

Non-linear, adaptive array processing for acoustic interference suppression

Elizabeth Hoppe^{a)} and Michael Roan

Department of Mechanical Engineering, Virginia Polytechnic Institute and State University, 131 Durham Hall, Blacksburg, Virginia 24060

(Received 9 December 2008; revised 9 April 2009; accepted 9 April 2009)

A method is introduced where blind source separation of acoustical sources is combined with spatial processing to remove non-Gaussian, broadband interferers from space-time displays such as bearing track recorder displays. This differs from most standard techniques such as generalized sidelobe cancellers in that the separation of signals is not done spatially. The algorithm performance is compared to adaptive beamforming techniques such as minimum variance distortionless response beamforming. Simulations and experiments using two acoustic sources were used to verify the performance of the algorithm. Simulations were also used to determine the effectiveness of the algorithm under various signal to interference, signal to noise, and array geometry conditions. A voice activity detection algorithm was used to benchmark the performance of the source isolation. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3126925]

PACS number(s): 43.60.Jn, 43.60.Fg [EJS]

Pages: 3835–3843

I. INTRODUCTION

Sound source localization is an important and widely studied problem in acoustics.^{1–4} One of the most fundamental and widely used spatial processing tools is phased array beamforming. Algorithms for beamforming range from simple, narrow-band, delay and sum beamformers,⁵ which are completely data independent and nonadaptive, to more complex adaptive algorithms, such as minimum variance distortionless response (MVDR) beamforming.^{6,7} One of the main challenges in any beamforming technique is to suppress sidelobe contributions to the beamformer output. MVDR, for example, attempts to accomplish this by minimizing the variance of the beamformer output under a distortionless constraint. In cases where the bearing angle to an interferer is known, a fixed null beamformer^{8,9} can be used to suppress contributions to the beamformer output from that specific direction. In most applications however, the locations of interfering sources are not known a priori. In 1969, Capon introduced a beamforming technique that adaptively suppresses non-look direction interferers based on the covariance of the received array data.⁶ Although a large set of algorithms exists to mitigate the effects of interferers, they mostly focus on adaptively placing nulls in the directions of interferers or minimizing sidelobe height.¹⁰

Another large body of work exists in the area of time or frequency domain blind source separation (BSS).^{11,12} These techniques do not typically exploit spatial information about the signals. Where spatial information has been used is in attempt to solve the BSS permutation ambiguity problem.^{13–16} In this paper, we bring together methods from BSS and phased array processing to create a unique approach to interference suppression. This method suppresses contributions to the spatial processor output from non-Gaussian

interferers using a combination of beamforming and independent component analysis (ICA) techniques. The new method is less sensitive to errors in the array manifold vector than the MVDR method. In addition, the new method does not depend on estimates of the signal covariance, which is often problematic in application of the MVDR beamformer. However, an estimation of the probability density function (pdf) of the signal amplitude distribution is required.

The high-level operation of the new algorithm is as follows: Beamforming is used to correct for time delays caused by the propagation of wavefronts arising from the spatial distribution of sources. The FastICA algorithm is applied to these delay-corrected signals. Spatial knowledge is also used to resolve the permutation ambiguity problem inherent in the ICA algorithm. Individual signals that are retrieved from the ICA algorithm are then filtered from the mixed signals in all channels and the remaining signals are processed spatially for display.

II. BACKGROUND

A. MVDR beamforming

The MVDR beamforming strategy is to minimize the variance of noise in the look direction (sometimes called the signal protection constraint¹⁷), so that the response of the array is constant in the signal look direction regardless of the choice of weights (i.e., 1). One limitation of addition of this constraint is that the beamformer is limited to nulling $m-2$ interferers, where m is the number of sensors. The weights are adaptively calculated, based on the signal covariance matrix, to coherently sum the signal of interest, while simultaneously minimizing the noise variance of the output. The derivation begins with a set of spatially compact acoustic sources in a free field with amplitudes $\mathbf{s}(t)=[s_1(t), s_2(t), \dots, s_n(t)]^T$. The sources are assumed to be narrow band and Gaussian distributed. The signals $x(t)$ at the output of the array at time t are

^{a)}Author to whom correspondence should be addressed. Electronic mail: ehoppe@vt.edu

$$\mathbf{x}(t) = \mathbf{a}(\theta)\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{x}(t)$ is the output of the array, $\mathbf{a}(\theta)$ is the array manifold vector, $\mathbf{s}(t)$ is the source signal vector that lies in direction θ_s , and $\mathbf{n}(t)$ is a vector of noise that represents the undesired signals including background noise and interferers. The combined beamformer output of the MVDR process is given by⁶

$$\mathbf{y}(t) = \mathbf{W}^*\mathbf{x}(t) = \mathbf{W}^*\mathbf{a}(\theta)\mathbf{s}(t) + \mathbf{W}^*\mathbf{n}(t), \quad (2)$$

with \mathbf{W} as a vector of array weights. The goal is to choose \mathbf{W} to extract $\mathbf{s}(t)$ by minimization of the effects of $\mathbf{n}(t)$ in Eq. (2). The variance of the noise is quantified by the noise covariance matrix, $\mathbf{R}_n = E\{\mathbf{nn}^*\}$, and the variance of the beamformer output when the beamformer is steered to θ_s is given by

$$\sigma_n^2 = E\{\mathbf{W}^*\mathbf{nn}^*\mathbf{W}\} = \mathbf{W}^*\mathbf{R}_n\mathbf{W}. \quad (3)$$

The distortionless constraint is expressed as the requirement that $\mathbf{W}^*\mathbf{a}(\theta) = 1$. Therefore, \mathbf{W} is given as the optimal solution to the LaGrange problem

$$\min \mathbf{W}^*\mathbf{R}_n\mathbf{W} \text{ subject to } \mathbf{W}^*\mathbf{a}(\theta) = 1. \quad (4)$$

In the MVDR method, the noise is typically unknown, and the matrix \mathbf{R}_n is replaced by the sample covariance matrix \mathbf{R}_y . Replacing \mathbf{R}_n with \mathbf{R}_y in Eq. (4) and solving produces the analytical MVDR solution⁶

$$\mathbf{W} = \frac{\mathbf{R}_y^{-1}\mathbf{a}(\theta)}{\mathbf{a}^*(\theta)\mathbf{R}_y^{-1}\mathbf{a}(\theta)}. \quad (5)$$

The MVDR equation (5) shows that the weights depend on the covariance matrix and the presumed array manifold vector. As a consequence, the beamformer is sensitive to inaccuracies in the manifold vector. This is often a problem for towed array sonar where the shape of the array is often unknown due to the motion of the submarine causing bending of the array.¹⁸ In some instances, the performance of the MVDR beamformer can degrade below the performance of the standard non-adaptive techniques. The advantage of the MVDR beamformer is that the locations of the interferers do not need to be known a priori, the beamformer suppresses contributions from all sources not lying along the steering direction θ_s up to $m-2$ degrees of freedom.

