Binaural sound source localization using the frequency diversity of the head-related transfer function

Dumidu S. Talagala,[a] Wen Zhang, and Thushara D. Abhayapala
Research School of Engineering, College of Engineering and Computer Science, Australian National University, Canberra, Australia Capital Territory 0200, Australia

Abhilash Kamineni[b]
Department of Electrical and Computer Engineering, University of Auckland, Auckland 1142, New Zealand

(Received 28 July 2013; revised 19 November 2013; accepted 20 January 2014)

The spectral localization cues contained in the head-related transfer function are known to play a contributory role in the sound source localization abilities of humans. However, existing localization techniques are unable to fully exploit this diversity to accurately localize a sound source. The availability of just two measured signals complicates matters further, and results in front to back confusions and poor performance distinguishing between the source locations in a vertical plane. This study evaluates the performance of a source location estimator that retains the frequency domain diversity of the head-related transfer function. First, a method for extracting the directional information in the subbands of a broadband signal is described, and a composite estimator based on signal subspace decomposition is introduced. The localization performance is experimentally evaluated for single and multiple source scenarios in the horizontal and vertical planes. The proposed estimator's ability to successfully localize a sound source and resolve the ambiguities in the vertical plane is demonstrated, and the impact of the source location, knowledge of the source and the effect of reverberation is discussed.
 [http://dx.doi.org/10.1121/1.4864304]

PACS number(s): 43.60.Jn, 43.66.Qp, 43.66.Pn [ZHM] Pages: 1207–1217

I. INTRODUCTION

A. Motivation and background

Determining the exact location of a sound source is critical for the performance of many military, robotic, and communications applications. Although many solutions to this problem have been proposed, high spatial resolution requires sensor arrays with a large number of elements. In contrast, the auditory systems of humans and animals can provide similar levels of performance using just two sensors. A localization technique that exploits the knowledge and diversity of the head-related transfer function (HRTF) can therefore lead to high precision source location estimates using compact sensor arrays.

A sound wave propagating from a source to a listener is transformed as it encounters the body and pinna of an individual. The scattering and reflections caused by the head, torso, and pinna are both frequency and direction dependent, and can be characterized using the head-related transfer function.[1,2] A human being exploits three localization cues described by the HRTF for sound source localization;[3–5] interaural time difference (ITD) caused by the propagation delay between the ears, interaural intensity difference (IID) caused by the head shadowing effect, and spectral cues caused by reflections in the pinna. Perceptual experiments have shown that any change to the physical structure of the ear can affect the source localization performance of humans,[6] and reaffirms the importance of the HRTF for binaural source localization.[7–10]

Given that the HRTF at each potential source location is known, the objective of a localization algorithm is to perform the inverse mapping of the perceived localization cues to a source location. Of the many localization mechanisms in existence, techniques that map ITD to a source location remain the most popular by far. This is mainly due to the fact that the time difference of arrival (TDOA) is a natural estimator of the source location for two spatially separated sensors. A number of techniques based on correlation analysis,[11] beamforming,[12] and signal subspace concepts[13,14] have been developed for broadband source localization using sensors arrays in free space. Although the change in ITD with the source location can be modeled such that these techniques can be applied,[15] the use of a two sensor array can still lead to complications in the localization process. They are primarily attributed to the approximately spherical shape of the human head, which results in regions of similar ITD, known as a "cone of confusion."[16] Thus, emphasis on ITD as the primary localization cue could lead to front-to-back confusions and poor performance distinguishing between locations on a sagittal (vertical) plane. This has been demonstrated in binaural localization experiments using artificial systems,[17,18] as well as in perceptual experiments on human subjects.[19–21] At higher frequencies, it is believed that IID acts as the primary localization cue, where it is much harder to make an accurate determination of ITD. Experimental results indicate that an accurate estimate of the elevation angle is possible when the ITD or IID cues are combined with the spectral cues generated by the pinna.[7–10,19,22,23] Hence, it is well established that any binaural source

[a]Author to whom correspondence should be addressed. Electronic mail: dumidu.talagala@anu.edu.au
[b]This paper includes work performed by the author while at the Australian National University from December 2012 to January 2013.

localization mechanism must exploit all three localization cues within the HRTF for accurate localization of a source, in both azimuth and elevation.

A number of algorithms that incorporate spectral cues have been proposed for sound source localization using the HRTF. Typically, these methods extract the relevant acoustic features in the frequency domain of the received signal, and identifies the source position through a pattern matching,[24,25] statistical[26] or parametric modeling approach.[27] Correlation based approaches[28,29] represent a subclass of these methods, where the correlation coefficient is used as a similarity metric to identify distinct source locations. However, each method is not without its own drawbacks, such as the training required by the system or the high uncertainty differentiating the actual source location from the adjacent locations.

B. Contribution and organization

In our previous work on direction of arrival estimation,[30,31] the possibility of exploiting the diversity in the frequency domain for high resolution broadband direction of arrival estimation was explored. It was shown through simulations that the use of the diversity in the frequency domain of the HRTF (caused by the scattering and reflections by the body) can provide clearer separation of closely spaced source locations, in comparison to the traditional localization methods. This work investigates the application of these concepts to binaural sound source localization using a Knowles Electronic Manikin for Acoustic Research (KEMAR) manikin. The contributions and organization of this work are summarized below.

In Sec. II, a summary of the signal model of the proposed MUSIC (MUltiple SIgnal Classification)[32] signal subspace method for sound source location is presented. Next, the process of extracting subbands signals, their relationship with the HRTF, and the composite estimator used in a binaural localization system are described. The experiment setup and the performance metrics used to evaluate the source localization performance of different estimators are presented in Sec. III. In Sec. IV, the localization performance is evaluated in the horizontal and vertical planes using calibrated and direct path HRTF measurements. The localization results imply that the HRTFs play a significant role in the single source localization scenario in both the horizontal and vertical planes. However, it is observed that the HRTF information must be combined with source knowledge for localization in multiple source scenarios, and that the actual source location and content dominate the localization performance. The localization performance using the direct path HRTF measurements suggests the proposed method could be used for source localization in mildly reverberant environments. Section IV discusses these results and their implications, and is followed by the concluding remarks in Sec. V.

II. SYSTEM MODEL: THE HRTF AS A SOURCE LOCATION ESTIMATOR

The location of a far field sound source in three-dimensional space can be described in terms of two angles; a lateral angle α and an elevation angle β. Thus, the location of the qth source shown in Fig. 1 is given by $\Theta_q \equiv (\alpha_q, \beta_q)$. This creates two localization regions; the horizontal plane at $\alpha \in [-\pi/2, \pi/2]$, $\beta = 0$, and vertical planes at $\beta \in [0, 2\pi)$ for some fixed value of α. In the case of binaural sound source localization by humans, the lateral angle α is determined by ITD or IID, which due to the biological time-coding limitations, dominate at lower and higher frequencies respectively. On the other hand, the elevation angle β is determined through the analysis of spectral cues[9,19] that exist from the mid to high frequencies, caused by the scattering and reflection of sound waves off the pinna, head, and torso.[2,7,8] The effect of the different localization cues on the localization performance in the horizontal and vertical planes can therefore be evaluated independently by varying the range of frequencies used for localization. The following section describes how these cues can be incorporated into a general framework for binaural source localization.

