

The Effects of Speech Cues on Long-term Memory

by

Gary Whitt

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

Teaching and Learning

Chair: Dr. John Burton

Dr. Glen Holmes

Dr. Mike Moore

Dr. Greg Sherman

Dr. Thomas Teates

October 24, 2000

Blacksburg, Virginia

keywords: speech cues, multi-modal assessment, context dependent memory, speech recall, speech recognition

# The Effects of Speech Cues on Long-term Memory

Gary Whitt

## (ABSTRACT)

This research examines a possible relationship between intentional memory and possible phonologic cues in the human voice. Specifically, if someone has told us something in the past, does hearing that same voice at the time of recall affect our ability to remember what was said? Also, if voice cues do affect memory, is the effect voice-specific? Since most standardized assessments of student learning and tests of human memory rest their conclusions about human learning solely on non-aural tests, it is necessary to determine if student performance changes with test modality.

Via a computer program, ninety-five adults each listened to a male voice read a one-minute story and were then randomly assigned to take one of three different tests consisting of multiple-choice and fill-in-the-blank items. In the first test, the male voice from the story read all questions and possible answers. The second test used a different male voice to read while the third test was text-only. All tests contained identical content and gave single-modality cues only, text or speech.

Results show no significant difference in long-term recall or recognition with respect to test-modality. Further research in this area is encouraged to determine if conclusions are generalizable to wider populations and hold for longer memory intervals.

## **Acknowledgements**

Thanks to my wife, Margaret, the truest heart I have known and the reason I know for a fact that God loves me more than most people. Thanks also to the golden girl, my daughter Hannah, who loves laughing with her daddy. These two are the reason I love my life.

Too, thanks to my mom and dad who taught me to never, ever give up. What I know of heart, integrity, humor, and love I have learned from them.

Finally, thanks to my committee. I could not have wished for better guidance or more kindness than they have shown.

## Table of Contents

(ABSTRACT).....	ii
Table of Contents .....	iv
List of Figures.....	vi
List of Tables .....	vii
Introduction.....	1
Need for the Study .....	3
Review of Literature .....	5
Introduction.....	5
The Uniqueness of Speech.....	9
Aural Acuity.....	13
Aural Information Processing.....	20
Relevant Studies.....	33
Research Questions.....	51
Methodology.....	54
Purpose.....	54
Experiment.....	54
Participants.....	59
Testing Facility and Equipment.....	59

Procedures.....	60
Data Analysis.....	64
Results.....	65
Recognition.....	65
Recall.....	67
Discussion.....	69
References.....	72
Appendix A.....	78
Appendix B.....	79
Appendix C.....	82

## List of Figures

Figure 1. Recognition item for the Same-voice and Different-voice tests. ....	62
Figure 2. Text-only test format for recognition items.....	62
Figure 3. Recall item from the oral tests.....	63
Figure 4. Recall item from the Text-only test.....	63

## List of Tables

Table 1. Mean Recognition Scores by Test Type .....	66
Table 2. Comparison of Group Mean Scores for Recognition Items .....	66
Table 3. Mean Recall Scores by Test Type .....	67
Table 4. Recall: Summary of Independent Samples t-tests .....	68

## The Effects of Speech Cues on Long-term Memory

When we consciously try to remember something from the past we are said to be using intentional memory (Schacter & Church, 1992). This research examines a possible relationship between intentional memory and human speech. Specifically, if someone has told us something in the past, does hearing that same voice at the time of recall affect our ability to remember what was said?

No one understands why we remember some things and forget others. During the course of any single day our senses dutifully record millions of pieces of information. These sensory perceptions may or may not register in our conscious thoughts but nevertheless, the brain continues receiving them: the look of a face, the tone of a voice, the smell and feel of a breeze. While some would argue that information not specifically attended to is quickly discarded, no one has shown that any information is ever really lost (Anderson, 1995). Improbable as it may seem, the human brain may permanently retain every sensory perception. Whether this is the case or not, it is certainly true that some pieces of information are not accessible on demand. When information is not consciously retrievable we may be said to have "forgotten" it but inability to retrieve information on demand does not necessitate its absence. To the contrary, it is a common human experience to have memories forcibly revived by unexpected stimuli in our environment. The smell of a particular perfume can bring back memories of a dance four decades in the past. The sound of a voice calls up a face not seen for years or the feel of a fabric recalls scenes from a picnic of long ago. Apparently, for each event recorded by the brain there are accompanying pieces of sensory data called "cues" that are recorded simultaneously. These cues: sights, sounds, tastes, and touches encoded at the time of the



event are subsequently linked to the event in memory (Anderson, 1995; Cann & Ross, 1989). When one or more of these corresponding sensory cues is presented to us we may remember that which is associated with the cue in our memory. This related sensory cue, referred to as a retrieval cue, is said to "activate" or bring to mind the specific memory. This propensity of retrieval cues to effect memory is a common human experience but what is not understood, and is therefore the impetus for this research, is the relationship of these retrieval cues to intentional memory. To what degree do sensory cues in our environment affect our ability to remember?

The research done in this area strongly suggests that sensory cues are key to the intentional recall of target memories, the memories we are trying to retrieve. Most informational theorists posit that failure to remember is as much a function of the number and quality of sensory retrieval cues as it is anything else ( Craik & Lockhart, 1972; Eysenck & Eysenck, 1979; Morris, Bransford, & Franks, 1977; Smith, Glenberg, & Bjork, 1978; Tulving, 1983). It may be that we do not remember because we are not given enough of the original environment to reference. Memory studies using olfactory stimuli have shown that smells serve as definite cues to intentional memory and significantly increase long term recall (Cann & Ross, 1989; Schab, 1990). The efficacy of visual cueing to intentional recall is also well documented (Moore, Burton, & Myers, 1996). What has not been examined with any great scrutiny is the ability of the human voice to serve as a sensory retrieval cue to intentional memory. The purpose of this research is to examine the relationship between the human voice and intentional long-term memory. It may be that the ability to remember what was said in the past largely depends on re-creating the aural environment at the time of encoding. Rephrased,

memory may rely on aurally consistent environments. Specifically, if voice retrieval cues are present in the environment at the time of attempted recall, they may change what we are able to remember.

### Need for the Study

Since learning can never be directly assessed it must be inferred from behavioral changes (Anderson, 1995). In academic settings, these changes in behavior usually take the form of written tests. While many schools are becoming more innovative in their use of assessment strategies, standardized tests in America are still predominantly written, multiple-choice tests. Those being evaluated are expected to remain silent until all tests are completed and the only relevant cues available to the participants are located on the paper in front of them. We test this way in America because, in the past, we have had few other economically viable choices. Paper and pencil tests provide the most affordable method of assessing learning.

The use of written assessment is almost equally common in behavioral science research related to intentional memory. Typically, participants are exposed to a prescribed treatment and their learning is assessed with some form of written test irrespective of treatment modality.

According to Gunter and Furnham (1986), same-mode assessments may favor the treatment modality and, to be certain "that this [does] not affect recall an experiment is required in which different methods of testing recall...are compared across different channels of communication" (p. 141). In other words, to increase the likelihood of accurately measuring learning, it may be necessary to test in more than one modality. Determinations of learning, both in the academic and research arenas, could conceivably

contain large errors due to the narrowness of the assessment methods, since demonstrated recall may change with the modality of the assessment.

Given the availability of personal computers and the ease with which aural assessments can be created, what was once impractical research becomes imperative. If some individuals perform better in the aural modality than the written and, if it is possible that assessment of their learning may change depending on the assessment modality, then it is necessary to establish the effect of aural tests on the perception of participant learning. Towards this end, this research may help to answer these basic questions concerning common assessment practice in learning environments. In the academic realm, if modality affects recall, then written assessments of learning may be biased and insufficient in their ability to measure changes in behavior. The corresponding applications of those assessments of learning, implied or expressed: academic readiness, achievement, and intelligence of students may also be based on flawed representations of student knowledge. In behavioral research, since all conclusions must necessarily rest on the accuracy of the data, any conclusions based on written-only assessments of learning may contain confounding margins of error. It, therefore, seems needful to determine if changes in test modality alter perceptions concerning the degree to which learning has occurred. This research, concerning aurally consistent test environments, is a small step in this direction.

## Review of Literature

### Introduction

Aural learning simply refers to the human ability to learn by hearing. The field has a deep history and has produced volumes of research in past years. Indeed, the field of aural learning was so well represented in the research community that several journals were dedicated solely to the topic of how humans hear and learn. However, technological advances in the presentation of visual images and text in recent years have led to decreased activity in the area of aural learning (Tripp & Roby, 1996). It is therefore not surprising that so few studies have been designed specifically to examine the role of aural speech cues in remembering. Experiments of aural learning involving similar aural environments in treatment and test conditions are extremely rare, especially when the aural stimulus is speech (Stankov & Horn, 1980). Most past experiments with speech focused either on the human ability to perceive speech or compared speech with other modes of information presentation (Tripp & Roby, 1996). As a result, much of the available literature in the aural learning field deals with the physiological limitations of the human ear or, more commonly, is related to the processing and recall ability of the aural encoding mechanism versus other modalities. These studies, however, are of limited relevance to speech-cued memory as the relative merits of visual processing and audio-visual processing are not in question. The viability of speech-cued memory must

necessarily rely most heavily on the human ability to perceive, process, and store speech information (Paivio, 1986). Whether another sense: sight, smell, or taste, perceives, processes, or retrieves information better than the aural sense is largely irrelevant. Instead, the ability to hear, process, and store individual speech must form the focus of this review of literature.

Speech carries two major types of information to the listener: semantic and phonologic (Prabu & Hirshman, 1998; Schacter & Church, 1992). Semantic information is concerned with the meaning of what is spoken. Phonologic information deals with the sounds of speech: the rhythm, pitch, and volume of the voice and can convey meaning apart from the semantic meaning of words (Schacter & Church, 1992). For example, a friend can use the word “crazy” to mean several things. The listener would necessarily need to take note of the phonological characteristics contained in the speaker’s voice to determine if he or she is being described as insane and dangerous or witty and fun to be with. While this review of literature must necessarily deal with both types of information contained in speech it is the phonologic information in speech that will form the focus. The terms “speech cues” or “speech characteristics” will refer to the non-semantic information contained in speech, the phonologic information contained in the sound of the words themselves and not in the meaning of the words.

If phonological characteristics of individual speech are to adequately cue memory there must be some representation of this particular speech stored in memory and it must be somehow linked to the target memory (Anderson, 1995). Remember, by definition, a cue is connected to a target memory record and, when attended to or in the environment, activates the associated target memory (Anderson, 1995). For the cue to activate an

associated target memory it must somehow be connected to that target memory. Also, it does not seem probable that the cue could be any speech at all or speech in general. If all speech served as a cue to the target memory then any speech would cue the target memory and the cue would lack utility (indeed such a cue would be extremely bothersome) (Anderson, 1995). Clearly, in order for speech to serve as a viable cue, the speech connected to a target memory must have some differentiable characteristics to set it apart from other speech in memory. The speech record in memory must be, if not unique, fairly detailed, if it is to serve as a link to a specific memory or small set of memories. Therefore, if the utility of a speech cue depends on differentiability from other speech cues, the brain must be able to keep records of characteristics that would differentiate one voice from another (Paivio, 1986). For the brain to adequately separate one voice from another in memory it must contain records of individual speech characteristics; properties of the voice that change significantly enough from person to person to allow differentiation (Prabu & Hirshman, 1998). Therefore, the unique properties of individual voices will be examined in some depth to determine if the individual voice produces enough variation for the brain to differentiate it from other voices.

Next, for memory records of individual voices to exist in the brain the ear must be able to perceive unique variations from voice to voice (Gelfand, 1998). It does not matter if speech characteristics significantly vary between voices if the ear is unable to differentiate them (Gelfand, 1998). It is necessary to determine to what degree the ear is able to differentiate the variations in the human voice. Therefore, the physiology of the

human ear and results of human hearing tests will be examined to decide the abilities of the human ear to perceive unique signals from the voice.

Then, if voices do have unique characteristics and the ear is able to perceive these differences, it is necessary to determine what the brain does with the information. Differences in voice signals must not only be perceived but must also be decoded in some manner for the speech to make sense (Pickett, 1999). And, even if the signals in the sound environment are successfully decoded, it is necessary to determine if the information is retained in the brain and available, or is somehow discarded or made unavailable. It would not seem plausible that aural information could be used as a cue to activate memory if the original aural memory record is not available. Therefore, the literature relevant to the processing of speech stimuli must be examined in an effort to determine what the brain does with unique information contained in the voice. It must be noted here that this determination will be inconclusive, as the topic of aural information processing is no more settled than the questions surrounding any other type of sensory information processing (Moore, Burton, & Myers, 1996). There is simply no consensus concerning what happens to aural information in the brain (Anderson, 1995). However, review of some dominant theories of information processing and several relevant experiments testing these theories may provide information that will allow a determination of the likelihood of speech-cued memory. If speech-cued memory fits logically within the parameters of what most current information processing models predict, then it may be a viable memory phenomenon.

Finally, though there are very few experiments testing for speech-cued recall effects specifically, some past experimental findings may relate closely to the effect

under scrutiny (Schacter & Church, 1992). Experiments concerning state, context, and music-dependent memory may prove particularly informative since they also focus on memory cues generated by some element or elements in the environment other than visual stimuli.

### The Uniqueness of Speech

Speech must be able to provide some means of differentiation. Without these variations in the physical characteristics of the voice, it would be impossible for the brain to differentiate speech (Gelfand, 1998). The greater the variation the more likely the brain will be able to distinguish sounds (Martin, 1997).

To understand the uniqueness of the voice, the nature of the sounds produced by the human speech mechanism must be briefly examined (Pickett, 1999). Speech is produced by the respiratory musculature pushing air through a series of anatomical structures in the throat and mouth (Gelfand, 1998; Pickett, 1999). If no part of the throat anatomy closes or constricts; the diaphragm produces ordinary breathing. However, if any of these structures constrict or close, the air pushed by the diaphragm has nowhere to go and creates a region of high-density air at the constriction (Pickett, 1999). Once the constriction opens, this compressed air is free to leave the mouth. The compressed air leaves the mouth much as a wave leaves the point of impact of a rock thrown into a lake. The number of compressed air regions leaving the mouth per second define the frequency of the sound created and are measured in regions per second or cycles per second called Hertz (Hz) (Gelfand, 1998). The frequency of the voice, high or low, is one its most distinctive characteristics.