MVDR and other nulling beamforming techniques perform interference suppression through minimizing the response of the array in non-look direction through the placing of nulls in those directions. Another widely used technique to extract signals from a mixture using multiple measurements is BSS. These techniques typically use information theoretic measures such as entropy maximization as criteria for source extraction. A brief review of the more widely used techniques is given in Sec. I B in order to provide a basic review and introduce notation.

B. BSS

BSS is an algorithm based on ICA that attempts to extract statistically independent sources from a set of mixed signals using multiple independent measurements. Bell and

Sejnowski¹⁹ developed a method for BSS based on minimizing the mutual information between sensor outputs using arguments based on entropy. The information content of the signal is represented by the entropy of the sensor outputs, given by

$$H(y) = -E\{\ln p_y(y)\}, \quad (6)$$

where $p_y(y)$ is the pdf of the sensor output y , and $E\{\cdot\}$ denotes expectation. Under the assumption of statistically independent sources, maximizing the joint entropy minimizes the mutual information between the channels. To maximize entropy in a constrained fashion (i.e., not just simply endlessly increasing the variance), a nonlinear ‘‘squashing function,’’ $g(x)$, is used to bound the pdf of the output signal. If $g(x)$ is a monotonically increasing function, the pdf of the array output can be written as a function of the pdf of the source signals x , according to

$$p_y(y) = \frac{p_x(x)}{|J|}, \quad (7)$$

where $|J|$ is the absolute value of the Jacobian of $g(x)$. Substituting Eq. (7) into Eq. (6) gives the final entropy function to be

$$H(y) = E\{\ln|J|\} - E\{\ln p_x(x)\}. \quad (8)$$

To minimize the mutual information, only the first term in Eq. (8) needs to be minimized. An online stochastic gradient descent learning rule for the Infomax method is used to determine the optimal weight set, W , to separate the sources. Other more robust methods using the natural gradient have also been developed.²⁰ The learning rule for Infomax is

$$\Delta W \propto \frac{\partial H(y)}{\partial W} = [W^T]^{-1} - 2yx^T. \quad (9)$$

Hyvärinen²¹ moved the mutual information minimization approach from the time domain to the frequency domain which resulted in a faster BSS algorithm called FastICA. The FastICA algorithm is based on the differential entropy, or negentropy, which is defined as

$$J(y) = H(y_{\text{Gauss}}) - H(y) \quad (10)$$

where J is the negentropy, $H(\cdot)$ is the entropy, and y_{Gauss} is a Gaussian random vector of the same covariance matrix as y . This provides a measure of non-Gaussianity that is always non-negative and only zero for a signal of Gaussian distribution. The FastICA algorithm seeks to find a direction, w , such that the projection, $w^T x$, maximizes the negentropy. An approximation of the negentropy of Eq. (10) is

$$J(w^T x) \propto [E\{g(w^T x)\} - E\{g(v)\}]^2, \quad (11)$$

where g is any non-quadratic function and v is a Gaussian variable of zero mean and unit variance. The FastICA algorithm chooses w as the solution to the optimization problem

$$\max [E\{g(w^T x)\} - E\{g(v)\}]^2 \text{ subject to } E\{(w^T x)^2\} = \|w\|^2 = 1. \quad (12)$$

Similar to the Infomax method, only the first term in Eq. (12) is dependent on updated values of w . Using Newton's methods and algebraic simplification, the stabilized fixed point algorithm can be solved as

$$w^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w. \quad (13)$$

An initial value of the weight vector, w , is chosen and w^+ is calculated. This updated value of the weight vector is normalized to unit length and compared to the previous weight vector. If the two vectors have a dot product of zero, they define the same direction and the algorithm has converged. A typical choice of nonlinear function is the hyperbolic tangent, which is ideal for separating super-Gaussian sources, such as voice signals.

One of the main limitations of these BSS techniques is that the mixing is assumed to be linear, a condition that does not exist in a convolutive mixing problem. When processing signals that are measured using multiple sensors, the geometry of the sources and sensors causes relative delays in the measured signals. The net result is that the mixing is convolutive mixing in most situations of interest. Therefore, some preprocessing must be performed on the array data to remove the convolutive mixing before it can be used in the FastICA algorithm.

C. Voice activity detection

An important application where both BSS and spatial processing can have a large impact is in voice activity detection^{22,23} (VAD) in the presence of interference. VAD algorithms are key tools for communications systems where conserving power is important (i.e., mobile telephones) and where transmitting noise (as in military communications) is to be avoided. The VAD algorithm used in this paper to quantify the signal enhancement capabilities of the proposed algorithm uses the higher order statistics of the signal to distinguish between active voice frames and noise frames.²⁴ Every 10 ms, the second-, third-, and fourth-order moments are calculated as

$$M_{k,x} = \frac{1}{N} \sum_{n=0}^{N-1} [x(n)]^k, \quad (14)$$

where $x(n)$ is the signal, k is the moment order, and N is the number of samples in $x(n)$. Using these moments, the unbiased normalized skewness (SK) and the unbiased kurtosis (KU) are calculated using

$$SK = \frac{M_{3,x}}{\sqrt{\frac{15\nu_g^3}{N}}}, \quad KU = \left(1 + \frac{2}{N}\right)M_{4,x} - 3(M_{2,x})^2, \quad (15)$$

where ν_g is the noise energy. The measure for SK and KU are normalized by the signal energy to give

$$\gamma_3 = \frac{SK}{M_{2,x}^{1.5}}, \quad \gamma_4 = \frac{KU}{M_{2,x}^2}. \quad (16)$$