A. Subband extraction of a broadband signal

A signal resulting from the convolution of a broadband signal and a channel impulse response can be characterized using a collection of narrow subband signals. In a previous study,[30] it was shown that these subband signals can be described as the sum of a weighted time varying Fourier series. For a binaural system using the HRTF, the subband expansion of a broadband signal can be expressed as follows.

The received signal at the left ear due to the qth source signal $s_q(t)$ is given by

$$y_q^L(t) = h^L(\Theta_q, t) * s_q(t), \tag{1}$$

where $h^L(\Theta_q, t)$ represents the channel impulse response between the source and the left ear in the direction Θ_q. This can be expanded further using a Fourier series approximation as

$$y_q^L(t) = \sum_{k=-\infty}^{\infty} H_q^L(k) S_q(k, t) e^{-jk\omega_0 t}, \tag{2}$$

where

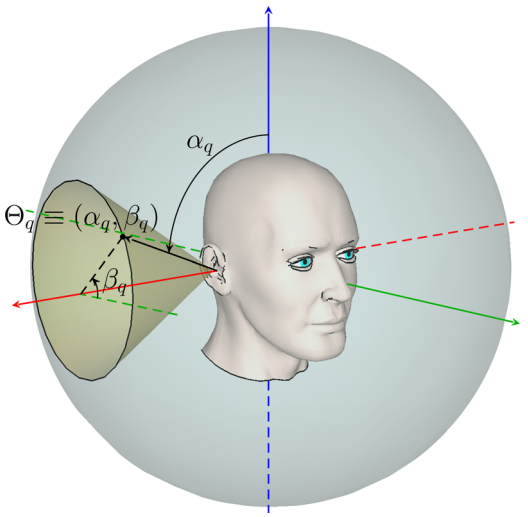


FIG. 1. (Color online) A source located on a "cone of confusion" in a sagittal coordinate system.

$$S_q(k,\,t)=\frac{1}{\sqrt{T}}\int_{t-T}^{t}s_q(\tau)e^{-jk\omega_0\tau}\mathrm{d}\tau$$

and

$$H_q^L(k)=\frac{1}{\sqrt{T}}\int_{0}^{T}h^L(\Theta_q,\,\tau)e^{-jk\omega_0\tau}\mathrm{d}\tau.$$

T is the length of the time limited channel impulse response $h^L(\Theta_q,\,t)$, $\omega_0=2\pi/T$ is the frequency resolution, $k\omega_0$ is the mid-band frequency of the kth subband signal, and $H_q^L(k)$, $S_q(k,\,t)$ represent the short-time Fourier transform coefficients of $h^L(\Theta_q,\,t)$ and $s_q(t)$ respectively.[30]

The k_0th subband signal in Eq. (2) is given by

$$y_q^L(k_0,\,t)=H_q^L(k_0)S_q(k_0,\,t)e^{-jk_0\omega_0t},\tag{3}$$

where $H_q^L(k_0)S_q(k_0,\,t)$ contains the location and source information at the frequency $k_0\omega_0$. Since the carrier term $e^{-jk_0\omega_0t}$ is devoid of any location information, it can be eliminated to obtain a set of focused subband signals (containing purely location and source information) for all $k_0=1,\ldots,K$. Conceptually this is the process of bandpass filtering and down-conversion, shown in Fig. 2, and can be implemented as a series of mixing and low pass filtering operations. Hence, the extracted k_0th subband signal becomes

$$\begin{aligned}\hat{y}_q^L(k_0,\,t)&\triangleq LPF\left\{y_q^L(t)e^{jk_0\omega_0t}\right\}\\&=LPF\left\{\sum_{k=-\infty}^{\infty}H_q^L(k)S_q(k,\,t)e^{-j(k-k_0)\omega_0t}\right\}\\&=H_q^L(k_0)S_q(k_0,\,t),\end{aligned}\tag{4}$$

where LPF is a low pass filter operation using a filter cut off bandwidth $\omega_c\le\omega_0/2$.

B. Signal subspace decomposition

Consider a binaural source localization scenario where the number of active sources is Q. The measured signal at the left ear is then given by

$$y^L(t)=\sum_{q=1}^{Q}y_q^L(t)+n^L(t),\tag{5}$$

where $n^L(t)$ is the noise measured at left ear. From Eqs. (4) and (5), the extracted subband signal at the frequency $k_0\omega_0$ can be expressed as

$$\hat{y}^L(k_0,\,t)=\sum_{q=1}^{Q}H_q^L(k_0)S_q(k_0,\,t)+\hat{n}^L(k_0,\,t),\tag{6}$$

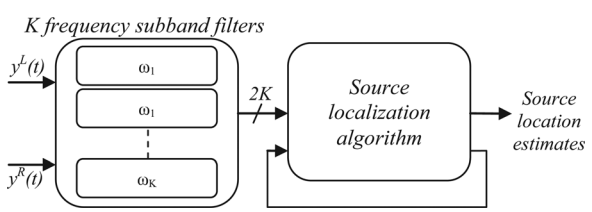


FIG. 2. Filter bank model of the binaural sound source localization system.

where $\hat{n}^L(k_0,\,t)$ represents the bandpass filtered noise signal $n^L(t)$ at the frequency $k_0\omega_0$. Thus, by separating each binaural signal into K subbands as seen in Fig. 2, a set of $2K$ subband signals can be extracted.

The set of extracted subband signals in Eq. (6) can be expressed using the vector notation

$$\hat{\mathbf{y}}=\sum_{q=1}^{Q}\mathbf{D}_q\mathbf{s}_q+\hat{\mathbf{n}},\tag{7}$$

where

$$\hat{\mathbf{y}}=\begin{bmatrix}\hat{y}^L(1,\,t) & \hat{y}^R(1,\,t) & \cdots & \hat{y}^R(K,\,t)\end{bmatrix}_{1\times 2K}^T,$$

$$\mathbf{D}_q=\begin{bmatrix}H_q^L(1) & 0 & \cdots & 0\\ H_q^R(1) & 0 & \cdots & 0\\ 0 & H_q^L(2) & \cdots & 0\\ 0 & H_q^R(2) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & H_q^L(K)\\ 0 & 0 & \cdots & H_q^R(K)\end{bmatrix}_{(2K\times K)},$$

$$\mathbf{s}_q=\begin{bmatrix}S_q(1,\,t) & S_q(2,\,t) & \cdots & S_q(K,\,t)\end{bmatrix}_{1\times K}^T,$$

and

$$\hat{\mathbf{n}}=\begin{bmatrix}\hat{n}^L(1,\,t) & \hat{n}^R(1,\,t) & \cdots & \hat{n}^R(K,\,t)\end{bmatrix}_{1\times 2K}^T.$$

Equation (7) is the familiar system equation used by signal subspace methods for DOA estimation,[13,32] whose signal and noise subspaces can be identified as follows. Reformulating the summation in Eq. (7),

$$\hat{\mathbf{y}}=\mathbf{D}\mathbf{s}+\hat{\mathbf{n}},\tag{8}$$

where

$$\mathbf{D}=\begin{bmatrix}\mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_Q\end{bmatrix}_{2K\times KQ}$$

and

$$\mathbf{s}=\begin{bmatrix}\mathbf{s}_1^T & \mathbf{s}_2^T & \cdots & \mathbf{s}_Q^T\end{bmatrix}_{1\times KQ}^T.$$