Another basic characteristic defining a human voice is the intensity of the voice. The density of the compressed air leaving the mouth determines the intensity of the sound, which relates directly to the loudness of the voice. Intensity is a physically derived parameter and refers to the power delivered by the voiced air per area (Gelfand, 1998). Loudness, on the other hand, is a psychological term, not a physical measure, and deals with human perception of a sound or, in other words, the acuity of the human ear (Durrant & Lovrinic, 1995). For example, if a man and a woman were speaking at the same intensity level, the woman would be perceived as speaking louder than the man. The human ear is more sensitive to intensity level at higher frequencies so, even though the man and woman were pushing the air with the same force, the woman is easier to hear; she is louder (Durrant & Lovrinic, 1995). Loudness is measured in decibels (dB) and defines the softest sound perceivable as 0 dB (Geisler, 1998). The human voice can speak as soft as 0 dB at its minimum and can cause physical pain at its maximum level occurring between 120 and 140 dB (Gelfand, 1998). The loudness varies as a function of ability of the diaphragm to push air.

So, voice frequency varies due to changes in the rates of constriction in the throat and voice intensity varies due to changes in the degree to which the diaphragm pushes air. Since particular characteristics of the human anatomy are responsible for frequency and intensity changes in the voice, it is not surprising that changes in anatomy produce profound changes in voice frequency and intensity. Babies tend to have very high, soft voices (crying not included); small children have lower frequency, greater intensity voices while adults tend to have the lowest frequency, greatest intensity voices. Adults speak with an average frequency of 1900 Hz but individual voices may vary from 50 to

10,000 Hz, a remarkable range (Pickett, 1999). Voice intensity can vary dramatically as well and can produce a large range of intensities.

Frequency and intensity are not the only characteristics of speech that may differentiate the voice, however. According to Pickett (1999),

The speech signal contains a wide variety of acoustic properties. In the right circumstances, most if not all of these properties can function as perceptual cues to the identity of some sound or feature. This variety, or redundancy, is one of the greatest strengths of natural speech because it makes it robust... (p. 199).

Many different qualities can characterize a voice and may serve to uniquely differentiate it (Durrant & Lovrinic, 1995). Glottal stops, Pickett explains (1999), which are produced by the opening and closing of the glottis in the back of the throat, occur at an average rate of 125 per second for men and 200 per second for women but may vary widely in rate from individual to individual. He says also that variations in the constrictions of the throat area cause aperiodic hissing sounds in the letters “s” and “f” which change in character from one person to another. Releases of air pressure after a build-up behind a constriction used to make the letters “p”, “k”, and “t”, also provide unique characteristics in the voice. Variations in rates of speech may also serve as unique cues to speech as the average rate of syllable pronunciation can vary from two to five syllables per second.

The timbre of the voice also deserves special notice as it forms one of the most unique qualities of the voice. Durrant (1995) describes timbre as a quality comprised of the frequencies present in a sound, the relative intensities of each of those frequencies, and then the phase of each frequency. Differences in timbre are easily demonstrated by a piano and a trumpet both playing a middle-C. The notes have the same fundamental

frequency and yet few would have difficulty determining that they were not from the same instrument. The differences in the sounds would be the result of the differing timbres created by the two instruments. Voices behave similarly along this parameter in that two voices with almost identical ranges of frequencies and intensities would be easily differentiable due to the relative phase shifts and frequency intensities of the separate voices (Durrant & Lovrinic, 1995).

While variations in just a few of these properties of the human voice would serve to discern one voice from another, voices are complex combinations of all of these characteristics (Pickett, 1999). Two voices may be similar along one parameter of speech but, as additional characteristics defining the voice are considered, the distinctiveness of each individual voice becomes more pronounced.

Therefore, it seems clear that the human voice may vary substantially with respect to several properties. Voice frequency has a range of 9,950 Hz and an equally impressive intensity range (Pickett, 1999). The frequency of glottal stops may change along with the character of the hissing during production of certain consonants (Gelfand, 1998). The rate at which speech is delivered may fluctuate widely, and the timbre of the voice may vary significantly.

Perhaps the best arguments for the uniqueness of the human voice is its ability to serve as a security measure in new software applications and its admissibility in a court of law as valid evidence of the identity of an individual. Macintosh's operating system, OS 9, advertises that the new software can distinguish the unique qualities of a human voice and determine if there is a match between the person attempting to use the machine and its owner (Apple.com, 1999). The voice must produce unique characteristics unlikely

to be duplicated if it can be used in this fashion as a suitable security device. And, in the Federal Court case United States versus Williams (1978) a stereoscopic voiceprint was allowed as admissible evidence of unique identity.

However, the human ear is not a computer and, while the voice may be unique, it is not clear yet that the human ear has the ability to differentiate the varying characteristics contained in the voice. Both abilities are crucial if speech is to be useful as a cue to memory (Paivio, 1986). The next section, therefore, examines the capabilities of the human ear in an attempt to discern its ability to differentiate voices.

### Aural Acuity

Clearly, the ear must be a remarkable sensory receiver if it is to perceive differences in the fast-paced environment of speech. The physical phenomena comprising speech occur in small fractions of a second, usually hundredths of a second or less (Pickett, 1999). To determine if the ear is equipped to make such minute differentiation among sounds it is necessary to examine the physiology of the ear. Such examination should provide the rationale for any statement concerning the ability of the human ear to discern individual speech. Following the review of the specific anatomy related to hearing, it is then necessary to examine the specific research findings as they relate to aural perceptual abilities in the laboratory. These studies should allow correlation of the physiological expectations generated by the impressive anatomical design of the ear with specific measurements of human performance.

According to Pickett (1999), consonants and vowels take about 100 milliseconds to pronounce and the time necessary to switch between the two is about 50 milliseconds. The glottis will open and close an average of 162 times every second. At its maximum,

the human voice can produce compressions of air that propagate at a speed of 1129 ft/sec resulting in 10,000 compressions striking the ear in one second. Clearly, the speech environment is extremely fast-paced and, if the ear is to serve as an adequate receptor, it must be able to do what all sensory organs do. The ear must absorb the stimulus energy, it must use the stimulus energy to bring about a change in the state of the receptor cell, and it must initiate electrical impulses in the nerve leading from the ear to the central nervous system (Durrant & Lovrinic, 1995). A brief look at the physiology of the human ear should allow determination of the degree to which it can perform all three functions

According to Durrant (1995) the peripheral auditory system begins with an external flap, the pinna, which serves to collect sound waves and focus them toward the ear canal. The sound waves are absorbed by the thin membrane forming the eardrum and cause it to vibrate rapidly. Thus absorbed, the sound energy from the sound waves is transferred to the bones of the middle ear, which oscillate sympathetically with the eardrum. These vibrations are transferred to the fluid of the cochlea and on to the organ of Corti, which cause complex movements of the stereocilia of the brittle inner and outer hair cells lining the interior walls of the cortical membranes. This change in the resting state of the cell produces an increase in the neural firing rate of the nerves. The nerves are always firing but it is this increase in the firing rate that carries information (Durrant & Lovrinic, 1995). The firing of an aural nerve is an all or nothing event taking approximately one millisecond. During this firing (spike) the nerve is absolutely unable to fire again and must return to near its rest potential before it is ready to repeat the process (Durrant & Lovrinic, 1995). This means that one nerve may fire about 1000 times every second. There are approximately 30,000 neurons in the human auditory nerve, so, if each nerve

were to fire at its maximum rate, the ear could theoretically generate 30 million signals to the brain each second (Gelfand, 1998 #34; Geisler, 1998). While it is currently impossible to determine the total number of concurrent auditory nerve cell spikes in one second, if the ear has a tenth the ability formulated above it would still be an impressive sound receptor. Then, the resulting nerve impulses are transmitted along the auditory nerve fibers, along the eighth cranial nerve, and enter the brain at the angle of the pons and cerebellum (Gelfand, 1998). The signal terminates in the brainstem at the cochlear nuclei in the central auditory system (Hernandez-Peon, 1961; Martin, 1997). Nerves then travel from the brainstem to the transverse temporal gyrus in the temporal cortex, a two-part region in the cortex of the brain consisting of the primary auditory cortex and the adjacent auditory association area (Gelfand, 1998). Short neural fibers allow the primary auditory cortex to communicate with the auditory associated cortex, which is connected to the rest of the brain via generous projections to other parts of the brain. There appears to be inter-hemispheric communication between the auditory associated cortex and the rest of the brain (Gelfand, 1998). The signal, once generated in the auditory nerve fibers takes approximately 1 to 2 milliseconds to travel from the periphery to the brainstem (Durrant & Lovrinic, 1995). The signal then takes another 10 milliseconds to travel to the cortex. Compared to the ear's neural firing rate, the travel time of the signal from auditory nerve fiber to the cortex of the brain is slow and could take as long as tenths of a second.

The preceding description of the physiology of the ear and auditory neural pathway to the brain show the ear to be extremely well adapted to receive and store speech stimuli. The highest frequency sounds produced in human speech are around 10,000 Hz, which

would mean that, at this frequency, ten thousand waves per second hit the ear (Geisler, 1998). The ear is, at least theoretically, capable of neural firings much faster than this, even by conservative estimates. Theoretical performances of the ear, however, do little to establish its efficacy as a speech receptor. It is necessary to examine the results of hearing performance experiments to determine if the ear performs as well as its physiology suggests it might.

Tests of the ear's perceptive abilities reveal a remarkable ability to differentiate sound, especially in the region of speech. The reader will recall that normal speech occurs between 200 and 10,000 Hz with average speech somewhere around 1900 Hz. Studies of hearing perception indicate that while the ear can perceive sounds anywhere between 20 to 20,000 Hz, it is most sensitive to changes in frequency and intensity between 100 and 2000 Hz ((Buser & Imbert, 1992; Durrant & Lovrinic, 1995; Geisler, 1998). The ear functions best in the region of speech (Geisler, 1998; Gelfand, 1998). In this range of speech the ear is able to distinguish a change in frequency as small as 3 Hz (Buser & Imbert, 1992). The human ear also has a remarkable ability to differentiate temporal differences and can distinguish individual clicks separated by 1 to 2 milliseconds (Durrant & Lovrinic, 1995). The ear is also extremely sensitive to changes in sound intensity and, while this ability varies with frequency, the most intense sound the ear can perceive is approximately 10 million times more intense than the least intense sound it can detect (Gelfand, 1998). The least intense sound the human ear can detect is 0 dB and the most intense sound the ear can differentiate is about 120 to 140 dB (Gelfand, 1998; Martin, 1997). Human hearing is most accurate near the middle of this range, between 40 and 50 dB (Geisler, 1998).

It is necessary to note before going further that tests of hearing typically use only one ear, as this methodology simplifies experimentation considerably. In tests involving both ears, hearing ability increases approximately by a factor of two and can be said to sum in measure of frequency differentiation, intensity variation, and temporal differences (Durrant & Lovrinic, 1995). Indeed one of the most impressive feats of the human ear involves the binaural ability to localize sound. Unless a sound source is directly in front of a listener there will be a slight difference in arrival times of a single sound wave at the right and left ears. This difference in arrival time is at its maximum when one of the ears is pointed directly at the sound source. In this orientation, with the listener's ear pointed directly at the sound source, the sound wave will strike the leading ear and register. To be perceived by the other ear it must travel around the head and strike the other ear. The difference in sound wave arrival time in this extreme condition is 0.67 milliseconds. The ear's ability to temporally differentiate arrival times of the sound waves is so acute that it can detect a shift of the sound source as small as  $2^\circ$  (Durrant & Lovrinic, 1995). The ability of the ear to localize sound within  $2^\circ$  simply serves to highlight its extreme temporal acuity.

Apparently, the ear is so well adapted to the perception of speech that it is possible to compress speech significantly without major loss of speech comprehension. Tripp and Roby's (1996) summary of the compressed speech research in the last 30 years indicates that while normal speech occurs between 12 and 150 words per minute, humans only experience an average comprehension loss of 6% between 225 and 325 words per minute. As the speech rate is increased to between 325 and 425 words per minute there is only a 14% drop in comprehension. In almost all of the studies quoted there was no significant



loss of comprehension for compressed speech slower than 282 words per minute: speech that is twice as fast as the fastest normal speech. These effects were constant across gender, content, and age. Abilities did drop for people over 60 and for children under 12 years old. However, it does seem clear that under normal circumstances the ear processes information at a rate far below its maximum capability.

The ability to hear and comprehend much more information than is normally present during single-person speech leads naturally to speculation concerning the ear's ability to differentiate signals from multiple sources presented simultaneously. While there is considerable debate concerning the ultimate capacity of the various modality channels to process information (Moore, Burton, & Myers, 1996) a study by Moray may offer insight regarding hearing capacity. He used split headphones to allow participants to hear two consonants per second on one, two, three, or four channels (Moray, Bates, & Barnett, 1965). While he, like many others, found a definite preference for single channel processing and a total capacity of about 8 letters, his study demonstrated an effect of aural memory similar to that recorded in studies of visual memory. After hearing a total of eight letters in two seconds on four separate channels, participants were asked to record in random order the letters they had heard. During additional trials with the same conditions Moray et al. decided, after the participants had heard all the letters, to quickly direct their attention to one specific channel and asked them to record the letters they had heard from this channel. Recall improved by 25 % and led the researchers to suggest that people were processing more information through the multiple channels than they had time to record. Without the specific channel prompt it appeared that participants had heard much less than they actually had processed. The experimenter's prompting allowed

the participants to demonstrate that they had actually encoded much more information from the multiple channel presentation than they were able to demonstrate with non-cued free recall. While agreeing with previous studies that the limit of the aural processor was about 8 items and hearing performance is best with one channel, his study does suggest that the ear can process multiple channels and that cued aural recall may surpass non-cued aural recall of speech.

Clearly, the ear is very capable of differentiating words and voices using changes in frequency, intensity, rhythm, nasality, voicing, frication (the breath associated with a letter), and place-of-articulation (Gelfand, 1998). All these speech phenomena occur well within the operational parameters of the human ear. There is also a well-established neural pathway from the ear to the cortex of the brain, which suggests that the brain is able to process these auditory stimuli.

Therefore, the research is very clear concerning the unique characteristics of speech and the ear's ability to successfully differentiate these signals. Not only does the United States Federal Court recognize a person's voice as a unique individual characteristic, it also recognized as permissible evidence, in the case of United States versus Smith (1989), a blind man's ability to accurately identify an individual by listening to his voice.

While debate may exist concerning how variations in speech are produced or how the ear receives such minute variations, there is little argument that these aural cues are produced, received, and monitored by the brain (Moray, Bates, & Barnett, 1965). What is not clear is what happens to these aural cues inside the brain. This topic has occupied researchers interested in information processing for decades and calls into question the entire nature of information storage and retrieval (Anderson, 1995; Paivio, 1986; Tulving,

1983). There is much debate, many theories attempting to describe how sensory information is processed, and only marginal consensus among researchers.