The noise power is estimated from any frame declared as non-speech. For initialization, the first three frames are automatically declared as non-speech and used to initialize the

noise energy. Whenever a frame is declared as non-speech, its energy is used to update the estimate for the noise energy according to an autoregressive averaging

$$v_g(k) = (1 - \beta)v_g(k-1) + \beta M_{2,x}, \quad (17)$$

where k is the iteration index, and β is 0.1 Prob[Noise] (the probability of a frame being noise). Prob[Noise] is based on the value of KU_b , which is the unit-variance version of the KU defined as

$$KU_b = \frac{KU}{\sqrt{\frac{3\nu_g^4}{N} \left(104 + \frac{452}{N} + \frac{596}{N^2}\right)}}, \quad (18)$$

where N is the number of samples in the frame. Using this value and the value of SK from Eq. (15), the probability of the frame being noise is

$$\text{Prob[Noise]} = \frac{\text{erfc}(SK) + \text{erfc}(KU_b)}{2}. \quad (19)$$

The final parameter calculated is the signal to noise ratio (SNR). The SNR of the frame is

$$\text{SNR} = \text{Pos} \left[\frac{M_{2,x}}{\nu_g} - 1 \right], \quad (20)$$

where $\text{Pos}[x] = x$ for $x > 0$ and 0 otherwise. The values of γ_3 , γ_4 , Prob[Noise], and SNR are used with thresholds to determine the state of the current frame. If the Prob[Noise] is below its threshold value, and the SNR is greater than its threshold value, the frame is declared as speech state. A noise state is indicated by a Prob[Noise] greater than its threshold, and values for γ_3 and γ_4 above their respective thresholds. For extensive details on the VAD algorithm, see Ref. 24.

In Section IV G the relative performances of the proposed algorithm, the MVDR beamformer, and unmodified FastICA are compared and provide a quantitative performance enhancement provided by the proposed algorithm. The structure of the proposed algorithm is the subject of Sec. III.

III. PROPOSED ALGORITHM

This section describes the proposed algorithm that uniquely combines spatial and BSS processing streams. The FastICA algorithm assumes instantaneous mixing; however, when multiple sources are spatially distributed, this condition does not exist. The spatial processing portion of the new algorithm aligns the signals in time. This allows application of FastICA (or other ICA techniques for linear/simultaneous mixing) because the mixing is no longer convolutive. As Secs. III A–III C will show, the new algorithm's performance is only limited by the SNR and signal to interferer ratio (SIR).

A. Strategy and motivation

A block diagram of the proposed new algorithm is provided in Fig. 1. Assuming that there are m sensors and n sources, the goal is to remove j sources from each array channel. In this way, signals such as loud interferers can be

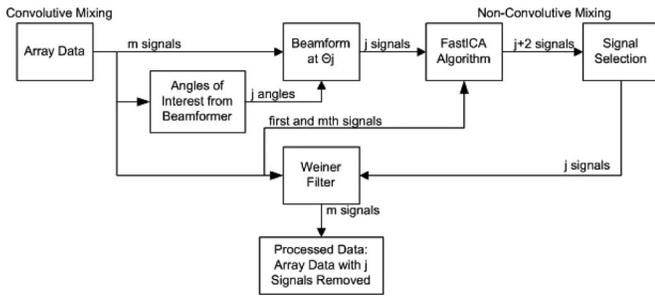


FIG. 1. Processing strategy for the combination of beamforming and ICA processing to suppress interferers.

removed, leaving behind the $(n-j)$ signals of interest. The inputs to the algorithm are the measured microphone array signals, and the output is the set of signals of interest. The m array signals are beamformed, which is used to estimate the direction of arrival (DOA) of the k signals. These angles are used to steer the array and extract signals from the k directions of interest. These signals, along with two unaltered array sensor signals, are passed to the FastICA algorithm, which separates the signals as explained in Sec. II B. A Weiner filter is used to cancel the interfering signals from the mixture of signals in each channel (least mean squares approaches can also be used).

One advantage of using this new method is that interferer rejection is not spatially based or null beamforming as in MVDR. As an interferer moves closer to the signal of interest, beamformers cannot distinguish the signal of interest from the interferer because they are spatially indistinct. This is especially true for compact arrays (i.e. low m), where the main beam lobe is very wide. In the new method, the higher order statistics of the signal are used for separation in addition to the spatial location and interfering signals can still be suppressed in spatially indistinct mixtures. While the algorithm's performance is degraded in this region, interferer suppression is still possible.

B. Correcting for delays

For simplicity of derivation, a linear array with m microphone elements is used. The n sound sources are all assumed to be in front of the array, with the source DOAs ranging from -90° to $+90^\circ$. Figure 2 provides the scenario geometry and a definition of the delays appearing in Eqs. (21) and (22). Each array sensor sees a mixture of delayed versions of each of the n sources according to the model

$$x_i(t) = \sum_{k=1}^n A_{ki} s_k(t - \tau_{ki}) + n_i(t), \quad \tau_{ki} \equiv \frac{d_k}{c},$$

for $i = 1, 2, \dots, m,$ (21)

where x_i is the response of the i th sensor of the array, A_{ki} is the linear mixing component, s_k is the k th source, τ_{ki} is the delay from the k th source to the i th sensor based on the distance between the two, and n_i is the noise associated with the i th sensor. The time delays caused by the sensor and source geometry must be removed before FastICA can be applied. Electronic steering of the array is used to compensate for the delays. Once corrected using beamforming, the

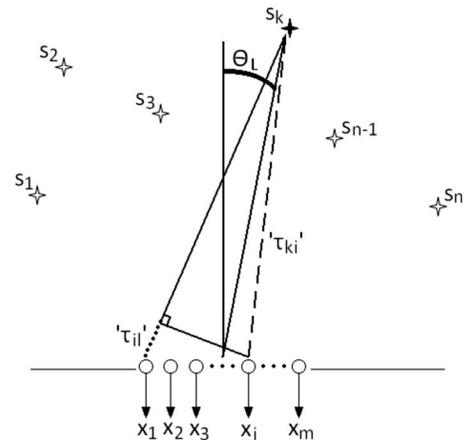


FIG. 2. The geometry of the sensors and sources assumed for algorithm development is shown. The array is a linear microphone array and all sources are in front of the array. The time delays introduced in Eqs. (21) and (22) are due to the corresponding distances indicated with the dashed and dotted line segments.

signals are linearly mixed and therefore can be processed by the FastICA portion of the algorithm.