For uncorrelated source and noise signals, this implies that the correlation matrix of the received signals can be expressed as

$$\mathbf{R}\triangleq E\{\hat{\mathbf{y}}\hat{\mathbf{y}}^H\}=\mathbf{D}E\{\mathbf{s}\mathbf{s}^H\}\mathbf{D}^H+E\{\hat{\mathbf{n}}\hat{\mathbf{n}}^H\},\tag{9}$$

where $E\{\cdot\}$ represents the expectation operator. Eigenvalue decomposition of Eq. (9) can now be used to identify the signal and noise subspaces of $\mathbf{R}$. Thus,

$$\mathbf{R} = \begin{bmatrix} \hat{\mathbf{D}}_\mathbf{S} & \hat{\mathbf{D}}_\mathbf{N} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda_S} & 0 \\ 0 & \mathbf{\Lambda_N} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{D}}_\mathbf{S}^H \\ \hat{\mathbf{D}}_\mathbf{N}^H \end{bmatrix}, \tag{10}$$

where $\mathbf{\Lambda_s}$, $\mathbf{\Lambda_N}$ are diagonal matrices containing the eigenvalues of $E\{\mathbf{ss}^H\}$, $E\{\hat{\mathbf{n}}\hat{\mathbf{n}}^H\}$, and $\hat{\mathbf{D}}_\mathbf{S}$, $\hat{\mathbf{D}}_\mathbf{N}$ contain the eigenvectors of the signal and noise subspaces, respectively. The orthogonal property of these subspaces,

$$\text{span}(\mathbf{D}_q) \perp \text{span}(\hat{\mathbf{D}}_\mathbf{N}), \tag{11}$$

can now be exploited to estimate the source location using the HRTF information contained in $\mathbf{D}_q$.

C. Source location estimation

The process of subbanding a broadband signal and collecting the information from multiple subbands to be used by a signal subspace localization technique was described in the previous subsection. In our previous work on direction of arrival estimation in multi-channel systems,[30] it was shown that the existence of a noise space is conditional, which leads to multiple localization scenarios. In the case of a binaural system, this implies that two localization scenarios may exist; single source localization and localizing multiple known sources. The source location estimates of each case can be determined as follows.

1. Single source localization

Consider the scenario of localizing a single sound source ($Q = 1$) whose subband signals are independent of each other [i.e., $\dim(\mathbf{\Lambda_S}) = K$, where $\dim(\cdot)$ is the dimension operator]. This results in a matrix $\mathbf{\Lambda_N}$ at the limiting case, where

$$\dim(\mathbf{\Lambda_N}) = 2K - \dim(\mathbf{\Lambda_S}) \geq K. \tag{12}$$

This represents the worst case scenario for the existence of $\hat{\mathbf{D}}_\mathbf{N}$, but the source position can still be estimated from

$$\hat{P}(\alpha_q, \beta_q) = \left\{ \sum_{k=1}^{K} \frac{|\mathbf{d}_q(k)^H \mathbf{P_N} \mathbf{d}_q(k)|}{|\mathbf{d}_q(k)^H \mathbf{d}_q(k)|} \right\}^{-1}, \tag{13}$$

where $\mathbf{d}_q(k)$ is the kth column of $\mathbf{D}_q$ and $\mathbf{P_N} = \hat{\mathbf{D}}_\mathbf{N}\hat{\mathbf{D}}_\mathbf{N}^H$ is the measured noise space.[30] The improvement in resolution is achieved due to the additional diversity obtained through the summation of subbands across $k = 1, \ldots, K$.

2. Multiple source localization

The process of localizing multiple sound sources localized in either time or frequency represents a variation on the single source problem described above. This scenario considers sound sources that are localized in neither time nor frequency. Equation (12) leads to the realization that a noise space will not exist if two or more sound sources are independent across subbands and each other. Hence, for multiple source localization, it is critical that the sources exhibit some correlation between its subband signals speech signals are good examples of such sources while remaining uncorrelated between each other.

The relationship between subbands of the qth source can be expressed using the source correlation matrix

$$E\left\{\mathbf{s}_q \mathbf{s}_q^H\right\} = \mathbf{U}_q \mathbf{\Lambda}_q \mathbf{U}_q^H, \tag{14}$$

where $\mathbf{\Lambda_q}$ contains the most significant eigenvalues [i.e., $\dim(\mathbf{\Lambda}_q) \ll K$] and $\mathbf{U}_q$ is a matrix of the corresponding eigenvectors. $\mathbf{D}$ in Eq. (9) can now be reformulated as

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_1\mathbf{U}_1 & \mathbf{D}_2\mathbf{U}_2 & \cdots & \mathbf{D}_Q\mathbf{U}_Q \end{bmatrix}_{2K \times LQ}, \tag{15}$$

where $L = \dim(\mathbf{\Lambda}_q)$. This implies that multiple source localization requires some knowledge of the source, and the location can be estimated for each source q' as

$$\hat{P}(\alpha_q, \beta_q, q') = \left\{ \sum_{l=1}^{L} \frac{|\tilde{\mathbf{d}}_q(l)^H \mathbf{P_N} \tilde{\mathbf{d}}_q(l)|}{|\tilde{\mathbf{d}}_q(l)^H \tilde{\mathbf{d}}_q(l)|} \right\}^{-1}, \tag{16}$$

where $\tilde{\mathbf{d}}_q(l)$ is the lth column of $\mathbf{D}_q\mathbf{U}_{q'}$.[30] Equation (16) assumes that

$$\text{span}(\mathbf{U}_q) \perp \text{span}(\mathbf{U}_{q'}), \tag{17}$$

and any correlation between sources can therefore result in ambiguous source location estimates

III. EXPERIMENTAL SETUP

A. Equipment and room setup

The experiments were conducted in a 3.2 m × 3.2 m × 2 m semi-anechoic audio laboratory at the Australian National University. The walls, floor, and ceiling of the chamber were lined with acoustic absorbing foam to minimize the effects of reverberation. The room is mildly reverberant, with a reverberation time RT_{60} of approximately 250–300 ms. A G.R.A.S. KEMAR manikin Type 45BA head and torso simulator was located at the center of the room, and used to measure the binaural signals of an average human subject. The manikin was fitted with the G.R.A.S KB0061 and KB0065 left and right pinna, while Type 40AG polarized pressure microphones were used to measure the received signals at the entrance to the ear canal. A G.R.A.S. Type: 26AC preamplifier was used to high pass filter the left and right ear signals before analog to digital conversion using a National Instruments USB-6221 data acquisition card at a sampling rate of 44.1 kHz.

The stimuli were delivered through a ART Tubefire 8 preamplifier coupled to a Yamaha AX-490 amplifier, which drives a set of Tannoy System 600 loudspeakers. The speakers were located at fixed positions 1.5 m away from the KEMAR manikin and a change in source location was simulated by rotating the manikin. This was achieved by mounting the KEMAR on a LinTech 300 series turntable connected to a QuickSilver Controls Inc. SilverDust D2 IGB servo motor controller, which allows the accurate positioning of the source in azimuth. Positioning of the source on a

vertical plane was carried out using a speaker mounted on an elevation adjustable hoop of 1 m radius. This equipment setup allows the simulation of sound sources in both the horizontal plane of the KEMAR manikin as well as in any vertical plane of interest as described in the experiment scenarios below.

