=2.97, *p*=0.003.

Table . Friedman test results demonstrating significant differences in perceived ratings among listening conditions for the trained and untrained groups at pretest, Perceived Accuracy.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**18.73**	**12**	**2**	**<0.000**
Untrained	**17.30**	**12**	**2**	**<0.000**

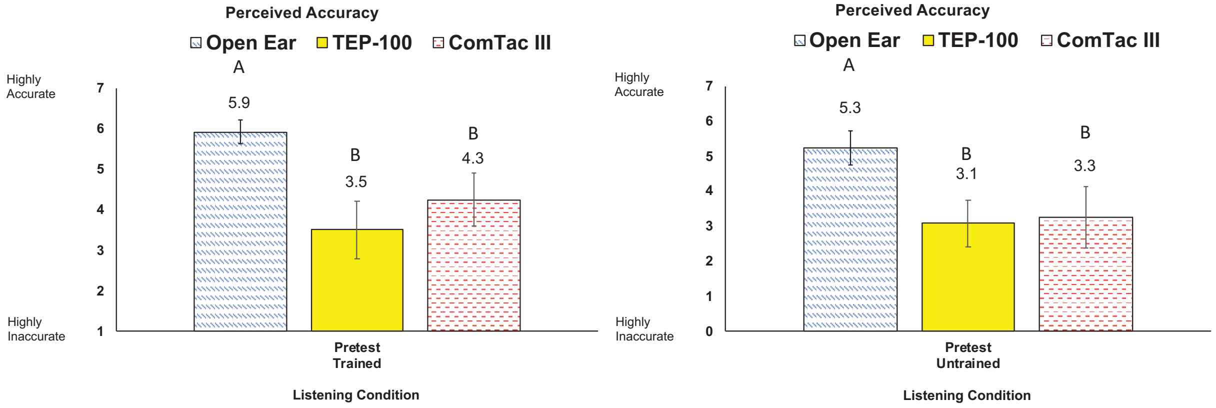


Figure . Mean ratings for each group at pretest for Question 2, Perceived Accuracy.

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the trained group at pretest for Question 2, Perceived Accuracy.

Listening Condition	*Z*	*p*
Open - TEP 100	**3.10**	**0.002**
Open - ComTac™ III	**2.84**	**0.005**
TEP-100 – ComTac™ III	2.07	0.038

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the untrained group at pretest for Question 2, Perceived Accuracy.

Listening Condition	*Z*	*p*
Open - TEP 100	**3.02**	**0.003**
Open - ComTac™ III	**2.96**	**0.003**
TEP-100 – ComTac™ III	0.24	0.810

Table . Friedman test results demonstrating significant differences in perceived accuracy ratings among listening conditions for the trained and untrained groups at posttest for Question 2, Perceived Accuracy.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**14.68**	**12**	**2**	**0.001**
Untrained	**18.53**	**12**	**2**	**<0.000**

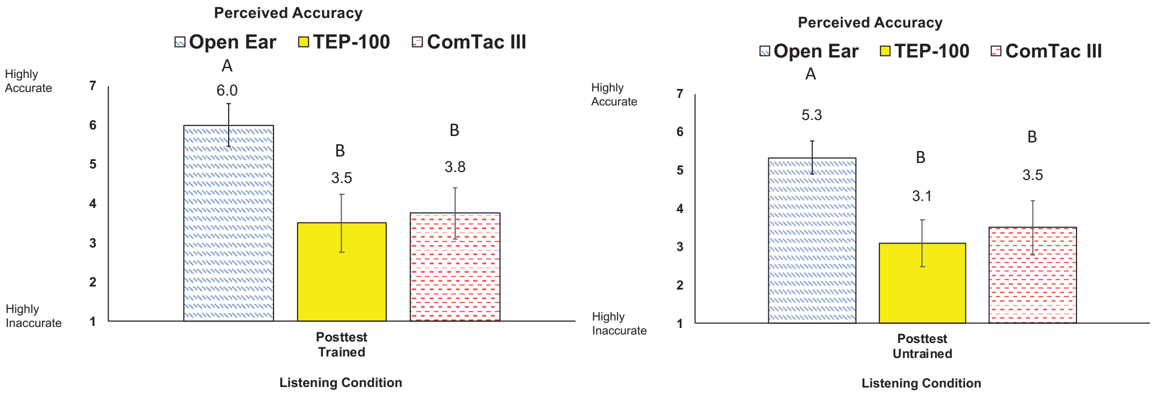


Figure . Mean ratings for each group at posttest for Question 2, Perceived Accuracy.

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the trained group at posttest for Question 2, Perceived Accuracy.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.87**	**0.004**
Open - ComTac™ III	**2.95**	**0.003**
TEP-100 – ComTac™ III	0.59	0.558

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the untrained group at posttest for Question 2, Perceived Accuracy.

Listening Condition	*Z*	*p*
Open - TEP 100	**3.11**	**0.002**
Open - ComTac™ III	**2.97**	**0.003**
TEP-100 – ComTac™ III	1.12	0.265

Question 3. Perceived Difficulty

Participants were asked the following: Rate how **difficult** it was to judge the **location** of the sounds under this listening condition from 1 (extremely difficult) to 7 (extremely easy). On ratings of difficulty, Wilcoxon signed-ranks tests showed no significant differences from pretest compared to posttest evaluated for each group and listening condition combination (Table 123). Mean ratings of difficulty for each listening condition, group, and stage of training are provided in Figure 129. To evaluate differences in perceived difficulty in the experimental versus control groups, separate Mann-Whitney *U* tests were conducted at pretest (Table 124) and posttest (Table 125) collapsed across all listening conditions. Results were not significant for group differences in ratings of perceived difficulty at pretest or posttest, collapsed across listening conditions. Throughout this discussion, it is important to note that *lower* ratings reflect *higher* difficulty.

Table . Wilcoxon signed-ranks results comparing difficulty ratings for pretest versus posttest for each listening condition and group for Question 3, Perceived Difficulty.

Listening Condition	Group	*Z*	*p*
Open ear	Trained	-1.00	0.317
Open ear	Untrained	-0.33	0.739
TEP-100	Trained	-0.26	0.792
TEP-100	Untrained	0.00	1
ComTac™ III	Trained	-1.03	0.305
ComTac™ III	Untrained	-0.91	0.366

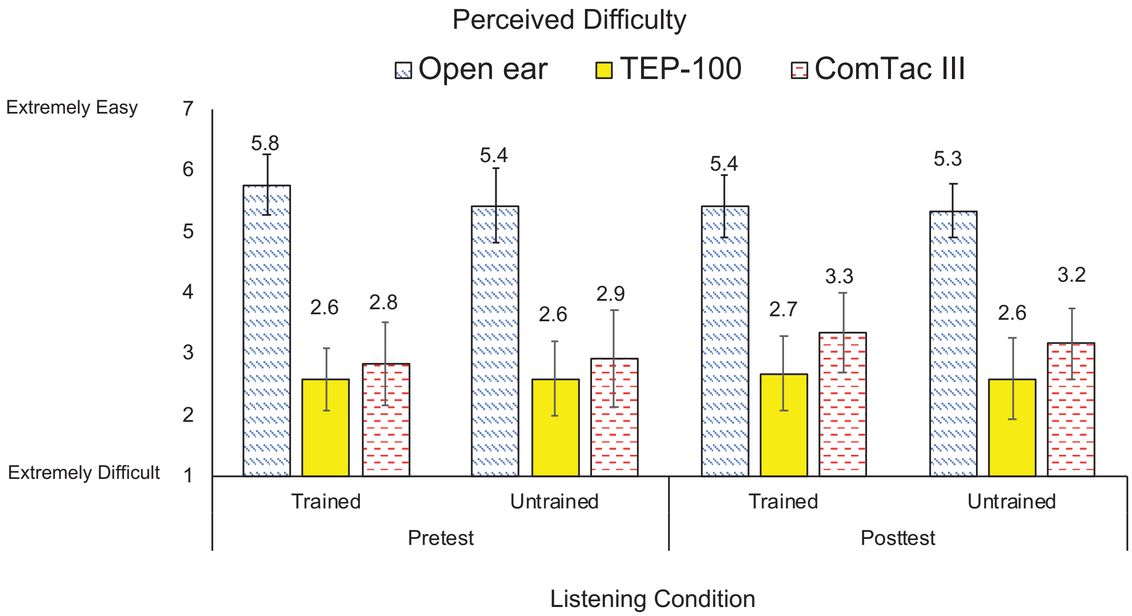


Figure . Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 3, Perceived Difficulty.

Table . Results of the Mann-Whitney *U* test assessing group differences at pretest collapsed across all listening conditions for Question 3, Perceived Difficulty.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	37	630	0.836
Untrained	36	36		

Table . Results of the Mann-Whitney *U* test assessing group differences at posttest collapsed across all listening conditions for Question 3, Perceived Difficulty.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	37.29	619.5	0.744
Untrained	36	35.71		

Mann-Whitney *U* tests were conducted to evaluate group differences in ratings of perceived difficulty at pretest (Table 126), and posttest (Table 127 and Figure 149, Appendix Q) for each listening condition. Results showed that at pretest and posttest, groups showed no significant differences in ratings of perceived difficulty by listening condition.

Table . Results of Mann-Whitney U tests comparing confidence ratings between training groups for each listening condition at pretest for Question 3, Perceived Difficulty.

Training Condition	*n*	Mean Rank	*U*	*p*
Open				
Trained	12	13.42	61.0	0.493
Untrained	12	11.58		
TEP-100				
Trained	12	12.79	68.5	0.829
Untrained	12	12.21		
ComTac™ III				
Trained	12	12.46	71.5	0.976
Untrained	12	12.54		

Table . Results of the Mann-Whitney *U* test evaluating group differences at posttest for the TEP-100 condition for Question 3, Perceived Difficulty.

Training Condition	*n*	Mean Rank	*U*	*p*
Open				
Trained	12	13.00	66.0	0.710
Untrained	12	12.00		
TEP-100				
Trained	12	12.96	66.5	0.742
Untrained	12	12.04		
ComTac™ III				
Trained	12	13.13	64.5	0.653
Untrained	12	11.88		

Within-subjects non-parametric analyses were conducted to compare perceived difficulty ratings among devices for each group at pretest and then for each group at posttest. Of note, lower ratings reflect increased difficulty. At pretest, Friedman tests showed significant differences in difficulty ratings among the listening conditions for the trained (χ^2[2]=19.24 , *p*<0.000) and untrained (χ^2[2]=17.64 , *p*<0.000) groups. Results are provided in Table 128 and Figure 130. Follow-up pairwise comparisons used α=0.016. For the trained group at pretest, follow-up pairwise comparisons using Wilcoxon tests, Table 129, showed significant differences in the open, *M*=5.8, and TEP-100 condition, *M*=2.6, *Z*=3.09, *p*=0.002, and between open and ComTac™ III, M=2.8, Z=3.10, *p*=0.002. For the untrained condition, Wilcoxon signed-rank tests, Table 130 showed significant differences at pretest in ratings of confidence between the open ear, $M=5.4$ and TEP-100, $M=2.6$, $Z=3.08$, $p=0.002$, and between the open ear and ComTac™ III, Mean=2.9, Z=2.90, $p=0.004$. Significant differences in difficulty ratings were also found at posttest among listening conditions for trained ($\chi^2[2]=16.31$, $p<0.00$) and untrained ($\chi^2[2]=20.31$, $p<0.00$) groups (Table 131 and Figure 131). For the trained group at posttest, Wilcoxon signed-rank tests, Table 132, showed significant differences in ratings of perceived difficulty between the open ear, $M=5.4$, and TEP-100, $M=2.7$, $Z=2.82$, $p=0.005$, and between the open ear and ComTac™ III, Mean=3.3, Z=3.10, $p=0.002$. For the untrained group at posttest, Wilcoxon signed-rank tests, Table 133, showed significant differences in ratings of difficulty between the open ear, $M=5.3$, and TEP-100, $M=2.6$, $Z=3.09$, $p=0.002$ and between the open ear and ComTac™ III, $M=3.2$, $Z=3.09$, $p=0.002$.

Table . Friedman test results demonstrating significant differences in perceived difficulty ratings among listening conditions for the trained and untrained groups at pretest for Question 3, Perceived Difficulty.

	χ^2	n	df	Asymp. Sig.
Trained	**19.24**	**12**	**2**	**<0.000**
Untrained	**17.64**	**12**	**2**	**<0.000**

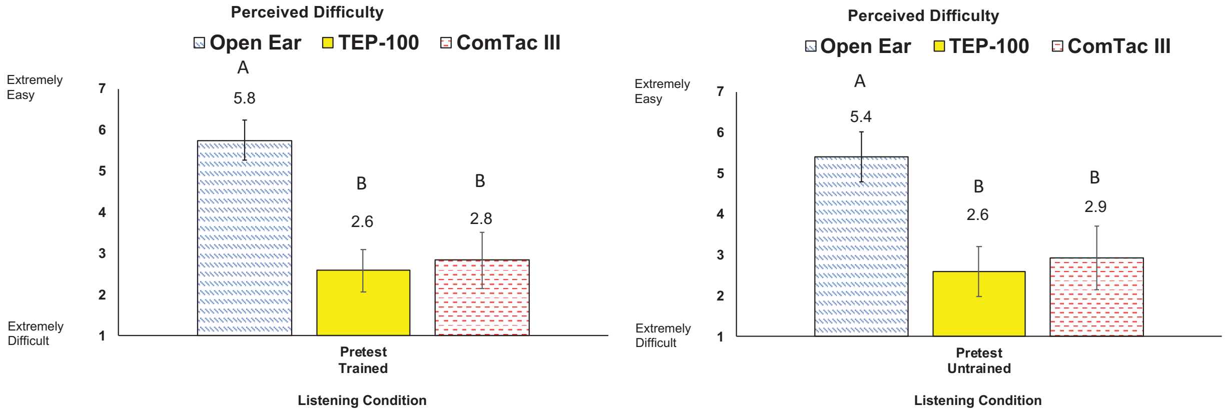


Figure . Mean ratings for each group at pretest for Question 3, Perceived Difficulty.

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the trained group at pretest for Question 3, Perceived Difficulty.

Listening Condition	*Z*	*p*
Open - TEP 100	**3.09**	**0.002**
Open - ComTac™ III	**3.10**	**0.002**
TEP-100 – ComTac™ III	0.81	0.417

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the untrained group at pretest for Question 3, Perceived Difficulty.

Listening Condition	*Z*	*p*
Open - TEP 100	**3.08**	**0.002**
Open - ComTac™ III	**2.90**	**0.004**
TEP-100 – ComTac™ III	0.72	0.473

Table . Friedman test results demonstrating significant differences in ratings among listening conditions for the trained and untrained groups at posttest for Question 3, Perceived Difficulty.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**16.31**	**12**	**2**	**<0.000**
Untrained	**20.31**	**12**	**2**	**<0.000**

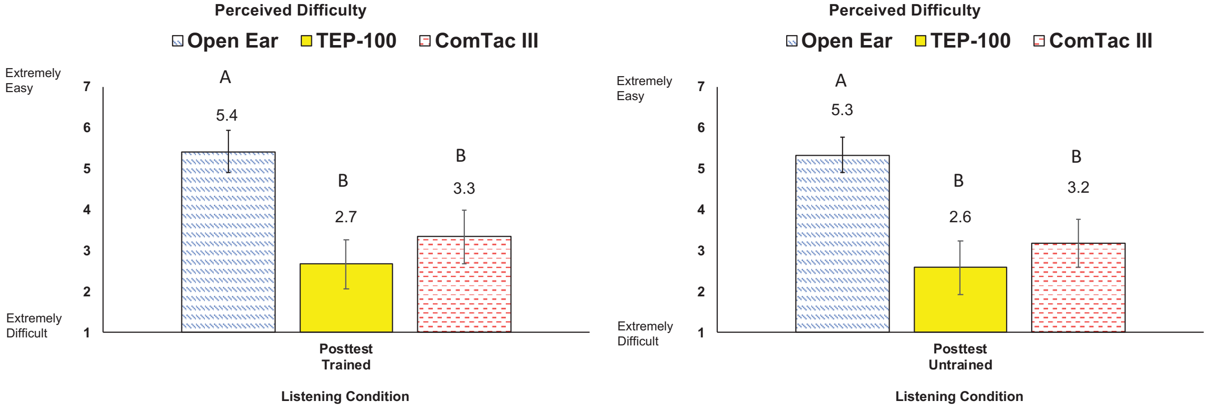


Figure . Mean ratings for each group at posttest for Question 3, Perceived Difficulty.

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the trained group at posttest for Question 3, Perceived Difficulty.

Listening Condition	Z	*p*
Open - TEP 100	**2.82**	**0.005**
Open - ComTac™ III	**3.10**	**0.002**
TEP-100 – ComTac™ III	1.27	0.203

Table . Wilcoxon signed-rank pairwise comparisons for the open, TEP-100, and ComTac™ III for the untrained group at posttest for Question, Perceived Difficulty.

Listening Condition	Z	*p*
Open - TEP 100	**3.09**	**0.002**
Open - ComTac™ III	**3.09**	**0.002**
TEP-100 – ComTac™ III	1.41	0.159

Question 4. Perceived Reaction Time

Participants were asked the following: Rate your perceived reaction time in determining the sound location under this listening condition" from 1 (extremely slow) to 7 (extremely fast). On ratings of perceived reaction time, Wilcoxon signed-ranks tests showed significant differences from pretest, *M*=4.1, compared to posttest, *M*=3.3, in the trained group for the TEP-100 listening condition, *Z*=-2.17, *p*=0.030 (Table 134 and Figure 132). Of note, a lower rating is indicative of slower perceived reaction times. Therefore, the aforementioned significant difference between pretest and posttest is consistent with participants' perception of feeling slower during the outdoor posttest. Significant differences also occurred in the trained group for the ComTac™ III listening condition from pretest, *M*=4.3, to posttest, *M*=3.3, Z=-2.49, *p*=0.031. The ComTac™ III, trained results are also consistent with a slower perceived reaction time during the posttest. To evaluate differences in perceived reaction time in the trained versus untrained groups, separate Mann-Whitney *U* tests were conducted at pretest (Table 135) and posttest (Table 136) collapsed across all listening conditions. Results were not significant for group differences in ratings of perceived reaction time at pretest or posttest for each listening condition (Figure 150, Appendix Q) for Question 4, Perceived Reaction Time.

Table . Wilcoxon signed-ranks results comparing response time for pretest versus posttest for each listening condition and group for Question 4, Perceived Reaction Time.

Listening Condition	Group	*Z*	*p*
Open ear	Trained	-1.4	0.161
Open ear	Untrained	-1.51	0.132
TEP-100	**Trained**	**-2.17**	**0.030**
TEP-100	Untrained	0.00	1.000
ComTac™ III	**Trained**	**-2.49**	**0.013**
ComTac™ III	Untrained	-0.58	0.564

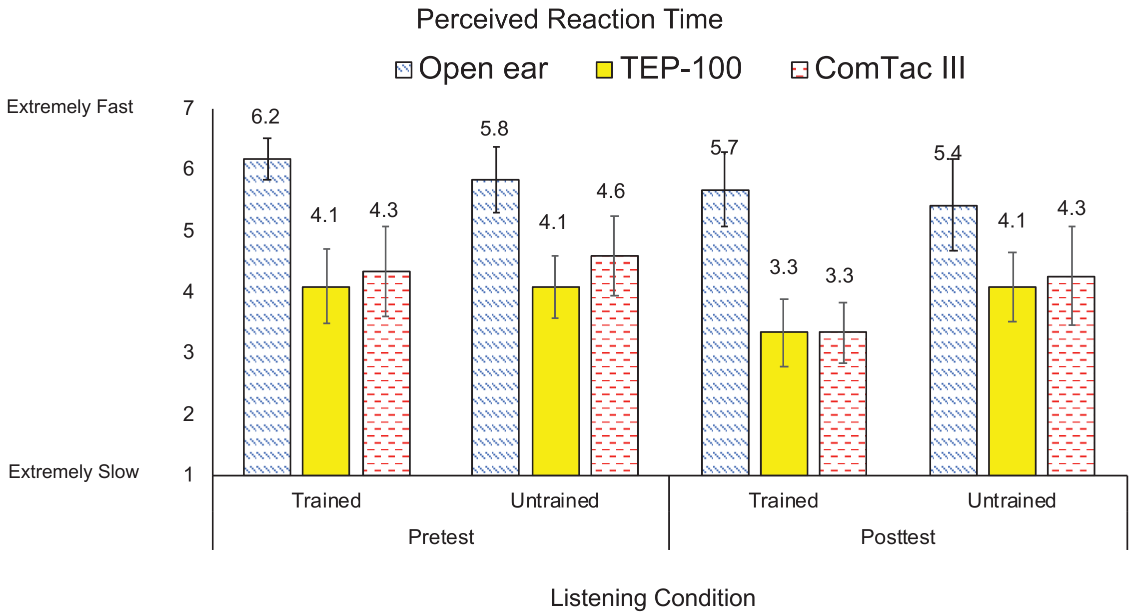


Figure . Plotted means and 95% confidence intervals for ratings at pretest and posttest for each group and listening condition for Question 4, Perceived Reaction Time.

Table . Results of the Mann-Whitney *U* test assessing group differences at pretest collapsed across all listening conditions for Question 4, Perceived Reaction Time.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	12	13.17	64	0.631
Untrained	12	11.83		

Table . Results of the Mann-Whitney *U* test assessing group differences at posttest collapsed across all listening conditions for Question 4, Perceived Reaction Time.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	12	10.08	43	0.078
Untrained	12	14.92		

Mann-Whitney *U* tests were conducted to evaluate group differences in ratings of perceived reaction time at pretest (Table 137) and posttest (Table 138) Results are displayed in Figures 151 (pretest) and 152 (posttest) in Appendix Q. At pretest and posttest, groups showed no significant differences in ratings of perceived reaction times between listening conditions.

Table . Results of the Mann-Whitney *U* test evaluating group differences at pretest for each listening condition for Question 4, Perceived Reaction Time.

Training Condition	*n*	Mean Rank	*U*	*p*
Open				
Trained	12	13.67	58	0.373
Untrained	12	11.33		
TEP-100				
Trained	12	12.50	72	1.000
Untrained	12	12.50		
ComTac™ III				
Trained	12	11.63	61.5	0.531
Untrained	12	13.38		

Table . Results of the Mann-Whitney *U* test evaluating group differences at posttest for each listening condition for Question 4, Perceived Reaction Time.

Training Condition	N	Mean Rank	*U*	*p*
Open				
Trained	12	13.17	64	0.631
Untrained	12	11.83		
TEP-100				
Trained	12	10.08	43	0.078
Untrained	12	14.92		
ComTac™ III				
Trained	12	9.92	41	0.061
Untrained	12	15.08		

Within-subjects non-parametric analyses were conducted to compare perceived reaction time ratings among devices for each group at pretest and then for each group at posttest. At pretest, Friedman tests showed significant differences in reaction times among the listening conditions for the trained ($\chi^2[2]=15.14$, $p< 0.001$) and untrained ($\chi^2[2]=15.57$, $p<0.001$) groups. Results are provided in Table 139 and Figure 133. Follow-up pairwise comparisons used α=0.016. For the trained group at pretest, Wilcoxon signed-rank tests, Table 140, showed significant differences in ratings of perceived reaction time between the open ear, *M*=6.2, and TEP-100, *M*=4.1, *Z*=2.96, *p*=0.003, and between the open ear and ComTac™ III, *M*=4.3, *Z*=2.83, *p*=0.005. For the untrained group at pretest, Wilcoxon signed-rank tests, Table 141, showed significant differences in ratings of perceived response time between the open ear, *M*=5.8, and TEP-100, *M*=4.1, *Z*=2.84, *p*=0.005 and between the open ear and ComTac™ III, *M*=4.6, Z=2.68, *p*=0.007. Significant differences in confidence ratings were also found at posttest among listening conditions for trained (χ^2[2]=14, *p*<0.001) and untrained (χ^2[2]=8.93, *p*<0.012) groups (Table 142 and Figure 134). For the trained group at posttest, Wilcoxon signed-rank tests, Table 143, showed significant differences in ratings of reaction time between the open ear, *M*=5.7, and TEP-100, Mean=3.3, *Z*=2.83, *p*=0.005, and between the open ear and ComTac™ III, *M*=3.3, *Z*=2.86, *p*=0.004. For the untrained group at posttest, Wilcoxon signed-rank tests, Table 144, showed significant differences in ratings of difficulty between the open ear, *M*=5.4, and TEP-100, *M*=4.1, *Z*=2.56, *p*=0.011.

Table . Friedman test results demonstrating significant among listening conditions for the trained and untrained groups at pretest Question 4, Perceived Reaction Time.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**15.14**	**12**	**2**	**<0.001**
Untrained	**15.57**	**12**	**2**	**<0.001**

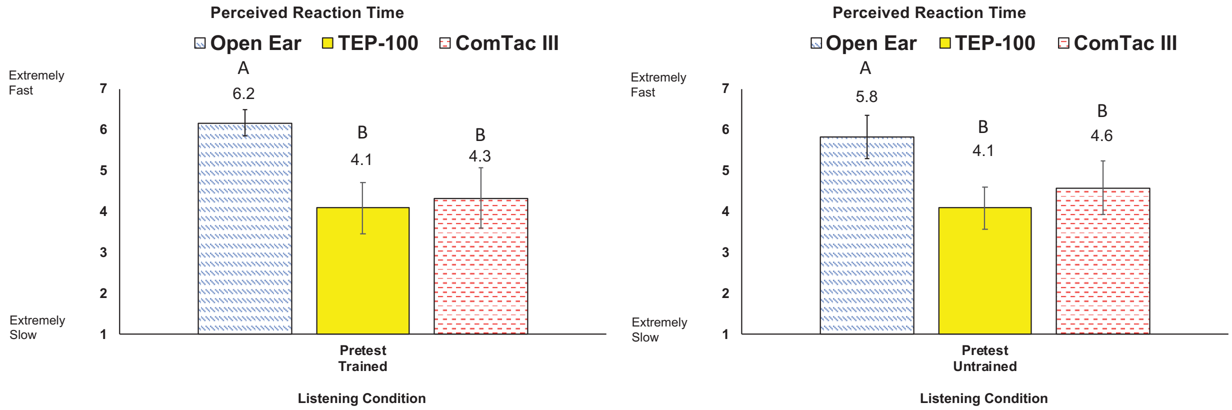


Figure . Mean ratings for each group and listening condition at pretest for Question 4, Perceived Reaction Time.

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III for Question 4 at pretest for the trained group at pretest, Perceived Reaction Time.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.96**	**0.003**
Open - ComTac™ III	**2.83**	**0.005**
TEP-100 – ComTac™ III	0.81	0.417

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III for Question 4 at pretest for the untrained group, Perceived Reaction Time.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.84**	**0.005**
Open - ComTac™ III	**2.68**	**0.007**
TEP-100 – ComTac™ III	1.73	0.083

Table . Friedman test results demonstrating significant differences in ratings among listening conditions for the trained and untrained groups at posttest, Question 4, Perceived Reaction Time.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**14**	**12**	**2**	**0.001**
Untrained	**8.93**	**12**	**2**	**0.012**

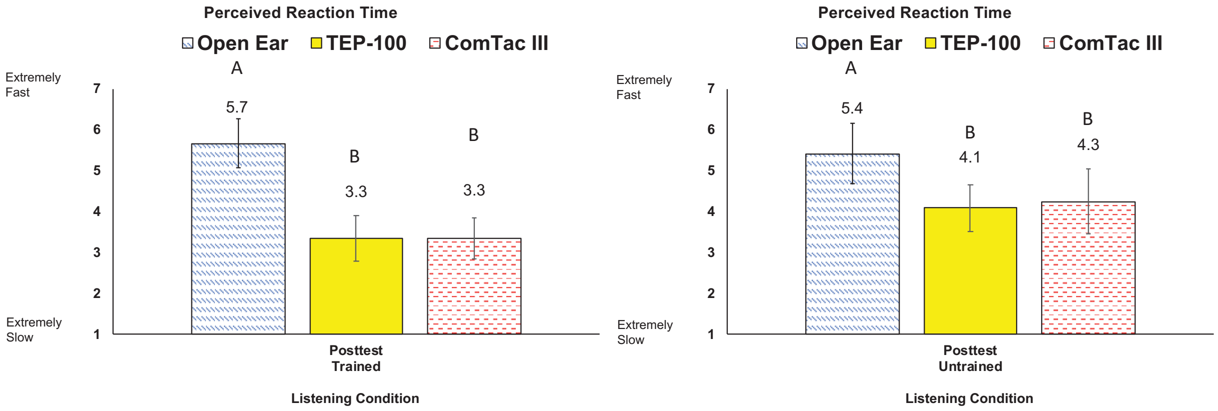


Figure . Mean ratings of perceived reaction time for each group and listening condition at posttest for Question 4, Perceived Reaction Time.

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III for Question 4 for the trained group at posttest, Perceived Reaction Time.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.83**	**0.005**
Open - ComTac™ III	**2.86**	**0.004**
TEP-100 – ComTac™ III	0.06	0.951

Table . Table Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III for the untrained group at posttest for Question 4, Perceived Reaction Time.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.56**	**0.011**
Open - ComTac™ III	2.26	0.024
TEP-100 – ComTac™ III	0.67	0.501

Question 5. Perceived Comfort

Participants were asked the following: Please rate how comfortable this hearing protection device condition (or open ear) was while wearing it during the experiment from 1 (extremely uncomfortable) to 7 (extremely comfortable). This rating thus was intended to apply to their comfort perception having worn the product during both in the office experiments and in the field tests. Wilcoxon signed-ranks tests showed significant differences in ratings of comfort in the untrained group for the ComTac™ III condition from pretest, *M*=5.08, to posttest, *M*=4.42, *Z*=-2.53, *p*=0.011 (Table 145 and Figure 135). Results are consistent with the perception of comfort decreasing pretest to posttest in the untrained group for the ComTac™ III. To evaluate differences in ratings of comfort in the experimental versus control groups, Mann-Whitney *U* tests were conducted at pretest (Table 146) and posttest (Table 147) collapsed across all listening conditions. Results were not significant for group differences in perceived comfort at pretest or posttest. Mann-Whitney *U* tests were conducted to evaluate group differences at pretest (Table 148 and Figure 153, Appendix Q) and at posttest (Table 149 and 154, Appendix Q) for each listening condition. Groups showed no significant differences in ratings of comfort according to listening condition at neither pretest nor posttest.

Table . Wilcoxon signed-ranks results comparing comfort ratings for pretest versus posttest for each listening condition and group for Question 5, Perceived Comfort.

Listening Condition	Group	*Z*	*p*
Open ear	Trained	-0.38	0.705
Open ear	Untrained	-0.82	0.414
TEP-100	Trained	-0.52	0.607
TEP-100	Untrained	-0.42	0.676
ComTac™ III	Trained	-1.19	0.234
ComTac™ III	**Untrained**	**-2.53**	**0.011**

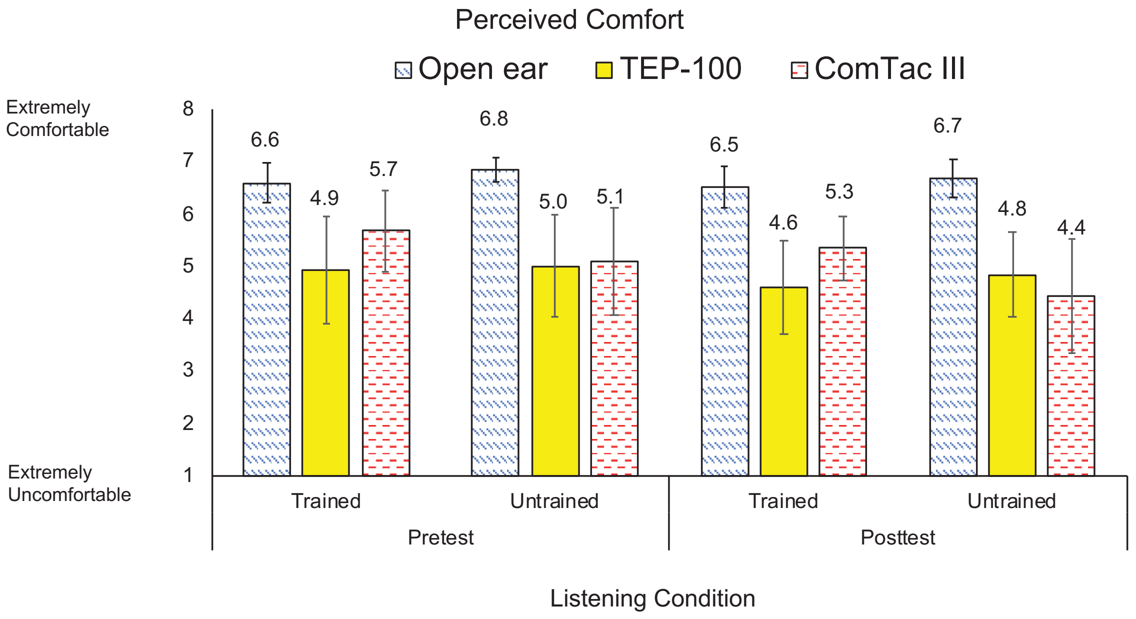


Figure . Plotted means and 95% confidence intervals at pretest and posttest for each group and listening condition for Question 5, Perceived Comfort.

Table . Results of the Mann-Whitney *U* test assessing group differences at pretest collapsed across all listening conditions for Question 5, Perceived Comfort.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	36.42	645	0.972
Untrained	36	36.58		

Table . Results of the Mann-Whitney *U* test assessing group differences at posttest collapsed across all listening conditions for Question 5, Perceived Comfort.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	36.88	634	0.876
Untrained	36	36.13		

Table . Results of the Mann-Whitney *U* test evaluating group differences at pretest for each listening condition for Question 5, Perceived Comfort.