### Aural Information Processing

An understanding of what happens to aural information once the brain receives it is key to understanding the brain's ability to reference aural memories when presented with an aural cue. It is clear that speech contains enough unique information to allow differentiation between individual words and individual voices. It is also clear that the physiology of the auditory system is well equipped to perceive this unique information. What is not clear is what happens to this unique aural information. Is it stored in the brain? If aural information is not stored in the brain it can be neither a cue to memory nor part of a memory (Paivio, 1986; Tulving, 1983). If stored in memory, is the aural information subject to decay? If it decays, at what rate does the information decay? Finally, if not decayed can the original aural information serve as a link to the target memory upon cue? The answers to these questions will help determine if aural signals are able to serve as memory cues and these may be found, if anywhere, in a brief examination of the research concerning information processing. The review of this research will not be exhaustive, as the specifics of information processing in the brain have been a focus of cognitive psychology for the past three decades (Tulving, 1983). Only a brief summary of some of the major theories will follow. This summary will also not provide definitive answers to the questions posed above, for the answers are still debated. What a review of the major information processing research will do is to help decide if it is reasonable to suspect that the unique characteristics of speech may serve as memory cues. In an effort to make this determination it is necessary to examine several

major theories of information processing: short-term memory, dual coding, depth of processing, encoding specificity, and activation theory. The main focus, as the theories are discussed, will be on the research findings concerning the processing of aural stimuli. Because information processing may not be observed directly, it must be inferred indirectly from changes in participant behavior, generally performance on some test of memory. It is these memory tests and the resultant conclusions made concerning aural learning that may prove most relevant to the discussion of aural memory.

There are several versions of the role of short-term memory in processing information (Deutsch, 1975; Moore, Burton, & Myers, 1996; Anderson, 1995). However, there appear to be some commonalities concerning the processing of aural stimuli and these will form the focus of the review of this theory. Short-term memory theory states that sensory information is moved into short-term memory and, if rehearsed sufficiently, is transferred to long-term memory. Short-term memory is verbal in nature while long-term memory is semantic in nature and information may only move from verbal-based short-term memory to semantic-based long-term memory if it is sufficiently rehearsed (Anderson, 1995; Deutsch, 1975). This rehearsal allows any nonverbal information to be coded verbally and sent to long-term memory. Information that is not rehearsed sufficiently in short-term memory is lost. “The key feature of this theory is the proposal that short-term memory is a necessary halfway station between sensory memory and long-term memory” (Anderson, 1995, p. 161). Obviously, if rehearsal is necessary for information to be incorporated into long-term memory then it is not reasonable to expect the brain to store acoustic information that is not specifically rehearsed. Characteristics differentiating one voice from another, while perceived by the ear, are not usually

rehearsed and, therefore, would not be stored in long-term memory. Clearly, if this is the manner in which short-term memory operates; long-term speech-cued memory is a doubtful phenomenon at best (Deutsch, 1975).

While this general view of the role of short-term memory accounts for several experimental findings, it fails to accommodate the data from others (Deutsch, 1975). Specifically, Anderson (1995) recalls the experiments of Glenberg, Smith, and Green, Craik and Lockhart, and Neisser to show that increased rehearsal does not improve long-term recall in all cases. In these experiments rehearsal did not significantly increase recall but instead recall seemed to depend on the attention given to the target items. Items rehearsed passively, regardless of frequency of the rehearsal, were not remembered significantly better than items that were not rehearsed at all. In other words, items that entered sensory memory without special attention had the same chance of being remembered as items that received repeated passive rehearsal (Tulving, 1983). Deutsch (1975) chronicles the experimental data that seem to indicate that the buffer for non-verbal information in short-term memory possesses “a systematization and specificity rivaling any found in verbal memory studies” (p. 108). Specifically, Cole (in Deutsch, 1975) found that memories for acoustic attributes of words were available at least eight seconds after word presentation, a full 7 seconds longer than predicted by most proponents of short-term theory. Murdock and Walker (in Deutsch, 1975) also concluded that pre-linguistic auditory information was retained in memory for at least five to ten seconds. Furthermore, Pollack (in Deutsch, 1975) found that acoustic information could prime recall of test words embedded in white noise as long as 15 seconds after initial presentation. Crossman (in Deutsch, 1975) found evidence of this acoustic memory as

much as 40 seconds after initial word presentation. Finally, Deutsch argues that if long-term memory is solely verbal in nature the phenomenon by which we remember unlabeled pieces of music for long periods of time is inexplicable. Often people can identify very short sequences taken from much longer pieces of music, and, if one note is played incorrectly, identify the mistake. She sights this ability as irrevocable proof that the mind stores highly specific acoustic information in long-term memory.

Therefore, the theory of short-term memory does accommodate the possibility that acoustic attributes of speech are stored in long-term memory. And, if it is possible that acoustic attributes are available in long-term memory, it is also possible that they may be used as cues to other memories.

While resolution of the different tenants held by short-term memory theorists are certainly beyond the scope of this review of literature, it is now apparent how questions concerning the encoding of sensory perceptions could lead to differences of opinion concerning pre-long-term processing. While elegant in its simplicity, original short-term memory theory has trouble explaining some of the experimental data concerning how humans process sensory information. Dual-coding theory attempts to explain many of these memory phenomena and is the next information processing model relevant to the discussion of the viability of speech-cued memory.

Dual coding, first posed by Alan Paivio in 1971, reacts to the view held by short-term memory theorists “that performance in memory and other cognitive tasks [is] mediated by processes that are primarily verbal or linguistic” (1986, p. vii). Paivio suggests people process sensory information in two basic ways. Visual images are processed through a channel dedicated to the processing of visual images while verbal or

linguistic information is processed in another separate channel. The two channels have different characteristics, different abilities and processing rates, and are designed to specifically encode their respective stimuli. Therefore, the ring of a telephone or the sound of laughter could have their own representations in permanent memory as separate sounds but would also have alternative visual representations in permanent memory as the image of a phone or the face of a person laughing is evoked. Individual characteristics of speech then would be coded in both processing channels and would be tied to the memory traces of the sounds of the voice and the image of the speaker (Paivio, 1986). “A basic assumption in the dual coding approach to retrieval is that trace contact in both cued recall and recognition is based on a similarity match between the pattern of information evoked by the retrieval cue and the information in the memory trace” (Paivio, 1986, p. 146). In other words, to best evoke the memory of the telephone ring or the laughter it would be necessary to reproduce as closely as possible the environmental cues present at encoding. Dual coding would seem then to accommodate the possibility of speech-cued memory exceptionally well. If dual coding accurately reflects the way humans process the unique characteristics contained in speech, speech characteristics unique to individual voices are coded in two ways and have twice the probability of serving as cues to target memories encoded at the same time.

A series of experiments by Pelligrino, Siegal, and Dhawan tested the plausibility of dual coding theory (1975). The researchers presented words or pictures in triads to participants for short periods of time. According to dual coding theory, the pictures are processed by both channels, image and verbal, while the words are processed in the verbal channel only. Theoretically, if the subjects are interfered with during the time

normally used for processing information, their recall is only affected if the interference occurs in the same channel as that responsible for processing the information. Pellegrino et al. used a backwards-counting task to interfere with the verbal channel reasoning that, because the participants were speaking the results of their mental computations, the speech would interfere with the verbal processing of words. They alternated the verbal interference task with an embedded-figures task designed to interfere with the visual channel only. In a later experiment they combined the interference tasks to occupy both processing channels (Pellegrino, Siegal, & Dhawan, 1976). Their results seemed to indicate the existence of two separate processing channels, one responsible for the processing of visual images and one for the processing of verbal information. Verbal recall had significantly decreased when participants had counted backwards before being asked to recall words and their recall of images had been much higher than their recall for words. Also, image recall had dropped while verbal recall remained unaffected during performance of the visual interference task. However, Pellegrino et al. failed to explain why visual recall had dropped 25% during the verbal-only interference task. According to Paivio's dual coding theory (1986), visual recall of images should have remained unaffected by the verbal interference task as the images were processed in the verbal and imaging channels. Since the imaging processor had not been interfered with, recall for images should have remained unaffected by the backwards-counting task.

Burton and Bruning (1982) refined these experiments to include a no-interference condition and changed the information triads to include speech as well as images and text. They also monitored recall for the presentation type of the nouns. Surprisingly, the researchers found that the dual-interference task (counting backward and finding



embedded figures) did not interfere with image recall as predicted by dual coding. They concluded that this “finding seriously undermines the dual coding interpretation of modality differences and favors a semantic-based interpretation” (p. 67) of the results. Burton and Bruning also called into question the ability of the interference tasks to interfere specifically in the target channel. Finally, they noted that the only data from the experiment supporting dual coding was the difficulty the participants had remembering the presentation type for printed words.

Dual coding experiments have proven inconclusive but have served to highlight some of the differences in the way the brain encodes images and verbal information (Gunter, Furnham, & Leese, 1986; Furnham, Benson, & Gunter, 1987; Steele, Lewandowski, & Rusling, 1996; Winters, 1996). If dual-coding theory is a plausible explanation of how people process verbal and pictorial information, then the resulting dual representation of the unique characteristics of human speech in memory would make speech-cued memory even more plausible.

Dual coding, however, is not the only information processing theory to offer a viable alternative to short-term memory models of remembering. The insistence of proponents of short-term memory that all sensory stimuli be verbally encoded encouraged the development of the dual coding model. Necessarily, dual coding could explain the non-verbal nature of some memories. In similar fashion, the experiments of short-term memory showing that recall was not especially dependent on rehearsal gave rise to the depth of processing theory.

Formulated by Craik and Lockhart (1972), depth of processing insists that "...the memory trace can be understood as a byproduct of perceptual analysis and that trace

persistence is a positive function of the depth to which the stimulus has been analyzed” (p. 671). Experiments show that the time spent rehearsing information is not always related to how well the information is recalled (Christianson, 1997). Indeed, Anderson (1995) relates the story of a Professor Sanford who estimated he had read the same mealtime prayer at least 5000 times over a 25-year period and still could not repeat it from memory. Craik and Lockhart attempted to explain this apparent discrepancy by proposing that it is the depth at which information is processed that determines the persistence of a memory. Depth is achieved by repetition at a certain level or by short intervals at very intense levels of processing. Depth of processing may also occur without specific attention being given to the processing information as the sensory stores are “pre-attentive” (Neisser in Anderson, 1995) and depend more on the maximum rate and capacity of the particular modality processor involved. Craik and Lockhart further suggest, contrary to short-term memory theory, that information is encoded almost literally into long-term memory. As a result, differences in retrieval performance are due not only to the quality of processing but also to differences in test type and the available cues given at time of recall. They reason that since people are able to recognize music and voices after long periods of time, specific phonemic information, as well as semantic and imaginal information, is available as long as it has not decayed or been replaced by subsequent information. The sensory information is not only available but can affect to a large degree the success of the retrieval effort. Therefore, differences in recall performance may not only reflect different levels of processing but could indicate insufficient sensory information present in the retrieval environment. The information readily accessible in long-term memory is not sufficiently activated by the test

environment to serve as cues to the target memory. It is at this point that depth of processing places real emphasis on the appropriateness of the retrieval environment and suggests that test type has as much to do with recall as the strength of the memory trace (Morris, Bransford, & Franks, 1977). This possibility will be examined later in the review of literature as the effects of testing environments on recall ability are discussed. For now, suffice it to say that the depth of processing theory of Craik and Lockhart easily accommodates the possibility of speech-cued memory.

While depth of processing theory may accommodate the possibility of speech-cued memory particularly well, it has some fairly severe criticism centering on the nebulous nature of the term “depth”. As a result, Craik and Lockhart have been accused of circular reasoning: information is deeply processed if it can be remembered for a long time and information can be remembered for a long time because it has been deeply processed (Eysenck, 1978). The inability of the theorists to adequately define what may constitute “depth” has led to a certain loss of momentum for those inclined to test the limits of the depth of processing model (Eysenck & Eysenck, 1979). Certainly, it is difficult to test limits where the theoretical parameters are not clearly defined. Regardless, testing of the depth of processing model of information processing has continued to the present in an effort to define what parameters or properties may be included in the definition of deep processing of information. According to the initial theoretical statement by Craik and Lockhart (1972), long-term records of the acoustic characteristics of the encoding environment, such as sensory memories of the voice, may constitute part of that depth.

Yet another theoretical framework of information processing needful of consideration is Endel Tulving’s theory of episodic memory. In this memory paradigm

several kinds of memory exist and interact: episodic memory records the actual experiences of everyday life while semantic memory holds the interpretations and meaning of those experiences (Tulving, 1983). According to Tulving, many of the confusing results of the preceding decades of memory research are due to a failure to realize that there are different kinds of memory and these memories respond in different ways to stimuli. Tulving proposes that

what a person remembers about an event, and how well, depends not only on the nature of the event and its encoding, but also on the conditions prevailing at the time of its attempted recollection, particularly that component of the conditions that we refer to as retrieval information...what a person recollects about an experience is not determined by the memory trace of that experience. The memory trace is only one important co-determinate of recollection; the other equally important one is the retrieval information that is used in the process of actualizing the trace (p. 4,5).

More than any other information processing theory mentioned heretofore, the theory of episodic memory postulates that recall may have as much to do with the quality and quantity of retrieval information as it does with the actual event memory. Encoding specificity, as he terms it, infers that characteristics of the environment at the time of encoding are always an important part of the event. Since the environmental perceptions are such an integral part of event memory it is critical that these environmental cues be present during recall. Failure to retrieve memories of an event therefore, do not necessarily imply an absence or weakness of the memory trace but may instead reflect the sparseness of the retrieval environment. As voice characteristics are an integral part of

any event involving speech, leaving them out during a retrieval attempt would constitute a breach of the encoding specificity principle. Tulving suspects that many of the discrepancies in memory research reflect the differences in the recall environment. Since the sensory stimuli contained in the initial learning environment are inextricably linked to the target memory, the degree to which the recall environment resembles the encoding environment determines what is remembered and how it is remembered. The theory of episodic memory then also accommodates the possibility of speech-cued memory and, indeed, seems to infer it.

Tulving's ideas concerning episodic memory and the importance of retrieval cues to recall imply a certain ability of retrieval cues to "actualize" the memory trace (Tulving, 1983). Other theorists have also suggested the ability of related memories, memories encoded simultaneously, to actualize or, more often, activate other memories. This paradigm permeates several theories discussed above and is prevalent enough in information processing theory to warrant separate discussion.