B. HRTF measurement

The HRTFs of the KEMAR manikin were computed indirectly, by first measuring its head-related impulse response (HRIR) in the specified directions. In this experiment, a 4.4 ms duration chirp signal with frequencies between 300 Hz and 10 kHz is output by the loudspeaker, and used as the stimulus for HRIR measurement. The duration of the stimulus was selected such that the direct path signal and any reflected signals (due to the scatterers within the measurement room) do not overlap at the receiver microphone. Ten chirp pulses were transmitted with 100 ms of silence between chirps, to ensure that the dominant reverberation signals of a previous pulse does not overlap with the adjacent direct pulse. The measured signals were then processed, by aligning the first peaks of the ten chirp signals, and averaged to obtain the received signal for a chirp input. Finally, the received signal was low pass filtered and equalized to obtain the measured HRIR of the KEMAR manikin in the specified direction.[33] The HRIRs are measured at 5° intervals in the horizontal and vertical planes of interest.

The measured HRIRs in the horizontal and vertical planes can now be used to calculate the HRTF in any direction using a discrete Fourier transform (DFT), which can then be used in place of the Fourier transform coefficients of the channel impulse response in Eq. (2). It should be noted that this measured HRTF is a calibrated HRTF for this particular reverberant measurement room, and is susceptible to change with different room conditions. However, the structure of the stimulus signal can be used to identify the direct path and the reverberant path contributions to the measured HRIR as shown in Fig. 3. Thus, two types of HRTFs can be considered; the calibrated HRTFs which include the reverberation effects, and the direct path HRTFs derived from the direct component of the measured HRIR. The truncation length of the direct path HRIR was determined by analysis of the measured HRIR and identifying the onset of the first reflections, for an average direct to reverberant path power ratio of approximately 11 dB.

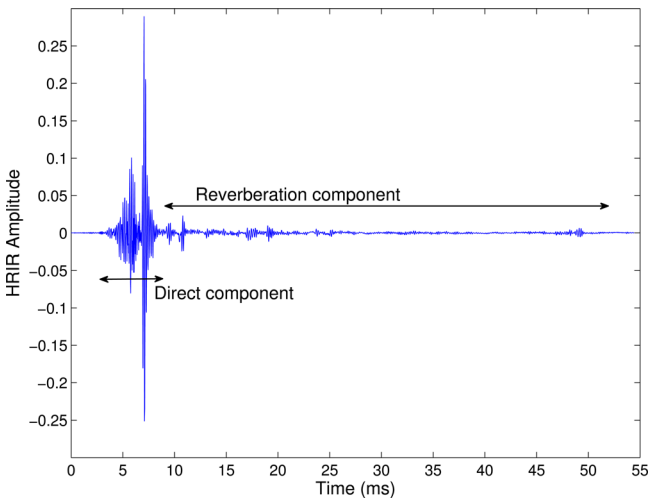


FIG. 3. (Color online) Direct and reverberant path components of the measured HRIR for the right ear of the KEMAR manikin in the azimuth direction 85°.

C. Stimuli

The characteristics of sound sources encountered in an everyday source localization scenario vary significantly depending upon the phenomenon generating the sound. As an example, the majority of the energy in human speech occupies a frequency bandwidth of 100−4000 Hz, and exhibits a high correlation between subbands. In contrast, motor vehicle or aircraft sounds occupy a much lower bandwidth, and may not be correlated across frequency subbands. Yet another source may be highly narrowband or uniformly distributed in frequency. Since the frequency bandwidth is a parameter of the source location estimator, the localization performance will naturally be impacted by the characteristics of the individual sound sources. Thus, from a practical standpoint, the performance of the source localization technique is best approximated by the average localization performance for a range of sound sources.

In this experiment, the stimuli are selected to satisfy a number of criteria, with respect to the frequency bandwidth of the source, inter-subband correlation and energy distribution in frequency. These included real-world sound sources such as speech, music, motor vehicle and aircraft noises, as well as a simulated white Gaussian noise source. The mechanical noise sources correspond to a frequency bandwidth of approximately 1500 Hz, while the speech and music sources include the frequencies up to 4500 Hz and above. The characteristics of an ideal source used to develop the localization algorithm is satisfied by the white noise source, and the different types of sound sources exhibit different degrees of inter-subband correlation. Thus, the stimuli is a collection of 10 sound sources between 2 and 3 s duration stored as 16 bit WAV files sampled at 44.1 kHz. Each sound source was separated by a 2 s silence period, and were reproduced in a single trial run for each source location.

D. Experiment scenarios

Two main experimental scenarios were considered; single and multiple source localization. For the single source scenario, the localization performance is investigated in the horizontal plane and a vertical plane, where the dominant localization cues are ITD and spectral cues, respectively. The received signals at the ears of the KEMAR manikin were recorded at a sampling rate of 44.1 kHz, and the sound sources were located at approximately 30° intervals on both planes. Evaluating the multiple source localization capability was restricted to two simultaneously active sound sources in the horizontal plane, where the sources were located in the combinations of the front, back, and sides of the KEMAR manikin.

The received signals were preprocessed as described in Sec. II with a 50 Hz bandpass filter located at 100 Hz

intervals above 300 Hz. The resulting subband signals were then used to evaluate the localization performance with increasing frequency (i.e., increasing spectral cues). Audio bandwidths of 1500 and 4500 Hz were selected to evaluate the impact of bandwidth on the localization performance. The 1500 Hz bandwidth broadly corresponds to the audio bandwidth of the low frequency noise sources (motor vehicle and aircraft noise), while the 4500 Hz bandwidth corresponds to the bandwidth of the speech and music stimuli. The upper bandwidth figure was selected based on previous simulations,[30] where it was found that the improvement in localization accuracy was marginal for audio bandwidths greater than 4000 Hz. The selected audio bandwidths also correspond to different localization cues; a low frequency region dominated by ITD and a high frequency region dominated by IID and spectral cues. Thus, the performance of the source location estimator in these conditions can also be used to evaluate the relative importance of the different localization cues.

In addition to the effects of varying the audio bandwidths described above, the effect of available knowledge of the HRTF has also been considered. In this context, the measured HRTFs in Sec. III B represent a calibrated HRTF dataset for the measurement room, while the direct path HRTFs represents a free field response to a sound source. The direct path HRTFs are independent of the acoustic environment (essentially a HRTF measurement in an anechoic chamber), and can be considered as a generic set of HRTF data to be used for source localization in any environment. Hence, the two sets of HRTFs can be used to evaluate the impact of reverberation on the source localization performance.

E. Localization performance metrics

The localization performance is compared with three other source localization techniques; wideband MUSIC[13] based on signal subspace estimation, GCC-PHAT[11] (Generalized Cross Correlation–PHAse Transform) based on cross-correlation analysis of the received signals (primarily an ITD based estimator) and matched filtering using the measured HRIR data. Figure 4 illustrates the "normalized localization confidence"[30] curve, i.e., $\hat{P}$ in Eqs. (13) and (16), of the different localization techniques for a source located on the horizontal plane at 339°. The localization confidence curve can be interpreted as a probability distribution of a source being present at a particular location (0 and 1 at the least and most probable source locations, respectively), and is normalized to ensure the different localization techniques remain comparable.