Training Condition	N	Mean Rank	*U*	*p*
Open				
Trained	12	11.42	59	0.320
Untrained	12	13.58		
TEP-100				
Trained		12.58	71	0.952
Untrained		12.42		
ComTac™ III				
Trained		13.58	59	0.438
Untrained		11.42		

Table . Results of the Mann-Whitney *U* test evaluating group differences at posttest for each listening condition for Question 5, Perceived Comfort.

Training Condition	*n*	Mean Rank	*U*	*p*
Open				
Trained	12	11.58	61	0.444
Untrained	12	13.42		
TEP-100				
Trained	12	12.42	71	0.953
Untrained	12	12.58		
ComTac™ III				
Trained	12	14.17	52	0.237
Untrained	12	10.83		

Within-subjects non-parametric analyses were conducted to compare perceived comfort ratings among devices for each group at pretest and then for each group at posttest. At pretest, Friedman tests showed significant differences in comfort ratings among the listening conditions for the trained ($\chi^2[2]=10.27$, $p<0.006$) and untrained ($\chi^2[2]=13.26$, $p<0.001$) groups. Results are provided in Table 150 and Figure 136. Follow-up pairwise comparisons used $\alpha=0.016$, for reasons explained previously. For the trained group at pretest, follow-up pairwise comparisons using Wilcoxon tests, Table 151, showed a significant difference between the open ear, *M*=6.6, and TEP-100 condition, *M*=4.9, Z=2.75, *p*=0.006. For the untrained condition at pretest, follow-up pairwise comparisons using Wilcoxon tests, Table 152, showed significant differences at pretest in ratings of comfort between the open ear, M=6.8 and TEP-100, M=5.0, Z=2.68, p=0.007, and between the open ear and ComTac™ III, M=5.1, Z=2.69, p=0.007. Significant differences in comfort ratings were also found at posttest among listening conditions for trained (χ^2[2]=12.4, p<0.002) and untrained (χ^2[2]=14.26, p<0.001) groups (Table 153 and Figure 137). For the trained group at posttest, Wilcoxon signed-rank tests, Table 154, showed significant differences in ratings of comfort between the open ear, M=6.5, and TEP-100, M=4.6, Z=2.69, p=0.007. For the untrained group at posttest, Wilcoxon signed-rank tests, Table 155, showed significant differences in ratings of difficulty between the open ear, M=6.7, and TEP-100, M=4.8, Z=2.83, p=0.005 and between the open ear and ComTac™ III, M=4.4, Z=2.67, p=0.008.

Table . Friedman test results demonstrating significant differences in perceived comfort ratings among listening conditions for the trained and untrained groups at pretest, Question 5, Perceived Comfort.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**10.17**	**12**	**2**	**0.006**
Untrained	**13.26**	**12**	**2**	**0.001**

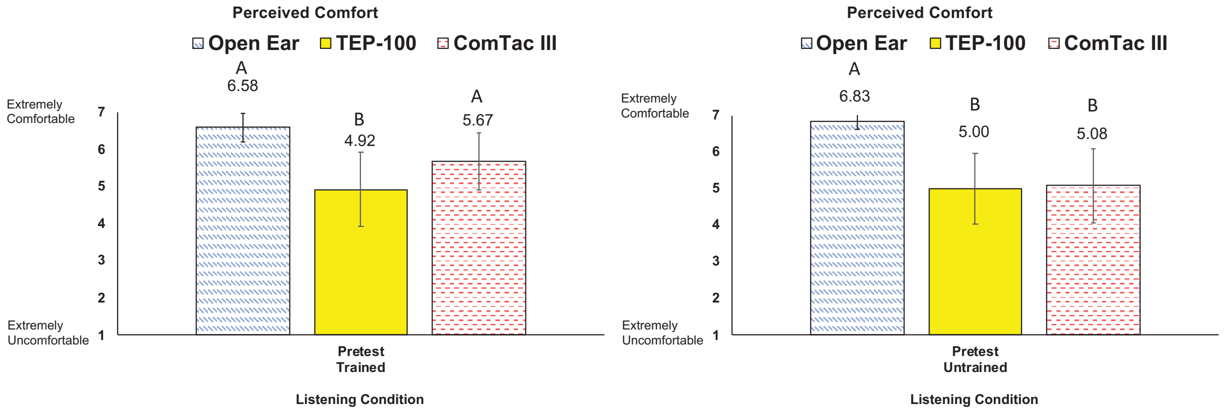


Figure . Mean ratings of perceived comfort for each group and listening condition at pretest for Question 5, Perceived Comfort.

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III for Question 5 at pretest for the trained group (α=0.016) for Question 5, Perceived Comfort.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.75**	**0.006**
Open - ComTac™ III	1.85	0.064
TEP-100 – ComTac™ III	1.38	0.169

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III for Question 5 at pretest for the untrained group (α=0.016) for Question 5, Perceived Comfort.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.68**	**0.007**
Open - ComTac™ III	**2.69**	**0.007**
TEP-100 – ComTac™ III	0.48	0.629

Table . Friedman test results demonstrating significant differences in comfort ratings among listening conditions for the trained and untrained groups at posttest for Question 5, Perceived Comfort.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**12.41**	**12**	**2**	**0.002**
Untrained	**14.26**	**12**	**2**	**0.001**

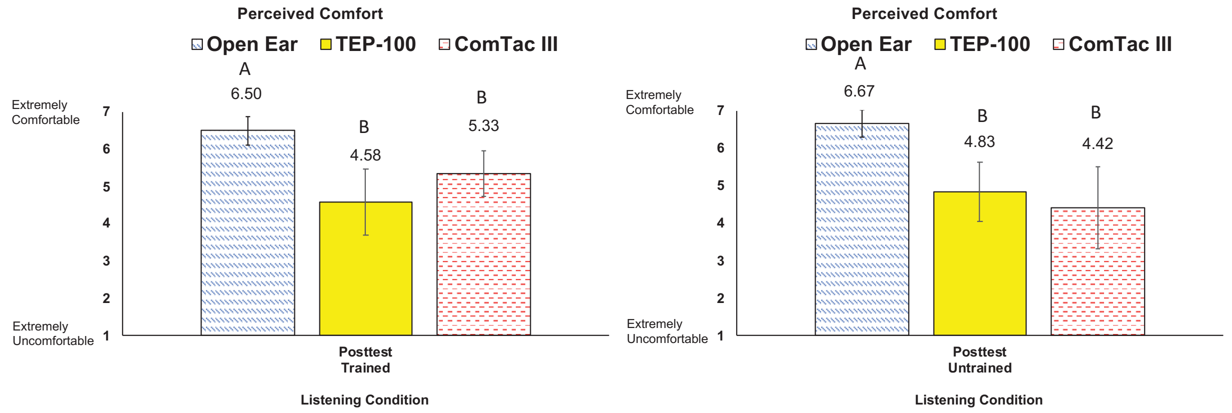


Figure . Mean ratings of perceived comfort for each group and listening condition at pretest for Question 5, Perceived Comfort.

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III for Question 5 at posttest for the trained group for Question 5, Perceived Comfort.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.69**	**0.007**
Open - ComTac™ III	2.26	0.024
TEP-100 – ComTac™ III	1.54	0.123

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III at posttest for the untrained group for Question 5, Perceived Comfort.

Listening Condition	Z	*p*
Open - TEP 100	**2.83**	**0.005**
Open - ComTac™ III	**2.67**	**0.008**
TEP-100 – ComTac™ III	0.71	0.478

Question 6. Likelihood of Wearing Device during a Sound Localization Task

Participants were asked the following: How likely would you be to wear this hearing protection device during a task similar to this experiment that required sound localization, if you had access to this hearing protection device [or open ear] from 1 (extremely unlikely) to 7 (extremely likely). Wilcoxon signed-ranks tests showed no significant differences in ratings of likelihood to maintain the current listening condition for the trained and untrained groups for each listening group from pretest to posttest (Table 156 and Figure 138). To evaluate differences in ratings of confidence in the experimental versus control groups, Mann-Whitney *U* tests were conducted at pretest (Table 157) and posttest (Table 158) collapsed across all listening conditions. Results were not significant for group differences in likelihood to wear the device at pretest or posttest collapsed across listening conditions. Mann-Whitney *U* tests were conducted to evaluate group differences at pretest (Table 159 and Figure 155, Appendix Q) and posttest (Table 160 and Figure 156, Appendix Q) for each listening condition. At neither pretest nor posttest, groups showed no significant differences in likelihood to wear the device under similar conditions according to device, or listening condition for Question 6, Likelihood of Wearing Device.

Table . Wilcoxon signed-ranks results comparing likelihood of maintaining the current listening condition given a similar task for pretest versus posttest for each listening condition and group for Question 6, Question 6. Likelihood of Wearing Device.

Listening Condition	Group	*Z*	*p*
Open ear	Trained	-0.28	0.783
Open ear	Untrained	-1.00	0.317
TEP-100	Trained	-0.26	0.796
TEP-100	Untrained	-0.18	0.857
ComTac™ III	Trained	-1.42	0.155
ComTac™ III	Untrained	-0.29	0.773

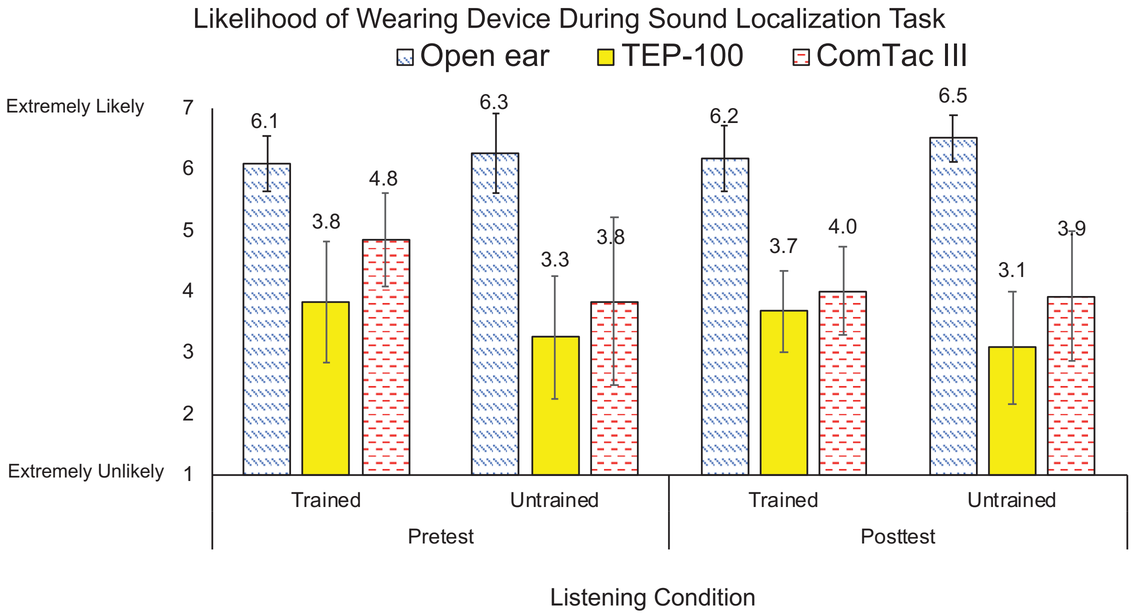


Figure . Plotted means and 95% confidence intervals for likelihood to maintain the same listening condition ratings at pretest and posttest for each group and listening condition for Question 6, Likelihood of Wearing Device.

Table . Results of the Mann-Whitney *U* test assessing group differences at pretest collapsed across all listening conditions for Question 6, Likelihood of Wearing Device.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	38.10	590.5	0.511
Untrained	36	34.9		

Table . Results of the Mann-Whitney *U* test assessing group differences at posttest collapsed across all listening conditions for Question 6, Likelihood of Wearing Device.

Training Condition	*n*	Mean Rank	*U*	*p*
Trained	36	36.57	645.5	0.977
Untrained	36	36.43		

Table . Results of the Mann-Whitney *U* test evaluating group differences at pretest for each listening condition for Question 6, Likelihood of Wearing Device.

Training Condition	*n*	Mean Rank	*U*	*p*
Open				
Trained	12	11.46	59.5	0.434
Untrained	12	13.54		
TEP-100				
Trained		13.67	58	0.412
Untrained		11.33		
ComTac™ III				
Trained		13.88	55.5	0.333
Untrained		11.13		

Table . Results of the Mann-Whitney *U* test evaluating group differences at posttest for each listening condition for Question 6, Likelihood of Wearing Device.

Training Condition	*n*	Mean Rank	*U*	*p*
Open				
Trained	12	11.33	58	0.373
Untrained	12	13.67		
TEP-100				
Trained	12	13.75	57	0.375
Untrained	12	11.25		
ComTac™ III				
Trained	12	12.79	68.5	0.837
Untrained	12	12.21		

Within-subjects non-parametric analyses were conducted to compare participant's likelihood of wearing the device, or keeping an open ear, given similar listening conditions for each group at pretest and then for each group at posttest. At pretest, Friedman tests showed significant differences in ratings among the listening conditions for the trained ($\chi^2[2]=9.95$, $p<$ 0.007) and untrained ($\chi^2[2]=12.76$, $p<0.002$) groups. Results are provided in Table 161 and Figure 139. Follow-up pairwise comparisons used $\alpha=0.016$ for reasons explained previously. For the trained group at pretest, follow-up pairwise comparisons using Wilcoxon tests, Table 162, showed a significant difference between the open, $M=6.1$, and TEP-100 condition, $M=3.8$, $Z=2.73$, $p=0.006$. For the untrained condition at pretest, follow-up pairwise comparisons using Wilcoxon tests, Table 163, showed significant differences at pretest in ratings between the open ear, M=6.3 and TEP-100, M=3.3, Z= 3.08, p=0.002, and between the open ear and ComTac™ III, M=3.8, Z=2.41, p=0.016. Significant differences in likelihood ratings were also found at posttest among listening conditions for trained ($\chi^2[2]=15.45$, $p<0.002$) and untrained ($\chi^2[2]=18.73$, $p<0.000$) groups (Table 164 and Figure 140). For the trained group at posttest, Wilcoxon signed-rank tests, Table 165, showed significant differences in ratings between the open ear, M=6.2, and TEP-100, M=3.7, Z=2.95, p=0.003 and between the open ear and the ComTac™ III, M=4.0, Z=2.7, p=0.007. For the untrained group at posttest, Wilcoxon signed-rank tests, Table 166, showed significant differences in likelihood ratings between the open ear, M=6.5, and TEP-100, M=3.1, Z=3.09, p=0.002 and between the open ear and ComTac™ III, M=3.9, Z=2.80, p=0.005.

Table . Friedman test results demonstrating significant differences among listening conditions for the trained and untrained groups at pretest for Question 6, Likelihood of Wearing Device.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**9.95**	**12**	**2**	**0.007**
Untrained	**12.76**	**12**	**2**	**0.002**

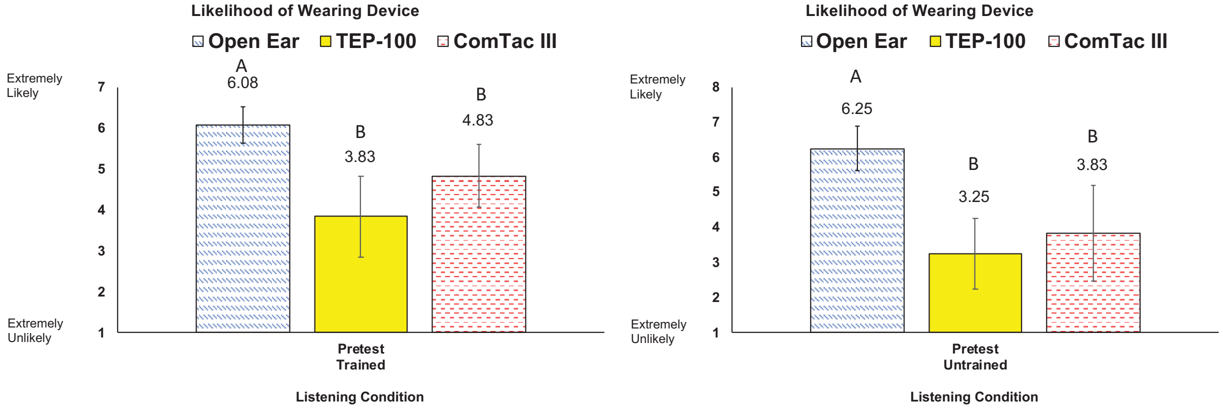


Figure . Mean ratings of likelihood of wearing device given similar conditions for each group and listening condition at pretest for Question 6, Likelihood of Wearing Device.

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III at pretest for the trained group, for Question 6, Likelihood of Wearing Device.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.73**	**0.006**
Open - ComTac™ III	2.20	0.028
TEP-100 – ComTac™ III	1.24	0.217

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III at pretest for the untrained group, for Question 6, Likelihood of Wearing Device.

Listening Condition	*Z*	*p*
Open - TEP 100	**3.08**	**0.002**
Open - ComTac™ III	**2.41**	**0.016**
TEP-100 – ComTac™ III	0.85	0.397

Table . Friedman test results demonstrating significant differences among listening conditions for the trained and untrained groups at posttest, for Question 6, Likelihood of Wearing Device.

	χ^2	*n*	*df*	Asymp. Sig.
Trained	**15.45**	**12**	**2**	**0.002**
Untrained	**18.73**	**12**	**2**	**<0.000**

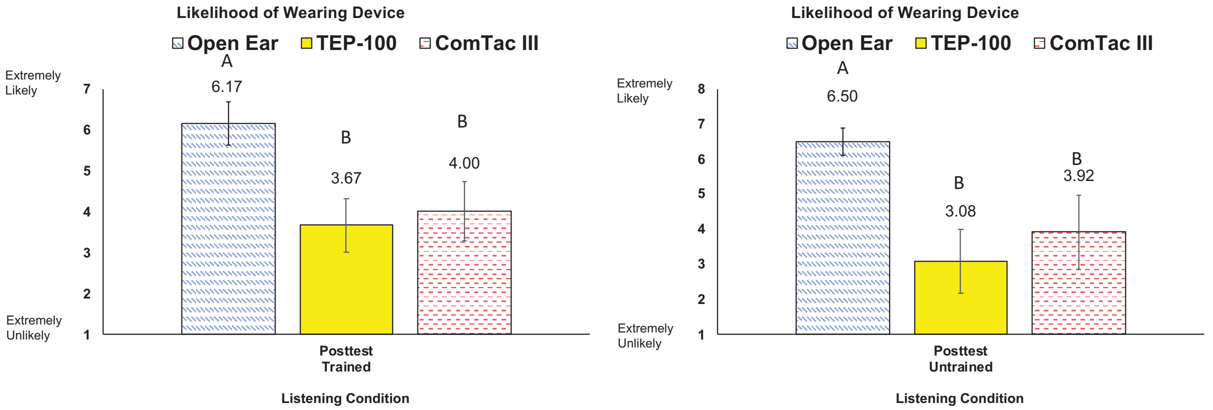


Figure . Mean ratings of likelihood of wearing device given similar conditions for each group and listening condition at posttest for Question 6, Likelihood of Wearing Device.

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III at posttest for the trained group (α=0.016) for Question 6, Likelihood of Wearing Device.

Listening Condition	*Z*	*p*
Open - TEP 100	**2.95**	**0.003**
Open - ComTac™ III	**2.70**	**0.007**
TEP-100 – ComTac™ III	0.50	0.618

Table . Wilcoxon signed rank tests, pairwise comparisons open, TEP-100, and ComTac™ III at posttest for the untrained group (α=0.016) for Question 6, Likelihood of Wearing Device.

Listening Condition	Z	p
Open - TEP 100	**3.09**	**0.002**
Open - ComTac™ III	**2.81**	**0.005**
TEP-100 – ComTac™ III	1.18	0.237

Question 7. Degree of Preparedness as Result of PALAT Training

Only the participants in the trained group were asked the following question at the conclusion of field testing for each listening condition: Rate the degree of preparedness you felt as a result of the training on the localization system (ring of loudspeakers) compared to the task of localizing .22 blank gunshot sounds from 1 (extremely unprepared) to 7 (extremely prepared). In order to compare ratings of the perceived degree of preparedness across listening conditions for the trained group in the field, a Friedman two-way analysis of variance by ranks was performed. Results showed a Chi-square value of 12.88, p=0.002 (Table 167). Follow-up testing for pairwise comparisons was conducted using Wilcoxon signed ranks tests. Testing yielded significant results for degree of preparedness for the open ear, M=5.9, compared to the TEP, M=4.5, Z=2.72, p=0.007, and for the open ear compared to the ComTac™ III, M=4.8, Z=2.27, p=0.010 (Table 168 and Figure 141). Results showed that the participants rated the open ear condition as the most likely they would use for a similar task and significantly more so than the TEP-100 or ComTac™ III condition.

Table . Friedman two-way analysis of variance results for listening condition for the trained group by ranks for Question 7, Degree of Preparedness.

χ^2	n	df	Asymp. Sig.
12.88	**12**	**2**	**0.002**

Table . Wilcoxon signed-ranks results comparing perceived degree of preparedness for each listening condition for Question 7, Degree of Preparedness.

Listening Condition	*Z*	*p*
Open ear - TEP-100	**2.72**	**0.007**
Open ear - ComTac™ III	**2.27**	**0.010**
TEP-100 - ComTac™ III	0.63	0.527

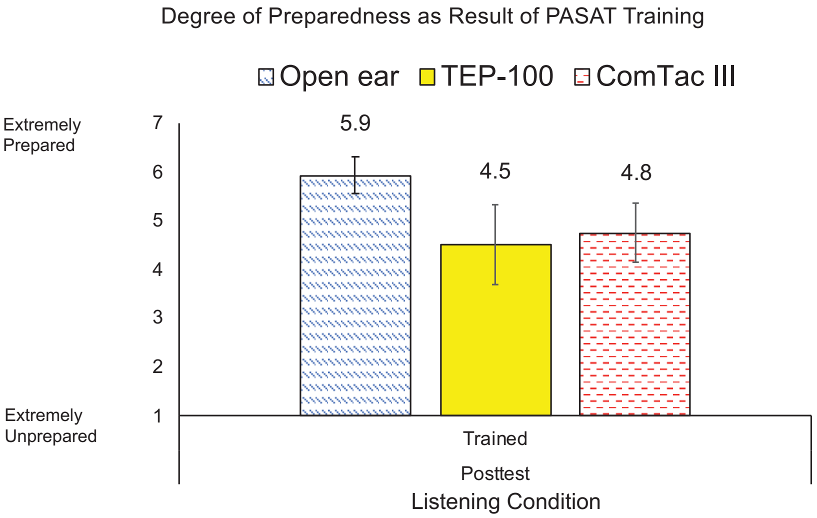


Figure . Plotted means and 95% confidence intervals for perceived degree of preparedness for each listening condition ratings at posttest for the trained group for Question 7, Degree of Preparedness.

Conclusions: In-Field Investigation of Transfer-of-Training

Overall, the transfer-of-training experiment supported the primary hypothesis that training in-office on the PALAT system transferred to in-field localization performance. The multivariate test only supported a significant main effect on the composite set of dependent variables for the listening condition variable. However, adjusting the alpha level for univariate testing to α=0.10 enabled a deeper analysis of the effects of listening condition, stage of training, and group on absolute score. As such, univariate results supported that auditory training in-office resulted in better absolute correct localization of gunshots outdoors. Univariate testing also revealed a significant increase in response time from pretest to posttest suggesting that the in-field localization test was a more difficult task.

In addition to supporting that training significantly improved in-field performance, results supported that in-office training and in-field testing were sensitive to differences in listening conditions. Multivariate and univariate testing both showed significant effects for listening condition. The secondary hypothesis that in-lab performance could predict in-field performance under all three listening conditions was only supported with statistical significance for trained subjects in the TEP-100 condition. However, comparing the slopes of the fitted regression lines for trained versus untrained participants for each listening condition showed significant differences in the open ear and TEP-100 conditions.

Listening Condition Conclusions

The open ear resulted in significantly higher performance measures of absolute correct score, ballpark correct score, response time and fewer front-back errors than in the TEP-100 and ComTac™ III conditions. These results were congruent with the in-lab, pretest findings of Casali & Robinette (2014) and Casali & Lee (2106a). The aforementioned studies reported that the open ear outperformed in-the-ear and over-the-ear hearing protection on the measure of absolute correct score. In-field, posttest results were also congruent with Talcott et al.'s (2012) field study where the open ear outperformed in-the-ear and over-the-ear HPDs on all accuracy and response time measures. Use of either TCAPS employed in this study resulted in at least a 50-percent degradation in absolute correct score when collapsed across training stage and training group. Therefore, results aligned with Abel (2008) and Bevis et al. (2014)'s qualitative evidence that reports of greatly reduced situation awareness hearing protection use among servicemembers. Training localization skills while using TCAPS may improve U.S. Military Service Members' confidence in their issued equipment, with strong likelihood of resulting in increased adoption rates.

Training Effect Conclusions

While the results of this and other studies quantified localization loss associated with TCAPS use, evidence supported that certain TCAPS and the open ear are susceptible to training effects. However, collapsed across all listening conditions, mean in-office performance was no different (*M*=46%) than mean absolute performance in the field (*M*=46%). Therefore, only certain listening conditions resulted in training transfer and only in the trained group. For the untrained group, compared to in-office performance, mean absolute performance in the field was, 13.8 % worse for the open ear, 2% worse for the TEP-100 and 1% better for the ComTac™ III. On the other hand, with in-office training, field performance was nearly equivalent (<1% better) for the open ear, 11% better for the TEP-100, and 2% worse for the ComTac™ III. Accuracy results illustrate that performance with the ComTac™ III was not susceptible to the effects of training administered in this study. However, in contrast, training effects, especially the significant results for the trained TEP-100 condition at posttest, suggest that training may improve performance and subsequent trust in device use. Radial plots (Figure 142) illustrate the training effects at each sound source location for each listening condition. A perfect score would result in a large circle along the 100% perimeter ring on the graph. The broader shape seen in the open ear and TEP-100 posttest plots for the trained group display the significant improvement with training compared to the untrained group.

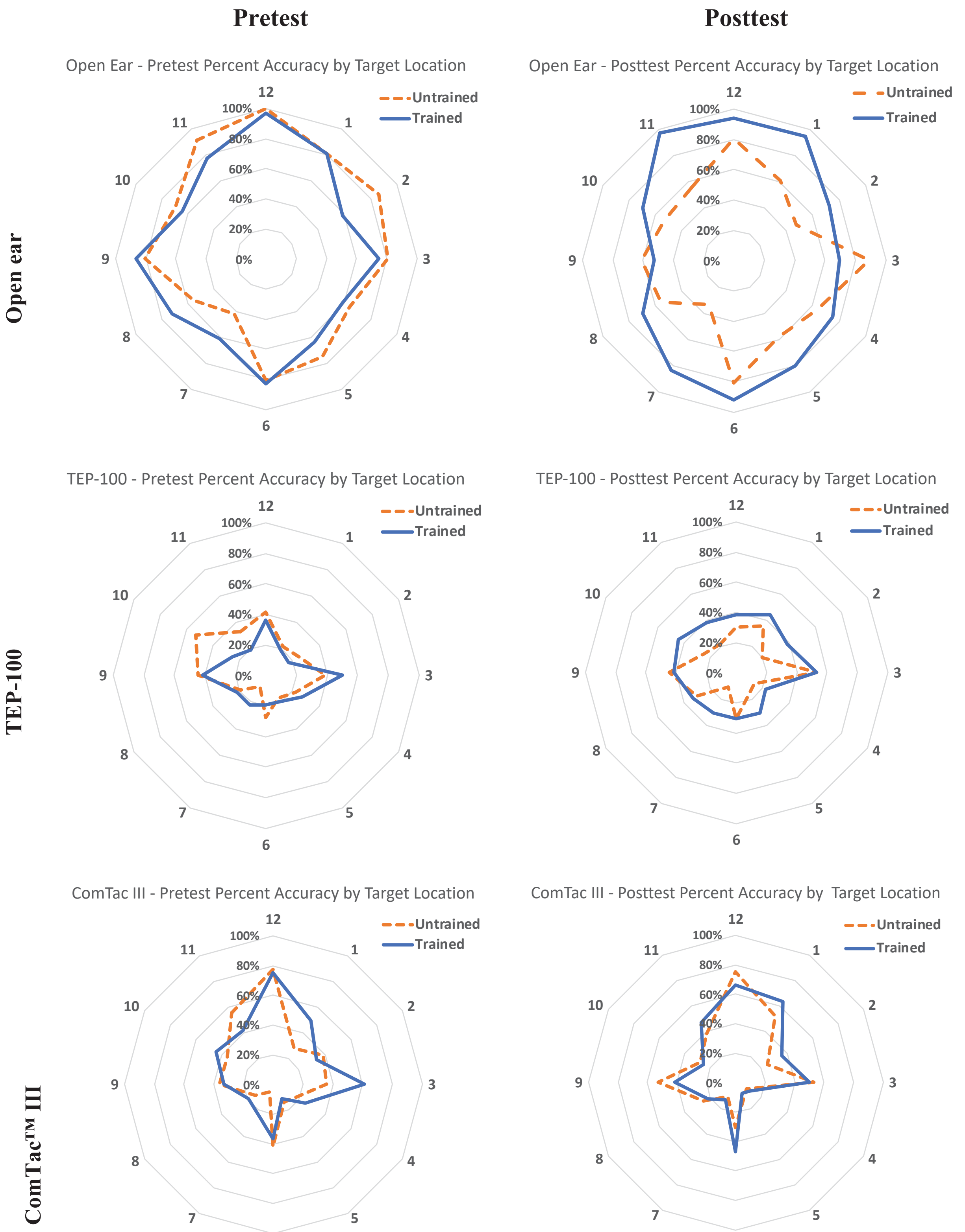


Figure . Radial plots of mean absolute correct accuracy percentage for each listening condition during pretest and posttest by group, trained (solid line) and untrained (dashed line).

Localization performance wearing TCAPS

This study was the first of its kind to train auditory localization ability in an office with loudspeakers, followed by testing in a field setting with live, blank gunshots. The TEP-100 listening condition was also unique to this investigation (to Phase II and Phase III of the broader research study). As a result, it is difficult to generalize the results of previous studies to the results of this investigation. However, in previous studies the in-the-ear style HPDs have consistently outperformed over-the-ear style HPDs in localization accuracy. Casali & Lee (2015) compared localization performance for several hearing protectors at unity gain given a 50 dBA dissonant signal amidst 40 dBA of pink noise. The EB15-LE BlastPLG®, an in-the-ear electronic hearing protector, resulted in approximately 46% localization accuracy compared to the ComTac™ III that resulted in approximately 36% accuracy (Casali & Lee, 2015). Performance in the open ear condition resulted in approximately 55% accuracy (Casali & Lee, 2015). The same listening conditions were tested using an 85 dBA signal and 75 dBA of pink noise. The open ear showed approximately 37% accuracy, compared to the 33% for the EB15-LE BlastPLG® and 26% for the ComTac™ III. Similarly, Talcott, Casali, Keady, and Killion (2012) evaluated localization accuracy for a .22 caliber blank stimulus, approximately 100-104 dB pSPL at the participant's ear, amidst 82 dBA of diesel truck noise and 45-50 dBA of rural noise. The investigators used the ballpark criterion to assess mean percent correct localization for the open ear, EB15-LE BlastPLG®, ComTac II over-the-ear electronic earmuff, and the EB1-BlastPLG® in-the-ear electronic hearing protector. Results for the open ear showed mean accuracy results of 88% and 81% for the rural and truck noise, respectively (Talcott, Casali, Keady, & Killion, 2012). The EB1- BlastPLG® and EB15- BlastPLG® resulted in 61-63% (rural noise) and 59-64% (truck noise) compared to 53% (rural noise) and 43% (truck noise) for the ComTac™ II. Again EB15- BlastPLG® resulted in 59-64% accuracy compared to 43-53% for the ComTac™ II (Talcott et al., 2012). Consistent with Phase II of this investigation, participants wearing the TEP-100 had, on average, achieved lower absolute accuracy compared to scores obtained while wearing the ComTac™ III. Training on the PALAT system was able to overcome this discrepancy, resulting in a significant difference in localization performance between the trained and untrained participants with the TEP-100 on the posttest. The TEP-100 also resulted in the highest mean absolute localization score for a TCAPS device at posttest. The significant training effect under the TEP-100 listening condition is promising for providing increased situation awareness while protecting Service Members hearing. However, as seen in the radial plots, additional training under the TEP-100 may be needed to achieve localization performance acceptable for most military operational needs.

A visual inspection of the radial plots under the ComTac™ III listening condition shows a concerning trend in poor performance from locations behind the participant, from 4 o'clock to 8 o'clock. Noticeably, the localization accuracy bias towards the frontal plane in the pretest condition using the ComTac™ III did not improve with training. Numerous participants voiced concerns during both the in-lab pretest and in-field posttest that all of stimulus signals seemed to be originating from in front of them while wearing the ComTac™ III. The investigator was unable to identify the primary contributing factor to the poor localization performance behind the listener with the ComTac™ III. However, it is hypothesized that the front-facing microphones of the ComTac™ III may be a contributing factor to its poor rear localization performance.

Front-Back Errors

The implications of the types of errors recorded in the ComTac™ III warrants further discussion. The ComTac™ III yielded significantly more front-back reversal errors than the open ear condition and slightly more front-back reversals than the TEP-100 condition. These results were consistent with previous study results where Talcott et al. (2012) found that earlier generation of the over-the-ear HPD (ComTac™ II) with similar forward-facing microphones resulted in a greater number of front-back reversals compared to the open ear and in-the-ear listening conditions (Etymotic EB1- BlastPLG® and EB15- BlastPLG®). As discussed previously, U.S. Military ground combat personnel are trained to orient in the direction they perceive the enemy threat signal. In a threat-detection scenario, localizing, and thus responding, to a perceived hazard in the opposite (i.e. wrong) plane is projected to have deleterious consequences. To mitigate the possibility of this occurring, a longer training session or training that specifically focuses on the dorsal plane may be needed to yield better localization performance.