In this paper, the steering is accomplished using a traditional narrow-band delay and sum beamformer. Other more sophisticated techniques could be used in this step; however, the narrow-band delay and sum beamformer is used for convenience due to its ease of implementation and processing speed. The weights for this beamformer are data independent and can be calculated in advance. The relative delays between sensors for each signal are given by

$$b_l(t) = \sum_{i=1}^m x_i(t - \tau_{il}) = \sum_{i=1}^m \sum_{k=1}^n A_{ki} s_k(t - \tau_{ki} - \tau_{il}) + n_i(t - \tau_{il})$$

for $l = 1, 2, \dots, n$

$$\tau_{il} = \frac{d_e}{c} (i - 1) \sin \theta_k, \quad (22)$$

where b_l is the l th beamformed signal, τ_{il} is the delay from the delay-and-sum beamformer, and θ_k is the DOA of the k th source. The beamformed outputs, combined with the unaltered first and m th array signals, form the set of mixed signals for the ICA algorithm. The first and m th array signals are included to introduce additional unaltered information (beamforming distorts signals away from the steering angle) about the mixed signals into the FastICA algorithm to aid in the separation process. Although not a requirement of FastICA, inclusion of these undistorted mixed signals improves the separation result.

C. ICA and signal selection

The mixed signal set generated by the beamforming process contains $(j+2)$ signals. The FastICA algorithm will also produce $(j+2)$ outputs, j of which will contain unmixed signals. The other two are a residual mixture of all the signals and Gaussian noise. The FastICA algorithm has a permutation ambiguity in the outputs, so additional postprocessing is necessary to identify the separated sources. The unmixed sig-

nals must be classified as either the signal of interest, interferers, or residual mixtures. This task is accomplished using the coherence between the beamformed signals and the FastICA outputs. The interferer signals are chosen as the FastICA output channel with the highest coherence with the beamformer output at the DOA of the interferers. The interferer signals are used as the “noise” sample input into a Wiener filter, which cancels these from each channel in the array. Using this approach, a technique can be implemented where, starting with the loudest, interferers can be canceled from displays such as bearing time recorder (BTR) displays leaving behind successively more quiet sources. Examples of this technique are provided via both simulation and experiment in Sec. IV.

IV. SIMULATIONS

A. Simulation parameters

For the simulations presented in Secs. IV B–IV F, the data contains two voice signals; one signal of interest and one interferer signal. The sources used are digitally recorded files of voices, which are artificially placed at various locations in the environment with respect to the array. The data in the simulation in Sec. IV G contains one voice signal and one broadband interferer signal. In order to simulate the DOA, the source signals are up-sampled to five times their sampling frequencies and then shifted by the appropriate number of samples corresponding to the DOA delay. This method is not frequency dependent and creates minimal distortion in the signal.

All of the simulations assume an anechoic environment, where the speed of sound is 343 m/s. A linear array is used in all cases, where the element spacing is constant. The number of elements and the elemental spacing varies depending on the simulation. For each case, the sampling frequency is 11025 Hz, which is a standard sampling frequency used in wav audio files. The frequency of interest used for simulations is 3250 Hz, which was chosen based on the frequency content of the simulated signals. Unless stated otherwise, the two signals are at equal power, equidistance from the array, and there is no additional non-correlated background noise.

B. Effect of the algorithm on unmixing

To quantify the performance of the new algorithm, a comparison was done to compare unmixing performance under the conditions of linear mixing, convolutive mixing, and corrected convolutive mixing. The signal of interest is located at $+5^\circ$, while the interferer is located at -10° . A 35 channel array is used with an element spacing of 0.035 m. For the convolutive mixing case, the first and m th array signal are passed directly to the FastICA algorithm. The corrected convolutive mixing case uses the beamforming technique to correct the time delays (as described above). Figure 3 shows the coherence between the original and unmixed signal for the three cases. While the linear mixing case has a coherence of nearly 1 for all frequencies, the other two cases show degradation of the signal after separation. However, the new algorithm shows much better performance for most of the frequencies considered.

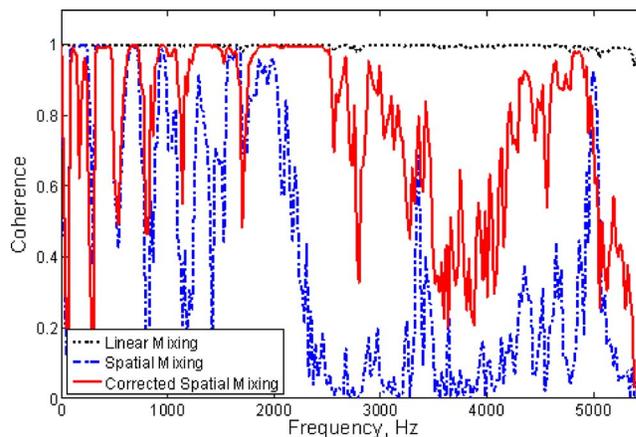


FIG. 3. (Color online) Coherence between the original signal and the unmixed signal for three cases: linear mixing with FastICA separation, convolutive mixing with FastICA separation, and convolutive mixing with time delay correction via beamforming and FastICA separation.

C. Two voices with equal signal strength

A simulation with two voice signals (one stationary, one moving) was done in order to examine the algorithm’s performance with respect to a variety of spatial situations. The interferer remains stationary at $+5^\circ$, while the signal of interest is moved from -20° to $+20^\circ$ at a rate of $1^\circ/\text{s}$. All sources are assumed to be in the far-field and have equal power. The data are simulated for a 35 channel linear array with an inter-microphone spacing of 0.035 m. Figure 4 shows the time-angle plot using a standard narrowband delay-and-sum beamformer. Each horizontal slice in Fig. 4 is the magnitude of the beamformer output calculated using one second of data. The two sources are clearly seen. The interferer appears as the vertical signature appearing at $+5^\circ$. The signal of interest is the signature that crosses diagonally from -20° to $+20^\circ$.