The performance metrics used in this study can therefore be described using the localization confidence curves as follows. For example, the secondary peak of GCC-PHAT at 200° indicates a false detection; specifically a front to back localization error corresponding to a position on the cone of confusion on the left side of the KEMAR manikin. The width of each peak above a specified localization confidence threshold indicates the uncertainty of the location estimate, where a wider peak suggests a source maybe located at one of many locations. Finally, the localization accuracy of a particular technique is described by the difference between

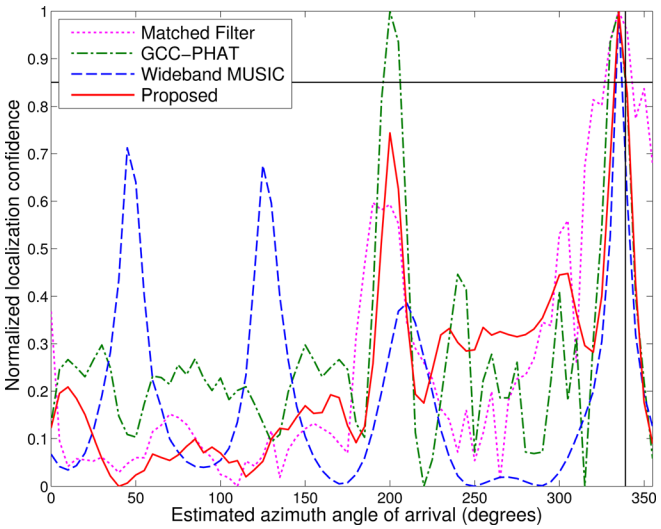


FIG. 4. (Color online) Normalized localization confidence curves for a sound source at the 339° azimuth position in the horizontal plane.

the estimated source location and the actual source location indicated by the solid vertical line. The localization confidence threshold is selected to achieve a balance between the number of false detections and the localization accuracy, and is fixed at 0.85 in this study. The overall performance of each source localization technique is obtained by averaging the normalized localization confidence curves for the different stimuli used in the experiment. The performance metrics can be defined as follows.

(1) Localization accuracy: Percentage of the total estimation scenarios where $|\hat{\Theta}_q - \Theta_q| \leq 5^\circ$. The acceptable estimation error of 5° is obtained to correspond to the spatial resolution of the HRIR measurements in Sec. III B, and the average horizontal plane localization accuracy of humans.[3,34]
(2) False detections: Average number of false source locations detected per experiment scenario. This includes a number of possible localization errors such as azimuth errors, elevation errors and front to back ambiguities (quadrant errors).
(3) Location uncertainty: Average uncertainty of an estimated source location in degrees. As described previously, for each region that exceeds the specified normalized localization confidence threshold, the localization confidence curve may span multiple source locations. A wider region suggests that the location of the peak is more uncertain, hence the width of the region can be considered a measure of the angular uncertainty of the source location estimate. Thus, the localization uncertainty acts as an indirect measure of the variance of the localization estimates.
(4) Median absolute error: The median error between the actual and estimated source locations. The median of the absolute error is selected to eliminate the skewness of the estimation error primarily introduced by front to back localization errors. The median error is therefore comparable with the localization performance of human beings.[3,34]

IV. RESULTS AND DISCUSSION

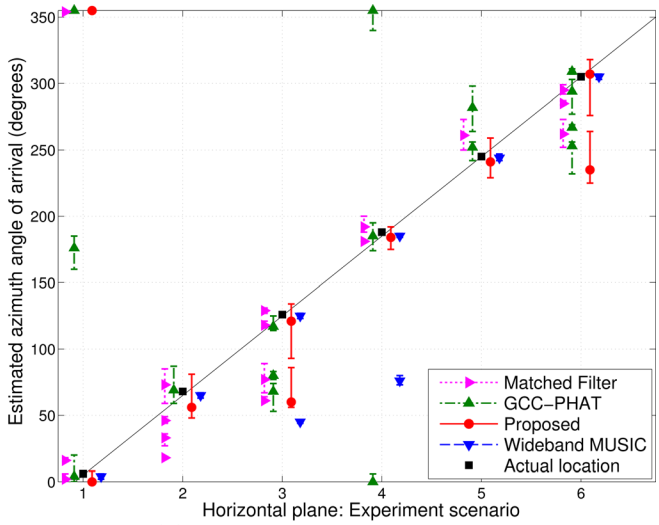
The previous section describes the experiment parameters and the source localization scenarios used to evaluate the performance of the proposed algorithm. The experiments for single and multiple sound sources represent two distinct localization scenarios. The impact of the audio bandwidth (number of subbands), actual source location, and the nature of the measured HRTFs on the localization performance for the two scenarios are discussed below.

A. Single source localization performance

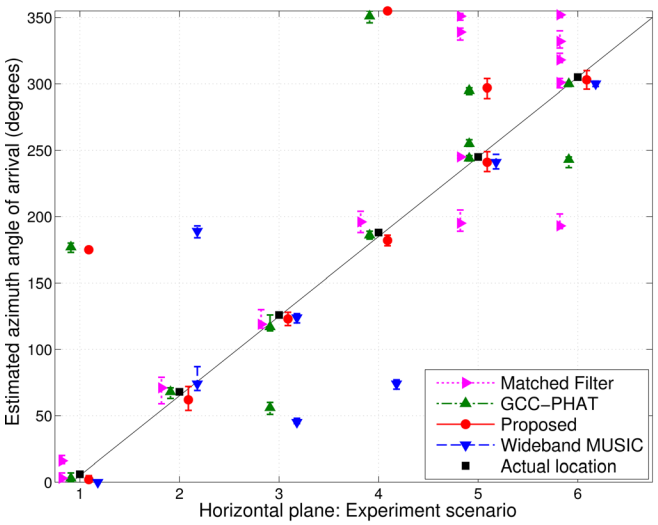
Although the localization of sound sources by humans generally occurs in the presence of background and other concurrent sources, the localization of a single active sound source represents the most common and widely evaluated localization scenario. Further, it represents the most basic estimation scenario for the two sensor system considered in this study, where the localization algorithms can be applied with no prior knowledge or restrictions on the time-frequency distribution of each source. Hence, the single source scenario provides a good baseline for the localization performance of each algorithm. In this experiment, the source localization performance is evaluated in the horizontal and 15° vertical planes, where the sources are located approximately 30° apart from each other. This corresponds to an azimuth angle between 0° and 360° for the horizontal plane (i.e., $\alpha \in [-\pi/2, \pi/2]$, $\beta = \{0, \pi\}$), and an elevation angle between −30° and 210° for the vertical plane (i.e., $\alpha = \pi/12$, $\beta \in [-\pi/6, 7\pi/6]$), respectively. The localization performance of the proposed method is compared with wideband MUSIC, GCC-PHAT, and the matched filtering algorithms. A source is detected when the normalized localization confidence curve exceeds a threshold of 0.85. For the sake of clarity, the source localization performance of a selected set of source locations is illustrated in Figs. 5 and 6 using the calibrated measured HRTFs in Sec. III B. A discernible difference is not observed between the localization performance of the proposed technique for different types of sources, likely due to the independence of the subband signals at the time scales being considered and the minimal differences in high frequency signal to noise ratios (SNRs). The overall localization performance metrics under different conditions are summarized in Tables I and II.