Response Time

A significant main effect occurred according to listening condition on response time. Collapsed across training stage, both the TEP-100 and ComTac™ III conditions had significantly higher mean response times than the open ear condition. Mean response times for the TCAPS devices were 0.4 seconds slower during the pretest and 0.5 seconds slower during the posttest. In addition, while not significant, the mean response time in the field environment with gunshot stimuli was higher for every condition. Half a second in military operations could have a detrimental impact on the ability to respond or locate the enemy threat signal. Additional training may be necessary to improve response time while wearing TCAPS devices. Pollastek and Rayner (1998) explain that reaction time can be used to delineate components of mental processing. For example, the authors explained that when items are perceived as similar, reaction time increases, reflecting increased processing time. Therefore, reaction time can reflect time to identify an event and make a decision: key components in situational awareness. (Pollatsek & Rayner, 1998).

While longer response times are consistent with decreased automaticity, or a more difficult task, longer response times also may reflect a difference in task demands. In the field study, participants were instructed to speak his or her response in order to provide a back-up written record. The investigator noted that the single participant who produced outlier data on the measure of response time spoke his response before responding. Given that the task of speaking the response was not part of the in-office experiment, and thus an unfamiliar task, this extra step could have contributed to the response time considerably. Additionally, the experimental apparatus in the field was far less tolerant to deviation, which may have led to participants being far more tentative and deliberate in their responses. Specifically, if the participant triggered the LabView software before he or she was ready, or if the device misfired, the participants knew that no re-starts were allowed. In the lab, participants in the trained group went through many practice sessions after the pretests, whereas the field left no margin of error.

Signal Duration and Head Movement

Adding to the complexity of localizing the in-field gunshot stimulus versus the in-office dissonant signal was the shorter duration of the in-field stimulus. In the lab, the dissonant signal duration was 1000 msec. However, the duration of gunshots from a pistol is less than 50 msec (Maher, 2006). The direct ray of the gunshot, containing broadband information, contained both ILD and ITD information. While not directly measured, gunshots are widely accepted as occurring for less than ½ a second, rendering them an impulse noise hazard. Some reflections of the lower frequency energy may increase the duration of the blank gunshots in the field via reverberance. However, the overall shorter duration of the originating signal creates a more acoustically challenging stimulus to localize. Scharine & Letowski (2005) reported that head movements, which can be particularly helpful in sound localization, are primarily beneficial for sounds of greater than 400 to 500 msec. Therefore, while reflected gunshots sounds should have provided these additional localization cues, but probably not with the stimulus integrity of the directly-emitted one second duration, broadband dissonant sound in the lab. Therefore, the impulse noise of the in-field gunshots contained both ILD and ITD information, but the shorter duration precluded some head-turn benefit available in the in-office environment.

Muller & Bovet (1999) found that the pinna effects and head movement had a synergistic effect on localization accuracy in the horizontal plane. Removing either the pinna or head movement effect resulted in a 10% degradation in localization acuity (Muller & Bovet, 1999). When just the pinna effects were reduced by filling the troughs of the pinnae with impression material, head movement displacements were larger. However, head movements did not fully compensate for the lost pinna effects (Muller & Bovet, 1999). The aforementioned study aligns with the results of the study. Specifically, at pre-test, the experimental and control groups in the open ear condition resulted in 73-75% absolute accuracy, whereas the ComTac™ III resulted in 33-36% absolute accuracy. While loss of pinna effects cannot entirely account for the degradation in performance recorded in the ComTac™ III pretest condition, the acoustical barrier the circumaural hearing protection causes is irrefutable. Similarly, in the field, the control group in the open condition showed a mean absolute accuracy of 62% whereas the ComTac™ III condition resulted in 34% absolute accuracy. Certainly, the signal processing strategy in the ComTac™ III and differences in stimuli spectra between the office and field environments can contribute to degraded performance. However, the deleterious effects on localization of the loss of pinnae filtering cannot be ruled out given the spectral shaping the pinnae provide.

Additionally, the 13% degradation in the open condition from pretest to the field for the control group aligns with loss of head turn cues. The TEP-100 only partially filled the concha and would have been expected to preserve at least some pinna-filtering cues. However, the effects of the device's signal processing, as previously discussed, are suspected to have interfered with this benefit. Absolute accuracy in the TEP-100 conditions were the same or worse than the ComTac™ III in the office and field environments for both groups. The TEP-100 showed poor absolute accuracy (29% for the control in the pretest and 28% for the experimental). Therefore, when participants had access to both pinna effects in the open ear condition and head movement of the lab, the open-ear in-lab resulted in the highest absolute localization scores (Muller & Bovet, 1999).

Signal Effect on Localization

Another challenge presented by the blank gunshot stimulus versus the dissonant signal was the narrower frequency spectrum in the blanks. The field stimuli generated signals with localizable spectral content from about 800 to 1500 Hz and from 3000-7000 Hz, as shown in Figure 88. The dissonant signal incorporated more low-frequency energy, particularly below 800 Hz, shown in Figure 85. Stevens and Newman (1936) found that frequency had a significant effect on error in horizontal localization, and front-back errors in particular. In their experiment, localization was most accurate below 1000 Hz, but stimuli between 2000 and 4000 Hz rendered the largest localization errors. In this study, the in-office signal incorporated energy at 104 and 295 Hz, spectral content not present in localizable levels in the in-field test. Therefore, the broader frequency signal content present in the in-office study rendered its signal easier to localize than the gunshot in the field, even though the gunshot did provide some ILD and ITD cues. Of course, in the interest of optimizing the training effect, the additional low frequency energy is appropriate for use in the PALAT system.

Inherent in over-the-ear (muff) hearing protection is the complete loss of pinna effect cues. Surprisingly, the TEP-100 creates less of an obstruction than the ComTac™ III, but the mean absolute performance for the TEP-100 was lower than the ComTac™ III. The investigator believes that the sound processing algorithm, the onset and ramp-up of the compression in the pass-through gain circuit, and/or the fidelity of processing and passing-through localization cues are the sources of lower performance in the TEP-100. The poor localization performance while using the TEP-100 was corroborated via the Phase II localization accuracy training and testing in two test facilities: the DRILCOM hemi-anechoic facility and the in-office PALAT system. The investigator acquired a frequency response curve chart depicting the compression curve from 3M^{TM} in order to analyze the effects of compression onset and ramp-up on localization performance. Figure 143 confirms that the TEP-100 begins compressing the auditory signal at 60 dBA as indicated by the knee-point in the minimum volume line (red line) using a broadband pink noise signal (Stergar, Fackler, & Hamer, 2019). At 70 dBA, the sound pressure level used in both the in-office and in-field experiments, the gain circuit using the minimum volume or unity gain setting on the TEP-100 compresses the signal to an output level below the input level. This sound level reduction in a single ear may not have a significant impact on localization in isolation. However, each TEP-100 earpiece acts independently with no communication with the earpiece in the opposite ear resulting in varying degrees of compression for the arriving signal.

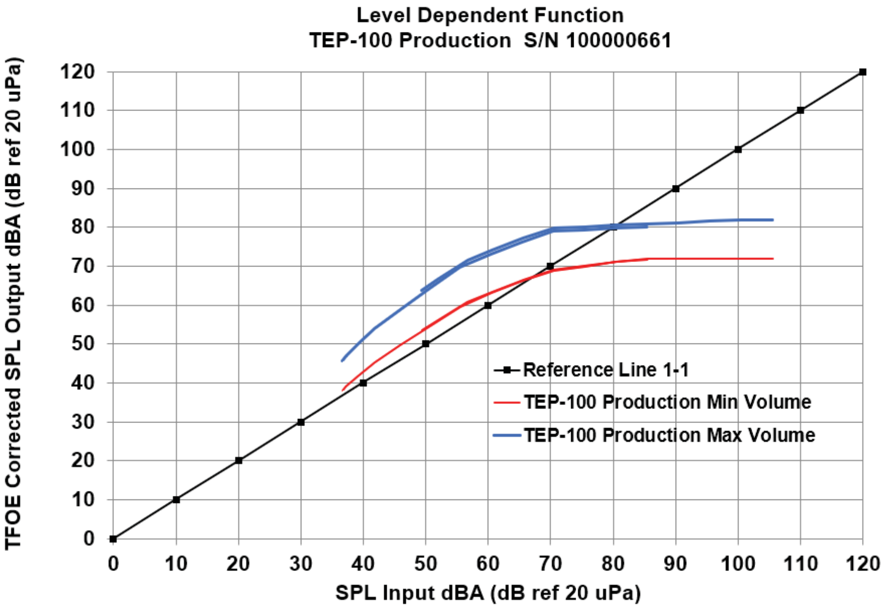


Figure . Level dependent function of the TEP-100 measured using a broadband pink noise signal (Stegar et al., 2019).

As a result of the independent or unsynchronized compression functions, the head shadow effect which create the interaural level difference localization cues could result in the earpiece on one side of the participant’s head to compress the signal while the opposite earpiece fails to reach the compression threshold. Altering or eliminating interaural level differences would significantly degrade localization performance. To test this scenario, the investigator measured the sound pressure level of the dissonant tonal signal at both 55 dBA and 80 dBA from the 12 loudspeakers in the DRILCOM facility using the KEMAR manikin. Figure 144 displays the sound pressure levels of the dissonant signal at both 55 dBA and 80 dBA at 800 Hz, 5000 Hz, and 8000 Hz 1/3 octave-band frequencies as measured using a Larson Davis® measurement microphone located inside the ear canal of the KEMAR manikin under the open ear and TEP-100 listening conditions. The resulting impacts of the compression settings are illustrated by comparing the radial plots at each 1/3 octave-band frequency between the 55 dBA and 80 dBA signals. At 800, 5000 and 8000 Hz, the difference in sound pressure level recorded with the open ear and the TEP-100 is larger in the near ear (right side) when the signal level reaches the knee point in the compression circuit. This attenuation of the signal in the near ear under the TEP-100 indicates that both the dissonant tone and blank gunshot signals set at 70 dBA were impacted by the TEP-100 compression circuit during the localization testing. In addition, the 80 dBA signal radial plots display the impacts of the independent earpiece circuitry. When the 80 dBA signal originated from the near ear, the compression circuit within the TEP-100 earpiece reduced the signal at all 1/3 octave-band frequencies, indicated by the gap between the open ear SPL, dashed blue line, and TEP-100 SPL, solid orange line, from loudspeakers 12 o'clock to 5 o'clock. Whereas, when the 80 dBA signal originated from the far ear, the compression circuit with the TEP-100 earpiece maintained unity gain presenting a similar SPL as presented with the open ear listening condition, indicated by the overlapping lines of the open ear SPL, dashed blue line, and TEP-100 SPL, solid orange line, from loudspeakers 6 o'clock to 11 o'clock. More testing is needed to confirm the impacts of the independent compression circuitry with each TEP-100 earpiece (of a pair) on localization performance. However, preliminary results indicate the poor localization performance under the TEP-100 listening condition may in part be attributable to the sound processing algorithm, the onset and ramp-up of the compression in the pass-through gain circuit, and/or the fidelity of processing and passing-through localization cues.

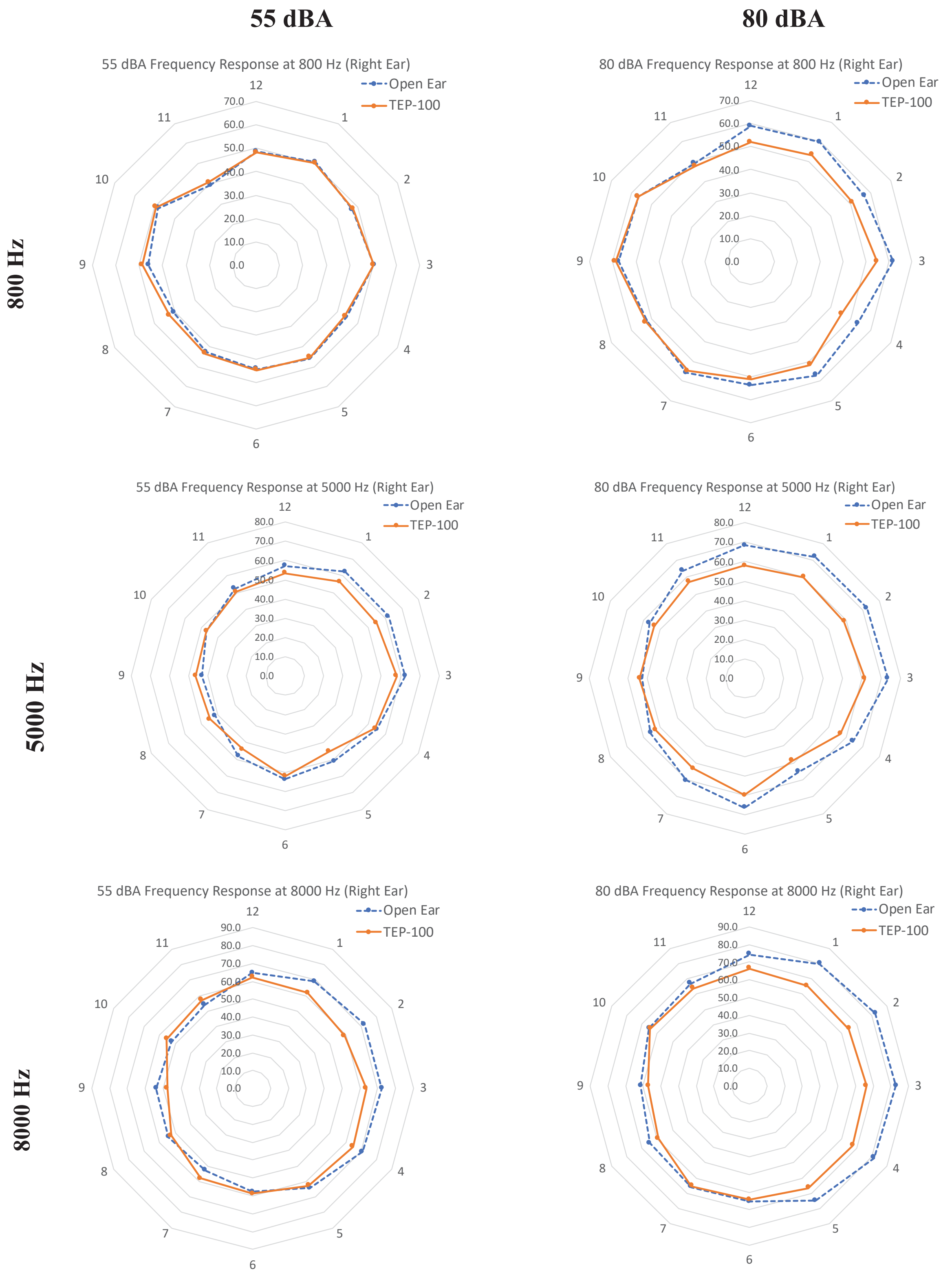


Figure . Frequency response radial plots of mean sound pressure level at a single 1/3 octave-band frequency. Measurements recorded using KEMAR manikin with open ear (dashed blue lines) and TEP-100 (solid orange line) in right ear. Dissonant tone set at 55 dBA (left column) and 80 dBA (right column).

Overall, those who received training demonstrated significant differences in performance compared to the untrained group for the open ear and TEP-100 conditions. Holding all other variables constant, the open ear condition resulted in the highest mean performance in the field. However, in the untrained, open ear condition, performance significantly decreased in the field as compared to the office. The trained group in the open ear condition improved, although not significantly. Therefore, results supported that the training overcame the loss in accuracy from the office to the field. The TEP-100 condition was the only condition that exhibited significantly improved performance from pretest in the office to posttest in the field. However, the *overall* mean performance for the TEP-100 condition was the worst compared to all other listening conditions in the office and the field, attesting to certain design issues that seem to exist with this TCAPS-Lite. The ComTac™ III was not susceptible to the transfer-of-training from the lab to the field.

While trends emerge for training effects in the field according to listening condition, the implications of these results remain unknown. Given that the degree of localization accuracy needed for certain duties remains to be defined, attaching meaning to the results of this experiment as to exactly what accuracy is advisable still presents a challenge. In other words, the degree of accuracy a service member requires to perform his or her tasks needs to be defined first, before any criteria for training performance is established. The localization capability may also vary considerably according to the mission or task. A scarcity of literature exists on this topic mainly because simulating poor localization ability is difficult to operationalize. However, Brungart & Sheffield (2016) were able to assess hearing loss and localization effects separately in simulated combat scenarios. A hearing loss simulator was used to either impose hearing loss or disrupt binaural cues. Participants with only a simulated hearing loss adopted compensatory strategies whereas those with degraded localization did not and performed considerably worse than those with hearing loss. Results were suggestive of those with localization loss not realizing the degradation of SA. Building on the work of Brungart & Sheffield (2016), hearing loss and localization loss simulators demonstrate potential in assessing localization requirements. In order to assess the effects of localization loss, the PALAT or DRILCOM apparatus could be used to quantify localization blur and the types of errors associated with the simulator settings. Soldiers could then be assessed while conducting standardized job-related tasks while using the simulator. From performance results, job-related performance could then be predicted given a certain degree of localization loss. The PALAT or DRILCOM system could then quantify localization blur with the use of various assigned HPDs or TCAPS. These products could then be used in the same job-related tasks in order to validate the localization apparatus results against real-world performance. If validated, then hearing protector performance given certain job-related tasks could be predicted and best matched to the service member's duty. While localization requirements are unknown, the results of this study provide data in order to render informed decision-making down to 30° acuity when that data becomes available. In other words, the current study serves as the initial study of additional field-validated experiments needed to determine specific localization requirements for Service Members that must function in different missions and tasks.

Before training is implemented in any setting, the extent of training extinction should be quantified. The temporal and monetary costs associated with this system is significant as Service Members already have onerous annual training requirements. Even a one-hour annual training requirement can be burdensome to Service Members, especially commanders who are responsible for training compliance. Therefore, use of this training system should include long-term benefits. Kraus et al. (1995) demonstrated that auditory discrimination skills persisted one-month post-experiment via electrophysiological responses. Not only should training extinction be assessed, but electrophysiological measurements should be collected to further quantify the long-term impact, if any, of this training. In other words, evidence of neural pruning given localization training could better quantify the benefits of auditory localization training. It is also recognized that assessment of the extent of training extinction is a large undertaking, well beyond the scope of the current research, and will entail considerable follow-on longitudinal field data collection.

Questionnaire Rating Scales

Regarding *confidence ratings,* training did not result in a significant increase in confidence from pretest to posttest for each listening condition. However, a trend emerged where the open ear was associated with the highest confidence ratings, regardless of stage of training or training group, followed by the ComTac™ III, and the TEP-100. While the differences between pretest and posttest ratings within each group were negligible, differences in confidence ratings among devices were evident. The open ear condition was consistently and significantly associated with the higher confidence ratings than the ComTac™ III and TEP-100. These results also align with the findings of Abel (2008) and Bevis et al. (2014) that perceived loss of situation awareness served as a barrier to wearing hearing protection. Fundamentally, if a user does not have confidence in the key environmental cues transmitted through a device, the likelihood of compliance and HPD usage decreases. Unlike Brungart & Sheffield's (2016) study, participants had other listening conditions by which to compare performance and via feedback given, were aware of the degraded performance imposed by a particular device.

Regarding *perceived accuracy,* trends generally aligned with those of confidence responses. Both trained and untrained participants, regardless of training stage, rated the open ear condition as consistent with the most accuracy, followed by the ComTac™ III and the TEP-100. As in responses regarding confidence, accuracy ratings were not significantly different from pretest to posttest according to listening condition and group. Similar to the confidence ratings, no significant differences emerged comparing the trained versus untrained groups at each training stage. However, significant differences were present for the trained and untrained group at pretest, suggestive of the trained group expressing greater perceived accuracy before any treatment was rendered. This intragroup difference did not persist into the posttest results. Just as in the ratings of confidence question, ratings of perceived accuracy showed significant differences among listening conditions at pretest and posttest. In particular, the open condition resulted in higher ratings of accuracy compared to the TEP-100 and ComTac™ III conditions. Results support that perceived accuracy aligned with actual accuracy on the objective measure of absolute and ballpark accuracy. Of note, participants were informed of their localization accuracy scores prior to populating the questionnaires, and therefore, feedback regarding results may have driven these ratings.

For the question regarding *perceived difficulty,* with a higher rating indicating an easier task, responses followed the same trend as the questions regarding confidence and accuracy. As with the confidence and accuracy questionnaire items, significant differences were present among listening conditions at pretest and posttest for the trained and untrained groups. Listening with the open ear condition was rated as significantly less difficult than the TEP-100 and ComTac™ III. The open ear condition resulted in ratings consistent with the least difficulty, followed by the ComTac™ III, and the TEP-100 was associated with the most difficult localization. One noteworthy trend was the stability of ratings within each condition and group from pretest to posttest. In other words, perceived difficulty did not change even after training, but ratings were consistent with perceived confidence and accuracy. The consistent difficulty ratings from pretest to posttest could have a profound effect on Service Members. The untrained group showed significantly poorer performance on the posttest compared to the pretest for the open ear condition. However, there was no perceived increase in difficulty by the untrained group. This important result potentially suggests that Service Members are not able to perceive the true difficulty associated with localizing various sounds in different environments, obviously posing a dangerous predicament. The PALAT and DRILCOM system could serve as a tool to inform Service Members of their actual localization performance in varying environments, providing them a more accurate situation awareness picture.

On ratings of *perceived reaction time,* an unexpected result emerged in the trained group. Trained group participants rated their posttest reaction time as slower using the TEP-100 and ComTac™ III compared to those in the in the untrained group. Slower response times may have reflected emerging skill acquisition, meaning that participants were more deliberate in their selections given the training. While not significantly different, every listening condition and group rated slower reaction times in the field with the exception of the TEP-100 for the untrained group. Given the lowest accuracy scores associated with this TEP-100 untrained group, the unchanged response time may have reflected a lack of motivation. In other words, given the difficulty of the task, participants may have decided that more time invested into the task did not result in more accurate results. As with previously discussed questionnaire items, reaction time ratings were significantly different for the trained and untrained participants at pretest and posttest. At pretest, both trained and untrained participants showed significantly faster ratings for the open ear than for the TEP-100 or ComTac™ III. At posttest, the trained participants showed significantly faster ratings for the open ear than the TEP-100 or the ComTac™ III. For the untrained condition, only the TEP-100 was rated as significantly slower than the open ear.

Participants' ratings of *comfort* offer another insight into barriers to compliance with hearing protection. Participants consistently reported the open ear as the most comfortable listening condition at each stage of training and for each group. Unexpectedly, wearers of the ComTac™ III in the untrained condition reported significantly less comfort in the posttest compared to the pretest stage. These results would have been more likely in the trained group, in view that these participants wore the devices longer. Otherwise, no significant differences occurred when comparing pretest to posttest comfort ratings. At pretest, the trained and untrained conditions showed significantly less comfort in the TEP-100 condition compared to the open ear. In the untrained condition at pretest, the ComTac™ III was also found to be significantly more uncomfortable than the open ear. The same statistically-significant results for the pretest carried through to the posttest. The data from the comfort question suggest that training with the ComTac™ III reduces perceptions of discomfort. In general, drawing inferences about the comfort of in-the-ear and over-the-ear TCAPS is difficult given that only two devices were used in this study. However, these particular devices demonstrate clear differences between devices. Additionally, an acceptable level of comfort was not addressed to serve as a basis of comparison. In other words, a lower rating of comfort does not necessarily imply that the level of discomfort would lead to non-use of hearing protection.

For the question regarding *likelihood of maintaining the same listening condition given a task that required sound localization,* participants clearly preferred the open ear. This study incorporated immediate feedback due to the training component. Additionally, because of risk to human subjects, hazardous background noise was not included. As such, participants demonstrated a clear preference for the open ear in non-noise hazardous environments and with knowledge about their performance. Had participants not known how good or poor their localization performance was while wearing these devices, they may not have realized their performance decrement as Brungart & Sheffield's (2016) study suggests. Moreover, the intended use of TCAPS is to protect the user from the physiological and psychoacoustic adverse effects of noise. In other words, this study did not fully incorporate all of the environments where TCAPS devices would be worn and the results should not necessarily be generalized to noise hazardous environments. However, Service Members must operate in quiet environments where unexpected noise hazards can occur, and the results have direct application for that. As such, TCAPS devices should be able to adequately address the need to localize in quiet. While not significantly different, participants rated the open ear higher in the posttest than pretest for both the trained and untrained conditions. As with all other rating scales, significant differences were found at pretest and posttest for the trained and untrained users among listening conditions. For the trained and untrained group at pretest, the open ear condition was rated significantly higher than the TEP-100. In the untrained group, ratings for the open ear were also significantly higher than the ComTac™ III. At posttest, the same significant findings were also recorded for the open ear versus TEP-100 conditions and for the open ear versus ComTac™ III conditions. Results reflect a lesser likelihood of wearing the devices in a quiet field environment for localization, as compared to having an unprotected ear. Results should be interpreted with the understanding that this evaluation only queried the localization aspect of auditory perception, and thus likelihood of use ratings may have differed given different environmental conditions (i.e., in noise).

An additional question was administered only to the trained group in the posttest stage. The question queried the *level of preparedness* the participant felt as a result of the in-lab training. Results showed a significant difference among listening conditions, with the open ear condition resulting in ratings consistent with the highest level of preparedness. Significant differences existed between the open ear versus TEP-100 and the open ear versus ComTac™ III, with the TEP-100 receiving the lowest ratings of preparedness. The mean ratings across listening conditions were all greater than 3.5, the middle rating, suggestive of feeling more prepared than not. In lieu of the training verbiage and substituting with the experience of the pretest, administering this question to untrained personnel as well might have provided a more comprehensive understanding of feeling of preparedness. In other words, administering a similar question to the untrained personnel may have provided an adequate basis of comparison. Otherwise, considerable limitations exist for drawing inferences regarding this question using only the training group.

IMPLICATIONS OF THE RESULTS

Limitations of the Research

As previously stated, the study did not evaluate training extinction. While at least a day lapsed between the final office session and field testing, the validity of this scenario of one day between training and field localization actually occurring remains unknown. Furthermore, Service Members generally conduct duty-related training with their issued hearing protection prior to localizing sounds in a field environment or in a deployed setting. The current in-field study served as the first step to converging on a clearer understanding of mechanisms employed in sound localization. Future investigations into the training effect on in-field localization should consider the influence of real-world experience with altered localization cues imposed by hearing protection. Hofman et al., (1998) demonstrated that without feedback regarding performance, listeners with impression-filled conchas adapted to the altered localization cues after three to six weeks. Thus, performance should be compared to those receiving formal in-office localization compared to those who only have experiential (i.e., on-the-job in-field) learning with the devices. While experiential learning may improve localization accuracy, it may also influence learning decay. Therefore, training extinction effects should be considered within the context of real world use of these devices.

Another limitation to the study was the validity of using a .22 caliber blank. This stimulus served as one of the few sound stimuli that was technically feasible given that live ammunition could not be used for obvious safety considerations. The quality control of the ammunition was previously discussed. The same brand of ammunition was used throughout the study, but that served as the only investigator-imposed means of quality control; but in any case, all blank shots were significantly suprathreshold and provided the opportunity for localization. In addition to quality control, the field study should be replicated using other types of real-world stimuli. Future investigations may incorporate higher caliber military weapon systems, whistle tube simulators to simulate incoming rocket propelled grenades, or mortar rounds that are relevant to ground-combat Service Members. The portable auditory localization and training (PALAT) system has already been updated to incorporate a variety of military relevant sounds for this purpose.

The terrain of the field environment represents one possible outdoor localization scenario amidst an endless possibility of scenarios that can be encountered in military environments. One environment that varies considerably from the environment used in this study, but which is commonly encountered by U.S. Service Members, is urban operations (UO). Ground combat

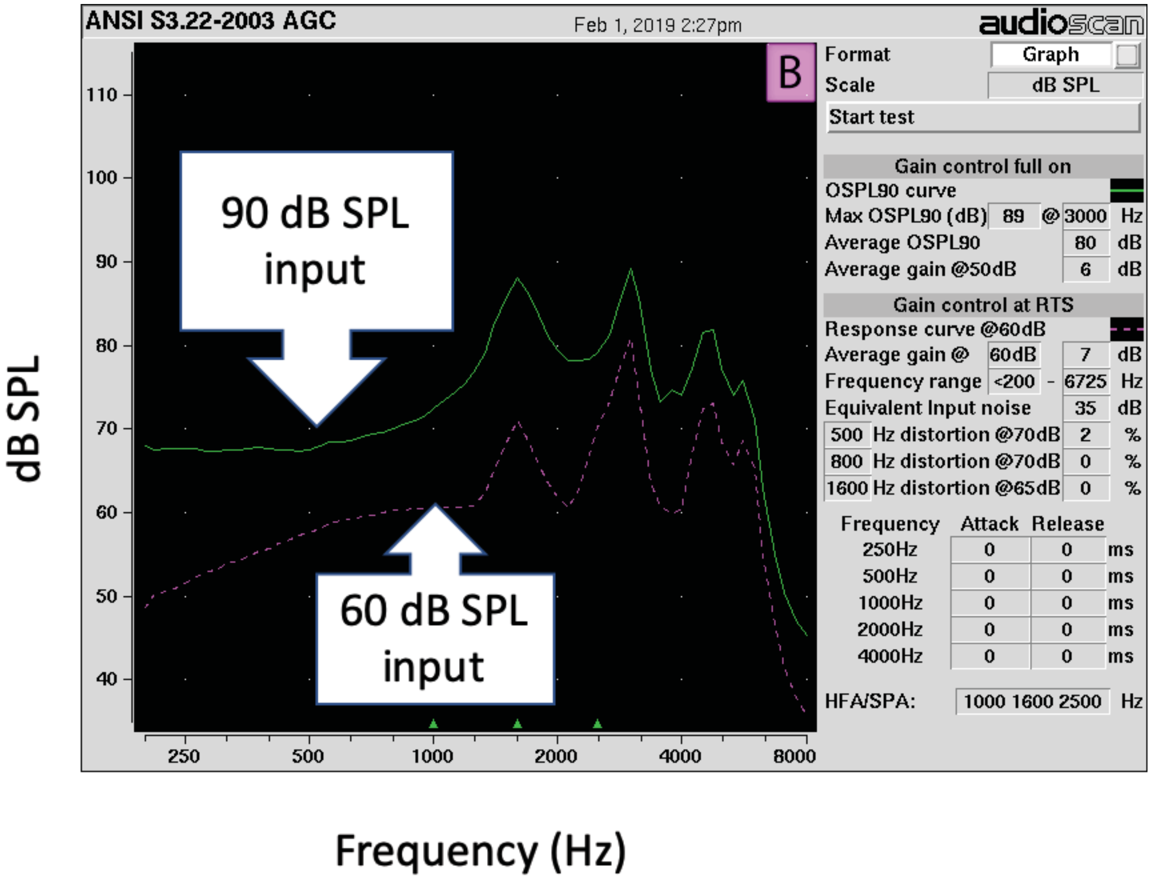
Service Members are required to conduct UO training. This environment presents many localization challenges due to reflections from buildings, ambient urban noise, and the presence of other non-target stimuli that can tax the listener during the detection/recognition phase of ASA. Clearly, even with its inherently more reflective environment, urban operations do require soldiers to localize sounds of various types, so its importance cannot be overlooked. Additionally, given the presence of buildings and subsequent threats from differing elevations in a UO environment, elevation cues remain paramount to threat localization. However, this experiment only addressed azimuth cues. In order to gain a broader understanding of how training impacts auditory localization, future studies should examine the impact in varied field environments. This study addressed one type of environment (arguably the most common for a field setting) and only employed azimuth cue training with its inherent limitation to the experiment's external validity. However, both the DRILCOM and PALAT systems incorporate elevation training capabilities and customizable training signals, thus they offer the capability for expanded environment investigations.

Results explained as a possible function of TCAPS design variables

The design of the ComTac™ III's forward-facing microphones most likely accounts for the frontal plane response bias discussed previously. In a forward-facing omnidirectional array, the microphones are most sensitive to sounds in front of the listener with null points at approximately 240° and 300°, or 4 o'clock and 7 o'clock, respectively (Dillon, 2001). The microphones on the ComTac™ III are also more distally mounted from the head than the ITE microphones of the TEP-100. This design conceivably offers greater sensitivity to environmental sounds due to proximity and larger microphone size of the ComTac™ III. Additionally, the passive sound isolation of the earcup may offer an advantage in the signal processing in the

ComTac™ III. The ComTac™ III earcups offer greater sound isolation, creating a more robust barrier between the noise source and the entrance to the canal. On the other hand, the TEP-100 receives a higher level of input due to pinnae funneling and resonances. The effect of higher input at the microphone is a more aggressive compression strategy that can distort or disrupt localization cues. Therefore, the closer proximity of the microphone to the source, larger surface area of the microphone, greater sound isolation of the earcup, and thus, less compression, may account for improved localization performance with the ComTac™ III compared to the TEP-100.

Another possible scenario is that the TEP-100's compression is triggered in the near-ear where the signal is louder (due to its relatively low threshold at compression onset), thus sending the near earplug into compression, rendering a lower level signal to the closer ear. Attenuating the closer signal disrupts the precedence effect where a closer signal is perceived as louder. Another alternative explanation for the poorer localization accuracy is gleaned from the input-output curves obtained in a hearing aid test box (Audioscan, Verifit) shown in Figure 145. In this measurement, the TEP-100 was set to unity gain and a pure tone frequency sweep was delivered at 60 and 90 dB SPL. These measurements were not in-situ measurements, which may better reflect how the outer ear transfer function would affect the frequency response of the output. The response curves show that the output varies considerably according to frequencies above 1600 Hz. For example, input at 1600 Hz is 68 dB for 60 dB input, but only 60 dB at 2000 Hz. Disparities at these frequencies given the same level could disrupt the ratio of frequencies employed in location perception. Thus, the poorer performance from the TEP-100 may be due to the smaller surface area of the microphones, the more aggressive compression strategy, an unequal compression algorithm applied at different frequencies, and an asymmetrical compression strategy with one the on-ear compressing first and applying more attenuation.