Activation theories consider memories to consist of individual memory records that have encoded in them information of past experiences, and cues that are somehow connected to these individual records (Anderson, 1995). While these theories may use different language to describe how memory works, the basic tenants of activation theories rest on the ability of cues to evoke or activate memory records. For a cue to activate a memory record it must either be present in the environment or in a rehearsal system and it must be connected strongly to only some memory records. A memory record may be activated by the presence of several weak cues in the environment or it may be activated by one very strong cue (Anderson, 1995). The Rescorla-Wagner theory attempts to

quantify cue strength and, although developed to estimate the strength of association between conditioned and unconditioned stimuli, is useful when discussing the parameters that define the strength of a cue (Anderson, 1995). One construct put forth by the theory is that of association strength. This parameter reflects how closely the cue relates to the specific memory record. For example, if it is permissible to borrow from Treasure Island (Stevenson, 1955), “earring” may serve as a very weak cue to the memory of Long John Silver for Jim Hawkins while “wooden-legged pirate” may be a very strong cue. The latter cue would have very high association strength to the memory record of Long John Silver. Another construct used by the Rescorla-Wagner theory to discuss cue strength is maximum level and has to do with the strength of the memory record (Anderson, 1995). For example, any traumatic event might have a very high maximum level and would be capable of having strong cues linked to it (Christianson, 1997). Too, any intensely emotional experience would have a good chance of becoming a strong memory. The last construct useful in defining the possible parameters of cue strength is rate of learning, or how quickly the cue becomes associated with the memory record. Studies have shown that some cues, like blood, are extremely fast acting (Christianson, 1997).

Therefore cue strength may depend on how closely the cue and memory record are linked together; the nature of the memory record to which the cue is tied (important memories have potentially strong cues), and the rate at which the cue is associated with the memory. As the cue becomes available in the environment or is somehow loaded into a rehearsal system the memory record becomes more available for recall. This process is called priming and, if the cue is of sufficient strength, significantly facilitates recall. Anderson (1995) describes a priming experiment of Kaplan’s in which subjects were

given a set of riddles to solve at home. One particular subject was stuck on the puzzle “What goes up a chimney down but cannot go down a chimney up?” Kaplan arranged a call to the subject’s home by an anonymous female caller who asked if she had left her umbrella in the subject’s office. The man reported the answer to the puzzle shortly after the call was made. Kaplan decided that placing relevant cues in the environment had primed the objects as solutions to the problems and had activated the objects in the subjects’ memories.

Activation theories, usually embedded in other theories such as short-term memory, dual coding, depth of processing, and episodic memory, provide an underlying theoretical framework for the existence of speech-cued memory. The human voice, from a behaviorist point of view, must certainly have great cue strength potential. In early development, the child must quickly learn to discriminate nurturing voices from voices not likely to provide nutrition and protection. Warnings of danger issued to pre-lingual children must be interpreted by tone of voice alone. Hearing is one of our first useful senses and must certainly shoulder the responsibility of interpreting the environment before the sight becomes keen or before the child can move about freely (Warren, 1982). The potential costs of failing to recognize warning sounds are very high, especially for young ones, and it is probable that they pay close attention to the voices of those responsible for their care to get cues concerning the relative safety of their environments. Later in life, speech continues to inform and protect as tones, pitches, and intensity levels are evaluated with incredible alacrity. Voice signals must be sent and received all day long and those failing to correctly interpret the signals contained in those voices often regret their inability or their inattention. Clearly, if cues do serve to prime and activate

memory records, speech cues may be among the earliest and perhaps the most predominant cues in use.

### Relevant Studies

Theoretical viability however is no substitute for empirical data. Having examined the creation, reception, and processing of speech cues it is necessary to examine the experiments relevant to the phenomenon in an effort to ascertain the plausibility of speech-cued memory effects. These experiments fall into three distinct categories: incriminating studies, state dependent experiments, and context dependent experiments.

Several incriminating studies seem to indicate that aural memory is poor and that aural cueing must be limited at best. The basic inferences of these studies are that aural abilities are significantly less than other sensory modalities and that aural information is usually subject to quick decay. However, these studies may contain basic flaws that render them unable to speak to aural ability in general and speech-cue strength in particular. State dependent experiments involve retrieval of information and changes in some internal state of being by chemical or natural means. While not specifically testing aural cueing, they may provide insight into cueing in general. Context dependent experiments focus on the effect of changes in the external environment on memory retrieval and may also provide information concerning the ability of the mind to encode environmental attributes for use as cues. Both types of experiments deal with the processing of environmental cues in some fashion and the utilization of those cues to activate memory. The most relevant experiments are those dealing with sound in general, the voice in particular, and form the final section of the experimental literature relevant to speech-cued memory.



It is not uncommon to find experiments decrying aural retrieval abilities (Furnham, Benson, & Gunter, 1987; Gunter, Furnham, & Leese, 1986; Winters, 1996; Pellegrino, Siegal, & Dhawan, 1975; Hill, 1994; Steele, Lewandowski, & Rusling, 1996). Usually in these experiments information is presented in some combination of modes: aural, visual, and audio-visual. Recall is then tested in some fashion and usually shows that aural recall is inferior to at least one of the other modalities. These experiments, however, often use different tests of memory: spoken free recall, written free recall or written cued recall. These tests of recall share two characteristics: they present no aural cues relevant to encoding and they tend to be semantic in nature. Most information processing theories (all of those examined previously) establish the importance of retrieval cues common to encoding and retrieval environments and yet most experiments attempting to establish the comparative strengths of the aural channel do not provide aural cues during retrieval. The encoded aural information is not able to cue target memories because the aural cues are not present in the recall environment and, if present, are wasted.

Perhaps Furnham and Gunter have been the most active in the study of the effects of modality on recall ability (Moore, Burton, & Myers, 1996). Through a series of experiments they have shown the superiority of written text over aural-only and audio-visual presentations of information (Furnham et al., 1987; Gunter et al., 1986). While very thorough in their experimental design they did comment that

“since the method of testing involved reading questions and writing responses, it conveyed an advantage to subjects in the print condition since they were already functioning in a written mode, whereas subjects in the other conditions had to switch from an audiovisual or audio-only mode to the written mode. To ensure

that this did not affect recall an experiment is required in which different methods of testing recall (e.g. spoken responses to spoken questions) are compared across different channels of communication” (Gunter, Furnham, & Leese, 1986, p. 141). Morris, Bransford, and Franks tested the effect of test type on recall performance in an experiment requiring different types of processing (1977). One group was asked to semantically process a list of words by deciding if a word was congruent with the meaning of a given sentence. Another group was asked to process words in a list phonologically by determining if a target word rhymed with part of a given sentence. When the participants were tested using a standard recognition test, those who had semantically encoded the information significantly outscored the other group. However, when the test was modified to have participants identify words that rhymed with the original words, not surprisingly, the phonological group scored significantly higher than the semantic group. The test type appeared to vary the inferences about which group had learned more. The phonologic task required use of encoded information that had not been required during the semantic test. Therefore, Morris et al. conclude that “different modes or levels of processing may simply allow people to acquire different sorts of information, each of which may have the potential for being equally strong and durable (as revealed by appropriate testing situations)” (p. 520). The tests, as Tulving (1983) suggests, may not measure learning as much as they measure the degree to which the testing environment matches the encoding environment.

Decades of experiments involving aural ability reveal that processing of aural information is somehow different than other types of information processing. Backward counting hampers aural processing more than image processing (Pellegrino et al., 1976;

Pellegrino et al., 1975; Burton & Bruning, 1982; Burton, 1982). Long-term aural memory differs from long-term image memory but they are both subject to decay with time and they both improve as relevant cues are given (Anderson, 1995). Aural processing may require more time than other types of processing and the capacity of the aural processor may be different. Clearly, there is ample evidence that aural processing is different in some aspects from other types of information processing. However, experiments that make statements concerning the relative abilities of the aural processor without ample consideration of the appropriateness of the testing environment should be examined with caution (Morris et al., 1977; Gunter et al., 1986). Clearly, an evaluation of the research concerning the efficacy of speech-cued memory must look to experiments with appropriate cues in the retrieval environment instead of examining the preponderance of visual-only experiments that are so common.

One set of potentially useful experiments concerning available cues at the time of encoding and retrieval are those examining state dependent learning. The state dependent experiments suppose two types of environments containing potentially useful retrieval cues: the external environment that provides information to the senses and the internal environment that provides its own cues to memory (Eich, Stillman, Weingartner, & Gillin, 1975). If either of these change from encoding to retrieval, the number of relevant cues available to activate a target memory decreases. Eich et al. effected a change in internal state to examine the effect of these changes on long-term memory. They proposed that

"Common sense and a wealth of empirical findings (e.g., Pan, 1926; Tulving & Thompson, 1973) lend support to the notion that one's memory for a given

perceptual event may be influenced by the contexts within which that event is initially experienced and subsequently recalled.... Few would seriously quarrel with the notion that completeness of recall depends to a large extent on the effectiveness of retrieval cues, however one defines an effective cue. But an effective cue must also be accessible in the context of the recall situation if it is to facilitate recall, and it is this process of accessing effective retrieval cues that might conceivably depend on the kind of contextual interaction noted above" (Eich et al., 1975, p. 416).

To test this notion, Eich et al. changed the internal states of some participants by having them smoke marijuana before attempting to memorize a list of words. The 48-word lists were divided into 12 different taxonomically related categories. The retrieval tests were given four hours after initial encoding and consisted of a written free-recall section and a written recall test with the 12 category types listed as cues. While the researchers concluded that marijuana seemed to reduce overall recall ability, both free-recall and cued recall of information were significantly improved by keeping participants in the same state they were in during encoding. Subjects intoxicated during encoding recalled significantly more when intoxicated than when sober. Eich et al. did not attribute their findings to specific cues but did propose that cue continuity between encoding and retrieval was key to recall.

Goodwin et al. (1969) devised a similar experiment to test the state-dependent effects of alcohol and met with similar findings. While drunken people remember less in general, they recall more in the state they were in when first exposed to the information to be remembered. The experimenters also reasoned, as did Eich et al., that learning

“depends for optimum expression on restoration of the original condition in which learning was acquired” (p. 1358). They attributed the state-dependent memory effects not to changes in registration and retention but to impairment in retrieval caused by the change in state. Of interest, the researchers also offered their clinical observations of alcoholics, as further evidence of state-dependent memory effects. Often, events occurring while the subjects were drunk could not be recalled when they were sober, even after detailed descriptions of the events were offered. However, memories returned (e.g. the location of hidden money or liquor) when they had returned to the intoxicated state.

The experiments of state-dependent memory may have much to say concerning cued memory. The researchers in both experiments however limit their conjectures concerning causality to suggestions that changes in the cue environments are somehow responsible for the state-dependent memory effects. Some cue available in one state is missing in another state. Whether people hear, see, smell, or taste differently when intoxicated, generating different cues than when sober, is a matter of mere speculation and is not attempted by the researchers. If cue continuity is important to retrieval, and a change in state alters that cue somehow rendering it unrecognizable, then recall depending on that cue would necessarily decrease. Regardless, the experiments do raise interesting questions concerning the process of cue generation and cue access. Research examining the specific effects of changing environmental cues on retrieval may also offer valuable insight and it is this area of the research that is examined next.

While state-dependent memory research focuses on changes in internal states, studies with context dependent memory are concerned with the effects of changing the

cues in the external environment. A number of experiments have been conducted in recent years that alter the quality and quantity of cues available in various modalities. Some of these studies change the total cue environment, the sum of the cues available in all modalities, while others purposefully mediate one modality alone. This review of context dependent memory will move from the broad to the specific, looking first at experiments changing more than one modality, next at non-aural single-modality studies, and concluding with the experiments that manipulate aural cues only. An examination of this research should allow a reasonable determination of the relative importance of cue-environment continuity to retrieval of information.

Any discussion of context dependent memory would be inadequate without at least a precursory examination of Godden and Baddeley's underwater memory tests (1975). In this classic study of recall nineteen divers memorized a list of words underwater or on land and attempted to recall the words in a same-environment context or in a different-environment context. While admitting that the less than rigorous experimental surroundings may have influenced their research, Godden and Baddeley took some steps to control for possibly confounding variables such as physical state during recall and aural acuity in different environments. The researchers found a significant state-dependent recall effect but were content to suggest that while the experiment did not conclusively demonstrate a context-dependent memory effect, it did raise questions concerning the importance of relevant environmental cues. Reacting to suggestions that the context dependent effects were due to a general effect accompanying any change in testing environment, Godden and Baddeley conducted a second experiment to test for this effect. All divers were asked to swim around in the water and dive to a depth of 20 feet

between encoding and retrieval to eliminate possible effects due to physical exertion. Again they found highly significant differences in recall performance that they attributed to memory context dependence.

One of the earliest experiments involving context dependent memory in an academic setting was Farnsworth's 1934 study of the effect of changing testing locations on student test scores (Farnsworth, 1934). On several instances he divided his classes and had them test in either a familiar or unfamiliar setting. While three of the four tests yielded better student scores for those left in familiar environments, Farnsworth judged the differences in test performances to be non-significant. It may be important to note that Farnsworth did not record the modality of the test or the type of test questions asked and it is therefore impossible to ascertain the objectivity of the test scores. Without this determination of grading reliability, his conclusions must necessarily remain in question. Regardless, Farnsworth's study, while non-rigorous in design, served to spur further research concerning context dependent memory effects.