1. Horizontal plane localization

Figures 5(a) and 5(b) illustrate the localization performance for a single source in the horizontal plane, using the calibrated measured HRTFs and audio bandwidths of 1500 and 4500 Hz, respectively. In general, from the results in Table I, an improvement in the localization accuracy of the proposed method is observed with the increasing audio bandwidth, while the localization uncertainty and false detections are reduced. A similar response is seen in the comparative methods, but they still suffer from a greater number of false detections. This improvement in performance with increasing bandwidth can be explained by considering the type of diversity present and the localization cues exploited by each



(a) 1500 Hz audio bandwidth

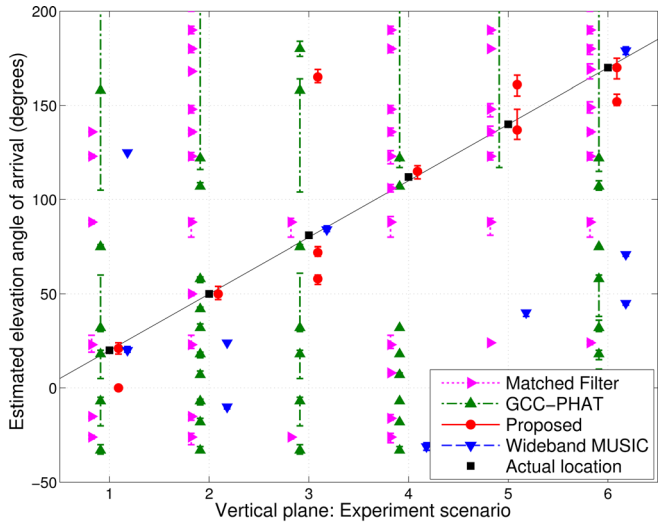


(b) 4500 Hz audio bandwidth

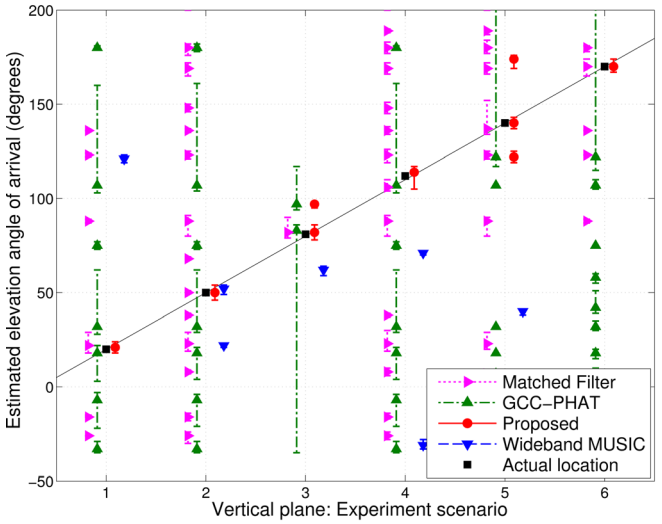
FIG. 5. (Color online) Source location estimates for a single source located at various positions in the horizontal plane. Results are averages of 11 experiments using different sound sources at 20 dB SNR and the calibrated measured HRTFs.

algorithm. For example, in Fig. 5(a) the experiment scenarios 3 and 6 indicate a false detection of a source at both the front and back positions on a cone of confusion. At the lower audio bandwidth ITD is dominant, thus the analysis of purely ITD information will identify a source at both locations. Yet as the audio bandwidth is increased, the IID and spectral localization cues provide additional information that resolves this ambiguity. However, the source location estimators that do not exploit this information will still identify two source locations at the higher audio bandwidth, as indicated by the GCC-PHAT localization results for the same experiment scenarios in Fig. 5(b).

The impact of the actual source location can be observed in experiments 2 and 4 in Fig. 5(a), representing a source at the side and back of the KEMAR, respectively. The higher localization uncertainty can be explained by considering the rate of change of the ITD (relatively smaller change in ITD with source location) as it approaches the sides of the KEMAR manikin. The impact is similar on each



(a) 1500 Hz audio bandwidth



(b) 4500 Hz audio bandwidth

FIG. 6. (Color online) Source location estimates for a single source located at various positions in the 15° vertical plane. Results are averages of 11 experiments using different sound sources at 20 dB SNR and the calibrated measured HRTFs.

algorithm, and a reduction in the localization uncertainty is observed with increasing audio bandwidth in Fig. 5(b). Once again this improvement can be attributed to the additional diversity information obtained through the IID and spectral localization cues. The results of the localization accuracy, false detections and localization uncertainty are mirrored in the median absolute error measurements in Table I. This measure enables the comparison with human localization capabilities,[3,34] and suggests the performance of the proposed technique is similar to humans using both calibrated and direct path HRTFs at the higher audio bandwidths.

The localization performance metrics for the same scenarios using the calibrated and direct path measured HRTFs are also summarized in Table I. It is seen that the exclusion of the reverberation effects in the HRTFs can have a significant effect on the performance of each algorithm, with up to a 10% performance penalty on the subspace source localization techniques at the higher audio bandwidth. However, the overall impact on the performance of the proposed method is minimal, which achieves an average source localization accuracy greater than 85% on the horizontal plane at an audio bandwidth of 4500 Hz. In the case of the matched filter, the drop in accuracy is expected due to the mismatch of the direct path and calibrated HRTFs. The impact on wideband MUSIC is somewhat less intuitive, but can be attributed to the misalignment of the coherent signal subspaces at each frequency of the actual and direct path HRTFs. Thus, the estimator essentially operates on ITD information, reducing the accuracy of the localizations estimates. In the case of the proposed localization approach, the increase in the median absolute error is minimal, possibly due to the higher direct to reverberant path power ratio at these audio bandwidths. Overall, the results suggest that the proposed technique achieves good horizontal plane source localization performance in mildly reverberant environments using only the direct path HRTFs, by exploiting the subtle directional information at the lower frequencies of the HRTF.

2. Vertical plane localization

Localization on a vertical plane typically presents a challenge for two sensor localization techniques. This is primarily due to the distribution of the source locations, where each potential source location is on the same cone of confusion, and therefore has the same ITD. Hence, estimating the source location by analyzing the ITD information will result in sources being detected at every possible position, and the accurate estimation would require the exploitation of the IID and spectral cues.

Figures 6(a) and 6(b) illustrate the localization performance for a single source located in the 15° vertical plane, using the calibrated measured HRTFs and audio bandwidths of 1500 and 4500 Hz, respectively. As expected, the ITD based location estimates GCC-PHAT and matched filtering identify large regions of potential source locations, while wideband MUSIC suffers from numerous false detections. In contrast, the proposed technique is capable of accurately localizing the sound sources. This is attributed to the exploitation of the diversity in the frequency domain, which is minimal in the case of wideband MUSIC, due to the dimensionality reduction of the focusing and summation processes. Naturally, increasing the audio bandwidth introduces more IID and spectral localization cues, which in turn improves the localization accuracy and reduces the localization uncertainty. The actual source location does not appear to have a significant impact on the localization performance in any given vertical plane, but a reduction in performance is expected for vertical planes on either side of the manikin (closer to a particular ear) due to the shrinking region of interest. However, unlike humans,[3,34] the median absolute error of the locations estimates are similar to the horizontal plane localization scenario at the higher bandwidth as shown in Table II, irrespective of the measured HRTFs used for localization.