ANSI S3.22-2003 AGC

Feb 1, 2019 2:27pm

Format: Graph

Scale: dB SPL

Start test

Gain control full on

OSPL90 curve

Max OSPL90 (dB) 89 @ 3000 Hz

Average OSPL90 80 dB

Average gain @50dB 6 dB

Gain control at RTS

Response curve @60dB

Average gain @ 60dB 7 dB

Frequency range <200 - 6725 Hz

Equivalent Input noise 35 dB

500 Hz distortion @70dB 2 %

800 Hz distortion @70dB 0 %

1600 Hz distortion @65dB 0 %

Frequency	Attack	Release	
250Hz	0	0	ms
500Hz	0	0	ms
1000Hz	0	0	ms
2000Hz	0	0	ms
4000Hz	0	0	ms

HFA/SPA: 1000 1600 2500 Hz

Figure . Input-output curve across frequencies using a pure-tone sweep at 60 and 90 dB SPL for the TEP-100 in the unity gain setting. Data was obtained with an Audioscan Verifit system.

Recommendations for Efficiency and Effectiveness in Training with the Portable Auditory Localization Training System (PALAT)

Recommendations for use of the PALAT in various room environments

Portable localization testing systems of the type developed and tested in this research also demonstrate promise as a tool to assess the adverse effects of hearing loss. Currently, auditory localization is not incorporated into hearing readiness metrics for U.S. Service Members.

Therefore, duty limitations are not aligned with the handicapping effects of impaired localization associated with hearing loss. To assess the impact of differing degrees of hearing loss on auditory localization, the portable PALAT (or the larger DRILCOM system) could quantify localization blur given a certain hearing loss. Assuming field validation of the hearing loss simulator used in Brungart & Sheffield's (2016), one application is that experiments could assess the impact of varying degrees of simulated loss on duty-related tasks. Secondly, advantages can be taken of the fact that the PALAT is designed to fit within a standard examination room inside an audiology clinic for possible inclusion in hearing readiness metrics. Service Members with hearing loss could thus be assessed with PALAT in order to predict the functional impact of their localization accuracy as part of a standard test battery. The goal of these evaluations would be for clinicians and leaders to better align duties given a service member's hearing loss. Given that a standard clinical encounter in audiological setting lasts approximately one hour, localization evaluations could reasonably be assessed using the PALAT in the open ear and with a few TCAPS within this timeframe.

The PALAT system and training protocol may be implemented as means of training to counter auditory localization loss associated with loss of auditory acuity. Undoubtedly, future research should address the effects of localization training in those with localization loss, either temporarily-imposed by hearing protection use or permanently from hearing loss. Localization remediation in hearing-impaired listeners through auditory training has yet to be thoroughly investigated. Furthermore, the possibility of restoring localization cues and the effect on localization accuracy through TCAPS has also yet to be evaluated. Therefore, the effects of localization loss remediation could be evaluated, and possibly treated, using the PALAT system and training protocol. The results of this study also support that training delivery of five learning units (LUs) can be administered within an hour, a timeframe that is currently the standard audiological evaluation time in most clinics. Also, training does not necessitate the presence of a clinician as the PALAT is intended to be executed independently by the listener, or trainee in this case.

Recommendations for use of the PALAT System as setup and deployed by a trainee

The PALAT system was specifically designed to be user operated by a trainee with no prior experience with localization training. The current portable system has a diameter measurement of seven feet. It is recommended that trainees select a room that allows for one meter of space between the perimeter of the system and any reflective surface. This will reduce the likelihood of reflected sound rays interfering with the training and testing. If possible, the trainee should try to place the PALAT system in the center of the room and remove or spread out any furniture in the room along the walls as far as possible from the portable system. Use of curtains, closing window blinds, and carpeting can mitigate reflections. However, the investigator understands that space is often limited and that the PALAT system may need to be operated in less than ideal environments. As such, in this research, the PALAT system was deployed in a semi-reverberant room with multiple reflective surfaces, during both Phase II and Phase III. In addition, the PALAT system was intentionally placed off-center within the room with several speakers located only two feet from the wall to mimic what was considered as "non-optimal" placement that would likely be established in actual military practice. As seen in the results of Phase II and Phase III, the PALAT system was still able to impart improvements in localization training from pretest to the last learning unit in all three listening conditions.

It is also recommended that the PALAT system be employed in quiet room setting free from any continuous or frequent loud noise sources. Ideally the ambient environmental noise should remain below the 55 dBA masking noise produced by the PALAT system. The trainee should test the ambient noise by turning on the pink noise or alternative masking noise source and listening to hear if any external noise sources can be heard. Training and testing sessions should be paused temporarily in the presence of any occasional loud noises. An additional feature is being incorporated in the next software generation of the PALAT system that will continuously measure the ambient noise and adjust the masking noise source and training signal to maintain the desired signal-to-noise ratio. This "automatic gain control" feature will need to be tested to identify the impacts on localization training.

Recommendations for use of the PALAT System given time likely available and time likely required

As previously discussed, this investigation did not measure the impacts of various training durations on localization performance. In addition, the military, or any other industry, has yet to define a standard for the minimum or desired level of localization performance. With that in mind, the training protocol developed in Phase I of the overarching investigation aimed to reduce the training time to a reasonable duration based on current U.S. service member training availability. The investigator equated localization training to rifle marksmanship training, based on their personal U.S. Army experiences. Typically, U.S. Service Members conduct basic rifle marksmanship at least twice a year and are provided with a few hours of training via simulator and practice ranges before being tested. The current training protocol employed during Phase II and Phase III was limited to three, one-hour training sessions. Participants were allowed to take a break at any point during their training. Based on results from Phase II and Phase III, one hour of training (5 LUs) for the open ear listening condition resulted in an asymptotic level of localization accuracy performance, with diminishing and negligible benefits of administering further LUs. Both the TEP-100 and ComTac™ III mean localization accuracy performance failed to reach the open ear performance levels with the same amount of training. As such, additional training may be needed if the goal is to achieve localization performance similar to the open ear while wearing certain TCAPS devices. Furthermore, more testing is needed to identify the frequency and duration of refresher training needed to maintain a desired level of localization performance.

Finally, the PALAT system can be used in a testing sense to determine when a TCAPS device places too high a training burden on users, and perhaps should be eliminated from consideration as a result. Toward this end, further research needs to be performed to determine at what point in the training process a TCAPS device should be eliminated if it requires an inordinate amount of training to bring the trainee up to a criterion level (or not at all). Casali & Lee (2016), found, for instance, that one particular in-the-ear prototype TCAPS never did asymptote in its learning curve even after 12 LUs, and never reached 86% absolute correct performance of the open ear performance level -- thus, it was recommended that its development be discontinued.

Implications for military implementation

Relevance to TCAPS design, selection, and procurement

As with any hearing protector, the best protector is the one most appropriate for the duty and environment. Acoustically and historically speaking, the major determining factor in hearing protection selection is how much attenuation is needed. Required attenuation and noise reduction ratings (NRRs) are generally known and routinely evaluated, and must be published for all HPDs sold in the U.S. The most important other determining factor in selection is the extent to which auditory components of situation awareness should be preserved to effectively perform job-related duties. Currently, how performance differs using a hearing protector compared to the open ear is mostly known, and that can be useful in communicating risk. For example, Table 84 demonstrated that given the effective range of a weapon, increasing localization blur, as is likely certain hearing protectors, can considerably increase the visual search area. However, the implications of the risks of degraded situation awareness cues can only be partially predicted, since they are different for every situation. Nonetheless, the lack of auditory situation awareness has been clearly evidenced as a causal factor in many accidents (Casali, 2019).

In the absence of a known need, the open ear's capability should serve as the gold standard in regards to assessing acceptable risk with TCAPS use from a products liability standpoint. However, this is most likely not necessary in all noise-hazardous situations, but it does ensure that the least amount of acceptable risk is adopted. Exceptions to where the open ear should not hold true as the standard is when Service Members operate in an enclosed environment. For example, those monitoring unmanned aerial vehicles may need to communciate with fellow operators, but do not need to localize particular threats in their environment, due to being enclosed in safe surroundings. Other examples of where hearing protection and communication are required, but not localization, are inside a tracked vehicle (e.g., military tank) or in an aviation setting. Weapons instructors at an outdoor range would not necessarily need open ear-equivalent localization capabilities, but would need to clearly hear incoming and outgoing communication. Therefore, from a liability perspective, quantifying localization loss, or any other degraded aspect of ASA, could better inform users of implied risk in using the device. Although not ideal, research can quantify the full scope of the effects of TCAPS use on ASA, and inform the user. Describing the risk and letting the stakeholder decide what device is most appropriate offers the best way to balance acceptable risk with safety requirements.

In the U.S. Military, Service Members do not typically have the opportunity to procure their own TCAPS. Commanders, acquisitions personnel, and occasionally clinicians render decisions regarding procurement of these devices. Employing a standardized testing system, especially a portable one, could better inform TCAPS stakeholds of the associated risks and benefits for these devices. Generally, one TCAPS is not well-suited for all scenarios where hearing protection would be needed. While the psychoacoustic needs may remain unknown, generally missions and duties have associated ASA requirements. Accordingly, a list of requirements generated by the end users and commanders thereof could be matched with capabilities requisite in an HPD or TCAPS to meet those requirements. Currently, a program of record exists for the Army for TCAPS, but the requirements are not disseminated.

Relevance to ground combat service member duties and mission

As policy, the U.S. Army does not release hearing loss metrics specific to units for security reasons. However, in 2016, hearing readiness metrics from an Army special operations unit showed signficantly lower incidence of hearing loss compared to conventional infantry units on the same installation (Klingseis, 2017). The difference in hearing thresholds between the two groups of units could not be accounted for due to differences to age or rank. One main difference between the two types of units was that in the special operations unit, ComTac™ III hearing protection was mounted on the helmet via rail attachment and integrated into the communication system. In other words, use of hearing protection was part of the standard personal protective equipment ensemble, and not a separate item. Conventional forces can use a variety of HPDs or TCAPS, but do not have such an ensemble requirement. Therefore, while a causal relationship between the use of the TCAPS integrated into standard equipment has not been established, audiometric data in this example and others supports a strong positive relationship between TCAPS use and hearing loss prevention.

Another implication of the finding of Klingseis (2017) is the importance of TCAPS not just restoring ASA cues and preventing hearing loss, but also providing U.S. Service Members with an operational, even tactical advantage. In other words, the goal of TCAPS should eventually be to achieve performance beyond that of the open ear and improving overall warfighter performance. For example, noise reduction algorithims integrated into hearing protection could enable improved speech transmission and understanding, more accurate localization, and improved threat detection. In addition to the psychoacoustic advantage afforded by improved designs, less noise exposure could also lead to positive implications of decreasing workload and fatigue. Gaining a strategic advantage through device use would certainly improve compliance, thus reducing hearing loss. While the ComTac™ III as evaluated herein may present challenges to certain aspects of ASA, compliance with wear of these devices is irrefutable, as demonstrated by Klingseis (2017). However, a current shortfall exists in communicating the associated risk of using a certain TCAPS devices to the user. Specifically, Service Members are not necessarily aware of the adverse, often lethal effects of not being able to localize well, especially when detection is improved with devices which provide amplification, such as the ComTac™ III. In other words, the service member perceives that they can hear better (the detection benefit) and thus views the TCAPS as a performance enhancer, not realizing that other aspects of hearing, such as localization, are compromised. While TCAPS offer a means to improve performance, in their current state they have associated risks that should be adequately communicated to the end users. Armed with this knowledge, stakeholders in military TCAPS programs could more accurately define the requirements for manufacturers.

Implications for NIHL reduction

Given the ever-present risk to U.S. Service Members of noise exposure due to training-related and unexpected exposures from hostile actions, compliance with hearing protection usage policies presents a unique set of challenges. The heightened risk of noise exposure is illustrated in the 30% greater likelihood of severe hearing loss in Service Members compared to non-veteran counterparts (Groenwold, Tak, & Matterson, 2011). While hearing protection is widely available to Service Members, compliance obviously lags the identified risk. All U.S. military personnel are required to undergo annual training that explains the risk of hearing loss and how to mitigate noise exposure. Despite these efforts, Service Members often choose not to wear hearing protection, but sometimes with good reason (Abel, 2008; Bevis et al. 2014). As the studies of Casali & Robinette (2014), Casali & Lee (2016a), Casali & Lee (2016b), Brown et al., (2015), Giguère et al, (2013) among many others, hearing protection use can degrade aspects of situation awareness. However, not all aspects of ASA are critical to each duty. Quantifying the risk inherent in each TCAPS application can assist Service Members in selecting devices that are best aligned with their operational needs, and will also help focus the pre-deployment training that may be advisable with a given product.

Not only does employing standardized testing of TCAPS devices improve device compatibility, but standardized training can better ensure confidence in associated TCAPS use. To overcome the tradeoff of choosing between sufficient protection or situation awareness, Service Members must have evidence-based confidence that TCAPS use will not compromise survivability or lethality. Training on aspects of ASA, including localization, serves as one method of instilling confidence. Establishing confidence in issued equipment is a common practice in the military. For instance, Service Members are required to test their gas masks in gas chambers to experience how the masks protects their breathing in the presence of CS (orto-chlorobenzylidene-malononitrile) gas. Likewise, Service Members test safety harnesses, parachutes, weapon systems, etc. to instill a sense of confidence in their equipment. However, TCAPS devices are typically stored in company supply rooms for accountability, and issued prior to training exercises or deployments without testing or training of the user. Training Service Members on the PALAT system while wearing their issued TCAPS device demonstrates strong potential as a means to improve localization performance and increase confidence in the fidelity of situation-awareness related cues. Conceivably, increased confidence would manifest as increased adoption rates of TCAPS devices and compliance, especially in noise-hazardous environments where detection and localization are critical. As the metrics illustrate, higher compliance with TCAPS use is associated with lower rates of hearing loss (Klingseis, 2017). Additionally, quanitfying the degradation to ASA with TCAPS use can assist manufacturers in better understanding the requirements that generate their designs, and improve future generations of their products. Standardized testing and training of TCAPS devices can improve device compatibility selection, confidence, and manufactuer design.

FINAL CONCLUSIONS

A series of convergent studies supported the presence of transfer of auditory localization training conducted in-lab to testing in-field using live gunshots. Both the lab and field environments demonstrated performance differences in the open ear versus two TCAPS. However, neither training nor testing environments showed significant differences in localization accuracy between the two TCAPS, although localization accuracy improved with training with

ComTac™ III use. Additionally, localization accuracy performance never reached that of the open ear even when training was conducted. The implications of degraded localization imposed by TCAPS use on actual duty-related performance remains unknown. A reasonable assumption is that achieving localization accuracy of the open ear while using TCAPS would improve confidence, and therefore use. In the interim, training offers a means to bridging the gap between localization accuracy in the open ear and while using certain TCAPS. Not all TCAPS are susceptible to training effects, however, as demonstrated by the ComTac™ III data. The training component to these studies also supported that training with a novel broadband DRILCOM dissonant tonal complex stimulus transferred to military-relevant broadband stimuli. While the whistle stimulus did not show transfer-of-training, a different exemplar and perhaps different component of that stimuli should be tested. Only a one-second tonal portion of an otherwise dynamic signal was used. The intentional but representative “non-optimal” PALAT office environment and other factors may have partially reduced achieved training benefits in the laboratory experiments. Results of the series of experiments illustrate the need for more training and testing to be conducted in military training scenarios. Therefore, evidence from a series of experiments conducted in the VT-ASL support, but have yet to confirm, that auditory localization training improves performance on military-related duties that impose auditory tasks in actual operational environments. Nonetheless, from these experiments, it is very clear that with the PALAT portable system as tested, individuals can significantly improve their open ear azimuthal localization performance with four LUs of training, and also exhibit training benefits with certain TCAPS more than with others.

REFERENCES

3M. (2016a). *TEP 100 Tactical Earplug Brochure.* Retrieved from www.3M.com: https://multimedia.3m.com/mws/media/1001819O/tep-100-tactical-earplug-brochure-single-pgs.pdf

3M. (2016b). *3M PELTOR Tactical Comm and Hearing Protection Brchure 2016.* Retrieved from www.3M.com: https://multimedia.3m.com/mws/media/1417830O/3m-peltor-tactical-comm-and-hearing-protection-brchure-2016.pdf

Abel, S., Boyne, S., & Roesler-Mulroney, H. (2009). Sound localization with an army helmet worn in combination with an in-ear advanced communications system. *Noise and Health, 11*(45), 199.

Abouchacra, K. S., & Letowski, T. R. (2001). Localization accuracy of speech in the presence of nondirectional and directional noise maskers. *17th International Congress on Acoustics.* Rome.

Alali, K. A., & Casali, J. G. (2011). The challenge of localizing vehicle backup alarms: Effects of passive and electronic hearing protectors, ambient noise level, and backup alarm spectral content. *Noise and Health Journal, 13*(51), 99-112.

Alali, K. A., & Casali, J. G. (2012). Auditory backup alarms: Distance-at-first detection via in-situ experimentation on alarm design and hearing protection effects. *Work: A Journal of Prevention, Assessment, and Rehabilitation, 41*, 3599-3607.

American National Standards Institute (ANSI). (2019). Methods for Measuring the Effect of Head-worn Devices on Directional Sound Localization in the Horizontal Plane. *ANSI/ASA S3.71-2019*. New York: Acoustical Society of America.

Blauert, J. P. (1997). *Spatial hearing: the psychophysics of human sound localization (Rev. ed.).* Cambridge: MIT Press.

Brown, A. D., Beemer, B. T., Greene, N. T., Argo, T., Meegan, G. D., & Tollin, D. J. (2015). Effects of active and passive hearing protection devices on sound source localization, speech recognition, and tone detection. *PloS Biology, 10*(8), e0136568.

Brungart, D., & Sheffield, B. (2016). The operational impacts of hearing impairment: The important roles that psychology and context play in determining the performance of military members with degraded hearing. *Paper presented at the meeting of the National Hearing Conservation Association.* San Diego.

Casali, J. G. (2010). Powered Electronic Augmentations in Hearing Protection Technology Circa 2010 including Active Noise Reduction, Electronically- Modulated Sound Transmission, and Tactical Communications Devices: Review of Design, Testing, and Research. *The International Journal of Acoustics and Vibration, 15*(4), 168-186.

Casali, J. G. (2010a). Passive augmentations in hearing protection technology circa 2010 including flat-attenuation, passive level-dependent, passive wave resonance, passive adjustable attenuation, and adjustable-fit devices: Review of design, testing, and research. *International Journal of Acoustics and Vibrations, 15*(4), 187-195.

Casali, J. G. (2010b). Powered electronic augmentations in hearing protection technology circa 2010 including Active Noise Reduction, electronically-modulated sound transmission, and tactical communications devices: Review of design, testing, and research. *International Journal of Acoustics and Vibrations, 15*(4), 168-186.

Casali, J. G., & Lee, K. (2015). *Objective metric-based assessments for efficient evaluation of auditory situation awareness characteristics of tactical communication and protective systems (TCAPS) and augmented hearing protective devices (HPDs) (Report No.*

W81XWH-13-C-0193). Virginia Polytechnic Institute and State University, Industrial Systems Engineering. Blacksburg: Auditory Systems Laboratory.

Casali, J. G., & Lee, K. (2016). An Objective, Efficient Auditory Situation Awareness Test Battery for Advanced Hearing Protectors and Tactical Communications and Protective Systems: DRILCOM (Detection- Recognition/Identification- Localization- Communication). *172nd Meeting of the Acoustical Society of America* (pp. 1-60). Honolulu: Acoustical Society of America.

Casali, J. G., & Lee, K. (2017, April 14). Innovative Portable Auditory Localization Acclimation-Test (PALAT) System for Military Applications, Including Full-Scale and Portable Versions, Leveraged on DRILCOM Test Battery and Validated via an In-Field Experiment. *Proposal BAA Number: N00014-17-S-B001, Department of Defense, Office of Naval Research*.

Casali, J. G., & Lee, K. (2019). Learning to localize a broadband tonal complex signal with advanced hearing protectors and TCAPS: the effectiveness of training on open-ear vs. device-occluded performance. *International Journal of Audiology, 58*(S1), S65-S73.

Casali, J. G., & Robinette, M. R. (2015). Effects of user training with electronically-modulated sound transmission hearing protectors and the open ear on horizontal localization ability. *International Journal of Audiology, 54*(Suppl), S37-45.

Casali, J. G., & Tufts, J. B. (2020). Auditory Situation Awareness and Speech Communication in Noise. In D. Meinke, *The Noise Manual (6th Edition).* Fairfax, VA: American Industrial Hygiene Association, in press.

Casali, J. G., Ahroon, W. A., & Lancaster, J. A. (2009). A field investigation of hearing protection and hearing enhancement in one device: For soldiers whose ears and lives depend upon it. *Noise and Health, 11*(42), 69-90.

Casali, J. G., Lancaster, J. A., Valimont, R. B., & Gauger, D. (2007). Headsets in the light aircraft cockpit: Speech intelligibility. *Proceesings of the 2007 International Congress of Noise Control Engineering*.

Casali, J. G., Lancaster, J. A., Valimont, R. B., & Gauger, D. (2007). Headsets in the light aircraft cockpit: Speech intelligibility. *Proceeding Proceedings of the 2007 International Congress of Noise Control Engineering.* Reno, NV.

Casto, K. L., & Casali, J. G. (2013). Effects of headset, flight workload, hearing ability, and communications message quality on pilot performance. *Human Factors, 55*(3), 486-498.

Cave, K. M., Cornish, E. M., & Chandler, D. W. (2007). Blast injury of the ear: Clinical update from the global war on terror. . *Military Medicine, 172*(7), 726-730.

Chandler, D. W., & Grantham, D. W. (1992). Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, soure azimuth, and velocity. *Jouranl of the Acoustical Society of America, 91*(3), 1624-1636.

Clasing, J. E., & Casali, J. G. (2014). Warfighter auditory situation awareness: Effects of augmented hearing protection/enhancement devices and TCAPS for military ground combat applications. *International Journal of Audiology, 53*(Sup2), S43-52.

Craig, J. (2001). Spectral analysis and resolving spatial ambiguities in human sound localization (Doctoral disseration). *The Sydney Digital Theses*. Retrieved from http://hdl.handle.net/2123/1342

Defense Manpower Data Center (DMDC). (2017, July). *DoD Personnel, Workforce Reports & Publications.* Retrieved October 14, 2017, from DMDC: https://www.dmdc.osd.mil/appj/dwp/glossary.jsp

Defense Manpower Data Center (DMDC). (2019, March 10). *DoD personnel, workforce reports & publications*. Retrieved May 2019, from DMDC: DoD Data Reports: https://www.dmdc.osd.mil/appj/dwp/dwp_reports.jsp

Department of Defense [DoD]. (2010). *DoD Instruction No. 6055.12, Hearing Conservation Program.* Washington, D.C.: U.S. Government Printing Office.

Department of the Army. (2017). *Ranger Handbook.* Alexandria, VA: Army Publishing Directorate.

Department of the United States Army . (2017). *AR 40-501, Standards of Medical Fitness*. Washington, D.C.: U.S. Army. Headquarters.

Department of the United States Army. (2015). *DA PAM 40-501, Army Hearing Program.* Washington, D.C.: U.S. Army Headquarters.

Dillon, H. (2001). Advanced signal processing schemes for hearing aids. In H. Dillon, *Hearing aids.* Turramurra, New South Wales, Australia: Boomerang Press.

Dixon, W. J. (1951). Ratios Involvling Extreme Values. *The Annals of Mathematical Statistics*, 22(1), 68-78.

DOEHRS-DR. (2016). *Defense Occupational Health Readiness System Data Repository (DOEHRS-DR).* Aberdeen Proving Ground: U.S. Army Public Health Command. Retrieved from https://doehrswww.apgea.army.mil/doehrsdr

Donahue, A. M., & Ohlin, D. W. (1993). Noise and the impairment of hearing. In D. P. Deeter, & J. C. Gaydos, *Occupational Health- The Soldier and the Industrial Base* (Vol. 2). Washington, D.C.: Office of the Surgeon General at TMM Publications Borden Institute.

Donahue, A. M., & Ohlin, D. W. (1993). Noise and the impairment of hearing. In D. P. Deeter, & J. C. Gaydos, *Occupational Health- The Soldier and the Industrial Base* (Vol. 2). Washington, D.C.: Office of the Surgeon General at TMM Publications Borden Institute.

Duda, O. R. (2011, February 25). *Psychoacoustics of spatial hearing.* Retrieved July 2016, from UC Davis electrical and computer engineering: http://interface.cipic.ucdavis.edu/sound/tutorial/psych.html#elev

Emanuel, D. C., Maroonroge, S., & Letowski, T. R. (2009). Auditory function. In C. E. Rash, M. B. Russo, T. R. Letowski, & E. T. Schmeisser, *Helmet-mounted displays: Sensation, perception and cognition issues* (pp. 279-306). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97-101). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (2012). Situation awareness. In G. Salvendy, *Handbook of human factors and ergonomics* (pp. 553-568). John Wiley & Sons.

Fligor, B. J. (2009). Risk for noise-induced hearing loss from use of portable media players: A summary of evidence through 2008. *Perspectives on Audiology, 5*(1), 10.

Gelfand, S. A. (1998). *Hearing: An introduction to psychological and physiological acoustics (3rd ed.).* New York: Informa Healthcare.

Giguère, C., Laroche, C., & Vaillancourt, V. (2013). Advanced hearing protection and auditory awareness in individuals with hearing loss. *Proceedings of Meetings on Acoustics. 19*, p. 040010. Montréal: Acoustical Society of America.

Giguère, C., Laroche, C., & Vaillancourt, V. (2013). Advanced hearing protection and auditory awareness in individuals with hearing loss. *Proceedings of Meetings on Acoustics, 19*, 1-6.

Grantham, D. W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *The Journal of the Acoustical Society of America, 79*(6), 1939-1949.

Groenwold, M. R., Tak, S., & Matterson, E. (2011). Severe hearing impairment among military veterans- United States, 2010. *Journal of the American Medical Association, 306*(11), 1192-1194.

Hajicek, J. J., Myrent, N., Li, Q., Barker, D., & Coyne, K. M. (2010). Protocols for improved understanding of situational awareness effects of head-borne PPE. doi:10.1109/. *2010 IEEE International Conference on Technologies for Homeland Security (HST).*, (p. doi:10.1109/ths.2010.5655076).

Hansen, J. T. (2014). Head and neck. In J. T. Hensen, *Netter's clinical anatomy* (p. 469). Saint Louis: Elsevier Health Sciences.

Hays, R. T., & Singer, M. J. (1989). Simulation Fidelity as an Organizing Concept. In *Simulation Fidelity in Training System Design* (pp. 23-46). New York: Springer-Verlag.

Helfer, T. M. (2011). Noise-induced hearing injuries, active component, U.S. Armed Forces 2007-2010. *Medical Surveillance Monthly Report, 18*, 7-10.

Helfer, T. M., Jordan, N. N., Lee, R. B., Pietrusiak, P., Cave, K. M., & Schairer, K. (2011). Noise-induced hearing injury and comorbidities among postdeployment U.S. Army Soldiers. *American Journal of Audiology, 20*(1), 33-41.

Henry, K. S., & Heinz, M. G. (2012). Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nature Neuroscience, 15*(10), 1362-1364.

Henshaw, H., & Ferguson, M. A. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence. *PLoS Biology, 8*(5), e62836.

Hétu, R., Getty, L., & Quoc, H. T. (1995). Impact of occupational hearing loss on the lives of workers. *Occupational Medicine, 10*(3), 495-513.

Hofman, P. M., Van Riswick, J. G., & Van Opstal, A. J. (1998). Relearning localization with new ears. *Nature Neuroscience, 1*(5), 417-421.

Humes, L. E., Allen, S. K., & Bess, F. H. (1980). Horizontal sound localization skills of unilaterally hearing-impaired children. *Audiology, 19*, 508-518.

Kacelnik, O., Nodal, F. R., Parsons, C. H., & King, A. J. (2006). Training-induced plasticity of auditory localization in adult mamals. *PLoS Biology, 4*(4), e71.

Kamm, C. A., Dirks, D. D., & Bell, T. S. (1985). Speech recognition and the articulation index for normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 7*(1), 281-288.

Kapralos, B., Jenkin, M. R., & Milios, E. (2008). Virtual audio systems. *Presence, 17*(6), 527-549.

Kendall, G. S. (1995, Winter). Directional hearing and stereo reporduction. *Computer Music Journal, 19*(4), 23-46.

Kiessling, J., Pichora-Fuller, M. K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., & Rosenhall, U. (2003). Candidature for and delivery of audiologist services: special needs of older people. *International Journal of Audiology, 42*(sup2), 92-101.

King, A. J., Hutchings, M. E., Moore, D. R., & Blakemore, C. (1988). Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. *Nature, 332*(6159), 73-76.

Klingseis, K. H. (2017). Hearing loss in infantry soldiers at Fort Benning. *Army Public Health Course.* Fort Dix: Army Public Health Command.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *Journal of Neuroscience, 29*(45), 14077-14085.

Lee, K., & Casali, J. G. (2017). Development of an auditory situation awareness test battery for advanced hearing protectors and TCAPS: Detection subset of DRILCOM (Detection-Recognition/Identification-Localization-Communication). *International Journal of Audiology, 56*, 22-33.

Lee, K., & Casali, J. G. (2019). Learning to localize with advanced hearing protectors and TCAPS: Importance and practicality of learning curves. *National Hearing Conservation Association.* Grapevine.

Letowski, T. R., & Letowski, S. T. (2012). *Auditory Spatial Perception: Auditory Localization (Report No. ARL-TR-6016).* Aberdeen Proving Ground: Army Research Laboratory.

Lorenzi, C., Gatehouse, S., & Lever, C. (1999). Sound localization in noise in hearing-impaired listeners. *Journal of the Acoustical Society of America, 105*(6), 3454-3463.

Maher, R. C. (2006, April 4). *Summary of gunshot acoustics.* Retrieved from Montana State University: R Maher Publications: http://www.montana.edu/rmaher/publications/maher_aac_0406.pdf

Maison, S. F., Usubuchi, H., & Liberman, M. C. (2013). Efferent feedback minimizes cochlear neuropathy from moderate noise exposure. *Journal of Neuroscience, 33*(13), 5542-5552.

Maroonroge, S., Emanuel, D. C., & Letowski, T. R. (2009). Basic Anatomy of the Hearing System. In *Helmet-Mounted Displays: Sensation, Perception and Cognition Issues.* Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Matterson, E. A., Bushnell, P. T., Themann, C. L., & Morata, T. C. (2016, April 22). Hearing impairment among noise-exposed workers- United States, 2003-2012. *Morbidity Mortality Weekly Report, 65*(15), pp. 389-394.

Matterson, E. A., Bushnell, P. T., Themann, C. L., & Morata, T. C. (2016, April 22). Hearing impairment among noise-exposed workers- United States, 2003-2012. *Morbidity Mortality Weekly Report*, pp. 389-394.

McMullen, K. A., & Wakefield, G. H. (2017). The effects of training on real-time localization of headphone-rendered, spatially processed sounds. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 1557-1561.

Mills, A. W. (1958). On minimum audible angle. *The Journal of the Acoustical Society of America, 30*(4), 237-246.

Mills, A. W. (1958). On the minimum audible angle. *The Journal of the Acoustical Society of America, 30*(4), 237-246.

Moore, B. C. (1997). Space perception. In B. Moore, *An Introduction to the Psychology of Hearing.* San Diego, CA: Academic Press.

Moore, B. C. (2014). Psychoacoustics. In T. Rossing, *Springer handbook of acoustics* (pp. 475-511). Berlin: Springer.

Muller, B. S., & Bovet, P. (1999). Role of pinnae and head movements in localizing pure tones. *Swiss Journal of Psychology, 58*(3), 170-179.

National Institute for Occupational Safety and Health [NIOSH]. (2012, August 6). *NIOSH, OSHA, and NHCA Establish Alliance on Workplace Hearing Loss Prevention.* Retrieved from https://www.cdc.gov/niosh/updates/upd-03-24-08.html

National Institute for Occupational Safety and Health [NIOSH]. (2019, January 18). Retrieved from Occupational Hearing Loss (OHL) Surveillance: https://www.cdc.gov/niosh/topics/ohl/default.html

National Institute for Occupational Safety and Health. (1998). *Occupational noise exposure: Revised criteria 1998.* Cincinnati: U.S. Department of Health and Human Services.

National Institute for Occupational Safety and Health. (2014). *Health hazard evaluation program: Measurement of exposure to impulsive noise at indoor and outdoor firing ranges during tactical exercises (Health Hazard Evaluation Report 2013-0124-3208).* Cincinnati: U.S. Department of Health and Human Services.

National Institute for Occupational Safety and Health. (2018, March 28). *About NIOSH*. Retrieved from https://www.cdc.gov/niosh/about/default.html

Nelson, I. N., Nelson, R. Y., Concha-Barrientos, C., & Fingerhut, M. (2005). The global burden of occupational noise-induced hearing loss. *American Journal of Industrial Medicine, 48*, 446-458.

Noble, W., Byrne, D., & Lepage, B. (1994). Effects on sound localization of configuration and type of hearing impairment. *The Journal of the Acoustical Society of America, 95*(2), 992-1005.

Occupational Safety and Health Administration [OSHA]. (1971). Occupational Noise Exposure. (29 CFR, Part 1910.95). Washington, D.C.: Office of Federal Register.

Peters, L. J., & Garinther, G. R. (1990). *The effects of speech intelligibility on crew performance in an M1A1 tank simulator.* Aberdeen Proving Ground: Human Engineering Laboratory.

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing.* San Diego, CA: Academic Press.