Eich also tested the effects of context change on recall ability in an experiment designed to test the cue strength of room objects (Eich, 1985). In two rooms, students heard 24 nouns spoken aloud one at a time. Half of the participants in each room were asked to attach the spoken words to objects in the room. The other half in each room were asked to picture the image represented by the test word. Eich found that only those participants who had been asked to attach an object in the room to the spoken noun showed the classic context dependency he expected. The other participants, those asked to imagine only the spoken test items, showed no significant change in recall ability when tested in a different room. Eich interprets his results to indicate that "context dependent

effects in recall were not simply enhanced by asking subjects to create integrated item/context images, but were clearly contingent on these instructions...” (p. 768). It is not clear why Eich determines that context dependent effects were contingent on his instructions since his experiment does not provide a no-cue condition. Subjects were instructed either to pick a room object to attach to the spoken image or imagine the item with no other option available to them. In view of the other experiments of context dependency it is logical to assume that, lacking specific instruction, people naturally use environmental characteristics to serve as memory cues. By speaking directly to the participants he is directing their attention in two ways. They must first attend to his voice (and perhaps his face) and must also imagine an image. The participants who link the spoken word to an object in the room have their attention directed to a specific item in the room. It is logical to assume that the characteristics of the room items are encoded at approximately the same time as the target item and are linked in memory. For those asked to merely imagine the image, to see it in their minds, their attention is directed to the experimenter’s voice (aural cues) and diverted from other objects in the room. It is probable that individuals look at the same place while imaging the spoken nouns, effectively linking all the images to one or at best a very few physical items in the room. Eich, by his methodology may have indirectly prevented his imaging-only group from attending to the physical environment. For the group using room objects as cues (the integrated group) he is causing their attention. If this is true, if environmental cues are necessary to memory and Eich inadvertently hinders acquisition of these cues for one group, that group should have decreased recall when compared to the integrated group no matter where they are. This is exactly what Eich finds. Regardless of which of the two



rooms they begin in, when the participants do not switch rooms, the integrated group remembers 45% of the words in the list while the imaging-only group can recall only 26% of the words. For the switched-room context, the integrated group falls to 31% recall (suggesting the importance of the room cues) but still outperform the imaging-only group by 7%. Eich appears to be correct concerning the lack of effect of room on recall for the imaging-only group. However, if it is true that his methodology may have prevented cue acquisition for the imaging-only group, then his interpretation of the data may be inaccurate. Rather than examine the effect of context dependency, he may have measured the importance of cue acquisition and found these environmental cues very important to recall indeed.

Also relevant to context dependent memory are the experiments investigating olfactory-cued memory. It is a common human experience to have an unsought memory forcibly recalled by a familiar scent (Cann & Ross, 1989). It is such a common experience that Cann and Ross suspected smells of serving as strong cues to memory. Male college students, in the presence of a pleasant or an unpleasant odor, were shown 50 slides of the faces of college females and asked to rank them for attractiveness. After a 48-hour delay they were asked to recognize the slides they had seen previously. This test of recognition was given in the presence of either of the same two odors. The researchers found that regardless of the attractiveness ratings, males recognized the faces significantly better when in the presence of the same odor as they had been exposed to during initial presentation of the slides. Cann and Ross concluded "...odors might operate effectively as context cues, providing a more general and diffuse association to the

material to be retrieved.... The characteristics of a context apparently can become associated with details encoded in that context" (p. 93).

Schab (1990), in a series of experiments, also tested the ability of olfactory stimuli to cue implicit memory. Male and female undergraduates were asked to generate antonyms to 40 adjectives in a list but did not explicitly try to remember the words in the list or the antonyms. While generating antonyms half the subjects were exposed to the smell of chocolate while half were exposed to only the smells in the lab. One day later, after returning to the lab for a second session, the subjects were asked to write down all the antonyms they could remember. Only half of the participants were exposed to the smell of chocolate in the recall condition. Regardless of the odor in the room, during initial information encoding and retrieval, the participants were asked to think about and imagine the smell of chocolate. Therefore the four groups had either not smelled chocolate at all, smelled it only during encoding or recall, or smelled it during both sessions. Schab found recall to be significantly greater when the olfactory context contained chocolate. Smelling chocolate during encoding only or during recall only produced no significant changes in recall. Only the group that had chocolate odor present at both sessions demonstrated significantly higher recall than the no odor-no odor condition. Imagining chocolate had no effect on recall. Schab decided "...contextual stimuli are encoded along with target information on learning and serve as memory cues to the target information at retrieval" (p. 654).

So, state-dependent recall studies conclude that recall may be affected by changes in cue perception. Studies of environmental context changes seem to also indicate a relationship between cue availability and recall performance while studies of olfactory

cueing efficacy conclude that a single modality cue, an odor, may serve to activate memory. The circumstantial evidence for the ability of the voice to serve as a cue to memory is not insubstantial. To make a better determination of this ability however it is necessary to review the few experiments concerned with memory in an aurally consistent environment. Specifically, it is necessary to review the research testing retrieval of information with background music or speech in the encoding and testing environments.

Smith conducted two experiments to determine if background music serves as a cue to memory (Smith, 1985). In the first experiment he tested 54 undergraduate students' ability to remember words written on index cards. The index cards were shown at five-second intervals and were accompanied by Mozart, jazz, or quiet in the background. Students were then asked to write all the words they could recall. Students, unaware that they would be asked to recall the same words, returned 48 hours later to the same laboratory to repeat the test. Smith found that students who heard the same music in the background during encoding and recall remembered significantly more words than did the other groups. Interestingly, when it was quiet during both sessions, recall dropped. The increase in recall, therefore, was not necessarily due to context dependency. If this had been the case recall with quiet during both sessions should have been the same as recall with the same music. Smith concluded that, rather than contextual continuity producing the observable increase in memory, the music must be serving as a cue to memory.

To make sure that the effect was not music dependent Smith designed a second experiment using music, white noise, and quiet and used two input modalities: speech and slide-projected words. He was interested in determining if the memory effect he had

observed in his first experiment was common to only one input modality and if it was dependent on a particular melody or style of music. Smith duplicated the results from the first experiment by finding significantly better recall in both modalities for the same context condition. Again, while this was true of music and white noise, when both sessions were quiet recall significantly decreased. Too, the music-dependent memory effect was even more pronounced in the aural mode. As a result, Smith concluded, “the result is a facilitative one caused by the reinstatement of contextual cues...” (p. 601). Seemingly, the aural cues were somehow tied to the words in memory and were being used to recall the information at time of retrieval. When these sounds were not present, especially when the words had been spoken instead of visually presented, recall decreased significantly.

Balch, Bowman, and Mohler (1992) also tested music dependent memory by playing background music to college students asked to judge the quality of 24 written words in a list. The experimenters were attempting to determine if music could serve as a cue to memory of the words. All 240 students heard either fast jazz, slow jazz, fast classical, or slow classical music while they were looking at the words for 250 seconds. During the written free recall tests (given 30 seconds and 48 hours after initial word presentation) the participants heard either the same music as they had before, different music, or no music. Students hearing the same music during encoding and retrieval recalled significantly more words than did students hearing different music. There were no differences in recall, however, for students who heard no music during the immediate recall test and no significant differences in recall for any group after 48 hours. The researchers noticed that the most pronounced differences in immediate recall were for

different tempo music, when the music changed from slow to fast or fast to slow. Changes in type of music, classical or jazz, had less pronounced effects. Following a second experiment to test the effect of changing music tempo on immediate recall the researchers concluded that there was a significant music-dependent effect. They reported that changing the tempo of background music between encoding and retrieval significantly reduces immediate recall of words. However, no music and same music contexts produce no significant differences in immediate recall and there are no significant effects of background music on long-term memory.

Both experiments find that music may serve as a cue to immediate memory and, while Smith (1985) reports long-term effects of background music on memory, Balch et al. (1992) do not. The experimenters, however, offer a possible explanation of the failure to duplicate Smith's findings. They reason that Smith's methodology affords the participants an opportunity to generate the words in the list in the initial presentation context by writing the words down. The participants generate the words while the original musical context is still present. Balch's experimental methodology affords no such additional processing of the words and, reasons the experimenters, may account for the lack of long-term music-dependent effects in their experiment. Smith, by allowing his participants to write the words down in the initial music condition, allows extra processing in that environment and more time to create music cues to the words in the list.

A second characteristic of the music-dependent memory experiments may bear strongly on the examination of the efficacy of speech-cued memory. In the experiments discussed above music is a secondary stimulus only and is not part of the information

contained in speech. If music, which is in the background of the aural environment, is able to serve as a cue to memory, it is not unreasonable to suspect that speech characteristics may serve as even stronger cues. It is this review of the research concerned specifically with speech cueing that forms the final part of this review of literature and is most relevant to the viability of speech as a cue to memory.

Smith, Glenberg, and Bjork (1978), interested in examining the effects of physical surroundings on recall, may have inadvertently demonstrated speech-cued memory. In the second of five experiments, 24 undergraduates, tested in pairs, listened as a female voice read a list of 45 word-pairs a total of four times (Room A). The words were weakly related semantically (e.g. car-body, smell-cabbage) and were presented via cassette recorder at 4-second intervals. The participants were instructed to remember the words and their cues, as they would be tested on them at the end of the day's session. After all the words had been heard, participants were given 15 cue words aurally and were asked to write down the associated nouns that accompanied the cues at presentation. The following day the same procedure was used except that a different list of 45 word pairs was presented via a slide projector at 3-second intervals. The visual presentation occurred in a different setting than the aural presentation (Room B) and used 30 new word pairs and 15 word pairs common to the aural presentation from the day before. At the end of the slide presentation the participants were shown 15 new cues and were asked to write the corresponding words. On the third and final day, participants were taken to one of three locations and given the 75 cue words (60 unique and 15 common to both lists) in different modalities. Participants taken to Room A, the room where they heard the cassette on the first day, were tested as cues were read by the same female voice via a

cassette player. Those tested in Room B, the room where they were shown slides on the second day, attempted to recall words as cues were projected via slide projector. The third group was taken to Room C, an environment distinctly different from the previous two. In Room C cues were shown on an index card and read simultaneously by a male experimenter.

While the experiment did show a definite context dependency, interpretation of main effects were difficult due to an interaction with presentation modality. Words from the aural list were remembered significantly better in the aural testing site. Words from the slide-projected list were remembered significantly better in the room where cues were projected. However, in Room C where the cues were projected and spoken by the experimenter, memory depended on modality. Recall of slide-projected words was high while memory for spoken words was significantly lower. While the researchers did not attempt to explain the modality-recall interaction, they did note that it was pronounced. By switching voices from the female of Room A to the male experimenter of Room C, the researchers may have removed most of the aural cues needed by the participants to remember the words. Recall of visually presented material was as good in room B as it was in room C. Recall of aural material in Room C was almost half that of aural recall in Room A. While the sample size is small and other factors may have influenced the data, Smith and his cohorts may have demonstrated the effect of switching voices on long-term purposeful recall. If this is the case, the experiment may suggest that speech cue generalization is limited and that speech cues may be effective in a narrow range defined by the encoding voice.

Testing this suggestion of minimal cue generalization are the experiments of Schacter and Church (1992). The PRS (pre-semantic perceptual representation system) view of aural perception holds that there are subsystems in place to process information about aural form and structure that are separate from those subsystems assigned to meaning and associative properties. Theoretically, semantic encoding tasks and structural tasks should produce different memory records subject to experimental differentiation. Accordingly, the researchers presented 24 spoken words to undergraduates using three male and three female voices. As each word was spoken the listeners were asked either to place the word in one of four categories (semantic encoding) or to rate the pitch of the speaker on a four-point scale of high to low (structural encoding). After presentation the subjects were distracted for approximately four minutes and were then asked to either recognize the word embedded in white noise or to identify the word as it was spoken clearly as one that had appeared in the list or a new word. While this test did find significantly higher recognition and identification for words that the participants had heard before and greater recall for semantic encoding over structural, it did not show any significant differences in recall for different voices. These results, according to the researchers, concur with similar findings of Jackson and Morton (in Schacter & Church, 1992) who also used words embedded in white noise. While speech cues were affecting retrieval ability, the gender of the voice did not seem to matter.

To determine if voice gender affected identification or recognition for incidental recall (since it had not for intentional recall), Schacter and Church conducted a second experiment duplicating the first except that participants were asked to record the first word that came to mind during the identification task. Results duplicated the first



experiment. Recall was significantly higher for previously heard words but voice did not seem to matter. Too, semantic encoding produced greater recall than structural encoding.

Suspecting that the encoding tasks may have interfered with each other, Schacter and Church designed yet another experiment to determine if different voices could produce different recall for incidental memory tasks. The semantic task was changed to require participants to determine the number of definitions a word might have. This task was deemed to have negligible structural processing. The structural task was to rate how clearly the speaker presented each word. Following a distractor task that typically lasted from three to four minutes, participants were given the first syllable of each of the 24 words and asked to complete the word. Surprisingly, recall for both types of encoding, semantic and structural, showed significantly better recall for the same-voice condition. When the gender of the voice changed, ability to complete the word stems dropped significantly.

Intrigued by the possible deleterious effects of white noise on voice-differentiated recall, Schacter and Church designed a last experiment to determine if they could duplicate their results of the previous experiment using white noise to disguise the voices. Predictably, as they had shown in their first two experiments, when the voices were embedded in white noise there were no differences in incidental memory due to voice changes. Incidental recall was still significantly higher for spoken items but there were no voice effects. The researchers concluded that speech-cued memory effects were only possible when additional distracting noise was held to a minimum.

Other factors may explain the relative rarity of speech-cued memory effects in the above experiments of incidental and intentional memory. As always the appropriateness

of the encoding task is critical and, until changed by the experimenters, the structural encoding tasks required less processing than the semantic tasks. Also, the maximum number of syllables heard at a time was less than a hardy sample for the encoding of a particular voice pattern. One is forced to speculate if the lack of voice effects is due to the meagerness of the sample. Rarely do people speak for mere syllables at a time only to be superceded by other voices speaking their few syllables. The experimental context does little to emulate reality and may not support generalization. At any rate, the experiments do show definite speech-cued memory effects for incidental and intentional memory.

These experiments, concerned expressly with the ability of the voice to activate memory, combined with the experiments examining state-dependent, context-dependent, and music-dependent memory studies, seem to suggest that information in the encoding environment is used to form cues to memory. And, these memories may then serve to activate memories encoded simultaneously with these environmental cues. Specifically, several experiments provide specific data that suggest the brain, if not distracted, naturally uses aural cues to remember.

### Research Questions

The literature is clear concerning the unique character of the voice, the competence of the ear to differentiate voices, and the ability of the brain to remember these individual voice characteristics. Almost all plausible information processing theories of the past thirty years recognize the importance of cue abundance to retrieval posing that, the more prevalent the environmental cues at time of encoding, the greater the probability of retrieval. The greatest body of the literature concerned with context dependency in general and speech-cued memory specifically seems to indicate that voice cues, if present

at encoding, do improve retrieval. In question, however, is the degree to which the brain uses individual speech characteristics as cues to activate specific target memories. Some researchers theorize that the brain uses only the semantic information from aural encoding as cues (Schacter & Church, 1992). Others hold that it is the unique characteristics of individual human voices that differentiate memories of speech and that speech cues, individual memories of the particularities of human voices, are used to activate memory. There is some research (albeit scarce) to indicate that individual voices may serve as cues to memory. According to researchers active in the area of information processing and aural learning, there is need of additional research to better determine the extent to which the brain uses individual voices to generate cues to target information. While the little research that has been done in this area suggests that the brain does not use individual structural phonological information to activate memory, the research may have limiting flaws. The most relevant experiments have usually embedded the voice in background noise minimizing its unique qualities and have used only a few syllables of any one voice to activate memory. The appropriateness of the experimental phonologic encoding tasks are also in question and may not adequately produce the aural processing of normal active listening environments. Also, much of the research has been aimed at incidental memory and has little to say concerning the efficacy of speech cues to activate memory in purposeful learning situations. Clearly, there is a need for additional research to determine if the brain uses voice characteristics in intentional retrieval tasks. Towards this end this research attempted to answer (for information presented via a single voice and for intentional long-term memory) these specific questions:

1. Does speech-cued recognition differ from text-cued recognition?

2. Do different voice-cues produce differences in recognition?
3. Does speech-cued recall significantly differ from text-cued recall?
4. Do different voice cues produce differences in recall?