The overall performance of the localization algorithms using the calibrated and direct path measured HRTFs for the 15° vertical plane is also summarized in Table II. In general,

TABLE I. Source localization performance using the calibrated and direct path measured HRTFs in the horizontal plane.

	Calibrated HRTFs				Direct path HRTFs		
Performance criteria	Matched Filter	GCC-PHAT	Proposed	Wideband MUSIC	Matched Filter	Proposed	Wideband MUSIC
Accuracy $\leq \pm 5°$: 1500 Hz	37.50%	54.17%	79.17%	91.67%	0.00%	37.50%	54.17%
Accuracy $\leq \pm 5°$: 4500 Hz	62.50%	66.67%	95.83%	58.33%	0.00%	87.50%	54.17%
Median absolute error: 1500 Hz	7.5°	51.5°	3.0°	1.0°	233.5°	41.0°	9.0°
Median absolute error: 4500 Hz	4.0°	9.0°	2.5°	6.0°	188.0°	3.0°	8.0°
Uncertainty: 1500 Hz	11.31°	21.96°	28.13°	6.37°	5.06°	25.36°	5.84°
Uncertainty: 4500 Hz	12.42°	8.42°	11.90°	8.53°	3.97°	11.29°	6.48°
False detections: 1500 Hz	1.54	1.62	0.50	0.29	6.17	1.58	1.00
False detections: 4500 Hz	1.33	1.33	0.29	0.83	3.00	0.83	1.17

a significant reduction in the localization performance is observed using the direct path HRTFs. Although the higher audio bandwidth improves the performance, the localization uncertainty and false detections have increased with respect to the horizontal plane scenario considered previously. These results can be explained by considering the localization cues being exploited, and the effect of reverberation on these localization cues. For example, IID and spectral cues dominate the localization process in the vertical plane, and present themselves as fluctuations in head-related transfer function in the frequency domain. However, reverberation (can be considered as the cumulative effects of multiple image sources) will significantly alter the profile of the transfer function in the frequency domain, and drastically distort the perceived IID and spectral cues. Since the perceived localization cues may be better correlated to another source location, false detections and localization uncertainty are expected to rise. Overall, the results suggest that the combination of the proposed technique and the direct path HRTFs could provide good localization performance in general, albeit at reduced localization accuracy (typically front to back ambiguities) and greater uncertainty in mildly reverberant conditions.

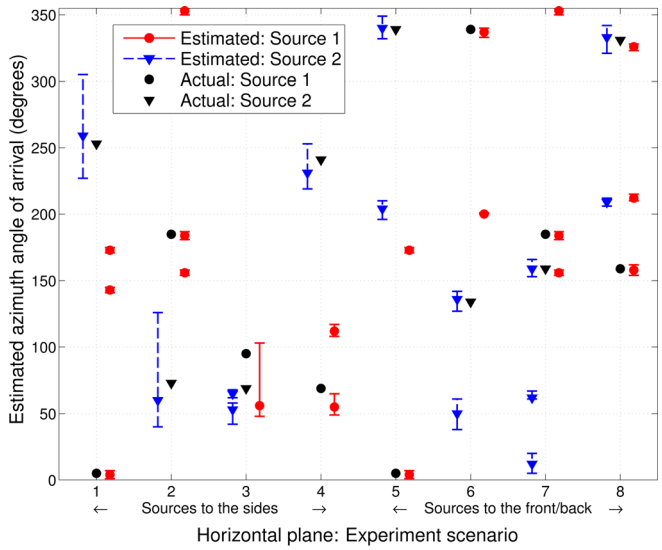
B. Multiple source localization performance

Localizing multiple simultaneously active sound sources is a challenge in binaural localization due to the availability of just two measured signals. Although the problem can be transformed into a single source localization problem if time-frequency constraints can be applied, knowledge of the source must be used to establish the locations of simultaneously active sources, as described in Sec. II C 2. It has been shown that human listeners exploit the inter-subband relationships for source segregation,[35] and we can therefore assume that similar information, for example inter-subband information relevant to a particular known speech signal, can be made available to the source localizer.

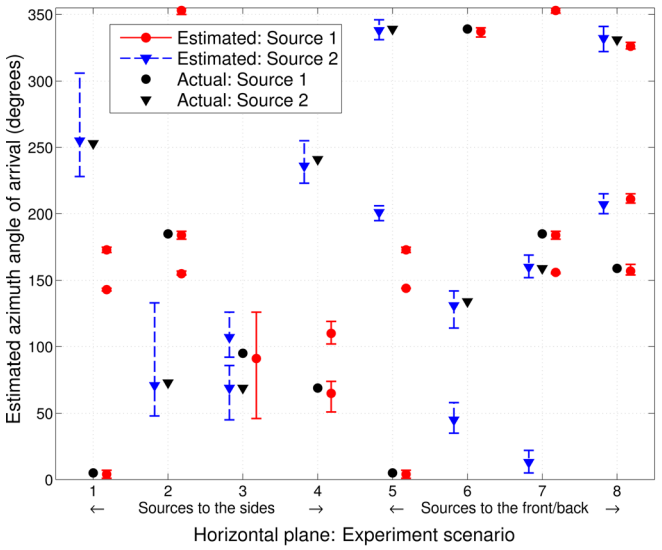
In this experiment scenario, two sound sources are arbitrarily positioned in the front, side and back regions of the KEMAR manikin in the horizontal plane. The localization performance of the proposed algorithm is evaluated in these regions, which correspond to the three localization regions known to exist in humans.[36] The assumed knowledge of the inter-subband correlation is imperfect, and includes the eigenvectors corresponding to the eigenvalues greater than 10% of the dominant eigenvalue. The source location estimates are calculated as described in Eqs. (14)–(16), and are illustrated in Fig. 7 for an audio bandwidth of 4500 Hz. A source is detected when the normalized localization confidence curve exceeds a threshold of 0.85. Figures 7(a) and 7(b) illustrate the localization performance using the calibrated and direct path measured HRTFs, respectively. For clarity, a selected set of source locations are presented, and are grouped into side-on source locations and front/back source locations that correspond to the experiment scenarios 1−4 and 5−8, respectively. The localization performance is summarized in Table III for the two main classes of stimuli, i.e., the noise like ideal source and the real-world sources. The impact of the partial knowledge of the inter-subband correlation is more acute on the real-world source localization performance, as can be expected.

TABLE II. Source localization performance using the calibrated and direct path measured HRTFs in the 15° vertical plane.

	Calibrated HRTFs				Direct path HRTFs		
Performance criteria	Matched Filter	GCC-PHAT	Proposed	Wideband MUSIC	Matched Filter	Proposed	Wideband MUSIC
Accuracy $\leq \pm 5°$: 1500 Hz	75.00%	12.50%	81.25%	31.25%	0.00%	12.50%	31.25%
Accuracy $\leq \pm 5°$: 4500 Hz	87.50%	68.75%	100.00%	25.00%	12.50%	50.00%	56.25%
Median absolute error: 1500 Hz	7.0°	75.0°	2.5°	55.5°	92.5°	57.0°	34.0°
Median absolute error: 4500 Hz	3.5°	68.0°	1.0°	59.5°	61.0°	6.0°	5.0°
Uncertainty: 1500 Hz	4.92°	20.14°	7.86°	3.79°	6.49°	9.70°	6.08°
Uncertainty: 4500 Hz	6.16°	17.39°	8.93°	4.57°	5.70°	11.16°	6.06°
False detections: 1500 Hz	6.19	4.38	0.88	1.62	7.19	5.25	2.75
False detections: 4500 Hz	5.00	5.50	0.62	1.06	8.56	2.88	1.75



(a) Calibrated measured HRTFs



(b) Direct path measured HRTFs

FIG. 7. (Color online) Source location estimates for two simultaneously active sources located at various positions in the horizontal plane. Results are averages of different sound sources at 20 dB SNR using the calibrated and direct path measured HRTFs at an audio bandwidth of 4500 Hz.