MVDR beamforming is used in this paper to provide a comparison or the new algorithm with spatial null forming techniques. The result of MVDR processing can be seen in

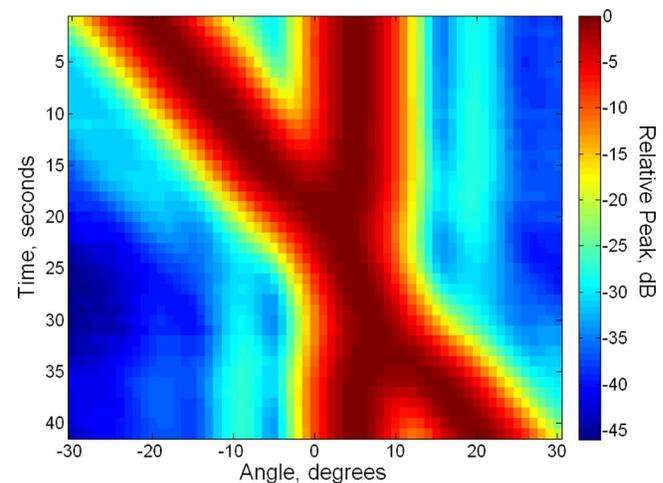


FIG. 4. (Color online) Bearing track recorder plot using a narrow-band delay and sum beamformer for two signals having equal power. One signal is located at $+5^\circ$, while the other signal moves from -20° to $+20^\circ$ at a rate of $1^\circ/\text{s}$.

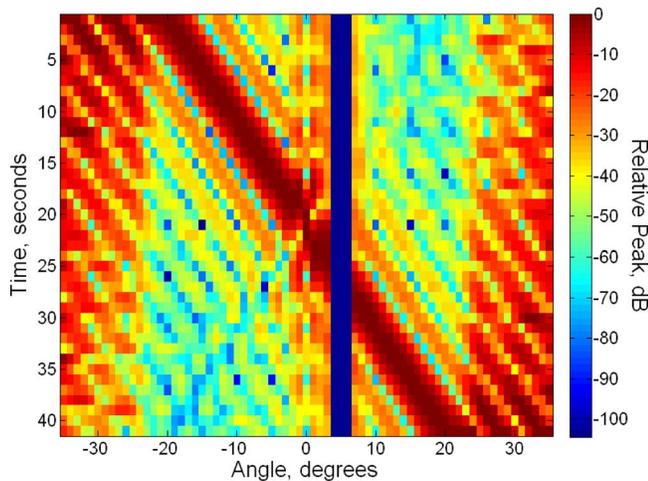


FIG. 5. (Color online) Bearing track recorder plot using MVDR beamforming. The undesired, stationary source at $+5^\circ$ has been nulled in all time steps, but numerous sidelobes in the beampattern obscure the bearing track recorder display.

the time-angle plot of Fig. 5. Over the entire duration of the simulation, the MVDR beamformer has placed a null in the direction of the interferer at $+5^\circ$. While the MVDR algorithm has good performance with regard to the interferer suppression at $+5^\circ$, as the signal of interest passes near the interferer, both signals are suppressed by the null at $+5^\circ$. This is one of the shortcomings of null based interference suppression, such as MVDR. The beampattern has numerous sidelobes that distort the bearing track of the signal of interest. Figure 6 shows the results of applying the new algorithm to the simulated data. The interferer at $+5^\circ$ is suppressed in all time steps, while the signal of interest is clearly visible in all the time steps.

The beamformer output shows the reduction in power from the interferer direction with the application of the new algorithm. In Fig. 7, the algorithm's effect on coherence between the original signal of interest and the output of the algorithm is shown. For each second of data, instead of plot-

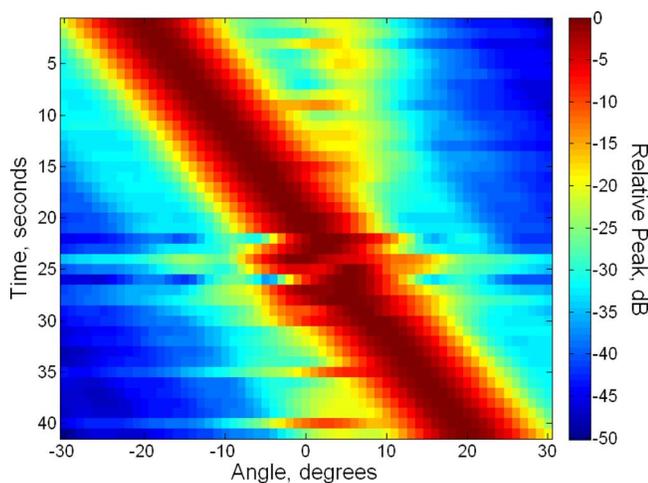


FIG. 6. (Color online) Bearing track recorder plot using a narrow-band delay and sum beamformer after application of the proposed algorithm. The undesired, stationary source at $+5^\circ$ has been suppressed in all time steps without significant suppression of the moving source.

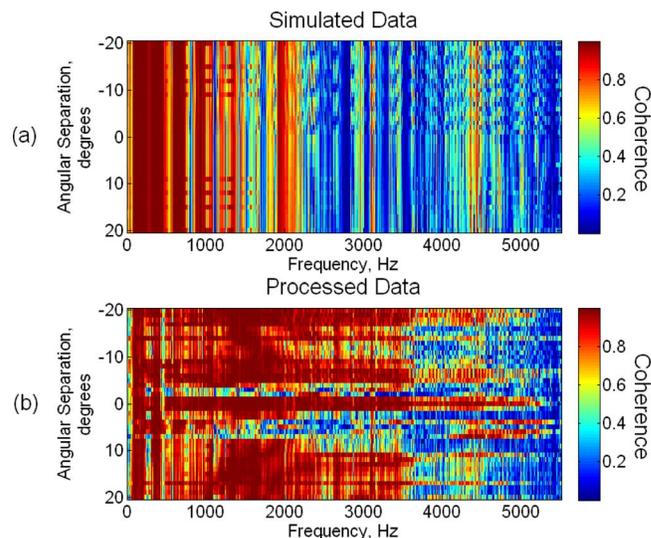


FIG. 7. (Color online) The algorithm's interference suppression performance as a function of the angular spacing between the interferer and the signal of interest is examined. (a) The coherence between the algorithm input data and the signal of interest. (b) The coherence between the algorithm output and the signal of interest.