## Methodology

### Purpose

This research sought to determine if the phonologic characteristics of the human voice serve as long-term memory retrieval cues to information encoded by voice. If this main effect existed for spoken information, long-term memory for those given voice-only cues should differ from long-term memory for those given text-only cues. Also, the study examined the effects of different speech patterns on long-term speech-cued memory. If the aural cues were given in a different voice than that used to encode the information, long-term memory should differ from that for the voice that provided the original cues.

### Experiment

This research used a true-experiment design with three different posttest groups. Participants each heard the same story read to them in a male voice. The story was of fixed duration and was followed by a fixed period of wait time to ensure that target information was in long-term memory. Next, participants were randomly assigned to one of three possible posttests. Each posttest was made of identical multiple choice and fill-in-the-blank questions. To maximize test reliability and validity, each question was taken verbatim from the story. The question omitted a key word, a fact relevant to the story, and required the participants to either choose the missing word from a list of five choices or type the missing word in a text box. One test presented the questions and possible answers using the same voice as that from the story. Another test used a different voice of the same gender to present the same set of questions and possible answers, while the last test again presented the same questions and possible answers in text-only format.

Explanatory variables were long-term recognition and long-term recall for aurally presented information while the independent variable was posttest modality. Both recognition and recall were tested: recognition required fewer cues than recall and should have been more sensitive to the effect of interest while recall would help determine the robustness of the main effect, if present (Anderson, 1995).

To maintain treatment integrity and maximize internal validity, all information: instructions, examples, information, and posttest questions were delivered via a computer program created with Macromedia's Authorware 4.0 in a controlled environment. This increased the probability that the original presentation of information was the same for all participants and that outside distractions were held to a minimum. The laboratory environment also allowed for the control of several other possibly confounding variables such as ambient noise level, light level, static and dynamic factors in the physical environment, and temperature.

Computer volume was pre-set to ensure that all auditory information was clearly heard but not so loud as to irritate or distract participants or distort sounds. Given the semantic and phonologic robustness of the human voice, no special augmentation or manipulation of the voices in the experiment was necessary (Gelfand, 1998; Pickett, 1999). Voices were recorded using Syntrillium's Cool Edit Pro at 44.1 kilohertz, 16-bit resolution (mono) and played back through Labtec stereo headsets. The relative high fidelity of the original recordings and testing of the playback devices by the experimenter helped to ensure audible, understandable auditory output. Each computer and headset combination used to collect data in the experiment was pre-tested each time by the experimenter before participants arrived to ensure sound audibility and program functionality.

Given the extreme acuity of the human ear, special attention and care were given to creation of the sound files used in the experiment. When creating text, consistent font, size, style, color, and placement will achieve uniformity. As long as normal rules of grammar and word usage are followed, there are limited non-semantic cues to differentiate one word from another. In the aural environment, any undue or unnatural emphasis on specific words or sounds might predispose it to memory and would have the same effect as using all capital letters or a bold font in a text environment. For this reason, care was taken to minimize the probability of inadvertently providing aural cues not contained in the treatment. For example, if one vocal choice was louder than the other four it would have predisposed a participant to select it. Too, if only one choice was without a "break", an aural glitch caused by splicing sound files, it would have unduly attracted the attention of the participants. The same care was taken with the recording of the story. No undue emphasis of words or phrases was allowed except those common to the normal speech patterns of the recorded voice.

The instructions, delivered via computer, were consistent across trials and negated variability of experimenter instruction to participants. To avoid possible variable practice effects, all participants were allowed to practice the tasks expected of them repeatedly before the test began.

The story and corresponding questions had a Flesch Reading Ease of 81.9 with a Flesch-Kincaid Grade Level of 4.9 as established by Microsoft Word's readability tools. Participants, therefore, experienced minimal effects due to unfamiliar grammar or vocabulary in the instructions, story, and questions.

Treatment time was held constant at one minute by the computer program so that participant exposure to the information was constant and controlled. The story the participants heard was delivered at the same rate, in the same voice for each participant and participants could not advance until the story was completed.

Since the experiment attempted to measure long-term recognition and recall, which both decay over time, it was necessary to control for time. From the moment the treatment began, pacing of the experiment was removed from the participant's control. The story lasted one minute. Following the story, a one-minute interval was built into the program followed by test questions that had identical timing characteristics across test type. While the participants taking the written tests may have experienced more wait time than those taking aural tests, they were not allowed to advance to the next question prematurely. In an attempt to relieve participant frustration levels at this possible inefficiency in the testing environment, the rationale for the wait was explained in the pre-treatment instructions. All participants, therefore, had to remember information for equal amounts of time. Without the timing constraints, it was entirely plausible that those participants taking the written posttests would have moved faster through each question and, therefore, would have been required to remember information for less time before attempting recognition or recall.

Response mechanisms were also held as constant as possible across test types. Each test type had identically phrased questions with identical answers (the written questions and responses were transcriptions of the aural questions and responses). The physical movements necessary to answer each question were also held constant: recognition questions required a single mouse click to answer and recall questions required a typed



response. Also, to negate any possible effects due to variation in the visual test environment, the visual response mechanisms, text boxes for typed responses and radio buttons for recognition questions, were held constant across test type and were placed in near-identical locations on the computer screen.

Same-gender voices were used to control for possible gender effects related to attention or motivation. Since it was possible that all participants might have found one gender or the other inherently more interesting or believable, male voices were used to ensure the same attention and motivation were generated by each voice and that any memory differences were due to posttest differences and not treatment effects.

Possible confounded results due to existing schema and knowledge were controlled by using a neutral schema and a fictional story with fictitious information.

Work by Furnham and Gunter (1987b) suggested a possible interaction between modality and time of day concerning recall with participants remembering certain kinds of information significantly better in the morning. Even though their work dealt with short-term memory, to alleviate any possible effects due to this possibly confounding variable, the experiment was performed at different times throughout the day.

Random assignment of subjects was also used to control for possible effects due to age, gender, ethnicity, typing skill, computer-use anxiety, and test anxiety.

Concerning motivation, it was necessary to note that different individuals respond differently to motivational tactics with no one technique ensuring maximum performance. While a competitive environment may have encouraged attentiveness of some participants, it may have raised test anxiety in others and inhibited performance. Therefore, participants were given optional access to the results of their test performance.

Those wishing to review the performance record did, while those intimidated by such a procedure were able to ignore it. Correspondingly, the possible negative effects of experimenter presence were minimized by the experimenter remaining seated in the corner of the testing facility while those participants encouraged by experimenter presence still received this stimulus. Too, participants were greeted warmly and thanked before their participation to encourage their attention to the procedures.

Finally, to decrease possible diffusion of treatment effects, participants were asked to refrain from discussing results of the experiments with each other. Those representing a single cluster were tested simultaneously or as near the same time as possible to reduce the amount of time available for discussion of the experiment among themselves.

### Participants

The sample population was a cluster sample of convenience consisting predominantly of graduate students at Virginia Tech taking summer courses on campus and university staff. Most of the 57 women and 38 men participating were between 20 and 30 years of age, were of European descent and were pre-service or practicing public school teachers. Sixty-four of the 95 participants considered themselves visual learners. Many of the subjects were attending the same class at the same time and were in close proximity to the testing facility.

### Testing Facility and Equipment

The experiment was conducted in the Center for Instructional Technology Solutions in Industry and Education in War Memorial Hall on the Virginia Tech campus. The lab had a single entrance with a heavy wooden door minimizing ambient sounds and housed

approximately 15 computers with Intel-based processors ( $f \geq 200$  MHz), various brands of 16-bit sound cards, and Labtec LVA 8520 stereo headsets. The computer monitors had 17-inch screens set to a resolution of 1024 by 768 pixels. All lab computers had Macromedia's Authorware 4.0 and ran the experiment in full-screen mode.

### Procedures

Participants were welcomed, asked not to discuss the experiment with anyone who might have been asked to participate, were told they could leave quietly when finished, and were shown to a seat in the computer lab. Up to six subjects participated at one time but they were seated at least ten feet from each other and were facing towards the perimeter of the lab. Subjects were then instructed to begin the experiment when they were ready by donning the headsets, adjusting them so that they felt comfortable, and double-clicking the experiment icon on their computer screen. The experimenter took a seat in an unused corner of the lab by the door to prevent intrusion or disturbance.

The Authorware program began with a welcome screen and displayed a text-only explanation of what the participants were to see, hear, and do (see Appendix C). They were asked to indicate name, age, gender, and ethnicity (which were saved to an external file for later reference). The rationale for the design of the experiment was explained to each participant so that all assessed learning was intentional and all facets of the interface were understood. The participants read that they would soon hear a story. The story would have no pictures or text, merely a blank screen. Participants also read that after the one-minute story they would be shown movable jigsaw pieces that were simply meant to give them something to do while a minute elapsed. At that time they would be invited to assemble the pieces of the puzzle into a single picture. Then they were instructed that

they would be asked different types of questions about the story they were to hear so they should pay careful attention and try to remember what was read to them in the story.

Thus informed, participants were allowed to practice the entire procedure ahead of them with a short story and example questions. They were instructed that, while they would not be given all the types of questions demonstrated in the examples, they would be given at least two types of the questions shown to them. Following the examples, participants indicated their readiness to begin the real story by clicking a button. From this point on, the program advanced automatically using pre-set timings. The story (see Appendix A), read to participants in a male voice, lasted one minute. Following the story, the jigsaw pieces appeared and were assembled at participant discretion. After one minute, the jigsaw puzzle disappeared and the 16-question test began.

The participants were randomly assigned to take one of three possible tests. These tests were identical in content (see Appendix B) and varied only in the modality of the test delivery: two were oral while the last was text-only. The Same-voice test read all questions and possible answers to the participants using the same voice as that heard in the story; there were no text cues provided (see Figure 1 and Figure 3). The Different-voice test, visually identical to the first test, read the same questions and answers but used a different male voice; again, no text cues were provided. The Text-only test presented the same questions and answers in text-only format, without the aural cues (see Figure 2 and Figure 4).

The 16 test questions alternated between recognition and recall items. Recognition items on the two oral tests appeared as shown in Figure 1.

Question 1

Repeat the Question

a

b

c

d

e

**Figure 1.** Recognition item for the Same-voice and Different-voice tests.

To hear questions repeated, participants clicked the "Repeat the Question" button.

Clicking the lettered buttons "a" through "e" played the possible answers. When participants thought they had identified the missing word from the story they clicked the radio button beside their choice. Changes in choice negated earlier responses.

Recognition items on the Text-only test emulated the oral tests as closely as possible (see Figure 2).

Question 1

She was \_\_\_\_\_ when I was born.

a. 45

b. 58

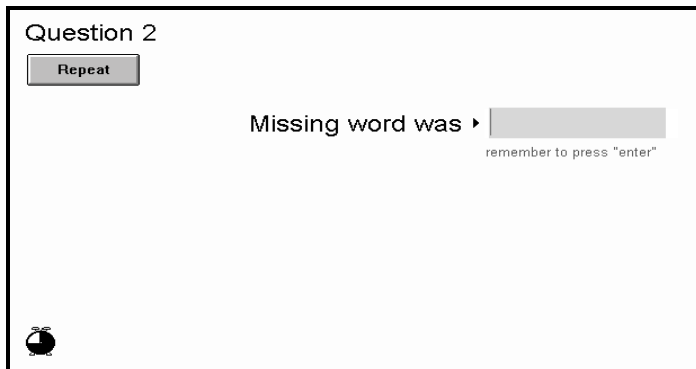
c. 63

d. 66

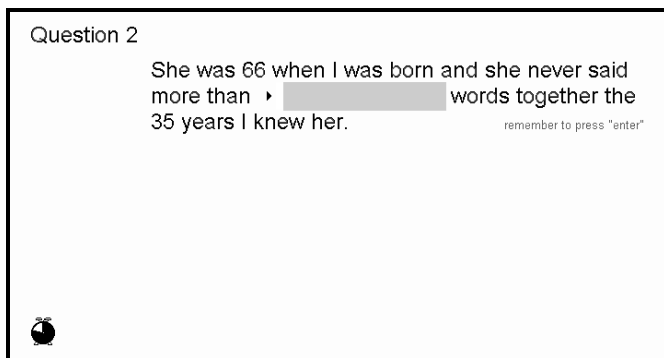
e. 71

**Figure 2.** Text-only test format for recognition items.

All recall questions consisted of statements taken verbatim from the story with a single word omitted. Participants were asked to type the omitted word into the text-box provided. For the two oral tests, the recall item format is shown in Figure 3 whereas recall items from the Text-only test are shown in Figure 4 .



**Figure 3.** Recall item from the oral tests



**Figure 4.** Recall item from the Text-only test.

The average time given to answer each question was 30 seconds. Once a question disappeared from the screen it could not be revisited and participants monitored their time constraints via the half-inch square sweep-second clock icon in the lower left part of the screen (see Figure 4 above).

After the sixteenth question, participants were thanked via a text screen, and were asked to indicate their study preference when preparing for a test: reading the class notes

or having the notes read to them. Next, they were given the option of viewing their performance scores on the multiple-choice portion of the test. Upon completion of the test participant descriptive information and each response to the 16 questions were written by the program to a spreadsheet for later analysis. All recognition items were automatically scored while recall items were manually scored by the experimenter.

### Data Analysis

Data were imported from the spreadsheet into SPSS 9.0 after variables were created for test type, total score for recognition, and total score for recall. Non-directional t-tests for independent group means ( to guard against Type II errors) were used to compare the total recognition and total recall scores for the three different posttest groups. An alpha level of .05 was used for all statistical measures.

Given the large sample size of each of the three different groups ( $n \geq 30$ ), the central limit theorem dictates that the assumption of normality was probably met. The assumption of homogeneity of variance was tested using Levene's statistic. The most tenuous assumption in the analysis of the data (since the samples were cluster samples of convenience), that of independence of observations, was checked by strict procedural guards to participant interaction: instructions to refrain from discussing the experiment with others, multiple-participant testing to limit the available population to discuss the experiment, and minimizing the time to test the entire cluster to reduce the amount of time that interactions could occur.