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS.* New York, NY: Routledge.

Pollack, I., & Rose, M. (1967). Effect of head movement on the localization of sounds in the equatorial plane. *Perception & Psychoacoustics, 2*(12), 591-596.

Pollatsek, A., & Rayner, K. (1998). Behavioral experimentation. In W. Betchel, & G. Graham, *A companion to cognitive science* (pp. 352-370). Malden: Blackwell.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research (4th ed.).* Upper Saddle River, NJ: Pearson Prentic Hall.

Price, G. R., Kalb, J. T., & Garinther, G. R. (1989). Toward a measure of auditory handicap in the Army. *Annals of Otology, Rhinology, and Laryngology, 98*, 42-52.

Quested, N. (Producer), & Junger, S. (Director). (2014). *Korengal* [Motion Picture]. United States.

Rakerd, B., & Harman, W. M. (1985). Localization of sound in rooms, III: Onset and duration effects. *Journal of the Acoustical Society of America, 80*(6), 1695-1706.

Rakerd, B., & Hartman, W. M. (1986). Localization of sound in rooms, III: Onset and duration effects. *Journal of the Acoustical Society of America, 80*(6), 1695-1706.

Rappaport, J., & Provencal, C. (2002). Neuro-otology for audiologists. In J. Katz, *Handbook of clinical audiology* (pp. 9-32). Philadelphia: Lippincott Williams & Wilkins.

Rash, C. E., Russo, M. R., Letowski, T. R., & Schmeisser, E. T. (Eds.). (2009). *Helmet-Mounted Displays: Sensation, Perception, and Cognition Issues.* Fort Rucker: U.S. Army Aeromedical Research Laboratory.

Robinette, M. B. (2012). Evaluation of an auditory fitness for duty test method for assessing actual and potential horizontal localization in the open-ear and when hearing protection enhancement device with electronic pass-through technology.

Sabin, A. T., Macpherson, E. A., & Middlebrooks, J. C. (2005). Human sound localization at near-threshold levels. *Hearing Research, 199*(1-2), 124-134.

Scharine, A. A., & Letowski, T. R. (2005). *Factors affecting auditory localization and situational awareness in the urban battlefield (Report No. ARL-TR-3474).* Aberdeen Proving Ground: Army Research Laboratory.

Scharine, A. A., Cave, K. M., & Letowski, T. R. (2009). Auditory perception and cognitive performance. In C. E. Rash, M. B. Russo, T. R. Letowski, & E. T. Schmeisser, *Helmet-mounted displays: Sensation, perception, and cognition issues* (pp. 391-490). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Schow, R. L., & Nerbonne, M. A. (2002). *Introduction to Audiologic Rehabilitation.* Boston, MA: Allyn and Bacon.

Shaw, E. A. (1974). The external ear. In *Handbook of sensory physiology* (Vol. 5). Berlin, Heidelberg, Germany: Springer.

Sheffield, B. M., Brungart, D. S., Tufts, J. B., & Ness, J. (2015). The relationship between hearing acuity and operational performance in dismounted combat. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 1346-1350.

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach, & A. Lesgold, *Adaptive technologies for training and education* (pp. 7-27). New York: Cambridge U Press.

Simpson, B. D., Bolia, R. S., McKinley, R. L., & Brungart, D. S. (2005). The impact of hearing protection on sound localization and orienting behavior. *Human Factors, 47*(1), 188-198.

Smith, A. T. (1991). A review of non-auditory effects of noise on health. *Work & Stress, 5*(1), 49-62.

Smith-Abouchacra, K. M. (1993). Detection and localization of a target signal in multiple-source environments. University Park, PA.

Soli, S. D. (2003). *Hearing and job performance.* Division of Behavioral and Social Sciences and Education of the National Research Council for the Committee on Disability Determination for Individuals with Hearing Impairment.

Stergar, M., Fackler, C., & Hamer, J. (2019). Correspondance with 3M Engineer. Indianapolis, IN.

Sweetow, R., & Palmer, C. V. (2005). Efficacy of individual auditory training in adults: A systematic review of the evidence. *Journal of the American Academy of Audiology, 16*(7), 494-504.

Tak, S., Davis, R. R., & Calvert, G. M. (2009). Exposure to hazardous workplace noise and use of hearing protection devices among US workers-NHANES, 1999-2004. *American Journal of Industrial Medicine, 52*(5), pp. 358-371.

Talcott, K. A., Casali, J. G., Keady, J. P., & Killion, M. C. (2012). Azimuthal auditory localization of gunshots in a realistic environment: Effects of open-ear versus hearing protection-enhancement devices (HPEDs), military vehicle noise, and hearing impairment. *International Journal of Audiology, 51*(Sup1), S20-30.

Talcott, K. A., Casali, J. G., Keady, J. P., & Killion, M. C. (2012). Azimuthal auditory localization of gunshots in a realistic field environment: effects of open-ear versus Hearing Protection-Enhancement Devices (HPEDS), military vehicle noise, and hearing impairment. *International Journal of Audiology, 51*(Suppl 1), S20-S30.

Tremblay, K. L. (2003). Central auditory plasticity: implications for auditory rehabilitation. *The Hearing Journal, 56*(1), 10-12.

Tufts, J. B., Vasil, K., & Briggs, S. (2009). Auditory fitness for duty: A review. *Journal of the American Academy of Audiology, 20*(9), 539-557.

U.S. Army. (2019). *Program Executive Office Soldier*. Retrieved from About Us: https://www.peosoldier.army.mil/aboutus/

U.S. Department of Defense. (2018). *2017 Demographics: Profile of the Military Community.* Alexandria, VA: Department of Defense (DoD), Office of the Deputy Assistant Secretary of Defense for Military Community and Family Policy.

U.S. Department of Veterans Affairs [VA]. (2016, August 12). Retrieved from VA Research on hearing loss: https://www.research.va.gov/topics/hearing.cfm

U.S. Department of Veterans Affairs [VA]. (2017, February 6). *Annual Benefits Report Fiscal Year 2016.* Retrieved from http://www.benefits.va.gov/REPORTS/abr/ABR-Compensation-FY16-0613017.pdf

U.S. Department of Veterans Affairs. (2018, August 6). *About VA*. Retrieved from History VA: https://www.va.gov/about_va/vahistory.asp

USMC. (2017). *Enemy Threat Weapons.* United States Marine Corps: The Basic School. Camp Barrett, VA: Marine Corps Training Command.

Vause, N. L., & Grantham, D. W. (1999). Effects of Earplugs and Protective Headgear on Auditory Localization Ability in the Horizontal Plane. *Human Factors, 41*(2), 282–294.

Viehweg, R., & Campbell, R. A. (1960). Localization difficulty in monaurally impaired listeners. *The Annals of Otology, Rhinology, and Laryngology, 69*(2), 622-634.

Virginia Tech. (2019). *Auditory Systems Laboratory*. Retrieved from https://www.ise.vt.edu/about/facilities/labs/auditory-systems.html

Vliegen, J., & Opstal, A. J. (2004). The influence of duration and level on human sound localization. *The Journal of the Acoustical Society of America, 115*(4), 1705-1713.

Walden, B. A., Montgomery, A. A., Schwartz, D. M., & Prosek, R. A. (1981). A comparison of the effects of hearing impairment and acoustic filtering on consonant recognition. *Journal of Speech and Hearing Research, 46*, 32-43.

Wanrooij, M. M., & Opstal, A. J. (2005). Relearning sound localization with a new ear. *Journal of Neuroscience, 25*(22), 5413-5424.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance.* Boston: Pearson.

Wightman, F. L., & Kistler, D. J. (2003). Sound localization. In W. A. Yost, A. N. Popper, & R. R. Fay, *Human psychophysics* (Vol. 3). New York, NY: Springer.

Wilson, R. H., & Cates, W. B. (2008). A comparison of two word-recognition tasks in multitalker babble: Speech Recognition in Noise Test (SPRINT) and Words-in- Noise Test (WIN). *Journal of the American Academy of Audiology, 19*(7), 548-556.

Wolfle, D. (1946). Military training and useful parts of learning theory. *Journal of Consulting Psychology, 10*(2), 73-75.

Wolfle, D. (1946). Military training and useful parts of learning theory. *Journal of Consulting Psychology, 10*(2), 73-75.

Wolfle, D. (1951). Training. In S. Stevens, *Handbook of Experimental Psychology* (pp. 1267-1286). New York, New York: Wiley.

Wright, B. A., & Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Sciences, 98*(21), 12307-12312.

Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions: Biological Sciences, 364*(1515), 301-311.

Yong, J. S., & Wang, D. (2015). Impact of noise on hearing in the military. *Military Medical Research, 2*(1), 1-6.

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica, 91*, 409-420.

Zimpfer, V., & Sarafian, D. (2014). Impact of hearing protection devices on sound localization performance. *Frontiers in Neuroscience, 8*(135), 1-10.

Appendix A. Email to Subject Matter Experts

Dear Sir/Ma'am,

Thank you for being willing to provide your input regarding what sounds you feel service members need to localize in a ground combat situation. I am seeking your input to ensure that I incorporate relevant sounds into an auditory localization training system in fulfillment of my dissertation, but ultimately for use by US service members.

Please provide sounds that you feel are important for a Service Member to hear in the space below. I am including an inventory currently available for the training apparatus, as you may find these helpful in generating your answer, but these are merely for guidance.

Sound clip name	
Car driving by	Children playing
Diesel truck idling	Arabic being spoken
Motorcycle	English being spoken
Military tank track noise	Dog growling
Heavy truck driving in and stopping	Bell in a tower
Truck Engine's compression "Jake Brake"	Bird singing
Helicopter passing by	Police car siren
Helicopter taking off	European emergency vehicle siren
Jet flying by	Fire truck siren and horn
Telephone ringing	AK-47 Rifle being cocked
Geiger counter	Bolt-action rifle being cocked
Vehicle's backup alarm	Jackhammer working
Footsteps in snow	M-16 Rifle single shots
Footsteps in leaves	AK-47 Rifle burst of shots
Footsteps in gravel	Handgun (Pistol) firing with a silencer
Car horn	Semi-automatic Pistol shots
Train horn	Incoming mortar shell
Railroad crossing bell	M-60 Machine Gun burst of shots

Please provide a list of at least five sounds you feel are important in rank order with 1 being the most important and 5 being less important

1.

2.

3.

4.

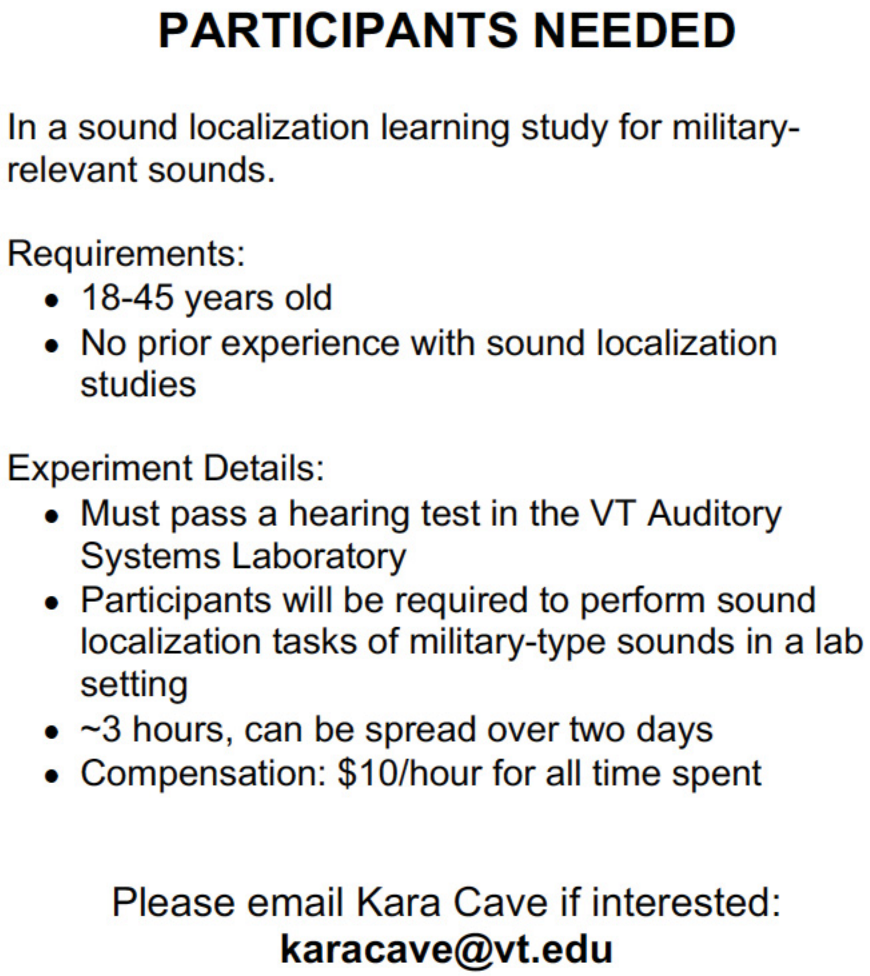
5.

Any other considerations in generating sound localization signals?

Thank you again for your time and consideration.

Respectfully,
Kara Cave
MAJ, US Army
Ph.D. Student | Department of Industrial & Systems Engineering Virginia Polytechnic Institute & State University
240-620-8069 | karacave@vt.edu

Appendix B. Participant Recruitment Flyer



PARTICIPANTS NEEDED

In a sound localization learning study for military-relevant sounds.

Requirements:

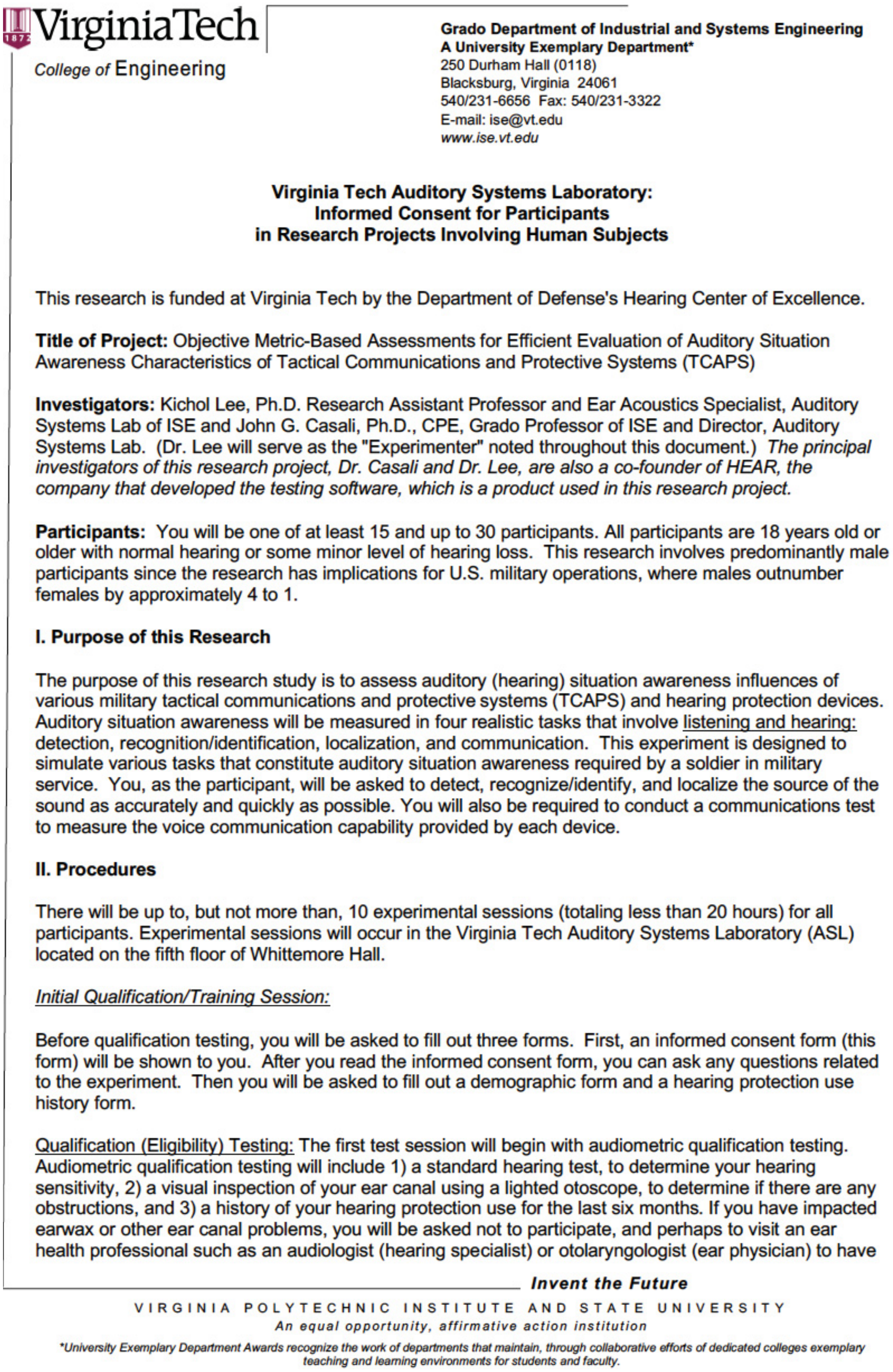
- 18-45 years old
- No prior experience with sound localization studies

Experiment Details:

- Must pass a hearing test in the VT Auditory Systems Laboratory
- Participants will be required to perform sound localization tasks of military-type sounds in a lab setting
- ~3 hours, can be spread over two days
- Compensation: $10/hour for all time spent

Please email Kara Cave if interested:
karacave@vt.edu

Appendix C. Phase I Informed Consent



VirginiaTech
College of Engineering

Grado Department of Industrial and Systems Engineering
A University Exemplary Department*
250 Durham Hall (0118)
Blacksburg, Virginia 24061
540/231-6656 Fax: 540/231-3322
E-mail: ise@vt.edu
www.ise.vt.edu

Virginia Tech Auditory Systems Laboratory: Informed Consent for Participants in Research Projects Involving Human Subjects

This research is funded at Virginia Tech by the Department of Defense's Hearing Center of Excellence.

Title of Project: Objective Metric-Based Assessments for Efficient Evaluation of Auditory Situation Awareness Characteristics of Tactical Communications and Protective Systems (TCAPS)

Investigators: Kichol Lee, Ph.D. Research Assistant Professor and Ear Acoustics Specialist, Auditory Systems Lab of ISE and John G. Casali, Ph.D., CPE, Grado Professor of ISE and Director, Auditory Systems Lab. (Dr. Lee will serve as the "Experimenter" noted throughout this document.) *The principal investigators of this research project, Dr. Casali and Dr. Lee, are also a co-founder of HEAR, the company that developed the testing software, which is a product used in this research project.*

Participants: You will be one of at least 15 and up to 30 participants. All participants are 18 years old or older with normal hearing or some minor level of hearing loss. This research involves predominantly male participants since the research has implications for U.S. military operations, where males outnumber females by approximately 4 to 1.

I. Purpose of this Research

The purpose of this research study is to assess auditory (hearing) situation awareness influences of various military tactical communications and protective systems (TCAPS) and hearing protection devices. Auditory situation awareness will be measured in four realistic tasks that involve listening and hearing: detection, recognition/identification, localization, and communication. This experiment is designed to simulate various tasks that constitute auditory situation awareness required by a soldier in military service. You, as the participant, will be asked to detect, recognize/identify, and localize the source of the sound as accurately and quickly as possible. You will also be required to conduct a communications test to measure the voice communication capability provided by each device.

II. Procedures

There will be up to, but not more than, 10 experimental sessions (totaling less than 20 hours) for all participants. Experimental sessions will occur in the Virginia Tech Auditory Systems Laboratory (ASL) located on the fifth floor of Whittemore Hall.

Initial Qualification/Training Session:

Before qualification testing, you will be asked to fill out three forms. First, an informed consent form (this form) will be shown to you. After you read the informed consent form, you can ask any questions related to the experiment. Then you will be asked to fill out a demographic form and a hearing protection use history form.

Qualification (Eligibility) Testing: The first test session will begin with audiometric qualification testing. Audiometric qualification testing will include 1) a standard hearing test, to determine your hearing sensitivity, 2) a visual inspection of your ear canal using a lighted otoscope, to determine if there are any obstructions, and 3) a history of your hearing protection use for the last six months. If you have impacted earwax or other ear canal problems, you will be asked not to participate, and perhaps to visit an ear health professional such as an audiologist (hearing specialist) or otolaryngologist (ear physician) to have

Invent the Future
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
An equal opportunity, affirmative action institution

**University Exemplary Department Awards recognize the work of departments that maintain, through collaborative efforts of dedicated colleges exemplary teaching and learning environments for students and faculty.*

your ear canals clinically checked and cleaned of earwax if needed, and perhaps return for a second screening for this experiment if you so desire.

Upon successful completion of all eligibility requirements, you will:
1. Receive training/orientation with the actual hearing protection devices (HPDs) (including HPEDs-hearing protection enhancement devices and TCAPS-tactical communications and protection systems) that will be used in the study, as well as training on the experimental procedures.
2. Undergo training sessions for the 4 situation awareness sub-tests: detection, recognition/identification, localization, and communication. You will be trained without any HPDs (i.e., with open ears) for the 4 tests. Once you are familiarized with each test, you will also practice using all HPDs.

Throughout the training sessions and the four sub-tests described below, your responses will always entail only a mouse click entry or an answer by your voice (speaking) At the conclusion of each sub-test, you will rate and rank your experiences and impressions with the various HPDs via rating and ranking scales, explained below.

The Experimenter will fit all HPDs on you to reduce any variability that might be due to individual differences in fitting and wearing HPDs. The purpose of this study is to determine each HPD's capability in maintaining auditory situation awareness, and *not* to assess the fit of HPDs.

On each of the 4 sub-tests below, where sound signals or noises are presented to you, these will always be kept to sound exposure levels that are less than half that allowed by the U.S. OSHA (1983) workplace health and safety laws for a U.S. worker in a single work day (8-hour shift), in order to minimize any risk of noise-induced hearing loss. In this experiment, the signal and noise levels are known to be at decibel and time duration values that do not pose a risk to your hearing, even when you are listening with open ears. Specific noises will be discussed below where they are to be applied under each sub-test.

Sub-test 1: Detection test

For the Detection test, you will be asked to listen to a test signal and press and hold a response switch whenever you can hear the test signal. Pressing and holding the button will lower the test signal volume and you should let go of the switch as soon as you cannot hear the sound. After several repetitions, the computer program will cause the test to advance to the next signal. You will repeat the test with open ears and several HPDs, and also with the signals coming from loudspeakers that are located at front, back, left, and right of you.

The Detection test will be similar to a standard Real Ear Attenuation at Threshold test (REAT) (already approved by Virginia Tech's Institutional Review Board (IRB) for Human Subjects as VT IRB-05-701) with modifications as follows: 1) the test signal will only be played via 1 loudspeaker located at front, back, left, and right of you, and 2) the sound signal will consist of 1/3-octave bands with center frequencies of 125Hz, 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz, 8000Hz (same as the standard REAT test signals), plus two military-relevant signals that are low-intensity gunshots and gun-cocking sounds. Because this will be a detection test at or very near auditory threshold (just barely heard), the sound levels of the gunshots and gun-cocking presented to the you will be at very quiet levels, and you will have to be extremely attentive and listen carefully to hear them. Your response (pressing a response switch and holding it pressed at any time when you can hear the signal) will be tracked by a computer program and recorded for later analysis. The intensity level of the background masking noise (pink noise, which includes all pitches of sound) will always be played at 85 dBA or below. (As mentioned above, note that OSHA (1983) allows U.S. workers to be exposed to an average of 90 dBA for a full 8-hour workday, and throughout the subtests described herein, all noise exposures will be less than half of that amount per day of this experiment.)

Sub-test 2: Recognition/Identification (Re/Id) test

In this test, you will be asked to listen and identify a particular (target) sound clip from the 3 sound clips presented in rapid sequence to you, and to identify that target sound as accurately and quickly as possible.

The Re/Id test will consist of listening to series of sound signals presented in groups of three and identifying the requested target sound. The sound signals will be played via 1 loudspeaker located in front and to the right of you. You will be asked to click one of the four buttons that identifies the target sound on your monitor as accurately and quickly as possible. You will also determine, through the button on your monitor, when the next set of sounds is to be played. The background masking noise (pink noise) level will always be played at 85 dBA or below, as described previously, which is considered a safe level. The test sound signals will always be played at 60-80 dBA throughtout the test. Some examples of test sound signals are: speech, truck engine idling, gun-cocking, footstep, rifle firing, heavy truck breaking.

Sub-test 3: Localization test

In this test, you will be asked to listen to test sound signals presented to you while sitting in the center of a circle of loudspeakers and to identify where the test signal is coming from, in other words, which loudspeaker is playing that signal.

The exact location and number of loudspeakers will be hidden from your view. Possible loudspeaker locations (targets) will be shown in your monitor and you will identify (click) the loudspeaker icon which corresponds to where you think the test signal came from. You will listen for a test sound signal and identify the absolute location in either 360˚ azimuth (horizontally around you) or in frontal elevation (level or upward in front of you) as accurately and quickly as possible. You will also signal, through the button on your monitor, when the next test sound signal is to be played.

Your response will be recorded by a computer program for later analysis. The background masking noise (pink noise) level will always be played at 85 dBA or below, as described previously. The sound signal for the localization test will be a sound that includes both low and high frequency ranges that are well-within the pitches of sound that can be heard, and localized, by the human ear. The test sound signal will be played at a level slightly higher than the background masking sound (not more than 10 dB higher) in order to be audible. The duration of each test signal will be very short, that is, 4 seconds or less, so the total noise exposure will be well below that allowed by OSHA for U.S. industrial workers, as described previously.

Sub-test 4: Communication test

In this test, you will be asked to listen to sentences of prerecorded speech and verbally repeat the sentences exactly as you heard them. The Experimenter will notify you before each prerecorded sentence is played.

The Communication test will be a modified version of the "QuickSIN" test, which is a standardized test used for testing people's ability to understand "Speech in Noise" (SIN). The modified QuickSIN test will be presented through loudspeakers in the test room instead of through a headphone which the original test was designed to use. Your responses will be entered into a computer by the Experimenter for later analysis. Because the QuickSIN test includes masking sound as part of test presentation, there will not be an additional background masking sound as will be used in the other 3 tests.

III. Risks

Experimental purpose: This experiment is designed to measure levels of auditory situation awareness afforded by each military hearing protection device (HPD) when a listener is required to detect, recognize/identify, and localize a sound source as well as to hear and repeat sentences spoken to him/her. The sound sources (test signals) and communication sentences are not loud enough to be hazardous, and there is no known bodily danger associated with this study. Furthermore, all of the signals are recorded sounds, and thus there are no "real" sources of the signals in the experimental room, such as gunshots. However, if you feel that this experiment would make you uncomfortable or cause you emotional distress during or after the experiment, you may freely decide not to participate.

Hearing Protection Devices (HPDs): HPDs are designed to have a tight fit and you may experience some minor discomfort while wearing them. If you experience more than minor discomfort, tell the Experimenter immediately and he will assist you in adjusting or removing the hearing protector or will provide a different size of eartip. Also, electronic protectors may emit a squealing or whistling noise if not properly sealed; while not dangerous it can be annoying, if this occurs please notify the Experimenter so that he can adjust the seal of the device. This can also occur if you accidentally place your finger over the microphone, so do not reach up and touch the device. In all cases, the Experimenter will fit the devices in or over your ears, and adjust the gain-amplification setting to help avoid the possibility of a squealing or whistling noise.

Other Risks: If you feel tired, or become thirsty during the tests, please inform the Experimenter and you will be allowed to rest and have something to drink.

Your Responsibilities: If you consent to participate, and later, if, in the unlikely event that you seek medical or counseling services that you feel are a result of your participation, you will be responsible for the costs of such services.

IV. Benefits

Your participation in this experiment will provide information on the level of auditory situation awareness afforded by advanced hearing protection devices. This information will *primarily* be used to help the U.S. military to determine what hearing protection devices that an individual should use for situations in which they need to hear and communicate. This information may also be of use for selection of protectors for certain law enforcement, industrial, construction, or hunting applications. No promise or guarantee of benefits has been made to encourage you to participate; however, you will receive the monetary compensation that is covered below in Section VI.

V. Extent of Anonymity and Confidentiality

Your identity will be kept confidential. If you choose to participate in the experiment, you will be identified by only a participant number. This number will be used in data collection and analysis. At no time will the researchers release your identity to anyone other than individuals working on the project without your written consent. This Consent Form as well as the raw data from the experiment (for example, your detection, identification, localization, and communications responses, and your experience ratings) will be solely in the possession of the investigators. Data will be analyzed on the investigators' Virginia Tech computers, which are password-protected.

It is possible that the Virginia Tech Human Participants Institutional Review Board (IRB), or a U.S. Military IRB, may review this study's collected data for auditing purposes. An IRB conducts the oversight of the protection of human subjects involved in research.

VI. Compensation

Participants will be monetarily compensated for participation in the study at the rate of $10 per hour during test/training sessions. For any fraction of time less than 1 hour, you will be paid for the closest ¼-hour period, rounded up (in your favor). You will also receive a $20 bonus for successful completion of all experimental sessions. You will be paid at the conclusion of each experimental session. Military participants (active duty, national guard, or reserve) are not eligible for study-related payment unless they are on leave status.

VII. Freedom to Withdraw

You are free to withdraw from this study at any time without penalty. If you choose to withdraw from the study, you will be compensated for the portion of your time spent in the study. There may be circumstances under which the investigator may determine that you should not continue as a subject, and while this is a rare occurrence, you must abide by that decision if it occurs. Again, you would be paid for the time that you have spent in the experiment under any circumstances.

VIII. Subject's Responsibilities

I voluntarily agree to participate in this study.
I have the following responsibilities:

- To listen for, detect, identify, and localize the signals, and to listen for and repeat the communications sentences in the experiment to the best of my ability, and to provide accurate ratings of my impressions about the listening conditions.
- To inform the Experimenter if a hearing protector, or any other aspects of the test condition, becomes overly uncomfortable.
- To inform the Experimenter if I become tired or thirsty and wish to rest.
- To avoid biasing other potential participants, to not discuss the study with anyone until 6 months after the day of my participation.

IX. Participant's Permission

I have read the Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent:

______________________ ________________________ _______
Participant's printed name Participant's signature Date

Age________

******Participant's Tear-Off Portion—Participant to Keep This******

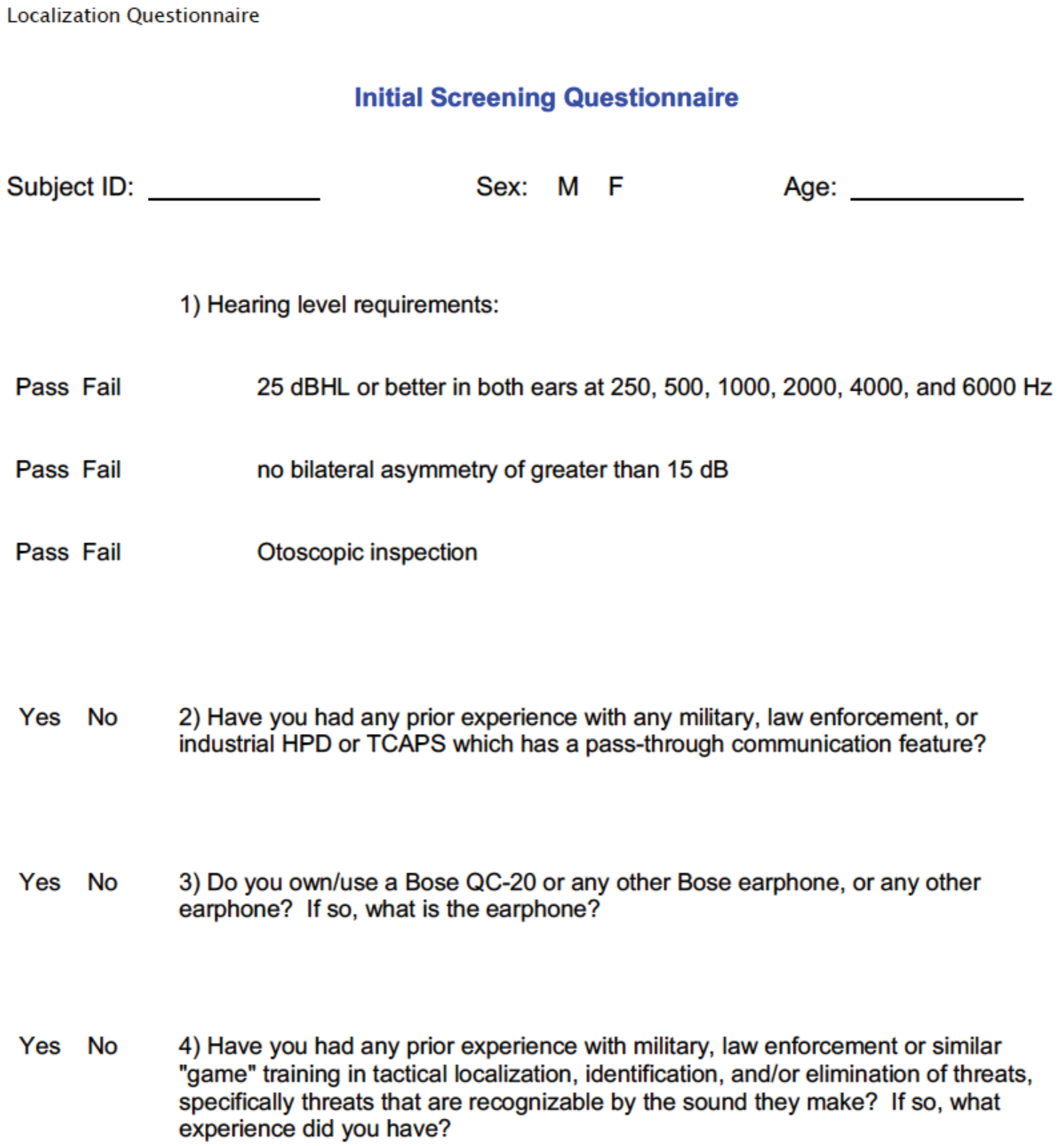
Contact information for investigators:

John G. Casali, Ph.D. (Principal Investigator)
(540) 231-5073 email: jcasali@exchange.vt.edu

Kichol Lee, Ph.D.
(540) 231-3294 email: kichol@exchange.vt.edu

Should you have any questions or concerns about the study's conduct or your rights as a research participant, or need to report a research-related injury or event, you may contact the Virginia Tech Institutional Review Board at irb@vt.edu or (540) 231-3732.