ting the beamformer output as in Fig. 6, the coherence is plotted. When the angular separation is less than 3° , the coherence across the frequencies of interest is greater for the algorithm output than for the unmixed array data. For angular separation less than 3° , the coherence increases with application of the new algorithm, but not as significant as for greater angular spacing. In these cases, the sources are not spatially distinct, and the beamforming stage does not contribute new information to the algorithm. In these cases, the mixing matrix tends to be ill conditioned and the FastICA stage cannot completely separate the sources.

D. Two voice signals with unequal signal strengths

The results presented previously are for the case where the signal of interest and the interferer have equal power. Next, in order to test the robustness of the new algorithm in terms of the SIR, cases were generated where the signal of interest and interferer powers are not equal. The SIR is given by

$$\text{SIR} = 10 \log_{10} \frac{P_{\text{signal}}}{P_{\text{int}}}, \quad (23)$$

where P_{signal} and P_{int} are the power of the signal of interest and interferer, respectively. For all cases, the signal of interest was positioned at $+5^\circ$, while the interferer was at -10° . Again, a 35 channel array with element spacing of 0.035 m is used. Figure 8 illustrates the impact of SIR on the new algorithm's performance. For all values of SIR greater than -15 dB, the signal of interest is clearly visible at $+5^\circ$ while the interferer at -10° is suppressed. When the SIR values falls below -15 dB, the signal of interest is completely masked by the interferer. However, even beyond -15 dB, the interfering signal sees nearly 20 dB of suppression as a result of the algorithm's application.

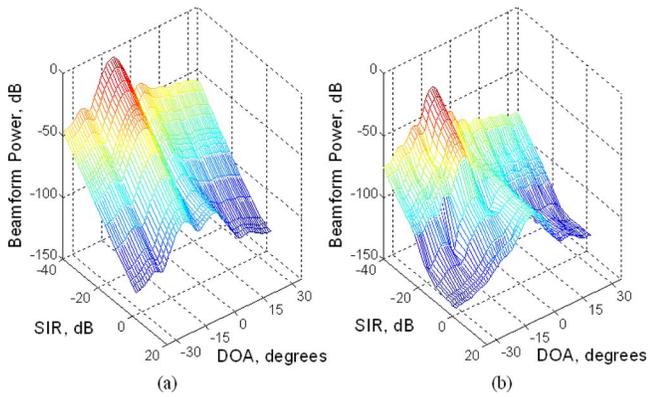


FIG. 8. (Color online) The effect of SIR on the performance of the new algorithm. (a) The beampattern of the algorithm input using a narrow-band delay and sum beamformer. This signal of interest is positioned at $+5^\circ$ and the interferer is at -10° . (b) The beampattern of the algorithm output using a narrow-band delay and sum beamformer. The interferer has been suppressed by 20 dB in all cases, and the signal of interest remains visible until SIR drops below -15 dB.

E. Two voices with equal signal strengths and the addition of noise

All of the cases previously presented contain no additional noise. To further test the performance of the new algorithm, cases were generated where the signal of interest and interferer are at equal power and additional white Gaussian noise is added to the system. The SNR is given by

$$\text{SNR} = 10 \log_{10} \frac{P_{\text{signal}}}{P_{\text{noise}}}, \quad (24)$$

where P_{signal} is the power of the signal of interest. Since the signal of interest and the interferer have the same power, the SNR could also be calculated using P_{int} instead of P_{signal} in Eq. (24). For all cases, the signal of interest was positioned at $+5^\circ$, while the interferer was positioned at -10° and a 35 channel array with 0.035 m spacing is used. Figure 9 shows a plot of the effect of SNR on the new algorithm's performance. A standard delay and sum beamformer was applied to the raw data, and the result is shown in Fig. 9(a), while Fig. 9(b) shows the results of applying the new algorithm.

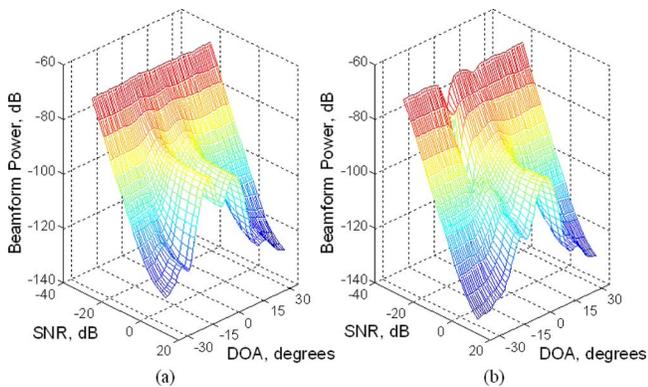


FIG. 9. (Color online) The effect of SNR on the algorithm. (a) The beampattern of the algorithm input using a narrow-band delay and sum beamformer. This signal of interest is positioned at $+5^\circ$ and the interferer is at -10° . (b) The beampattern of the algorithm output using a narrow-band delay and sum beamformer. The interferer is suppressed in all cases and the signal of interest remains visible until SNR drops below -18 dB.

Peaks in the beamformer output are identified as sources. Any peak that rises above a selected threshold on the beamformer output is considered a source. This threshold value is dependent on the data processed, especially on the background noise. For Fig. 9(a), two distinct signal peaks are apparent in the beamformer output when the SNR value is above -18 dB. After the application of the new algorithm, the signal of interest is a distinct peak for all values of SNR greater than -18 dB. When the SNR value falls below -18 dB, the signal of interest is not distinguishable from the background noise, but there is still some suppression at -10° . The breakdown of the algorithm beyond a SNR of -18 dB is comparable to the performance of other spatial techniques.