## Results

This experiment posed four research questions, two which dealt with recognition of spoken information and two which were concerned with recall of spoken information. Accordingly, results of the experiment are grouped along these two specific parameters and are preceded by the research questions to which they correspond.

### Recognition

The first research question asked if speech-cued recognition differed from text-cued recognition. Comparisons of the Same-voice group mean to the Text group mean showed no significant differences in recognition ability according to cue type: speech or text. The Same-voice group correctly recognized 5.94 words on the average which varied from the Text group mean of 5.48 by a mere 0.46 questions (see Table 1 and Table 2 below). Changing the voice used on the test also produced no significant differences in recognition performance. The mean number of words correctly recognized on the Different-voice test, 5.77, was not significantly different from recognition on the Text test, 5.48, and differed by only .29 questions. Therefore, the null hypothesis for the first research question was accepted and it was concluded that speech-cued recognition did not differ from text-cued recognition.

The second research question asked if different voice-cues produced differences in recognition. Mean comparisons of the Same-voice group and the Different-voice group were not significantly different and differed by 0.17 questions (see Table 2). Changing the voice used on the test did not alter recognition performance and the null hypothesis was accepted once again.



In summary, group means from Table 1 showed no significant variation in recognition performance with respect to test type and, therefore, allowed us to answer both research questions negatively. The type of cues provided on the tests: same-voice, different-voice, or text had no significant effect on the ability of participants to recognize key words from the story.

**Table 1**  
Mean Recognition Scores by Test Type

Test type	N	Mean	Std. Deviation	Std. Error Mean
Same-voice	34	5.94	1.48	.25
Different-voice	30	5.77	1.07	.20
Text	31	5.48	1.26	.23

**Table 2**  
Comparison of Group Mean Scores for Recognition Items

Compared Groups	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig.	Mean Difference (2-tail)	Std. Error Difference
Same-voice vs. Text	.181	.67	1.34	63	.19	.46	.34
Different-voice vs. Text	.595	.44	.94	59	.35	.29	.30
Same-voice vs. Different	1.158	.29	.54	62	.59	.17	.33

Levene's statistic was insignificant for each group comparison (see Table 2) and, therefore, all assumptions of homogeneity of variance between groups were met ( $p > .05$ ) as were assumptions of normality ( $n \geq 30$ ).

Participant feedback tended to validate the assumption that the test questions were indeed measuring long-term memory and were not confounded by test construction or

procedural design. No participants, when questioned, indicated that questions or answers were ambiguous, misleading, or confusing. Rather, they tended to note their inability to remember what had been said in the story.

### Recall

The third research question asked if speech-cued recall significantly differed from text-cued recall. Relevant data were summarized in Tables 3 and 4 below. The Same-voice group managed to correctly recall 3.82 words from the story (see Table 3). When compared to the Text group mean of 3.45, they differed by .37 word; hardly a significant difference (see Table 4). Switching voices on the test also failed to produce any significant variation in recall. The Different-voice group mean of 3.86 was only .41 larger than the Text group mean. Clearly, speech-cued recall scores did not significantly differ from text-cued recall scores and the null hypothesis was accepted.

The last research question asked if different voice cues produced differences in recall. The Same-voice group mean of 3.82 words recalled correctly differed only slightly from the Different-voice group mean of 3.86, leading to acceptance of the null hypothesis once again. Clearly, different voice cues produced no significant differences in recall.

**Table 3**

#### Mean Recall Scores by Test Type

Test-type	N	Mean	Std. Deviation	Std. Error Mean
Same-voice	34	3.82	1.59	.27
Different-voice	29	3.86	1.53	.28
Text	31	3.45	1.52	.27

**Table 4**Recall: Summary of Independent Samples t-tests

Compared Groups	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tail)	Mean Difference	Std. Error Difference
Same-voice vs. Text	.140	.71	.96	63	.34	.37	.39
Different-voice vs. Text	.092	.76	1.04	58	.30	.41	.39
Same-voice vs. Different	.438	.51	-.10	61	.92	-.04	.39

To be counted as correct, the typed answers had to approximate the actual words taken from the story to within a few letters. Gross misspellings (more than 3 letters) were counted incorrect as were synonyms of correct responses.

Due to the nature of the test, the reliability and the validity of the questions to measure recall ability necessarily rested on the method of test construction. Each question was taken verbatim from the story and required participants to supply the missing word. As with the recognition questions, participants indicated little confusion concerning what was expected or how to answer but rather attributed their failure to respond correctly to their inability to remember what was said during the story. Also, since the questions were difficult to confuse with other material from the story, wrong answers tended to reflect wild guessing rather than the interspersing of one story fact with another.

All assumptions of homogeneity of variance between groups were tested and met using Levene's statistic (see Table 4). Tests of normality were assumed based on the tenants of the central limit theorem ( $n \geq 30$ ).

## Discussion

Neither recognition nor recall of semantic speech varied significantly with test modality. For long-term recognition, the average number of correct responses, of eight possible, was 5.94 for the Same-voice group, 5.77 for the Different-voice group, and, 5.48 for the Text-only group. Measurement of long-term recall yielded relatively similar results: the Same-voice group averaged 3.82 correct of eight questions, the Different-voice group 3.86, and the Text-only group averaged 3.45 correct. For spoken information, long-term recognition and recall on tests with speech-only cues did not differ significantly from recognition and recall on tests with text-only cues.

It is also important to note that neither recognition nor recall was voice-specific. Mean recognition scores for the Same-voice and Different-voice tests differed by only .17 questions while mean recall scores on these two tests differed by a mere .04 questions. Changing the voice used on the test did not significantly change test recognition or recall performance.

Paivio (1986), Tulving (1983), Craik and Lockhart (1972), and Eysenck (1978), all theorized that the retrieval environment is crucial to memory and that what is remembered may depend on the recall environment as much as the encoding environment. Smith (1985) corroborated this by demonstrating that certain music, played in both encoding and retrieval environments, aided recall. The work of Schab (1990) and Cann and Ross (1989) showed that the same was true of smells. However, these cues merely *accompanied* the target information at time of encode and did not make up part of the target itself. While these aural and olfactory cues may be associated with the target in

memory, and may serve to activate the memory, they were not the target memory. In this experiment, the aural cues were actually part of the target memory. The data clearly show, for recognition and for recall, that retrieval when only given aural cues was virtually the same as retrieval of the information with text-only cues. From this information, it can be concluded that the brain was accessing the target information, which was aurally encoded, as efficiently with text cues as it did with the aural cues. It cannot be concluded from this experiment that the brain made a dual representation of the word at time of encode: one aural and one textual. It can be stated, however, that both sets of cues, aural and visual, provided approximately equal activation of the memory when presented in the retrieval environment.

The results of this experiment have some fairly far-ranging ramifications. One area that is impacted by this study concerns past research using text-only assessments of memory. Much of what we know concerning memory rests alone on research conducted using text-only assessments of recognition and recall. Typically, memory for lists of common words would be tested using a test with no aural cues at all, only visual cues. This experiment indicates that conclusions concerning recall and recognition would not have changed had aural cues, similar to those used in this experiment, been provided during assessment. Providing aural tests would not have changed the conclusions regarding memory and, fortunately, past experimentation is not inviolate based on procedural oversight with respect to test modality.

Another area possibly impacted by this research is that of common assessment practice in education and training. Standardized tests and classroom assessments generally provide text-only cues. This research implies that, for recognition and recall of

common speech, this method of assessment is as accurate as an aural-only test and that test results will not change with test modality.

An additional application of the results of this experiment is that current instructional designers, creating instructional modules for computer-based environments, may not need to spend valuable time and resources creating aural-only assessments when text-only assessments produce the same results.

Finally, this research indicates that recall and recognition of spoken information can be tested as satisfactorily using aural-only cues as with a written test. Within the parameters of this experiment, aural assessments yield approximately the same results as written tests and should accordingly guide instructional practice.

## References

- Anderson, J. R. (1995). Learning and Memory: An Integrated Approach. New York: John Wiley & Sons, Inc.
- Apple. (1999). OS 9, [on-line]. Apple Computer, Inc. Available: [http://www.apple.com/macos/pdf/MacOS9\\_DS-a.pdf](http://www.apple.com/macos/pdf/MacOS9_DS-a.pdf) [Nov 21, 1999].
- Balch, W. R., Bowman, K., & Mohler, L.A. (1992). Music-dependent memory in immediate and delayed recall. Memory & Cognition, 20 (1), 21-28.
- Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 5, 325-337.
- Burton, J. K. (1982). Dual coding of pictorial stimuli by children. Journal of Mental Imagery, 6, 159-168.
- Burton, J. K., & Bruning, R. H. (1982). Interference effects on the recall of pictures, printed words, and spoken words. Contemporary Educational Psychology, 7, 61-69.
- Buser, P., & Imbert, M. (1992). Audition (R.H. Kay, Trans.). Cambridge, MA: Massachusetts Institute of Technology.
- Cann, A., & Ross, D. A. (1989). Olfactory stimuli as context cues in human memory. American Journal of psychology, 102 (1), 91-102.
- Christianson, S.-A. (1997). On emotional stress and memory: we need to recognize threatening situations and we need to "forget" unpleasant experiences. In D. G. Payne & F. G. Conrad (Eds.), Intersections in Basic and Applied Memory Research (pp. 133-156). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Deutsch, D. (1975). The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (Eds.), Short-Term Memory (pp. 107-151). New York: Academic Press, Inc.

Durrant, J. D., & Lovrinic, J. H. (1995). Bases of Hearing Science. (3rd ed.). Baltimore: Williams & Wilkins.

Eich, E. (1985). Context, memory, and integrated item/context imagery. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11 (4), 764-770.

Eich, J. E., Stillman, R.C., Weingartner, H., & Gillin, J.C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. Journal of Verbal Learning and Verbal Behavior, 14, 408-417.

Eysenck, M. W. (1976). Extraversion, activation and the recall of prose. British Journal of Psychology, 67 (1), 53-61.

Eysenck, M. W. (1978). Levels of processing: a critique. British Journal of Psychology, 69, 157-169.

Eysenck, M. W., & Eysenck, M. C. (1979). Processing depth, elaboration of encoding, memory stores, and expended processing capacity. Journal of Experimental Psychology, 5 (5), 472-484.

Farnsworth, P. R. (1934). Examinations in familiar and unfamiliar surroundings. The Journal of Social Psychology, 5, 128-129.

Federal Reporter. (1978). United States vs Williams, 583 F. 2d 1194 (CA2 1978).

Federal Reporter. (1989). The United States vs Smith, 869 F. 2d 353 (CA7 1989).



Fleishman, E. A., Roberts, M. M., & Friedman, M. P. (1958). A factor analysis of aptitude and proficiency measures in radiotelegraphy. Journal of Applied Psychology, 42 (2), 129-135.

Furnham, A., Benson, I., & Gunter, B. (1987). Effects of time of day and medium of presentation on immediate recall of violent and non-violent news. Applied Cognitive Psychology, 1, 255-262.

Furnham, A., Benson, I., & Gunter, B. (1987). Memory for television commercials as a function of the channel of communication. Social Behaviour, 2, 105-112.

Geisler, C. D. (1998). From Sound to Synapse: Physiology of the Mammalian Ear. New York: Oxford University Press Inc.

Gelfand, S. A. (1998). Hearing: an introduction to psychological and physiological acoustics. New York: Marcel Dekker, Inc.

Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? British Journal of Psychology.

Godden, D. R., & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: on land and underwater. British Journal of Psychology, 66, 325-331.

Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., & Stern, J. (1969). Alcohol and recall; state-dependent effects in man. Science, 163, 1358-1360.

Gunter, B., Furnham, A., & Leese, J. (1986). Memory for information from a party political broadcast as a function of the channel of communication. Social Behaviour, 1, 135-142.

Hernandez-Peon, R. (1961). Reticular mechanisms of sensory control. New York: Wiley.

Hill, M. (1994, 1993). A word in your ear: to what extent does hearing a new word help learners to remember it? Paper presented at the Annual International Language in Education Conference, Hong Kong.

Humes, L. E., Nelson, K. J., Pisoni, D. B., & Lively, S. E. (1993). Effects of age on serial recall of natural and synthetic speech. Journal of Speech and Hearing Research, 36, 634-639.

Janowitz, J. M. (1992, 16 August, 1992). Response modality variations affect determinations of children's learning styles. Paper presented at the American Psychological Association Centennial Convention, Washington D.C.

Kroll, N. E., & Kelicutt, M. H. (1972). Short-term recall as a function of covert rehearsal and of intervening task. Journal of Verbal Learning and Verbal Behavior, 11, 196-204.

Martin, F. N. (1997). Introduction to Audiology. (6th ed.). Needham Heights, MA: Allyn & Bacon.

Moore, D. M., Burton, J. K., & Myers, R. J. (1996). Chapter 29: Multiple-channel communication: the theoretical and research foundations of multimedia. In D. H. Jonassen (Ed.), Handbook of Research for Educational Communications and Technology (pp. 851-875). New York: Simon & Schuster Macmillan.

Moray, N., Bates, A., & Barnett, T. (1965). Experiments on the four-eared man. Journal of the Acoustical Society of America, 38, 196-201.

Moreno, R., & Mayer, R.E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. Journal of Educational Psychology, 91 (2), 358-368.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519-533.

Paivio, A. (1986). Mental Representations: A Dual Coding Approach. (Vol. 9). New York: Oxford University Press, Inc.

Pellegrino, J. W., Siegal, A.W., & Dhawan, M. (1975). Short-term retention of pictures and words: evidence for dual coding systems. Journal of Experimental Psychology: Human Learning and Memory, 104 (2), 95-102.

Pellegrino, J. W., Siegel, A. W., & Dhawan, M. (1976). Short-term retention of pictures and words as a function of type of distraction and length of delay interval. Memory & Cognition, 4 (1), 11-15.

Pickett, J. M. (1999). The Acoustics of Speech Communication. Needham Heights, MA: Allyn & Bacon.

Prabu, D., & Hirshman, E. (1998). Dual-mode presentation and its effect on implicit and explicit memory. American Journal of Psychology, 111 (1), 77-87.

Schab, F. R. (1990). Odors and the remembrance of things past. Journal of Experimental Psychology: Learning Memory and Cognition, 16 (4), 648-655.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: implicit and explicit memory for words and voices. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18 (5), 915-930.