Appendix D. Demographic Form



Localization Questionnaire

Initial Screening Questionnaire

Subject ID: ____________ Sex: M F Age: ____________

1) Hearing level requirements:

Pass Fail 25 dBHL or better in both ears at 250, 500, 1000, 2000, 4000, and 6000 Hz

Pass Fail no bilateral asymmetry of greater than 15 dB

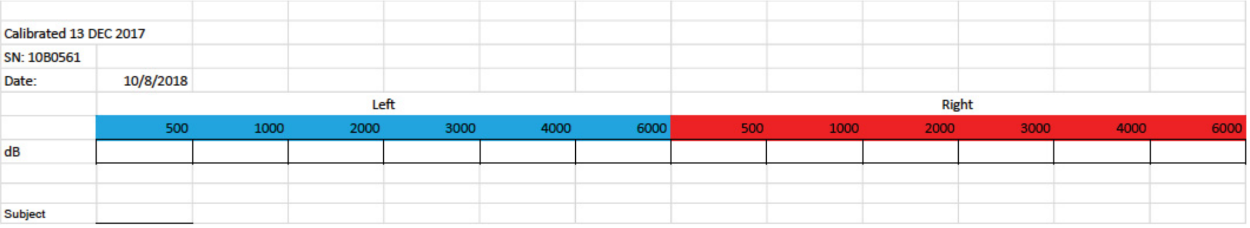
Pass Fail Otoscopic inspection

Yes No 2) Have you had any prior experience with any military, law enforcement, or industrial HPD or TCAPS which has a pass-through communication feature?

Yes No 3) Do you own/use a Bose QC-20 or any other Bose earphone, or any other earphone? If so, what is the earphone?

Yes No 4) Have you had any prior experience with military, law enforcement or similar "game" training in tactical localization, identification, and/or elimination of threats, specifically threats that are recognizable by the sound they make? If so, what experience did you have?

Appendix E. Participant Audiogram



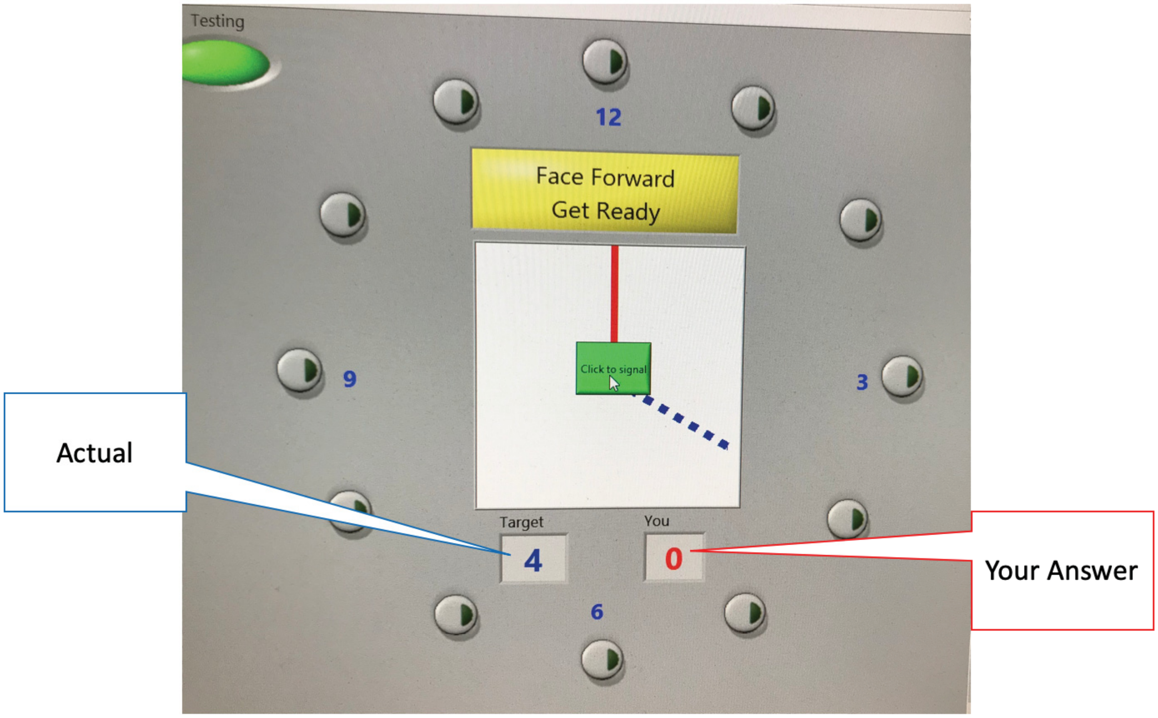
Calibrated 13 DEC 2017												
SN: 10B0561												
Date:	10/8/2018											
	Left						Right					
	500	1000	2000	3000	4000	6000	500	1000	2000	3000	4000	6000
dB												
Subject												

Appendix F. Pilot 1 Instructions

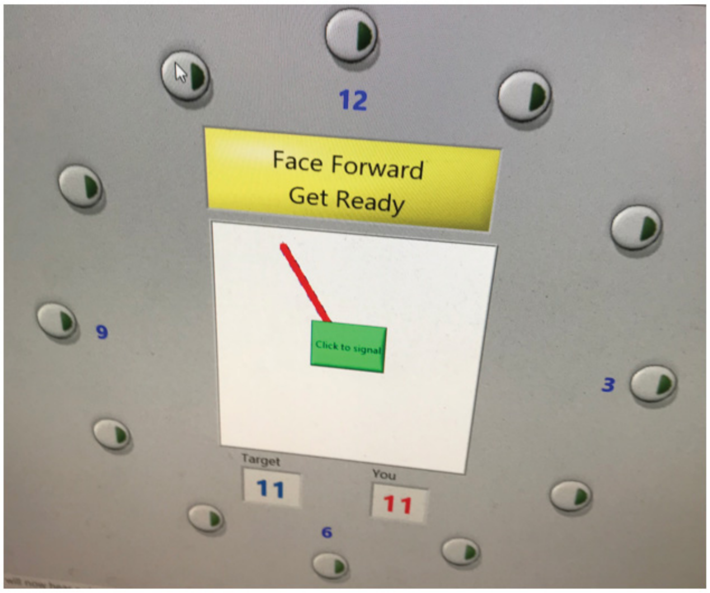
This purpose of the study is to better understand learning strategies used in determining sound location. You are going to hear an unusual sound coming from one of 12 different speakers arranged similar to a clock face. Your basic task is to use a mouse and a screen display to indicate where you heard the sound. You will undergo 8 blocks. Each block will be comprised of a 3 mini blocks. In the first mini block, you will know where the sound will be coming from, you'll respond, and then you'll tell the computer when you are ready to hear the next sound. In the next block you won't know where the sound is coming from, the display will tell you what your result was and what the answer was. In the last mini block, you'll hear the sound, you will respond, but you won't know your score. In the last session, you will hear sounds you haven't heard before. Some of the sounds will be weapon-related. All of the signals have been tested today, none are at the level that could damage your hearing. You can take a break or withdraw at any time.

When you are ready, select the 'click signal.' You'll hear a signal coming from 12, 3, 6 and 9 o'clock, the red line will indicate your response, the blue dashed line is the actual answer. Your task throughout the experiment is to respond as **quickly** and **accurately** as possible. Try to use the feedback given on the display to hone your responses. [Participant shown displays below and re-instruct just before 9 o'clock instruct to answer incorrectly].

The **red solid line** will indicate **your** answer and the **blue dashed line** is the **actual answer.**



When your response is correct, the two lines will be on top of each other.



Sequential Instructions:

“Now the first mini-block. The sound will start at 12 o’clock and will go all the way around clockwise” *how many sequential presentations they hear will depend on condition*
-Then 9 o’clock counterclockwise
-Then 3 o’clock clockwise
-Then 6 o’clock counterclockwise

“When you have finished, the bright green button under “test” will go from lime green to dark green, at that time, let me take control of the mouse.”

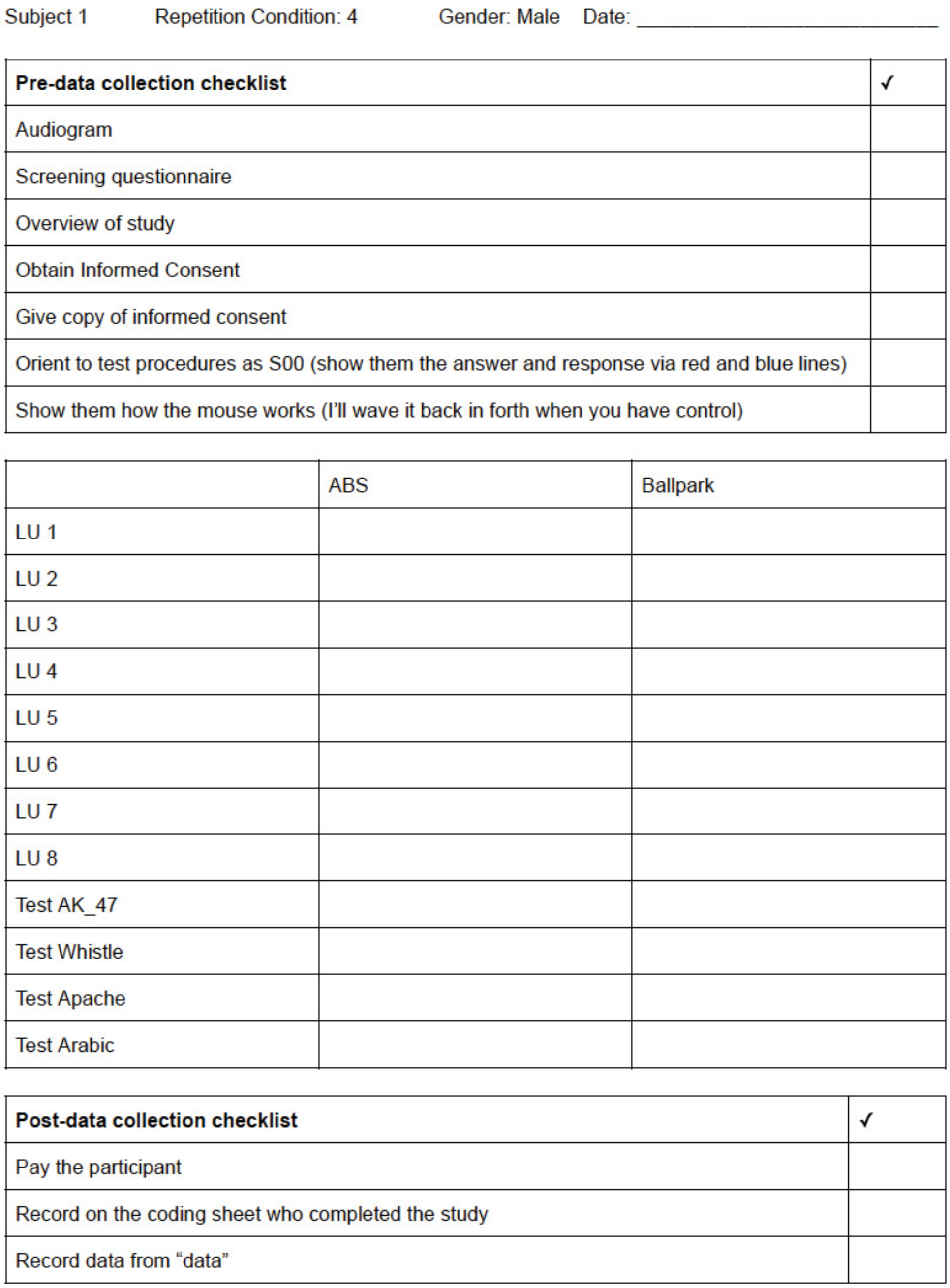
Random Instructions:

“Now you won’t know where the sound will be randomly presented, meaning it won’t be heard in a CW or CCW manner, just respond as quickly and accurately as possible.”

Test Instructions:

“For this next mini-block, you won’t know where the sound is coming from or if you answered correctly, so it’s really important to answer as quickly and accurately as possible.”

Appendix G. Example of Experimenter Checklist for Pilots 1 and 2



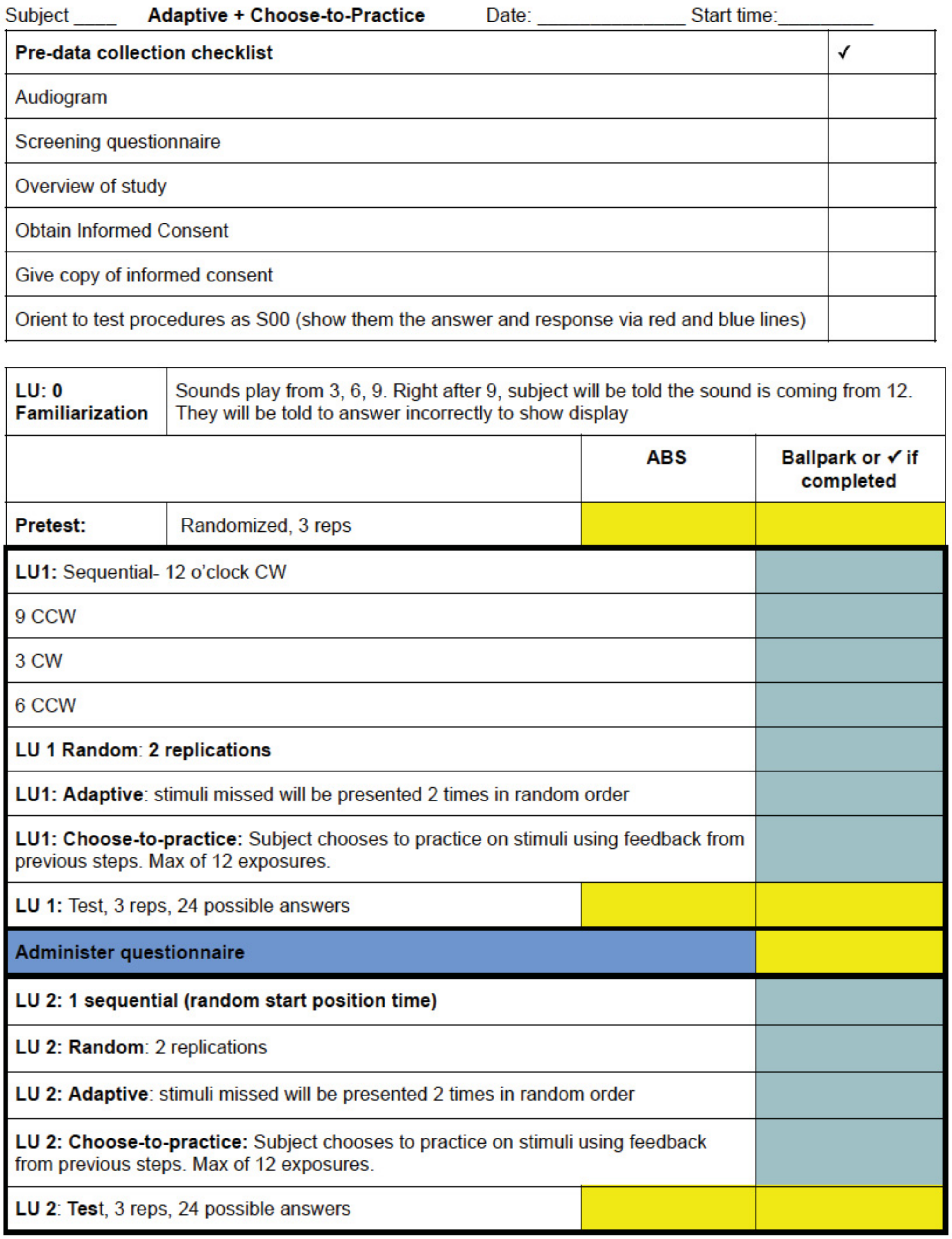
Subject 1 Repetition Condition: 4 Gender: Male Date: ___________________________

Pre-data collection checklist	✓
Audiogram	
Screening questionnaire	
Overview of study	
Obtain Informed Consent	
Give copy of informed consent	
Orient to test procedures as S00 (show them the answer and response via red and blue lines)	
Show them how the mouse works (I'll wave it back in forth when you have control)	

	ABS	Ballpark
LU 1		
LU 2		
LU 3		
LU 4		
LU 5		
LU 6		
LU 7		
LU 8		
Test AK_47		
Test Whistle		
Test Apache		
Test Arabic		

Post-data collection checklist	✓
Pay the participant	
Record on the coding sheet who completed the study	
Record data from "data"	

Appendix H. Example of Experimenter Checklist for the Main Experiment



Subject ____ **Adaptive + Choose-to-Practice** Date: ______________ Start time:_________

Pre-data collection checklist	✓
Audiogram	
Screening questionnaire	
Overview of study	
Obtain Informed Consent	
Give copy of informed consent	
Orient to test procedures as S00 (show them the answer and response via red and blue lines)	

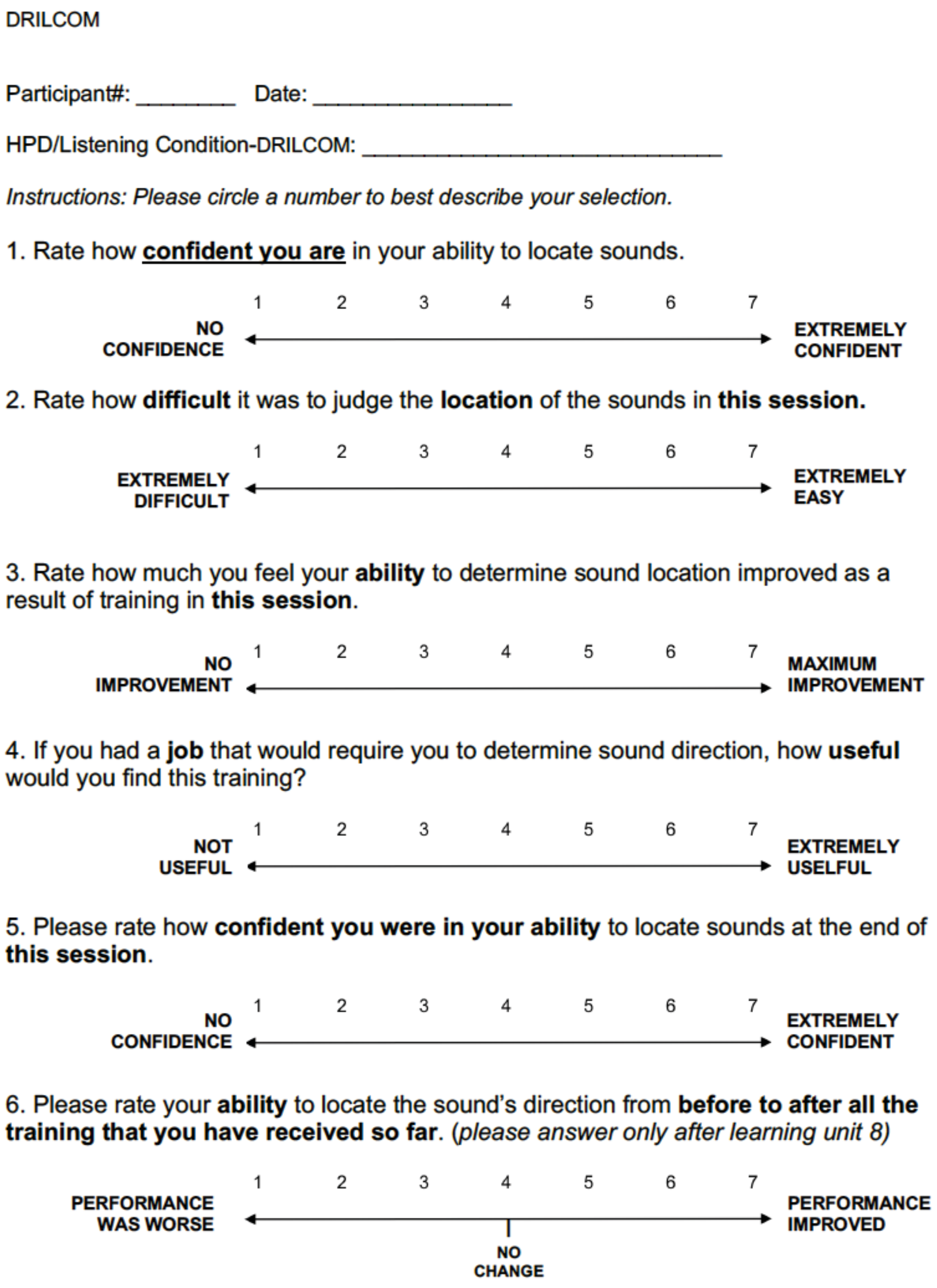
LU: 0 Familiarization	Sounds play from 3, 6, 9. Right after 9, subject will be told the sound is coming from 12. They will be told to answer incorrectly to show display		
		ABS	**Ballpark or ✓ if completed**
Pretest:	Randomized, 3 reps		
LU1: Sequential- 12 o'clock CW			
9 CCW			
3 CW			
6 CCW			
LU 1 Random: **2 replications**			
LU1: Adaptive: stimuli missed will be presented 2 times in random order			
LU1: Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.			
LU 1: Test, 3 reps, 24 possible answers			
Administer questionnaire			
LU 2: 1 sequential (random start position time)			
LU 2: Random: 2 replications			
LU 2: Adaptive: stimuli missed will be presented 2 times in random order			
LU 2: Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.			
LU 2: **Test**, 3 reps, 24 possible answers			

LU 3: 1 sequential (random start position time)		
LU 3: Random: 2 replications		
LU 3: Adaptive: stimuli missed will be presented 2 times in random order		
LU 3 : Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.		
LU 3: **Tes**t, 3 reps, 24 possible answers		
LU 4: 1 sequential (random start position time)		
LU 4: Random: 2 replications		
LU 4: Adaptive: stimuli missed will be presented 2 times in random order		
LU 4: Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.		
LU 4: **Tes**t, 3 reps, 24 possible answers		
LU 5: 1 sequential (random start position time)		
LU 5: Random: 2 replications		
LU 5 : Adaptive: stimuli missed will be presented 2 times in random order		
LU 5: Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.		
LU 5: **Tes**t, 3 reps, 24 possible answers		
LU 6: 1 sequential (random start position time)		
LU 6: Random: 2 replications		
LU 6: Adaptive: stimuli missed will be presented 2 times in random order		
LU 6: Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.		
LU 6: **Tes**t, 3 reps, 24 possible answers		
LU 7: 1 sequential (random start position time)		
LU 7: Random: 2 replications		
LU 7: Adaptive: stimuli missed will be presented 2 times in random order		
LU 7 : Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.		

LU 7: **Test**, 3 reps, 24 possible answers		
LU 8: 1 sequential (random start position time)		
LU 8: Random: 2 replications		
LU 8: Adaptive: stimuli missed will be presented 2 times in random order		
LU 8: Choose-to-practice: Subject chooses to practice on stimuli using feedback from previous steps. Max of 12 exposures.		
LU 8: **Test**, 3 reps, 24 possible answers		
Administer questionnaire		

Pay the participant	
Record on the coding sheet who completed the study	
Record data from “data”	
Ending time	

Appendix I. Questionnaires for the Main Experiment



DRILCOM

Participant#: ________ Date: ________________

HPD/Listening Condition-DRILCOM: ______________________________

Instructions: Please circle a number to best describe your selection.

1. Rate how **confident you are** in your ability to locate sounds.

NO CONFIDENCE ← 1 2 3 4 5 6 7 → **EXTREMELY CONFIDENT**

2. Rate how **difficult** it was to judge the **location** of the sounds in **this session.**

EXTREMELY DIFFICULT ← 1 2 3 4 5 6 7 → **EXTREMELY EASY**

3. Rate how much you feel your **ability** to determine sound location improved as a result of training in **this session**.

NO IMPROVEMENT ← 1 2 3 4 5 6 7 → **MAXIMUM IMPROVEMENT**

4. If you had a **job** that would require you to determine sound direction, how **useful** would you find this training?

NOT USEFUL ← 1 2 3 4 5 6 7 → **EXTREMELY USELFUL**

5. Please rate how **confident you were in your ability** to locate sounds at the end of **this session**.

NO CONFIDENCE ← 1 2 3 4 5 6 7 → **EXTREMELY CONFIDENT**

6. Please rate your **ability** to locate the sound's direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

PERFORMANCE WAS WORSE ← 1 2 3 4 5 6 7 → **PERFORMANCE IMPROVED**

NO CHANGE

7. Please rate your **reaction time** in determining sound direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

	1	2	3	4	5	6	7	
SLOWER	←						→	**FASTER**
				NO CHANGE				

8. Please rate the extent that the amount of **time** allotted for training was appropriate for the amount of improvement?

	1	2	3	4	5	6	7	
NEEDED MORE TIME	←						→	**TRAINING TOOK TOO MUCH TIME**
				TRAINING TOOK THE RIGHT AMOUNT OF TIME				

On each of the following aspects of your training, rate how beneficial you felt the particular part of training was in learning where sounds were located:

9. The display revealing where the sound would be located and then the sound was played.

	1	2	3	4	5	6	7	
NO BENEFIT	←						→	**MAXIMUM BENEFIT**

10. Sounds played randomly, you answered, and then the correct location was revealed.

	1	2	3	4	5	6	7	
NO BENEFIT	←						→	**MAXIMUM BENEFIT**

11. Test phase where the sound was played, you provided a response and no feedback was given.

	1	2	3	4	5	6	7	
NO BENEFIT	←						→	**MAXIMUM BENEFIT**

12. At any point(s) in the training did you feel yourself becoming disinterested or disengaged? You can check one or more options below or describe this point(s) in time in the space provided.

____a) The display revealed where the sound would be located and then the sound was played.

____b) Sounds played randomly, you answered, and then the correct location was revealed.

____c) Test phase where the sound was played, you provided a response and no feedback was given.

____

____

____

Choose-to-practice

Participant#: ________ Date: ________________
HPD/Listening Condition- Choose-to-Practice: __________________

Instructions: Please circle a number to best describe your selection.

1. Rate how **confident you are** in your ability to locate sounds.

NO CONFIDENCE	1	2	3	4	5	6	7	**EXTREMELY CONFIDENT**

2. Rate how **difficult** it was to judge the **location** of the sounds in **this session.**

EXTREMELY DIFFICULT	1	2	3	4	5	6	7	**EXTREMELY EASY**

3. Rate how much you feel your **ability** to determine sound location improved as a result of training in **this session**.

NO IMPROVEMENT	1	2	3	4	5	6	7	**MAXIMUM IMPROVEMENT**

4. If you had a **job** that would require you to determine sound direction, how **useful** would you find this training?

NOT USEFUL	1	2	3	4	5	6	7	**EXTREMELY USELFUL**

5. Please rate how **confident you were in your ability** to locate sounds at the end of **this session**.

NO CONFIDENCE	1	2	3	4	5	6	7	**EXTREMELY CONFIDENT**

6. Please rate your **ability** to locate the sound's direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

PERFORMANCE WAS WORSE	1	2	3	4	5	6	7	**PERFORMANCE IMPROVED**
				NO CHANGE				

7. Please rate your **reaction time** in determining sound direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

	1	2	3	4	5	6	7	
SLOWER								**FASTER**
				NO CHANGE				

8. Please rate the extent that the amount of **time** allotted for training was appropriate for the amount of improvement?

	1	2	3	4	5	6	7	
NEEDED MORE TIME								**TRAINING TOOK TOO MUCH TIME**
				TRAINING TOOK THE RIGHT AMOUNT OF TIME				

On each of the following aspects of your training, rate how beneficial you felt the particular part of training was in learning where sounds were located:

9. The display revealing where the sound would be located and then the sound was played.

	1	2	3	4	5	6	7	
NO BENEFIT								**MAXIMUM BENEFIT**

10. Sounds played randomly, you answered, and then the correct location was revealed.

	1	2	3	4	5	6	7	
NO BENEFIT								**MAXIMUM BENEFIT**

11. Choose-to-practice at your own pace.

	1	2	3	4	5	6	7	
NO BENEFIT								**MAXIMUM BENEFIT**

Choose-to-Practice

12. Test phase where the sound was played, you provided a response and no feedback was given.

1 2 3 4 5 6 7

NO BENEFIT ←——————————→ **MAXIMUM BENEFIT**

13. At any point in the training did you feel yourself becoming disinterested or disengaged? You can circle one option below or describe this time in the space provided.

____a) The display revealed where the sound would be located and then the sound was played.

____b) Sounds played randomly, you answered, and then the correct location was revealed.

____c) Choose-to-practice at your own pace.

____d) Test phase where the sound was played, you provided a response and no feedback was given.

____

____

____

Adaptive

Participant#: ________ Date: ________________

HPD/Listening Condition-Adaptive: __________________

Instructions: Please circle a number to best describe your selection.

1. Rate how **<u>confident you are</u>** in your ability to locate sounds.

NO CONFIDENCE ← 1 2 3 4 5 6 7 → **EXTREMELY CONFIDENT**

2. Rate how **difficult** it was to judge the **location** of the sounds in **this session.**

EXTREMELY DIFFICULT ← 1 2 3 4 5 6 7 → **EXTREMELY EASY**

3. Rate how much you feel your **ability** to determine sound location improved as a result of training in **this session**.

NO IMPROVEMENT ← 1 2 3 4 5 6 7 → **MAXIMUM IMPROVEMENT**

4. If you had a **job** that would require you to determine sound direction, how **useful** would you find this training?

NOT USEFUL ← 1 2 3 4 5 6 7 → **EXTREMELY USELFUL**

5. Please rate how **confident you were in your ability** to locate sounds at the end of **this session**.

NO CONFIDENCE ← 1 2 3 4 5 6 7 → **EXTREMELY CONFIDENT**

6. Please rate your **ability** to locate the sound's direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

PERFORMANCE WAS WORSE ← 1 2 3 4 5 6 7 → **PERFORMANCE IMPROVED**

NO CHANGE

7. Please rate your **reaction time** in determining sound direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

SLOWER 1 2 3 4 5 6 7 **FASTER**

NO CHANGE

8. Please rate the extent that the amount of **time** allotted for training was appropriate for the amount of improvement?

NEEDED MORE TIME 1 2 3 4 5 6 7 **TRAINING TOOK TOO MUCH TIME**

TRAINING TOOK THE RIGHT AMOUNT OF TIME

On each of the following aspects of your training, rate how beneficial you felt the particular part of training was in learning where sounds were located:

9. The display revealing where the sound would be located and then the sound was played.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

10. Sounds were played randomly, you answered, and then the correct location was revealed.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

11. The items that were answered incorrectly were presented three additional times.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

12. Sound was played, you provided a response and no feedback was given.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

Adaptive

13. At any point(s) in the training did you feel yourself becoming disinterested or disengaged? You can check one or more options below or describe this point(s) in time in the space provided.

___1) Test phase where the sound was played, you provided a response and no feedback was given.

___2) Sounds played randomly, you answered, and then the correct location was revealed.

___3) The items that were answered incorrectly were presented three additional times.

___4) Test phase where the sound was played, you provided a response and no feedback was given.

____

____

____

Adaptive + Choose-to-Practice

Participant#: ________ Date: ________________

HPD/Listening Condition- Adaptive + Choose-to-Practice Condition: __________________

Instructions: Please circle a number to best describe your selection.

1. Rate how **<u>confident you are</u>** in your ability to locate sounds.

1 2 3 4 5 6 7

NO CONFIDENCE ← → **EXTREMELY CONFIDENT**

2. Rate how **difficult** it was to judge the **location** of the sounds in **this session.**

1 2 3 4 5 6 7

EXTREMELY DIFFICULT ← → **EXTREMELY EASY**

3. Rate how much you feel your **ability** to determine sound location improved as a result of training in **this session**.

1 2 3 4 5 6 7

NO IMPROVEMENT ← → **MAXIMUM IMPROVEMENT**

4. If you had a **job** that would require you to determine sound direction, how **useful** would you find this training?

1 2 3 4 5 6 7

NOT USEFUL ← → **EXTREMELY USELFUL**

5. Please rate how **confident you were in your ability** to locate sounds at the end of **this session**.

1 2 3 4 5 6 7

NO CONFIDENCE ← → **EXTREMELY CONFIDENT**

6. Please rate your **ability** to locate the sound's direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

1 2 3 4 5 6 7

PERFORMANCE WAS WORSE ← → **PERFORMANCE IMPROVED**

NO CHANGE

7. Please rate your **reaction time** in determining sound direction from **before to after all the training that you have received so far**. (*please answer only after learning unit 8)*

SLOWER 1 2 3 4 5 6 7 **FASTER**

NO CHANGE

8. Please rate the extent that the amount of **time** allotted for training was appropriate for the amount of improvement?

NEEDED MORE TIME 1 2 3 4 5 6 7 **TRAINING TOOK TOO MUCH TIME**

TRAINING TOOK THE RIGHT AMOUNT OF TIME

On each of the following aspects of your training, rate how beneficial you felt the particular part of training was in learning where sounds were located:

9. The display revealing where the sound would be located and then the sound was played.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

10. Sounds were played randomly, you answered, and then the correct location was revealed.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

11. The items that were answered incorrectly were presented three additional times.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

12. Choose-to-practice at your own pace.

NO BENEFIT 1 2 3 4 5 6 7 **MAXIMUM BENEFIT**

Adaptive + Choose-to-Practice

13. Sound was played, you provided a response and no feedback was given.

NO BENEFIT ← 1 2 3 4 5 6 7 → MAXIMUM BENEFIT

14. At any point(s) in the training did you feel yourself becoming disinterested or disengaged? You can check one or more options below or describe this point(s) in time in the space provided.