F. VAD for one voice with one interferer

The VAD algorithm was used to further quantify the performance of the algorithm and its ability to separate signals. For this simulation, the signal of interest is a voice signal at $+5^\circ$, while the interferer is a white Gaussian noise source at -10° . Three different SIRs were tested: 5, 0, and -6 dB. A 35 channel array with element spacing of 0.035 m is used. The VAD technique was applied to the data before and after it was passed through the proposed algorithm as well as after MVDR processing. Figure 10 shows the results of the application of VAD to the data. Each box encloses a time segment that the VAD identified as an active voice frame. The left column shows the frames identified as active voice frames in the original simulated data for the three SIR cases. The middle column shows the active voice frames after the data were processed using MVDR, while the right column shows the active voice frames after the data have been passed through the separation algorithm. As the SIR drops, the VAD algorithm is unable to correctly identify the active speech sections of the signal in the unprocessed data. Application of the MVDR algorithm increases the VAD algorithm's ability to identify the active voice frames, especially in the lower SIR cases. However, for the lower SIR cases, some of the active voice frames are not identified. After application of the new separation algorithm, the VAD algorithm is able to identify all the active voice frames. Even in the lowest SIR case, it is able to identify the active voice frames.

V. EXPERIMENTS

A. Two voices with equal signal strength

Experiments were performed in an anechoic chamber. A 24 element linear array was used to collect the data. The array had an inter-elemental spacing of 0.75 in. and was sampled at 50 kHz. The signals were two voices. The signal of interest is a voice located at 0° and the interferer is moved from -20° to $+20^\circ$. Both sources were located 15 ft away from the center of the array at all time steps. The frequency of interest for the experiments is 3250 Hz. Figure 11 shows the time-angle plot of the beamformed data collected during in-chamber testing. The signal of interest was a male voice, while the interferer is a female voice. As seen in Fig. 11, the male voice is barely visible in most instances, while the interferer is always easily observable.

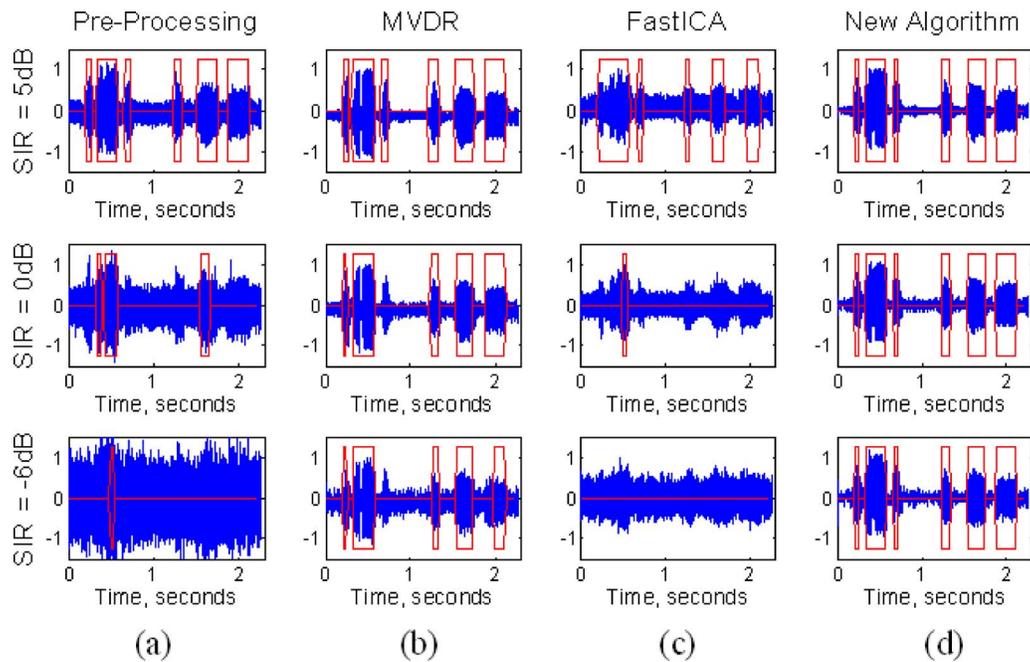


FIG. 10. (Color online) VAD on data containing one voice signal of interest and one white Gaussian interferer. The three rows show the results at varying SNR values. The active voice frames are enclosed with boxes. Column (a) shows the VAD results before the application of the new algorithm, column (b) shows the VAD results after MVDR, column (c) shows the VAD results after FastICA, and column (d) shows the VAD results after the new algorithm application.

The new algorithm was used to process the experimental data. To illustrate the performance of the algorithm in removing a slowly moving interferer, the crossing signal was chosen to be the interferer (opposite to the simulation cases). The result of the application of the new algorithm is shown in Fig. 12. In almost every time step, the signal of interest is now visible while the interferer is significantly suppressed when compared to Fig. 11. This supports the results from the simulation testing.

VI. CONCLUSION

This paper introduces a new method for interferer suppression based on a combination of spatial and ICA-based techniques. The new method has the capability to suppress

non-Gaussian interferers, ideal for telecommunication applications speech processing applications. The new algorithm allows for the separation of signals that are not simultaneously mixed. In addition, the introduction of spatial processing allows the permutation ambiguity of the output of the FastICA algorithm to be solved. Using the location of a source, we can associate that source with an output of the FastICA algorithm using cross-correlations.

Simulations showed the capabilities of the new algorithm in suppressing interferers in a variety of conditions. During the simulations, the interferer was suppressed while the signal of interest saw little suppression for 0 dB SIR. Even when the signals become spatially indistinct, the algorithm was still able to partially suppress the interferer. This is not possible using spatial nulling techniques. Anechoic ex-

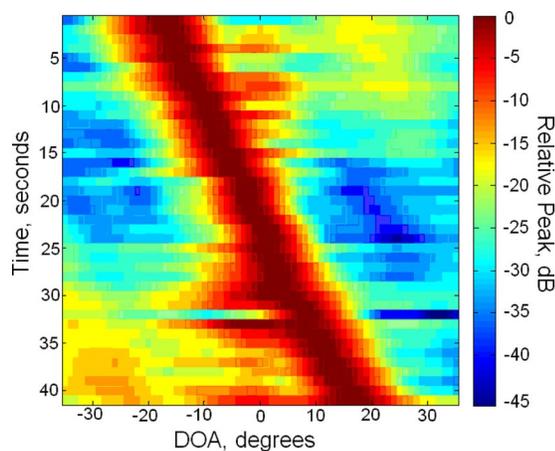


FIG. 11. (Color online) Bearing track recorder plot using a narrow-band delay and sum beamformer for the anechoic chamber testing. One signal is located 0° , while the other signal moves from -20° to $+20^\circ$ at a rate of $1^\circ/\text{s}$.