Smith, S. M., Glenberg, A., & Bjork, R.A. (1978). Environmental context and human memory. Memory & Cognition, 6 (4), 342-353.

Smith, S. M. (1985). Background music and context-dependent memory. American Journal of Psychology, 98 (4), 591-603.

Stankov, L., & Horn, J. (1980). Human abilities revealed through auditory tests. Journal of Educational Psychology, 72 (1), 21-44.

Steele, E., Lewandowski, L., & Rusling, E. (1996, March 12-16). The effectiveness of bimodal text presentation for poor readers. Paper presented at the Annual Convention of the National Association of School Psychologists, Atlanta, GA.

Stevenson, R. L. (1955). Treasure Island, Racine, WI: Whitman Publishing Company

Travers, R. M. W. (1964). The transmission of information to Human Receivers. AV Communication Review, 12, 373-385.

Tripp, S. D., Roby, Warren B. (1996). Chapter 28: Auditory presentations and language laboratories. In D. H. Jonassen (Ed.), Handbook of Research For Educational Communications and Technology (pp. 821-850). New York: Simon & Schuster Macmillan.

Tulving, E. (1983). Elements of Episodic Memory. (Vol. 2). New York: Oxford University Press.

Warren, R. M. (1982). Auditory Perception: A New Synthesis. Elmsford, NY: Pergamon Press Inc.

Winters, C. A. (1996). Adult learning and multisensory teaching. ERIC., ED 393 966.

## Appendix A

### The Story

She was 66 when I was born and she never said more than 20 words together the 35 years I knew her. Everybody called her Aunt Patty. She had 8 kids of her own and about 37 grandchildren, but since she had 12 brothers and sisters in her own family she was Aunt Patty a lot more often than she was anything else. She was a midwife at a time when there were no doctors and she could cure colds with yellow root and croup with melon rind. I saw her separate two men about to kill each other and laughingly take away the Feldon pistol. She loved the color violet, and she liked to play the banjo. Her favorite song was "Handsome Mabel" and she had never heard of Superman. And she didn't teach me a lot of things. She didn't show me how to be angry and I never learned from her how to hate. I had to learn prejudice from schoolmates. When she turned 102 she started to talk to people who had been dead for 40 years but she still knew me. And when she died that December, I lost my grandmother, my friend, and my hero.

## Appendix B

### Posttest Content: Multiple Choice and Fill-in-the-blank

*Italicized words are spoken during the test. Un-italicized words represent possible answers.*

1. She was \_\_\_\_\_ when I was born.
  - a. *She was 45 when I was born.*
  - b. *She was 58 when I was born.*
  - c. *She was 63 when I was born.*
  - d. *She was 66 when I was born.*
  - e. *She was 71 when I was born.*
  
2. She was 66 when I was born and never said more than \_\_\_\_\_ words together the 35 years I knew her.
  
3. She had \_\_\_\_\_ kids of her own.
  - a. *She had 2 kids of her own*
  - b. *She had 4 kids of her own*
  - c. *She had 6 kids of her own*
  - d. *She had 8 kids of her own*
  - e. *She had 12 kids of her own*
  
4. She had 8 kids of her own and about \_\_\_\_\_ grandchildren.
  
5. She had 8 kids of her own and about 37 grandchildren but since she had \_\_\_\_\_ brothers and sisters in her own family...
  - a. *but since she had 5 brothers and sisters*
  - b. *but since she had 9 brothers and sisters*
  - c. *but since she had 10 brothers and sisters*
  - d. *but since she had 12 brothers and sisters*
  - e. *but since she had 14 brothers and sisters*
  
6. Everybody called her Aunt \_\_\_\_\_.

7. She was midwife at a time when there were no doctors and she could cure colds with \_\_\_\_\_.
- Yellow root
  - Red turnip
  - Dandelions
  - Sassafras
  - Mint tea
8. ...and she could cure colds with yellow root and croup with \_\_\_\_\_.
9. I saw her separate two men about to kill each other and laughingly take away the \_\_\_\_\_ pistol.
- Feldon
  - Johnson
  - Murthen
  - Rugen
  - Deelon
10. She loved the color \_\_\_\_\_ and she liked to play the banjo.
11. Her favorite song was Handsome \_\_\_\_\_.
12. I had to learn \_\_\_\_\_ from schoolmates.
- I had to learn* hatred
  - I had to learn* racism
  - I had to learn* intolerance
  - I had to learn* prejudice
  - I had to learn* jealousy
13. She didn't show me how to be angry and I never learned from her how to \_\_\_\_\_.
14. When she turned \_\_\_\_\_ she started to talk to people who had been dead for 40 years.
15. And when she died that \_\_\_\_\_ I lost my grandmother, my friend.....
- and when she died that* January
  - and when she died that* September
  - and when she died that* March
  - and when she died that* November
  - and when she died that* December

16. I lost my grandmother, my friend, and my \_\_\_\_\_.

- a. *and my Heart*
- b. *and my Hero*
- c. *and my Love*
- d. *and my Innocence*
- e. *and my youth*



## Appendix C

### Experiment Script:

*Italicized words are spoken whereas un-italicized text appears as text on screen.*

- Screen 1- Hello and welcome. Thanks for coming to participate in this activity. I appreciate it. *Would you type in your first name please? I won't remember it but it will let me call you something besides "you".* (text box)
- Screen 2- I need to find out some things about you if you don't mind {Username}.
- Screen 3- Which gender are you? ( male female)
- Screen 4- What is your age? (text box)
- Screen 5- Where were the majority of your ancestors living 500 years ago? (buttons on a map with these choices: Africa, Australia, North America, South America, Pacific Islands, Europe, West Asia, East Asia, Middle East)
- Screen 6- Thanks {Username}! Soon you will hear a short story that lasts about a minute. There are no pictures or text to go along with the story because I'm interested today in how you remember what you hear.
- Screen 7- After you hear the story you'll have to wait exactly 1 minute. I'll give you a picture to watch while you wait.
- Screen 8 Then you will be asked 16 questions about the story that was read to you. You'll have 30 seconds to answer each question.
- Screen 9 Now 30 seconds may seem like an eternity to those of us who work fast. To those of us who work more slowly 30 seconds may be pushing it. You'll have a little clock {Username} in the lower left corner to tell you how much remaining time you have to answer the question.
- Screen 10 I set up the timing this way because it's important that everyone has to remember the story the same amount of time. The fast folks wouldn't have to remember the story nearly as long as the slow people if I didn't set a fixed time for each question.
- Screen 11 After the clock runs out, the program advances. You can't make it wait and you can't speed it up. I told you this so you'll know that you're waiting for the "good of science" and not just because I'm a lousy experimenter.

- Screen 12 The questions will be multiple choice and fill in the blank and require you to remember what you hear during the story.
- Screen 13 You may have to read the questions or the questions may be read to you. It just depends on which test you get.
- Screen 14 When you finish the questions, you'll be able to see how you did if you want. If you don't want to know, you won't be shown your score.
- Screen 15 Let me give you two examples of the types of questions you could be asked after you hear the story.
- Screen 16 Just click the button to hear a short practice story.  
*Last winter a big brown bear came walking through my living room on the way to the refrigerator. He looked up at me on the way by and said, "Hey Leroy. Your curtains don't even come close to matching your sofa."*
- Screen 17 That was the story.
- Screen 18 Next you'll hear a short part of the story with one word left out.
- Screen 19 Example A: *"Last winter a big \_\_\_\_\_ bear came walking through my living room..."*
- Screen 20 You should have heard a brief blank space between the words "big" and "bear".
- Screen 21 Clicking on buttons lets you hear phrases. Your job is to choose the word or phrase that was actually in the story.
- Screen 22 *"Last winter a big \_\_\_\_\_ bear came walking through my living room..."*  
*a. black*  
*b. brown*  
*c. grizzly*  
*d. Kodiak*  
*e. Teddy*
- Screen 23 If you chose 'b' for brown {FirstName} you are correct.
- Screen 24 Sometimes {FirstName} you might get a fill in the blank question instead of a multiple-choice question. They look like this.

- Screen 25 *"Last winter a big \_\_\_\_\_ bear came walking through my living room..."*
- Screen 26 Missing word was (text box).
- Screen 27 If you typed in "brown", again you are correct. You may also see written versions of the questions. They look like this.
- Screen 28 *"Last winter a big \_\_\_\_\_ bear came walking through my living room..."*  
a. black  
b. brown  
c. grizzly  
d. Kodiak  
e. Teddy
- Screen 29 ...and this
- Screen 30 Last winter a big (text box) bear came walking through my living room on the way to the refrigerator.
- Screen 31 Of course, "brown" is the correct answer for both of these questions.
- Screen 32 Let's try one more example without all the screen prompts so you can get used to this stuff {FirstName}.
- Screen 33 The story...
- Screen 34 Question B (multiple choice)  
*Last winter a big brown bear came walking through my living room on the way to the refrigerator. He looked up at me on the way by and said, "Hey Leroy. Your curtains don't even come close to matching your sofa."*  
a. *hey Leroy*  
b. *hey Big Bob*  
c. *hey Billy*  
d. *hey Jimmy*  
e. *hey William*
- Screen 35 The correct answer was choice "a", "Hey Leroy". Let's look at a fill in the blank with sound.
- Screen 36 Question B (fill in the blank)  
*He looked up at me on the way by and said, " \_\_\_\_\_ . Your curtains don't even come close to matching your sofa."*

- Screen 37 Missing phrase was (text box).
- Screen 38 If you typed "Hey Leroy" in the blank {FirstName}, you are correct.
- Screen 39 Now let's look at one more multiple-choice question with no sound. It looks like this.
- Screen 40 Last week a big brown bear came walking through my living room on the way to the refrigerator. He said \_\_\_\_\_ your curtains don't even match your sofa.
- a. hey Leroy
  - b. hey Big Bob
  - c. hey Billy
  - d. hey Jimmy
  - e. hey William
- Screen 41 Great {FirstName}, you've seen all the types of questions that might follow the story.
- Screen 42 When you click continue you will hear a story that lasts about one minute. Listen carefully.
- Screen 43 the story...  
*All right. Everyone listen carefully starting now.*
- She was 66 when I was born and she never said more than 20 words together the 35 years I knew her. Everybody called her "Aunt Patty". She had 8 kids of her own and about 37 grandchildren, but since she had 12 brothers and sisters in her own family she was "Aunt Patty" a lot more often than she was anything else. She was a midwife at a time when there were no doctors and she could cure colds with yellow root and croup with melon rind. I saw her separate two men about to kill each other and laughingly take away the Feldon pistol. She loved the color violet, and she liked to play the banjo. Her favorite song was "Handsome Mabel" and she had never heard of Superman. And she didn't teach me a lot of things. She didn't show me how to be angry and I never learned from her how to hate. I had to learn prejudice from schoolmates. When she turned 102 she started to talk to people who had been dead for 40 years but she still knew me. And when she died that December, I lost my grandmother, my friend, and my hero.*
- Screen 44 That was the story...
- Screen 45 And now...that picture I promised you.

Screen 46 Can you fit the pieces together? (It's just to help you pass the time until the questions start {FirstName}).

Screen 47 (movable pictures)

Screen 48 ...and now the test.

Screen 49 Question 1

*She was \_\_\_\_\_ when I was born.*

- a. She was 45 when I was born.*
- b. She was 58 when I was born.*
- c. She was 63 when I was born.*
- d. She was 66 when I was born.*
- e. She was 71 when I was born.*

Screen 50 Question 2

*She was 66 when I was born and never said more than \_\_\_\_\_ words together the 35 years I knew her.*

Screen 51 Question 3

*She had \_\_\_\_\_ kids of her own.*

- a. She had 2 kids of her own*
- b. She had 4 kids of her own*
- c. She had 6 kids of her own*
- d. She had 8 kids of her own*
- e. She had 12 kids of her own*

Screen 52 Question 4

*She had 8 kids of her own and about \_\_\_\_\_ grandchildren.*

Screen 53 Question 5

*She had 8 kids of her own and about 37 grandchildren but since she had \_\_\_\_\_ brothers and sisters in her own family...*

- a. but since she had 5 brothers and sisters*
- b. but since she had 9 brothers and sisters*
- c. but since she had 10 brothers and sisters*
- d. but since she had 12 brothers and sisters*
- e. but since she had 14 brothers and sisters*

Screen 54 Question 6

*Everybody called her Aunt \_\_\_\_\_.*

Screen 55 Question 7

*She was midwife at a time when there were no doctors and she could cure colds with \_\_\_\_\_.*

- a. Yellow root*
- b. Red turnip*
- c. Dandelions*
- d. Sassafras*
- e. Mint tea*

Screen 56 Question 8

*...and she could cure colds with yellow root and croup with \_\_\_\_\_.*

Screen 57 Question 9

*I saw her separate two men about to kill each other and laughingly take away the \_\_\_\_\_ pistol.*

- a. Feldon*
- b. Johnson*
- c. Murthen*
- d. Rugen*
- e. Deelon*

Screen 58 Question 10

*She loved the color \_\_\_\_\_ and she liked to play the banjo.*

Screen 59 Question 11

*Her favorite song was Handsome \_\_\_\_\_.*

Screen 60 Question 12

*I had to learn \_\_\_\_\_ from schoolmates.*

- a. I had to learn hatred*
- b. I had to learn racism*
- c. I had to learn intolerance*
- d. I had to learn prejudice*
- e. I had to learn jealousy*

Screen 61 Question 13

*She didn't show me how to be angry and I never learned from her how to \_\_\_\_\_.*

Screen 62 Question 14

*When she turned \_\_\_\_\_ she started to talk to people who had been dead for 40 years.*

Screen 63 Question 15

*And when she died that \_\_\_\_\_ I lost my grandmother, my friend.....*

- a. and when she died that January*
- b. and when she died that September*
- c. and when she died that March*
- d. and when she died that November*
- e. and when she died that December*

Screen 64 Question 16

*I lost my grandmother, my friend, and my \_\_\_\_\_.*

- a. and my Heart*
- b. and my Hero*
- c. and my Love*
- d. and my Innocence*
- e. and my youth*

Screen 65 That's all! Thank you very much {FirstName} for your help in this. You've been great!

Screen 66 {FirstName}, if you had to pick one, which do you tend to remember better. Information you have read or information you have heard. (i.e. when you study for a test, do you need to see it written on a page or is it better if you read it aloud or someone else reads it aloud to you?)

Screen 67 Would you like to see how you did on the test? (see results) (no thanks)

Screen 68 (If they choose "see results")  
Well {FirstName}, you got {total} multiple-choice questions correct out of 8. Not too shabby for this experiment.

Screen 69 Thanks again {FirstName}. You were great!