___a) The display revealed where the sound would be located and then the sound was played.

___b) Sounds were played randomly, you answered, and then the correct location was revealed.

___c) The items that were answered incorrectly were presented three additional times.

___d) Choose-to-practice at your own pace.

___e) Test phase where the sound was played, you provided a response and no feedback was given.

____

____

____

Appendix J. Experimenter Script for the Main Experiment

DRILCOM instructions

Introduction to the study:
The purpose of the study is to better understand how people determine sound location. You are going to hear an unusual sound similar to a buzz coming from one of 12 speakers. The speakers are arranged in a clock face, so the one directly in front of you is 12 o’clock. Your basic task is to use the mouse and indicate on the screen where you think you heard the sound coming from. You will undergo one familiarization task, a pretest, and then 8 learning blocks with three sub-blocks in each.

Show them the hard copy of what a correct and incorrect answer looks like.

You may take a break or withdraw at any time.

Familiarization

You’re going to hear sounds coming from 3, 6, 9, and 12 in order. Go ahead and begin by selecting the green button in the center of the screen. This is self-paced, so the task today will pace according to when you select the button. Please respond as quickly and accurately as possible. The first sound you will hear is coming from 3 o’clock
*3, 6, after 9, tell them that the last one will answer incorrectly to demonstrate what an incorrect answer will look like.

Pretest

Good, now we are going to do a pre-test. You won't know where the sound is coming from and you won't receive feedback regarding your results. Just respond as quickly and accurately as possible.

LU1

Now you are going to do what we call sequential presentations. You'll hear the sound coming from 12 o'clock then it will move in clockwise order to 11 o'clock.
-Then, the sound will come from 9 o'clock and move to 10 in a counterclockwise order.
-Then the signal will start at 3 o'clock and move clockwise to 2.
-Then the signal will start at 6 and move CCW to 7.

Random

For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results. The screen is going to look a little different as you will have 24 instead of 12 choices that will look like this *(show them the hard copy)*

LU2-LU8 Sequential

Now you are going to do one sequential presentation. Other than the first presentation, you will know where it's coming from and you will receive feedback on your results.

Random

For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

Adaptive

Introduction to the study:
The purpose of the study is to better understand how people determine sound location. You are going to hear an unusual sound similar to a buzz coming from one of 12 speakers. The speakers are arranged in a clock face, so the one directly in front of you is 12 o'clock. Your basic task is to use the mouse and indicate on the screen where you think you heard the sound coming from. You will undergo one familiarization task, a pretest, and then 8 learning blocks with four sub-blocks in each.

Show them the hard copy of what a correct and incorrect answer looks like.

You may take a break or withdraw at any time.

Familiarization

You're going to hear sounds coming from 3, 6, 9, and 12 in order. Go ahead and begin by selecting the green button in the center of the screen. This is self-paced, so the task today will pace according to when you select the button. Please respond as quickly and accurately as possible. The first sound you will hear is coming from 3 o'clock
*3, 6, after 9, tell them that the last one will answer incorrectly to demonstrate what an incorrect answer will look like.

Pretest

Good, now we are going to do a pre-test. You won't know where the sound is coming from and you won't receive feedback regarding your results. Just respond as quickly and accurately as possible.

LU1

Now you are going to do what we call sequential presentations. You'll hear the sound coming from 12 o'clock then it will move in clockwise order to 11 o'clock.
-Then, the sound will come from 9 o'clock and move to 10 in a counterclockwise order.
-Then the signal will start at 3 o'clock and move clockwise to 2.
-Then the signal will start at 6 and move CCW to 7.

Random

For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

Adaptive

In this sub-block, stimuli that you missed in the previous sub-block will be presented four times in random order. Just answer as quickly and accurately as possible.

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

LU2-LU8 Sequential

Now you are going to do one sequential presentation. Other than the first presentation, you will know where it's coming from and you will receive feedback on your results.

Random

For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

Adaptive

In this sub-block, stimuli that you missed in the previous sub-block will be presented four times in random order. Just answer as quickly and accurately as possible.

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

Choose

Introduction to the study:

The purpose of the study is to better understand how people determine sound location. You are going to hear an unusual sound similar to a buzz coming from one of 12 speakers. The speakers are arranged in a clock face, so the one directly in front of you is 12 o'clock. Your basic task is to use the mouse and indicate on the screen where you think you heard the sound coming from. You will undergo one familiarization task, a pretest, and then 8 learning blocks with four sub-blocks in each.

Familiarization

You're going to hear sounds coming from 3, 6, 9, and 12 in order. Go ahead and begin by selecting the green button in the center of the screen. This is self-paced, so the task today will pace according to when you select the button. Please respond as quickly and accurately as possible. The first sound you will hear is coming from 3 o'clock
*3, 6, after 9, tell them that the last one will answer incorrectly to demonstrate what an incorrect answer will look like.

Pretest

Good, now we are going to do a pre-test. You won't know where the sound is coming from and you won't receive feedback regarding your results. Just respond as quickly and accurately as possible.

LU1

Now you are going to do what we call sequential presentations. You'll hear the sound coming from 12 o'clock then it will move in clockwise order to 11 o'clock.
-Then, the sound will come from 9 o'clock and move to 10 in a counterclockwise order.
-Then the signal will start at 3 o'clock and move clockwise to 2.
-Then the signal will start at 6 and move CCW to 7.

Random
For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

Choose-to-practice
This next session is for you to choose to re-play some of the sound locations you wish to hear again. You'll see a counter in the upper left corner that will tell you how many trials you have remaining (18 total).

Test
For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

LU2-LU8 Sequential
Now you are going to do one sequential presentation. Other than the first presentation, you will know where it's coming from and you will receive feedback on your results.

Random

For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

Choose-to-practice

This next session is for you to choose to re-play some of the sound locations you wish to hear again. You'll see a counter in the upper left corner that will tell you how many trials you have remaining (18 total).

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

Choose + Adaptive

Introduction to the study:
The purpose of the study is to better understand how people determine sound location. You are going to hear an unusual sound similar to a buzz coming from one of 12 speakers. The speakers are arranged in a clock face, so the one directly in front of you is 12 o'clock. Your basic task is to use the mouse and indicate on the screen where you think you heard the sound coming from. You will undergo one familiarization task, a pretest, and then 8 learning blocks with five sub-blocks in each.

Familiarization

You're going to hear sounds coming from 3, 6, 9, and 12 in order. Go ahead and begin by selecting the green button in the center of the screen. This is self-paced, so the task today will pace according to when you select the button. Please respond as quickly and accurately as possible. The first sound you will hear is coming from 3 o'clock
*3, 6, after 9, tell them that the last one will answer incorrectly to demonstrate what an incorrect answer will look like.

Pretest

Good, now we are going to do a pre-test. You won't know where the sound is coming from and you won't receive feedback regarding your results. Just respond as quickly and accurately as possible.

LU1

Now you are going to do what we call sequential presentations. You'll hear the sound coming from 12 o'clock then it will move in clockwise order to 11 o'clock.

-Then, the sound will come from 9 o'clock and move to 10 in a counterclockwise order.

-Then the signal will start at 3 o'clock and move clockwise to 2.

-Then the signal will start at 6 and move CCW to 7.

Random

For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

Adaptive

In this sub-block, stimuli that you missed in the previous sub-block will be presented four times in random order. Just answer as quickly and accurately as possible.

Choose-to-practice

This next session is for you to choose to re-play some of the sound locations you wish to hear again. You'll see a counter in the upper left corner that will tell you how many trials you have remaining (12 total).

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

LU2-LU8 Sequential

Now you are going to do one sequential presentation. Other than the first presentation, you will know where it's coming from and you will receive feedback on your results.

<u>Random</u>

For this block, you will not know where the sound is coming from, but you will receive feedback regarding your results. Please respond as quickly and accurately as possible.

<u>Adaptive</u>

In this sub-block, stimuli that you missed in the previous sub-block will be presented four times in random order. Just answer as quickly and accurately as possible.

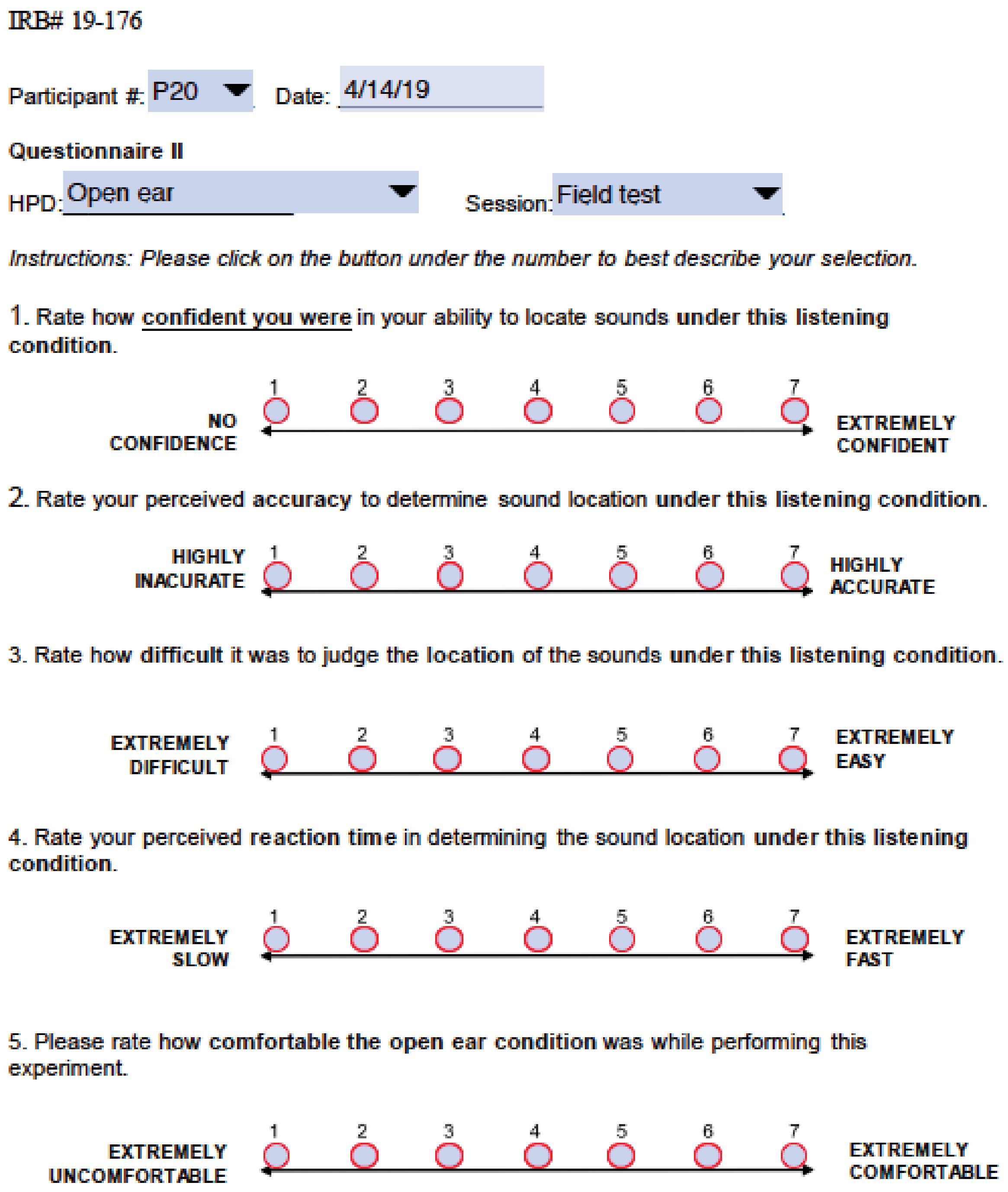
<u>Choose-to-practice</u>

This next session is for you to choose to re-play some of the sound locations you wish to hear again. You'll see a counter in the upper left corner that will tell you how many trials you have remaining (12 total).

<u>Test</u>

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

Appendix K. Phase III Questionnaire



IRB# 19-176

Participant #: P20 Date: 4/14/19

Questionnaire II

HPD: Open ear Session: Field test

Instructions: Please click on the button under the number to best describe your selection.

1. Rate how <u>confident you were</u> in your ability to locate sounds under this listening condition.

	1	2	3	4	5	6	7	
NO CONFIDENCE	○	○	○	○	○	○	○	**EXTREMELY CONFIDENT**

2. Rate your perceived accuracy to determine sound location under this listening condition.

	1	2	3	4	5	6	7	
HIGHLY INACURATE	○	○	○	○	○	○	○	**HIGHLY ACCURATE**

3. Rate how difficult it was to judge the location of the sounds under this listening condition.

	1	2	3	4	5	6	7	
EXTREMELY DIFFICULT	○	○	○	○	○	○	○	**EXTREMELY EASY**

4. Rate your perceived reaction time in determining the sound location under this listening condition.

	1	2	3	4	5	6	7	
EXTREMELY SLOW	○	○	○	○	○	○	○	**EXTREMELY FAST**

5. Please rate how comfortable the open ear condition was while performing this experiment.

	1	2	3	4	5	6	7	
EXTREMELY UNCOMFORTABLE	○	○	○	○	○	○	○	**EXTREMELY COMFORTABLE**

6. How likely would you be to **keep your ears open (no hearing protector)** during a task similar to this experiment that required sound localization if you had access to a hearing protection device.

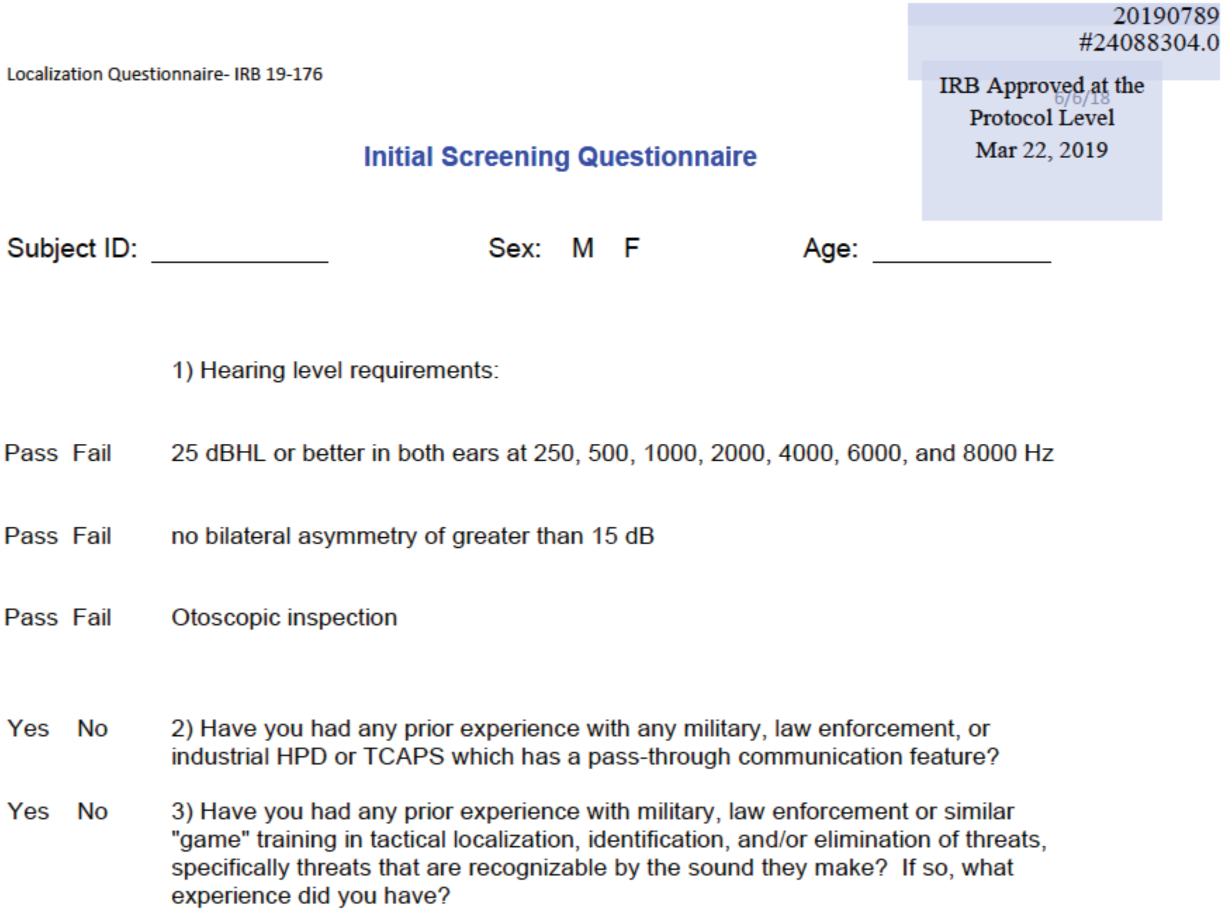
	1	2	3	4	5	6	7	
EXTREMELY UNLIKELY	○	○	○	○	○	○	○	**EXTREMELY LIKELY**

7. Rate the degree of preparedness you felt as a result of the training on the localization system (ring of loudspeakers) compared to the **task of localizing .22 blank gunshot sounds**.

	1	2	3	4	5	6	7	
EXTREMELY UNPREPARED	○	○	○	○	○	○	○	**EXTREMELY PREPARED**

Please provide any additional comments about your localization testing using the localization system.

Appendix L. Phase III Screening Form



20190789
#24088304.0

IRB Approved at the
6/6/18
Protocol Level
Mar 22, 2019

Localization Questionnaire- IRB 19-176

Initial Screening Questionnaire

Subject ID: ____________ Sex: M F Age: ____________

1) Hearing level requirements:

Pass Fail 25 dBHL or better in both ears at 250, 500, 1000, 2000, 4000, 6000, and 8000 Hz

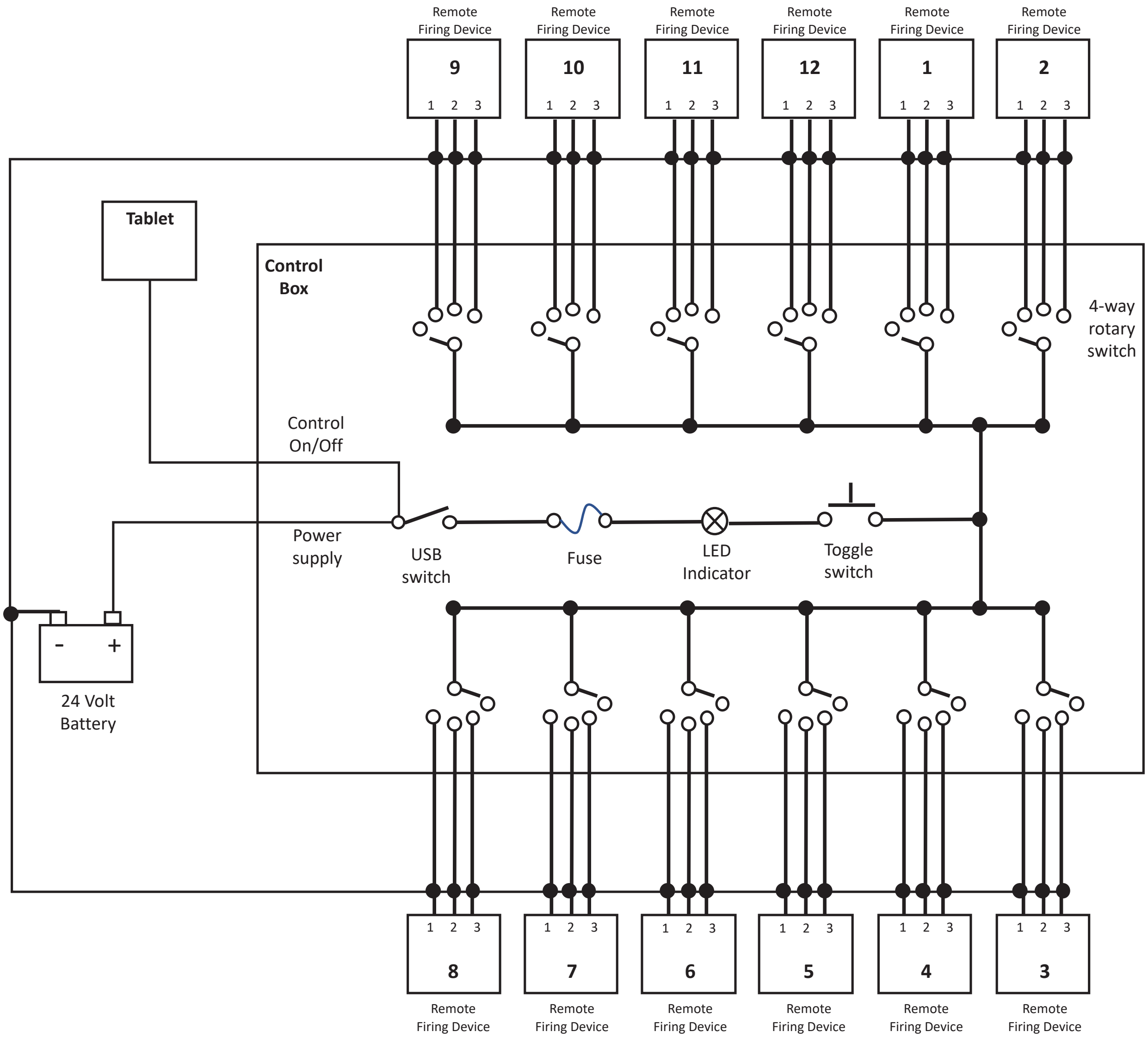
Pass Fail no bilateral asymmetry of greater than 15 dB

Pass Fail Otoscopic inspection

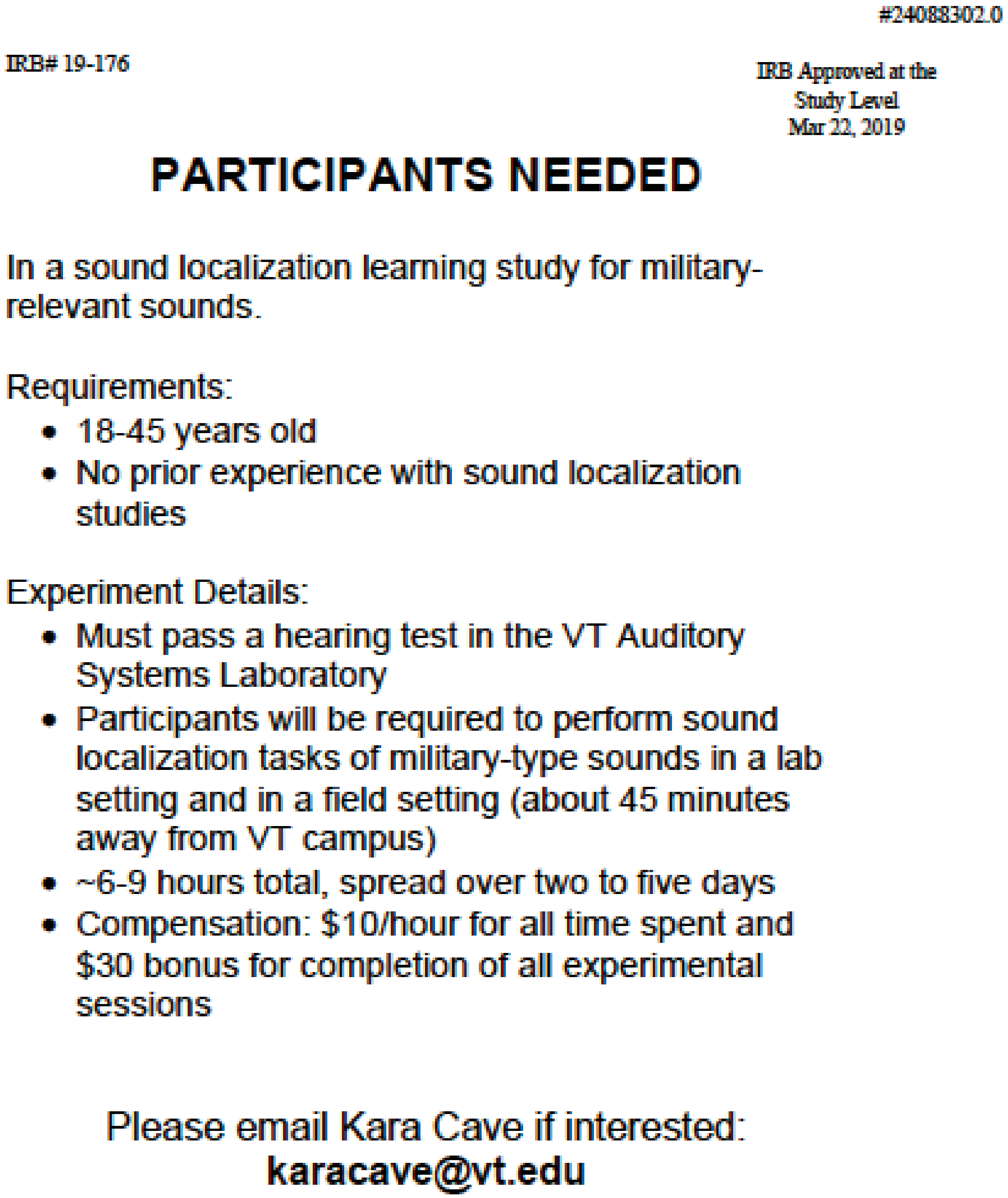
Yes No 2) Have you had any prior experience with any military, law enforcement, or industrial HPD or TCAPS which has a pass-through communication feature?

Yes No 3) Have you had any prior experience with military, law enforcement or similar "game" training in tactical localization, identification, and/or elimination of threats, specifically threats that are recognizable by the sound they make? If so, what experience did you have?

Appendix M. Remote Firing Device Wiring Diagram



Appendix N. Phase III Participant Flyer



#24088302.0

IRB# 19-176

IRB Approved at the
Study Level
Mar 22, 2019

PARTICIPANTS NEEDED

In a sound localization learning study for military-relevant sounds.

Requirements:

- 18-45 years old
- No prior experience with sound localization studies

Experiment Details:

- Must pass a hearing test in the VT Auditory Systems Laboratory
- Participants will be required to perform sound localization tasks of military-type sounds in a lab setting and in a field setting (about 45 minutes away from VT campus)
- ~6-9 hours total, spread over two to five days
- Compensation: $10/hour for all time spent and $30 bonus for completion of all experimental sessions

Please email Kara Cave if interested:
karacave@vt.edu

Appendix O. Phase III Informed Consent

IRB APPROVED
Mar 22, 2019

RESEARCH SUBJECT CONSENT FORM

Title:	Innovative Portable Auditory Localization Acclimation-Test System (PALAT) for Tactical Communications and Protective Systems (TCAPS) in Military and Other Applications, including Full-Scale and Portable Versions, Leveraged on DRILCOM Test Battery and Validated via an In-Field Experiment (VT OSP #450489)
Protocol No.:	19-176 WIRB® Protocol #20190789 19-176
Sponsor:	Virginia Polytechnic Institute and State University
Investigator:	John G. Casali, PhD, CPE 250 Durham Hall (0118) Blacksburg, Virginia, 24061 USA
Daytime Phone Number:	540-231-5073
24-hour Phone Number:	(904) 307-8144

RESEARCH CONSENT SUMMARY

You are being asked for your consent to take part in a research study. This document provides a concise summary of this research. It describes the key information that we believe most people need to decide whether to take part in this research. Later sections of this document will provide all relevant details.

What should I know about this research?

- Someone will explain this research to you.
- Taking part in this research is voluntary. Whether you take part is up to you.
- If you don't take part, it won't be held against you.
- You can take part now and later drop out, and it won't be held against you
- If you don't understand, ask questions.
- Ask all the questions you want before you decide.

How long will I be in this research?

We expect that your taking part in this research will last up to approximately nine hours spread over five sessions, spread out over the span of a week (seven days). The experiment could be as short as approximately six hours spread over two days, spanning a week.

Why is this research being done?

The purpose of this research is to assess the benefits of training using a portable auditory, or hearing, localization training system on localization performance in an outdoor environment. The experiment is designed to simulate a scenario where a military service member is listening for gunshots from a long distance. You, the participant, will need to locate, via hearing, where the simulated gunshot is located. Another purpose to the experiment is to determine if training affects localization accuracy without hearing protection (that is, with open ears) and with certain types of hearing protection.

What happens to me if I agree to take part in this research?

If you decide to take part in this research study, the general procedures include at least two stages and possibly a third. In the first stage, a screening session will occur that involves filling out a demographic questionnaire, looking in your ears with an ear microscope, and taking a brief hearing test. The first session will also involve a test to determine how well you locate sounds. The sound used in this session will seem similar to a beep, played at a moderate (not loud) level. Some participants will be asked to take part in three additional training sessions in the same location as the first session. All participants will be asked to participate in a sound location test in a field environment about 45 minutes from the Virginia Tech Campus. This field session will take place five to seven days after the first session. The sound used in the field session is a blank simulated gunshot. No "live" ammunition will be used. The blanks will be fired from a fabricated device specifically made for this purpose and located at least 150 feet away from you. Should the fabricated device fail, a starter pistol, used at sporting events to mark the start of a competition such as a foot race, will be used. You will be asked to listen with your open (uncovered) ears and with two different types of hearing protection.

Could being in this research hurt me?

The most important risks or discomforts that you may expect from taking part in this research include discomfort from hearing blank simulated gunshots. The experiment is designed to simulate a scenario where a soldier must listen for shots coming from a long distance away. If you feel this scenario would make you uncomfortable, please do not participate. Exposure to the noise from the blanks will not be loud enough to create a risk to your hearing, in that they will be well below that which is governed by the Occupational Safety and Health Administration (OSHA). OSHA allows peak exposures to be up to 140 dB peak sound pressure level. The blanks in this study will not exceed 113 dB peak sound pressure level. OSHA also allows average sound pressure levels to be up to 90 dBA for up to an 8-hour day, at which level the use of hearing protection becomes mandatory. The average levels in this experiment are at less than 85 dBA, and will not be presented for more than 4 hours per day. You will be given an opportunity to observe the blank ammunition and the firing device (starter pistol or fabricating device as applicable).

Hearing protectors are designed to have a tight fit and you may experience some minor discomfort while wearing them. If you experience more than minor discomfort, please tell the experimenter and he/she will assist you in adjusting or removing the hearing protector.

If you feel tired or become thirsty during the test, please inform one of the experimenters. You may take a break at any time. Water is will be made available to you.

Will being in this research benefit me?

The most important benefits that you may expect from taking part in this research include information on the ability to learn to localize sounds with the open ear and while wearing different hearing protectors. No promise or guarantee of benefits have been made to encourage you to participate. It is not expected that you will personally benefit from this research.

Possible benefits to others include assisting the military and law enforcement to determine the effects of auditory localization training and testing using hearing protection devices. The data in this study will also be used in fulfillment of two dissertations in human factors engineering at Virginia Tech.

What else should I know about this research?

Other information that may be important for you to consider so you can decide whether to take part in this research is that one session will take place in an outdoor field, located about 45 minutes away in Pulaski County. The session that will take place in the field will span about 3-4 hours, including transportation time.

DETAILED RESEARCH CONSENT

You are being invited to take part in a research study. A person who takes part in a research study is called a research subject, or research participant.

What should I know about this research?

- Someone will explain this research to you.
- This form sums up that explanation.
- Taking part in this research is voluntary. Whether you take part is up to you.
- You can choose not to take part. There will be no penalty or loss of benefits to which you are otherwise entitled.
- You can agree to take part and later change your mind. There will be no penalty or loss of benefits to which you are otherwise entitled.
- If you don’t understand, ask questions.
- Ask all the questions you want before you decide.

Why is this research being done?

The purpose of this research is to assess training effects of an indoor Portable Auditory Localization Acclimation-Test System (PALAT) by testing participants’ ability to determine the directions from which sounds are coming (sound localization) in an outdoor field setting. Furthermore, the experimenters are attempting to determine if sound localization ability is affected by the wearing of certain types of hearing protection. The experiment was designed to evaluate a training method for sound localization for later use in a military population.

About 24 subjects will take part in this research.

How long will I be in this research?

We expect that your taking part in this research will last approximately 9 hours. Each participant will complete at least two sessions. The first session will last approximately two hours. All participants will be asked to complete the field session that will last from three to four hours. Some participants will be randomly selected to complete training. Training take place in three sessions lasting approximately an hour each for each session, totaling three hours spread over no more than five days. After no more than seven days from the first session, a field test will be conducted (three-four hours). Therefore, the time commitment is expected not to exceed a week.

What happens to me if I agree to take part in this research?

Before audiometric testing, you will be asked to review and sign an informed consent form. Two copies of the informed consent form (this form) will be provided to you upon arrival at the initial screening and training session (as applicable): one signed copy will be maintained by the researchers and one copy is for you, the participant. This informed consent form is the same form that was provided to you, and each participant, via email upon volunteering for the study. After you read the informed consent form, you can ask any questions related to the experiment. You will be put into a study group by chance (like a coin toss/ like drawing straws). You have a 1 out of 2 chance of being placed in each group. You cannot choose your study group.

Audiometric Qualification (Eligibility) Testing

The screening and training stage will begin with audiometric qualification testing. Audiometric qualification testing will include 1) a standard hearing test, to determine your hearing sensitivity, 2) a visual inspection of your ear canal using a lighted otoscope, to determine if there are any obstructions, and 3) a history of your hearing protection use and localization study/training experience. If you have impacted earwax or other ear canal problems, you will be asked not to participate, and perhaps to visit an ear health professional such as an audiologist (hearing specialist) or otolaryngologist (ear physician) to have your ear canals clinically checked and cleaned of earwax if needed, and perhaps return for a second screening for this experiment if you so desire. The audiometric test and visual inspection will be conducted by an Active Duty U.S. Army Audiologist. You will be informed if you met the hearing eligibility requirements to complete the experiment.