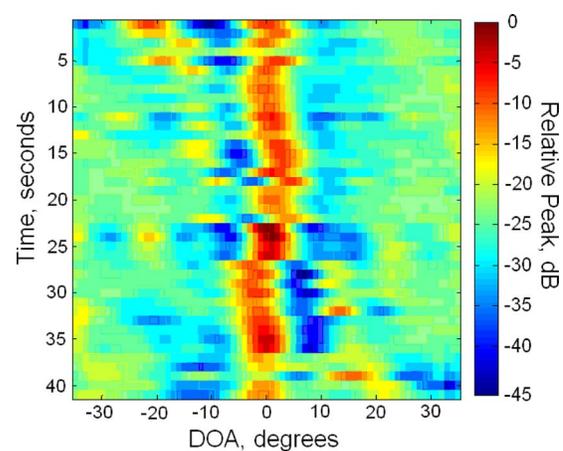


FIG. 12. (Color online) Bearing track recorder plot using a narrow-band delay and sum beamformer after the anechoic chamber data has been processed using the new algorithm. The moving interferer has been removed, and the signal of interest at 0° is visible.

perimental results confirmed the results obtained in the simulations. Although the signal of interest was not visible in many of the time steps of the original BTR, the new algorithm was able to increase its visibility by removing the interferer. The algorithm showed 20 dB of suppression of broadband interferers, even when the SIR dropped to -30 dB. In addition, the algorithm was shown to have good performance for SNR values above -18 dB, which is comparable to other spatial techniques. The performance of the new algorithm as a front-end to a VAD processor was shown to be superior to MVDR in the cases studied; however, future work will include a statistical performance on this enhancement using many types of voice signals and many more geometries.

- ¹J. C. Chen, K. Yao, and R. E. Hudson, "Acoustics source localization and beamforming: Theory and practice," *EURASIP J. Appl. Signal Process.* **4**, 359–370 (2003).
- ²L. T. Fialkowski, M. D. Collins, J. S. Perkins, and W. A. Kuperman, "Source localization in noisy and uncertain ocean environments," *J. Acoust. Soc. Am.* **101**, 3539–3545 (1997).
- ³Steven J. Spencer, "The two-dimensional source location problem for time differences of arrival at minimal element monitoring arrays," *J. Acoust. Soc. Am.* **121**, 3579–3594 (2007).
- ⁴V. Best, S. Carlile, C. Jin, and A. van Schaik, "The role of high frequencies in speech localization," *J. Acoust. Soc. Am.* **118**, 353–363 (2005).
- ⁵H. L. Van Trees, *Optimum Array Processing: Detection, Estimation, and Modulation Theory* (Wiley, New York, 2002), Part IV.
- ⁶J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE* **57**, 1408–1418 (1969).
- ⁷R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.* **53**, 1684–1696 (2005).
- ⁸P. J. Kootsookos, D. B. Ward, and R. C. Williamson, "Imposing pattern nulls on broadband array responses," *J. Acoust. Soc. Am.* **105**, 3390–3398 (1999).
- ⁹P. Jordan, J. A. Fitzpatrick, and C. Meskill, "Beampattern control of a microphone array to minimize secondary source contamination," *J. Acoust. Soc. Am.* **114**, 1920–1925 (2003).

- ¹⁰P. J. Bevelacqua and Constantine A. Balanis, "Minimum sidelobe levels for linear arrays," *IEEE Trans. Antennas Propag.* **55**, 3442–3449 (2007).
- ¹¹B. C. Ihm and D. J. Park, "Blind separation of sources using higher-order cumulants," *Signal Process.* **73**, 267–276 (1998).
- ¹²J. F. Cardoso, "Blind signal separation: statistical principles," *Proc. IEEE* **86**, 2009–2025 (1998).
- ¹³H. Saruwatari, S. Kurita, and K. Takeda, "Blind source separation combining frequency domain ICA and beamforming," *IEEE Trans. Audio, Speech, Lang. Process.* **14**, 666–678 (2006).
- ¹⁴C. M. Coviello and L. H. Sibul, "Blind source separation and beamforming: algebraic technique analysis," *IEEE Trans. Aerosp. Electron. Syst.* **40**, 221–234 (2004).
- ¹⁵S. Araki, S. Makimo, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures," *EURASIP J. Appl. Signal Process.* **11**, 1157–1166 (2003).
- ¹⁶L. C. Parra and C. V. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. Speech Audio Process.* **10**, 352–362 (2002).
- ¹⁷S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Upper Saddle River, NJ, 1996).
- ¹⁸M. E. G. D. Colin, J. Groen, and B. A. J. Quesson, "Experimental comparison of bearing estimation techniques for short passive towed sonar arrays," in *OCEANS*, '04 Vol. **2**, pp. 608–612.
- ¹⁹A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.* **7**, 1129–1159 (1995).
- ²⁰S.-i. Amari, "Natural gradient works efficiently in learning," *Neural Comput.* **10**, 251–276 (1998).
- ²¹A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.* **10**, 626–634 (1999).
- ²²K. Li, M. N. S. Swamy, and M. Omair Ahmad, "An improved voice activity detection using higher order statistics," *IEEE Trans. Speech Audio Process.* **13**, 965–974 (2005).
- ²³S. G. Tanyer and H. Ozer, "Voice activity detection in nonstationary noise," *IEEE Trans. Speech Audio Process.* **8**, 478–482 (2000).
- ²⁴Project Group 841, "Robust Voice Activity Detection and Noise Reduction Mechanism Using Higher-order Statistics," Department of Communication Technology, Institute of Electronic Systems, Aalborg University, 2005.