Screening session

All participants will be asked to complete this session. Upon successful completion of all eligibility requirements, you will:

1. Receive familiarization training/orientation with the actual hearing protection devices (HPDs) known as Tactical Communication and Protective Systems (TCAPS) that will be used in the study.
2. Undergo a auditory localization pretest using the Portable Auditory Situation Awareness Training (PALAT) system. During the pretest, you will be asked to listen to a series of 36 sound signals (beeps) presented to you while sitting in the center of a circle of 12 loudspeakers and to identify and respond as accurately and quickly as possible with the direction you perceived the sound. Your responses will be recorded on a computer tablet by a

computer program for later analysis. The background masking noise (pink noise) level will always be played at 55 dBA or below, which is a quiet level. The sound signal for the localization test will be a sound that includes both low and high frequency ranges that are well-within the pitches of sound that can be heard and localized by the human ear. The test sound signal will be played at 70 dBA for one second in length (this is not loud, but well below sound levels allowed by the Occupational Safety and Health Administration (OSHA) for U.S. industrial workers).

Training sessions

Half of the participants will undergo training sessions. The investigator will tell you if you are assigned to this group. Auditory localization training will occur over a period of 3 sessions, each session lasting approximately 1 hour. You will be asked to perform the auditory training and testing under three listening conditions: open ear (no hearing protection device), wearing an in-the-ear TCAPS, and an over-the-ear TCAPS. The experimenter will fit the hearing protectors in (earplugs) or on (earmuffs) your ears. After each training session you will be asked to fill out a questionnaire rating your confidence in ability to localize the sound signal.

Field session:

All participants will be asked to complete this session. The experimental stage will take place in a rural field in which you will stand in surrounded by a wood forest in which two to three experimenters will be located. The experimenters will initiate blank gunshots from at least 150-feet away, that is, ammunition that is not "live" and has no bullet, from a device designed specifically for this purpose -- that is, it is not an actual weapon or gun. This equipment is similar to perimeter alarms used to contain livestock. You will be able to inspect the device that fires the blanks, as well as the blanks, before the start of the experiment. After each shot, you will be asked to verbally identify one of 24 numbered signs that corresponds most closely to the direction (location) you think the shot was fired from. There will be three listening conditions: open ear (no hearing protection device), wearing an in-the-ear TCAPS, and an over-the-ear TCAPS. The experimenter will fit the hearing protectors in (earplugs) or on (earmuffs) your ears. After each localization test you will be asked to fill out a questionnaire rating your confidence in ability to localize the blank gunshot signals and your impressions about the TCAPS.

What are my responsibilities if I take part in this research?

If you take part in this research, you will be responsible to:

- Listen for and localize the signals in the experiment to the best of my ability
- Furnish accurate ratings of my impressions about my ability to localize under all listening conditions
- Inform the Experimenter if a protector, or any other aspects of the test condition becomes uncomfortable
- Schedule multiple sessions within the allotted time and adhere to scheduled appointments with the experimenter
- Inform the Experimenter if you are unable to make your scheduled time
- Inform the Experimenter if you become tired, thirsty or wish to rest
- Inform the Experimenter if you wish to withdraw from the study

Could being in this research hurt me?

- This experiment involves localizing blank simulated gunshot sounds. No live ammunition or weapons will ever be present or used. This scenario could pose an emotional risk to the participant. If think this sound could be distressing you are asked decline participation.
- You will be exposed to the impulse (pop) sounds from blanks that will be moderately loud, but not of a level that is hazardous to hearing, even in conditions where hearing protection is not applied (i.e., the open ear condition).
- You will be wearing hearing protection as part of this study. They are designed to be snug-fitting and may occasionally emit a whistle or squeal if the microphones are covered. These are not hazardous conditions, but any reports of discomfort will be met with an offer of a rest breaks, re-inspection, and re-fitting of the devices.
- You may become thirsty during this study, especially in the field. Please let the experimenter know and water and rest breaks will be offered.
- Given the unknown availability obstetric emergency care in the field location located 45 minutes from Virginia Tech, this research has an unknown risk for pregnant females. Due to this distance, and because taking part in this research may harm a pregnancy in unknown ways, pregnant females cannot participate in this research. Please notify the investigator if you think you may be pregnant.

Will it cost me money to take part in this research?

No, it will not cost you money. Should you choose to drive yourself to the field location in Pulaski County, you will be reimbursed for mileage at the rate Virginia Tech currently uses. Additionally, you will be paid for your travel time.

Will being in this research benefit me?

We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits to you include information on the ability to learn to localize sounds while wearing different hearing protectors.

What other choices do I have besides taking part in this research?

This research is not designed to diagnose, treat or prevent any disease. Your alternative is to not take part in the research.

What happens to the information collected for this research?

Your private information will be shared with individuals and organizations that conduct or watch over this research, including:

- The investigators listed on this study
- The Institutional Review Board (IRB) that reviewed this research
- Representatives of the Department of Defense
- Your personal information will not be shared with the research sponsor, but the data you produce will be shared

We may publish the results of this research. However, we will keep your name and other identifying information confidential and you cannot be identified in any manner in these publications.

We protect your information from disclosure to others to the extent required by law. We cannot promise complete secrecy.

Data collected in this research will be deidentified and used for future research or distributed to another investigator for future research without your consent.

Who can answer my questions about this research?

If you have questions, concerns, or complaints, or think this research has hurt you or made you sick, talk to the research team at the phone number listed above on the first page.

This research is being overseen by an Institutional Review Board ("IRB"). An IRB is a group of people who perform independent review of research studies. You may talk to them at (800) 562-4789, help@wirb.com if:

- You have questions, concerns, or complaints that are not being answered by the research team.
- You are not getting answers from the research team.
- You cannot reach the research team.
- You want to talk to someone else about the research.
- You have questions about your rights as a research subject.

What if I am injured because of taking part in this research?

If you are injured or get sick because of being in this research, you will be responsible for the medical care and costs incurred. If an emergency arises, emergency medical care (911) will be called. Your insurance may be billed for this treatment.
If you are injured as a result of this study, you do not give up your right to pursue a claim through the legal system.

Can I be removed from this research without my approval?

The person in charge of this research can remove you from this research without your approval. Possible reasons for removal include:

- You are unable to keep your scheduled appointments

We will tell you about any new information that may affect your health, welfare, or choice to stay in this research.

What happens if I agree to be in this research, but I change my mind later?

If you decide to leave this research, contact the research team so that the investigator can reimburse you for your time. No adverse consequences will exist if you withdraw.

Will I be paid for taking part in this research?

For taking part in this research, you may be paid up to a total of $ 120.00. Your compensation will be broken down as follows:

- $10.00 per hour during training/test sessions/travel time
- For any fraction of time less than 1 hour, you will be paid for the closest ½ hour rounded up in your favor
- You will be paid at the conclusion of each screening, training, and experimental session
- You will be paid a $30.00 bonus for successful completion of all experimental sessions

Statement of Consent:

Your signature documents your consent to take part in this research.

Signature of adult subject capable of consent Date

Signature of person obtaining consent Date

Appendix P. Phase III Participant Instructions

IC process:

1) Furnish participant with 2 copies of the Informed Consent (one is for participant to keep)

2) Instruct participant to review the informed consent.

Highlight the following:

Purpose

Introduction to the study:
The purpose of the study is to better understand how people determine sound location. You will be asked to locate a sound with and without hearing protection. Two types of electronic hearing protection will be used. If at any time the hearing protector becomes uncomfortable, please let the experimenter know. If the hearing protector is moved, you may hear a squeal. Please let the experimenter adjust the device to avoid this.

In the first session today, you are going to hear an unusual sound similar to a buzz coming from one of 24 speakers. The speakers are arranged in a clock face, so the one directly in front of you is 12 o'clock. Your basic task is to select on the screen via touchscreen to indicate which loudspeaker emitted the sound. You will undergo one familiarization task and a pretest. We are attempting to learn how individuals localize sound without any previous training.

(show graphic of display)

Everyone in the study will be asked to complete field testing. Please note that gunfire-like sounds will be used. These sounds have been tested repeatedly and pose no risk to your hearing. Blanks will be used, but no weapons will be firing these blanks. You will have an opportunity to examine the firing device should you wish to do so.

For the training condition
In between the pretest and the field test, some personnel will be asked to conduct training. The task is very similar to the pretest, only practice sessions will be incorporated to the training session. Each training session will take about an hour, spread over three sessions.

You may take a break or withdraw at any time.

Procedures section: Pending signature of this form, the next step would be to look in your ears and then conduct a hearing test. Provided the results are acceptable for continuation in this study, we will proceed with the pretest to see how well people localize sounds. Some personnel in this study will proceed to a training session comprised of 3 sessions of about an hour each. All personnel will take a second test in a field environment 45 min away from here in an open field in Pulaski County (show picture of site). The field site will take about 3-4 hours to complete including transportation. The field session will occur in the next 3-5 days.

Show them the hard copy of what a correct and incorrect answer looks like.

Please let the experimenter know if you would like to take a break, water and coffee is available in room 513. At the field site as bathroom facilities are located in a cabin about ½ mile from the testing site and we can drive you to the location.

Before they sign, ask:

___Do you understand the information provided?
___Do you feel like you are deciding without the pressure of time or other factors to make a decision?
___Do you understand that there is a voluntary choice to make?
___Are you capable of making and communicating an informed choice?

What are your questions?

Familiarization

You're going to hear sounds coming from 12, 3, 6, and 9 in order. Go ahead and begin by selecting the green button in the center of the screen. This is self-paced, so the task today will pace according to when you select the button. Please respond as accurately and quickly as possible. The first sound you will hear is coming **from 12 o'clock**
*3, 6, after 9, tell them that the last one will answer incorrectly to demonstrate what an incorrect answer will look like.

Pretest

**change ear condition (ITE, OTE, open) on tablet*
Now we are going to do a pre-test. You won't know where the sound is coming from and you won't receive feedback regarding

your results. Just respond as accurately and quickly as possible. I emphasize that it is very important that you are <u>accurate</u>, but also respond as <u>quickly</u> as you can.

Administer the questionnaire (PASAT desktop folder, open subject's folder, then condition, press "save" in the upper left, and submit in the upper right).

LU1

Now you are going to do what we call sequential presentations. You'll hear the sound coming from 12 o'clock then it will move in clockwise order to 11 o'clock.
-Then, the sound will come from 9 o'clock and move to 10 in a counterclockwise order.
-Then the signal will start at 3 o'clock and move clockwise to 2.
-Then the signal will start at 6 and move CCW to 7.
-Once the sequential presentations have finished, the program will auto-advance and the screen will change slightly.
-You will move into the random session.

Random

For this block, you will not know where the sound is coming from like in the sequential, but you will receive feedback regarding your results. Please respond as accurately and quickly as possible.
-Once this block is over, the screen will change slightly and you will proceed to the "choose" session.

Choose

The choose session is for you to choose to re-play some of the sound locations you wish to hear again. You'll see a counter in the upper left corner that will tell you how many trials you have remaining (18 total). Your task is to touch the black circle of the speaker where you would like to hear more presentations.

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results

LU2-LU5 Sequential

Now you are going to do one sequential presentation. You will know where it's coming from (listed on the screen) and you will receive feedback on your results. Please respond as accurately and quickly as possible.

Random

For this block, you will not know where the sound is coming from like in the sequential, but you will receive feedback regarding your results. Please respond as accurately and quickly as possible.

-Once this block is over, the screen will change slightly and you will proceed to the "choose" session.

Choose

of the sound locations you wish to hear again. You'll see a counter in the upper left corner that will tell you how many trials you have remaining (18 total). Your task is to touch the black circle of the speaker where you would like to hear more presentations.

Test

For this block you will not know where the sound is coming from and you will not receive feedback regarding your results.

*Administer questionnaire after LU5

Field test

Familiarization

We are going to walk through four examples from 12, 3, 6, and 9 o’clock. This task is a little different from back in the lab. The experimenter will tell you when to start, you will select the “click to start” button. The sound will play, you will respond on the tablet, and then the screen will prompt you to speak your response so we can write it down. Then the experimenter will tell you it’s okay to start the experiment.

The tablet response is the same as before. You have 24 options, even the in-between responses are considered valid. For example, when the sound is between 1 and 2 o’clock, please say that. We will tell you when to select the green button, a sound will play.

If at any time the hearing protector becomes uncomfortable, please let the experimenter know. If the hearing protector is moved, you may hear a squeal. Please let the experimenter adjust the device to avoid this.

If you need a break at any time, please let the experimenter know. Water, sunblock, and bug spray are available. Also, bathroom facilities are located about ½ mile down the hill. One of the experimenters can drive you to the facility.

Appendix Q. Figures of statistically non-significant findings included in qualitative analysis.

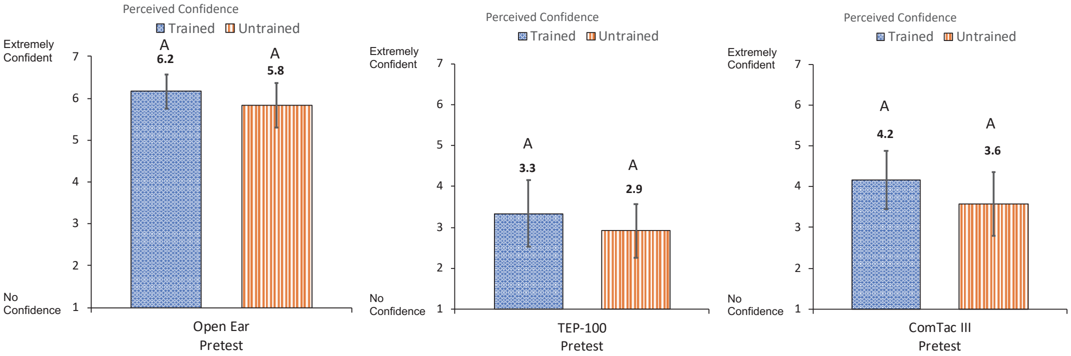


Figure . Mann-Whitney U results comparing mean ratings of confidence for each listening condition at pretest for trained versus untrained groups for Question 1, Perceived Confidence.

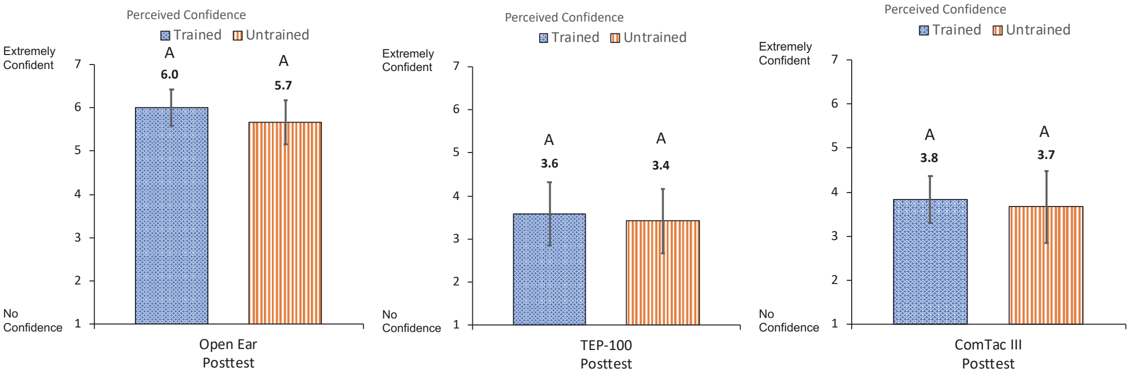


Figure . Mann-Whitney U results comparing mean ratings of confidence for each listening condition at posttest for trained versus untrained groups for Question 1, Perceived Confidence.

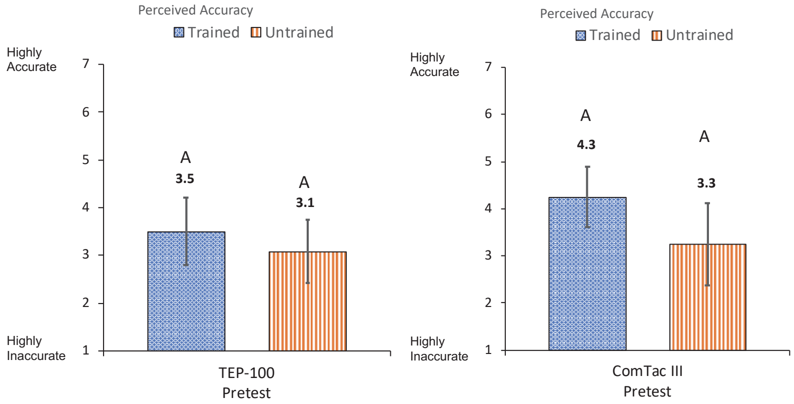


Figure . Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest in the TEP-100 and ComTac™ III conditions for Question 2, Perceived Accuracy.

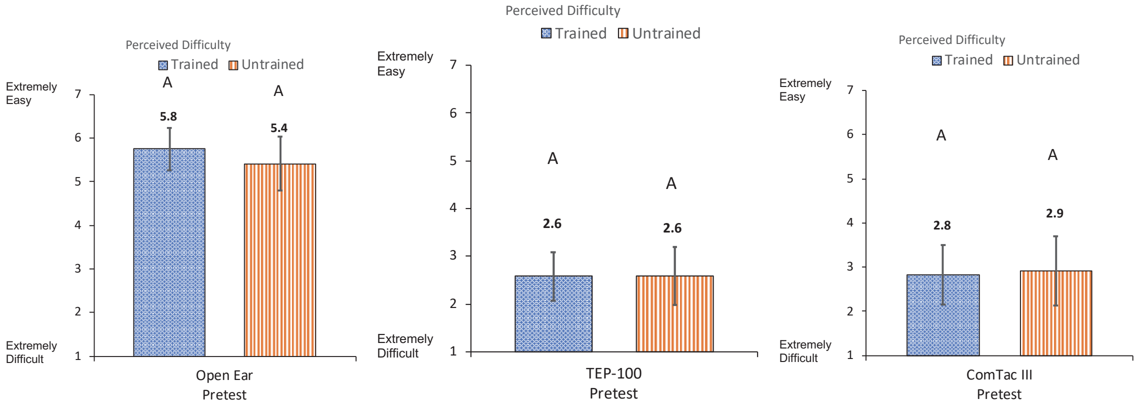


Figure . Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest for all listening condition for Question 3, Perceived Difficulty.

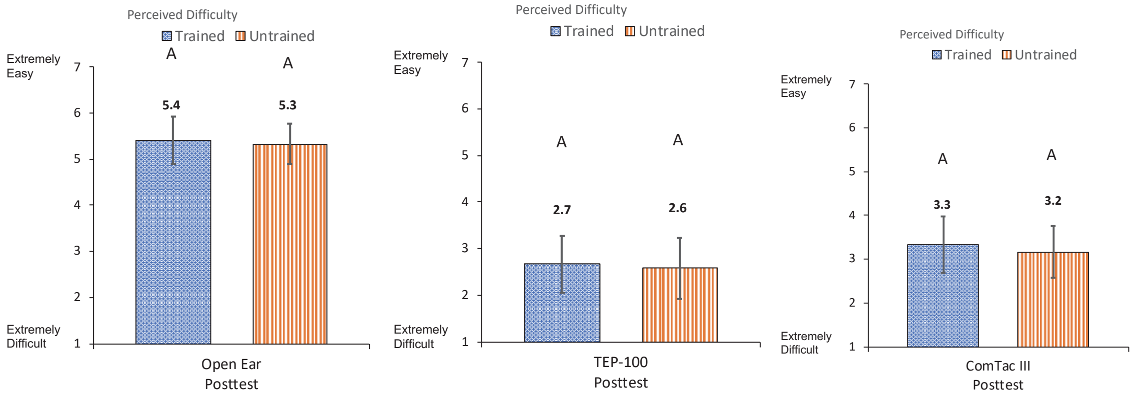


Figure . Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at posttest for all listening condition for Question 3, Perceived Difficulty.

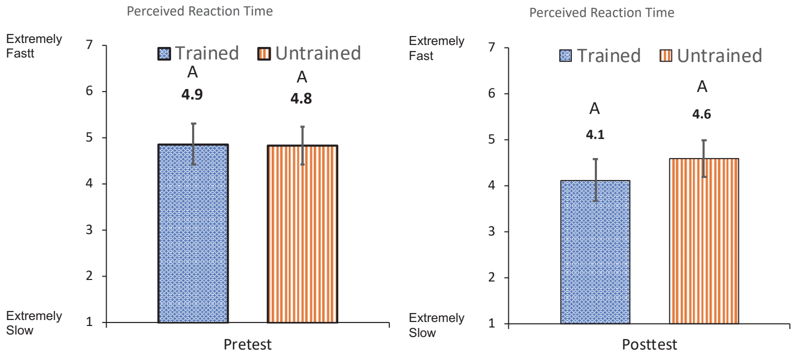


Figure . Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest and posttest collapsed across listening conditions for Question 4, Perceived Reaction Time.

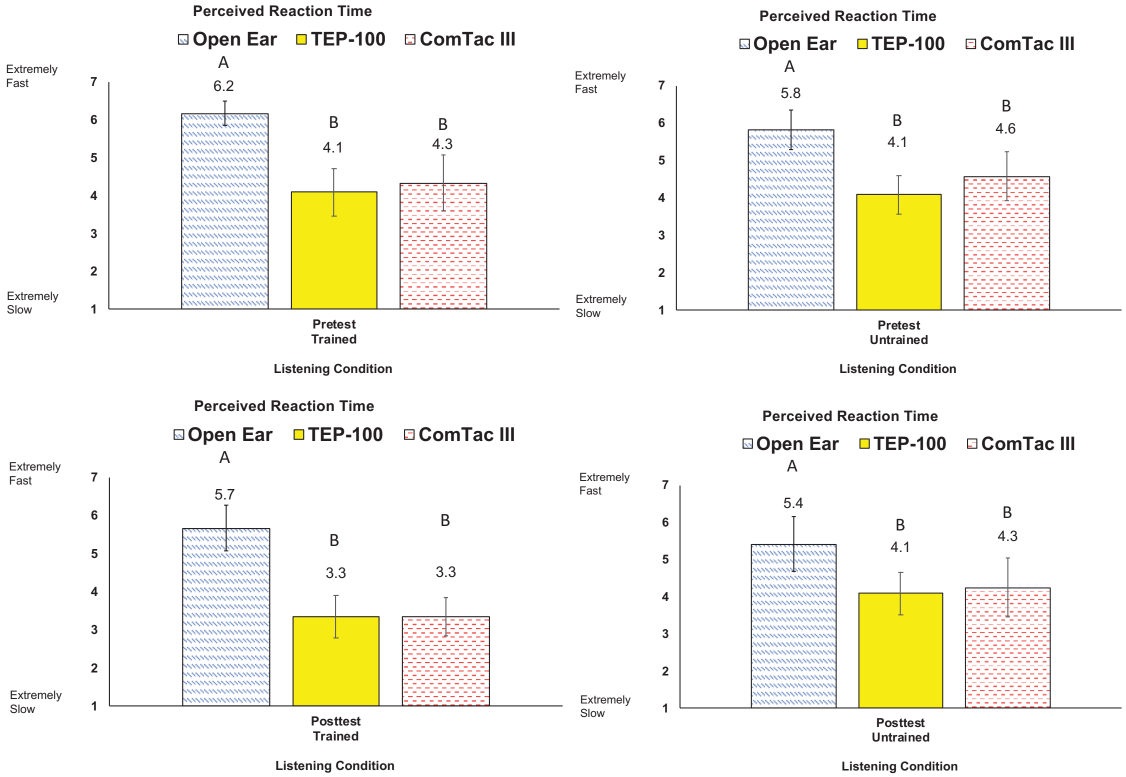


Figure . Mann-Whitney *U* results comparing trained and untrained ratings of perceived reaction time at pretest for all listening condition for Question 4, Perceived Reaction Time.

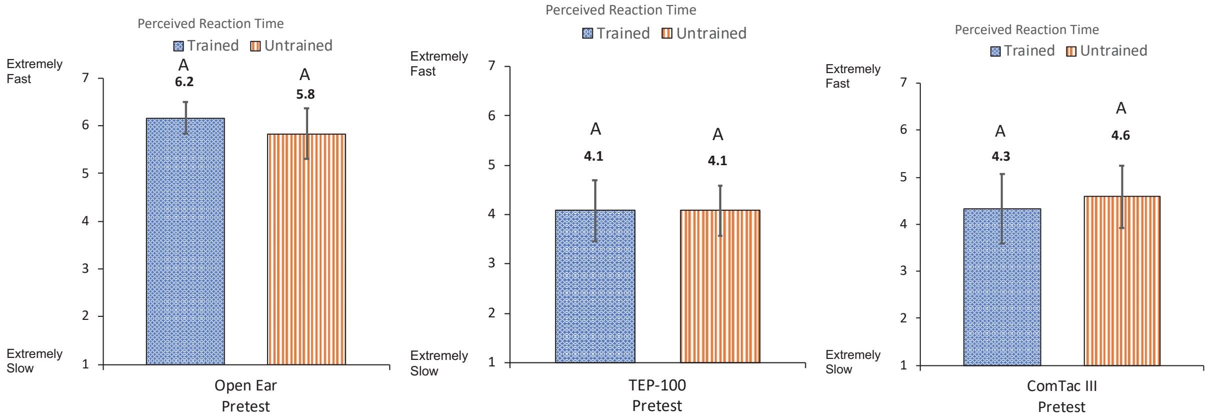


Figure . Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at pretest for all listening condition for Question 4, Perceived Reaction Time.

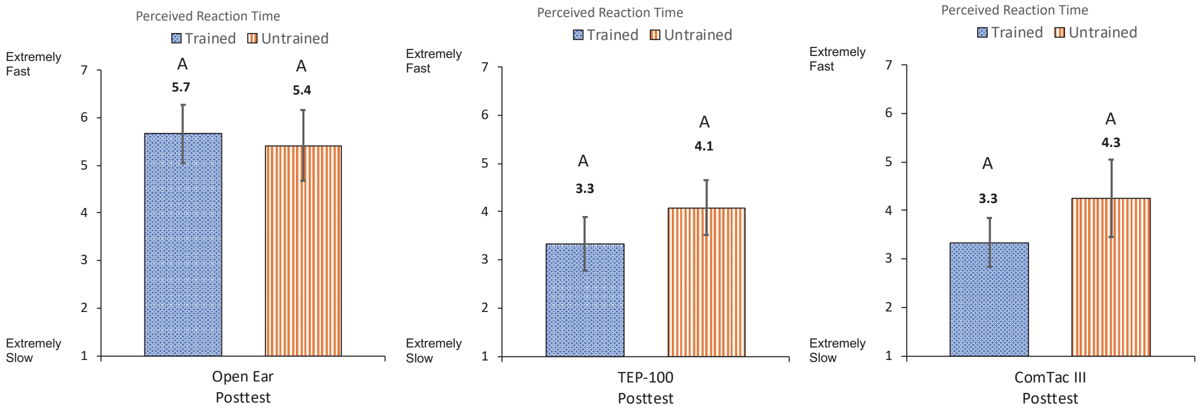


Figure . Mann-Whitney *U* results comparing trained and untrained ratings of perceived accuracy at posttest for all listening condition for Question 4, Perceived Reaction Time.

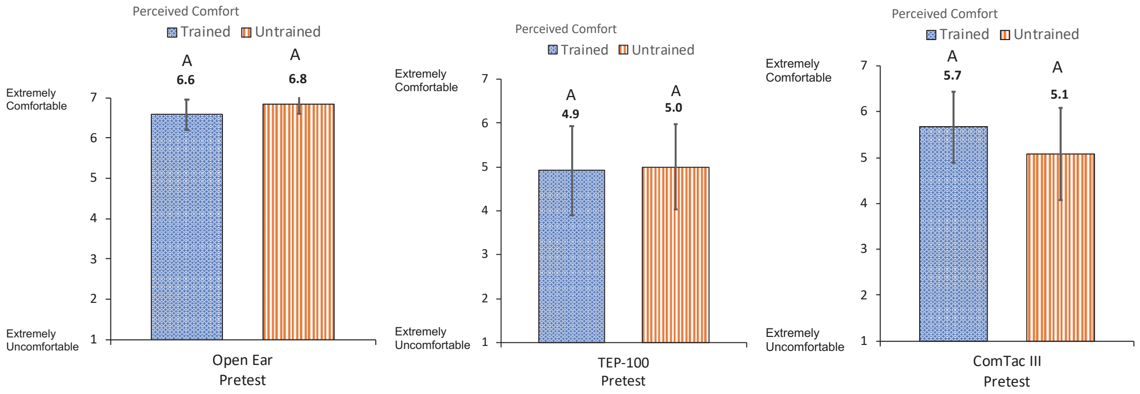


Figure . Mann-Whitney U results comparing mean ratings of confidence for each listening condition at pretest for trained versus untrained groups for Question 5, Perceived Comfort.

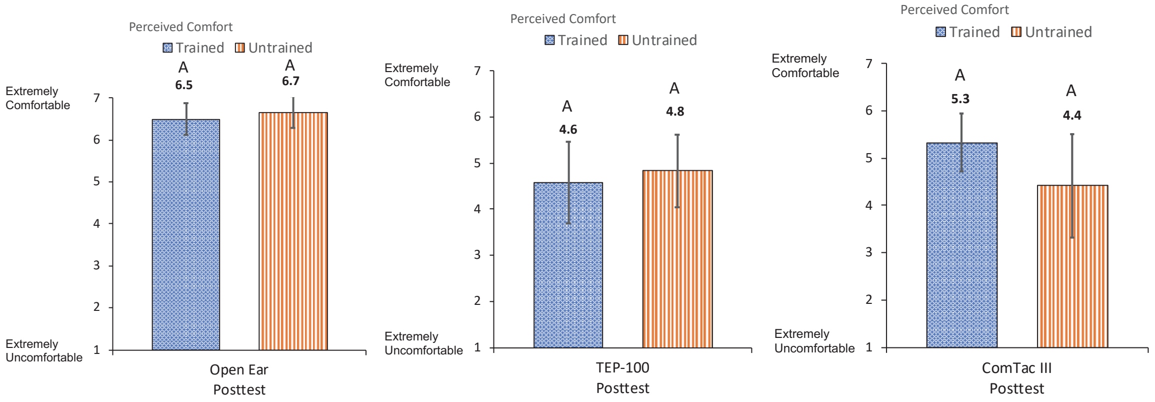


Figure . Mann-Whitney U results comparing mean ratings of confidence for each listening condition at posttest for trained versus untrained groups for Question 5, Perceived Comfort.

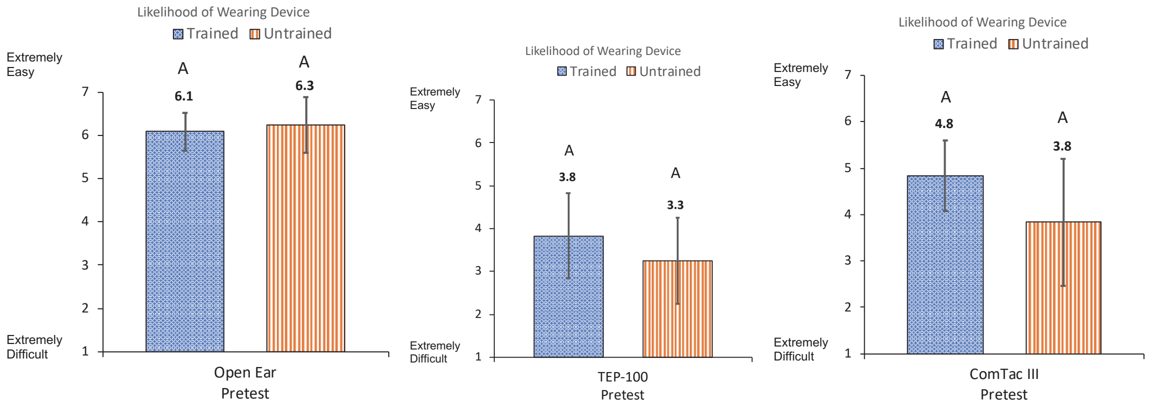


Figure . Mann-Whitney U results comparing mean ratings of likelihood of wearing device for each listening condition at pretest for trained versus untrained groups for Question 6, Likelihood of Wearing Device.

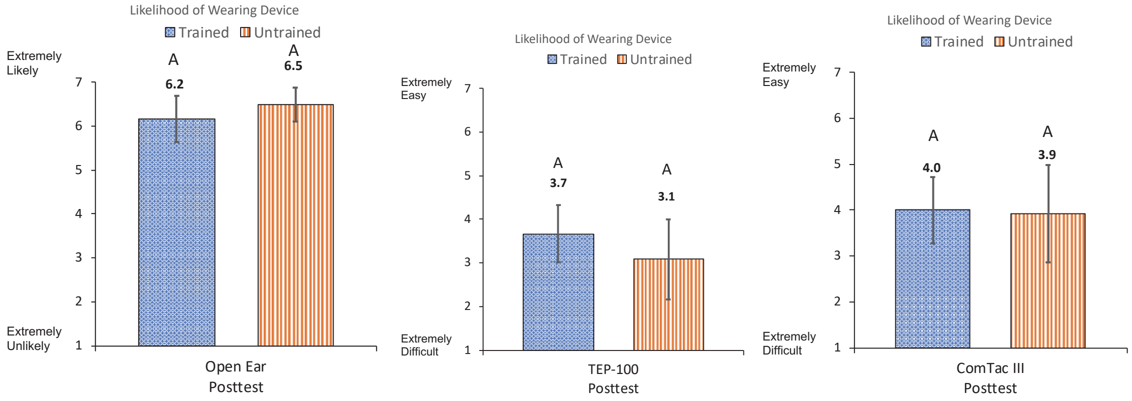


Figure . Mann-Whitney U results comparing mean ratings of likelihood of wearing device for each listening condition at posttest for trained versus untrained groups for Question 6, Likelihood of Wearing Device.