More importantly, unlike the single source localization scenarios in the previous subsection, it can be observed that the higher audio bandwidth has not improved the localization uncertainty of the sources located at the sides of the KEMAR manikin in experiment scenarios 1−4 in Fig. 7(a). Similarly, false detections are observed on the cone of confusion in the front and back locations in experiment scenarios 5−8 in Fig. 7(a). Both effects are well known[36] and can be explained in terms of missing diversity information. For example, for a source located on the side, the SNR at the contralateral ear is greatly reduced due to the head shadowing effect. This, together with the inherently low signal power of the high frequency subbands, reduces the fidelity of the perceived spectral localization cues. In addition, the model of the inter-subband correlation (i.e., the knowledge of the source) is also imperfect at high frequencies, due to the low SNR of the high frequency subband signals. This loss of spectral localization cues results in an inability to separate closely spaced source locations, and is reflected in the high localization uncertainty of sources at the sides and the larger number of false detections for sources at the front and back in Fig. 7(a). Similarly, a greater median localization error is observed in the sides, which can be attributed to the same loss of fidelity of the perceived spectral localization cues, and is consistent with what is known of the general localization abilities of humans.[34,36]

The localization performance using the direct path measured HRTFs is illustrated in Fig. 7(b). In general, comparing the localization performance in Table III, a small rise in the localization uncertainty is observed over using the calibrated measured HRTFs. The likelihood of false detections on the cone of confusion is also increased. Both can be attributed to the distortion of the perceived high frequency spectral localization cues described above. However, the degradation in the performance due to the use of the direct path HRTFs is not as significant as in the single source localization scenarios in Tables I and II. This suggests that the HRTF plays a secondary role in multiple source localization, whereas the source location and knowledge of the source are more significant. Intuitively, this can be attributed to the higher prevalence of inter-subband relationships at lower frequencies, which are less affected by reverberation due to the higher direct path to reverberation power ratios at lower frequencies in mildly reverberant conditions. Overall, the results suggest that the performance advantages of using the calibrated measured HRTFs are negligible for multiple source localization, and that the direct path HRTFs may achieve reasonable localization performance in different mildly reverberant environments.

TABLE III. Multiple source localization performance of the proposed technique using the calibrated and direct path measured HRTFs in the horizontal plane.

	Calibrated HRTFs			Direct path HRTFs		
Performance criteria	Overall	Ideal	Real-world	Overall	Ideal	Real-world
Accuracy $\leq \pm 5°$: Sides	41.66%	33.33%	66.66%	75.00%	66.66%	100.00%
Accuracy $\leq \pm 5°$: Front/Back	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Median absolute error: Sides	13.5°	14.0°	10.0°	4.5°	4.0°	5.0°
Median absolute error: Front/Back	1.0°	1.0°	0.5°	1.5°	1.0°	2.0°
Uncertainty: Sides	50.81°	44.00°	87.00°	55.53°	39.60°	59.66°
Uncertainty: Front/Back	7.22°	5.50°	15.00°	7.83°	7.40°	16.00°
False detections: Sides	0.33	0.66	0.33	0.42	0.66	0.33
False detections: Front/Back	1.58	1.66	0.75	1.42	1.55	0.50

V. CONCLUSION

This study has experimentally evaluated the performance of a source location estimator that exploits the diversity in the frequency domain of the head-related transfer function for binaural sound source localization. Experiments localizing sources in the horizontal and vertical planes show the single source localization performance of the proposed method approaches the localization abilities of humans.[3,34] It was observed that the diversity in the frequency domain played a crucial role in the localization of a source in the vertical plane. Increasing audio bandwidth and calibrated measured HRTFs achieved good localization performance irrespective of the source locations. In addition, it was shown that the direct path HRTFs could still be used to localize sources using the proposed method, without any dereverberation of the measured binaural signals, in mildly reverberant conditions. The localization accuracy and error in multiple source scenarios was dominated by the actual source location and the knowledge of the source, while the head-related transfer function played a secondary role. In conclusion, it was demonstrated that the subtle frequency domain diversity of the head-related transfer function at relatively low audio bandwidths can be exploited for higher resolution source location estimates in single source localization scenarios. However, in the case of multiple concurrent sources, the diversity gained through the HRTF was overshadowed by the actual source location and the available knowledge of the nature of the sources themselves.

[1]S. Mehrgardt and V. Mellert, "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576 (1977).

[2]N. Gumerov, R. Duraiswami, and Z. Tang, "Numerical study of the influence of the torso on the HRTF," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 2, 1965–1968, Orlando, FL, (2002).

[3]J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159 (1991).

[4]C. Trahiotis, L. R. Bernstein, R. M. Stern, and T. N. Buell, "Interaural correlation as the basis of a working model of binaural processing: An introduction," in *Sound Source Localization*, edited by A. N. Popper and R. R. Fay (Springer, New York, 2005), Chap. 7, pp. 238–271.

[5]H. S. Colburn and A. Kulkarni, "Models of sound localization," in *Sound Source Localization*, edited by A. N. Popper and R. R. Fay (Springer, New York, 2005), Chap. 8, 272–316.

[6]P. M. Hofman, J. G. A. van Riswick, and A. J. van Opstal, "Relearning sound localization with new ears," Nat. Neurosci. **1**, 417–421 (1998).

[7]B. Rakerd, W. M. Hartmann, and T. L. McCaskey, "Identification and localization of sound sources in the median sagittal plane," J. Acoust. Soc. Am. **106**, 2812–2820 (1999).

[8]V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122 (2001).

[9]M. Morimoto, K. Iida, and M. Itoh, "Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences," Acoust. Sci. Technol. **24**, 267–275 (2003).

[10]V. 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).

[11]C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **24**, 320–327 (1976).

[12]D. Ward, Z. Ding, and R. Kennedy, "Broadband DOA estimation using frequency invariant beamforming," IEEE Trans. Signal Process. **46**, 1463–1469 (1998).

[13]H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," IEEE Trans. Acoust., Speech, Signal Process. **33**, 823–831 (1985).

[14]H. Hung and M. Kaveh, "Coherent wide-band ESPRIT method for directions-of-arrival estimation of multiple wide-band sources," IEEE Trans. Acoust., Speech, Signal Process. **38**, 354–356 (1990).

[15]V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," J. Audio Eng. Soc. **49**, 472–479 (2001).

[16]B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco, "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636 (2000).

[17]C. Lim and R. Duda, "Estimating the azimuth and elevation of a sound source from the output of a cochlear model," in *Proc. Twenty-Eighth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, CA (1994), Vol. 1, pp. 399–403.

[18]L. Calmes, G. Lakemeyer, and H. Wagner, "Azimuthal sound localization using coincidence of timing across frequency on a robotic platform," J. Acoust. Soc. Am. **121**, 2034–2048 (2007).

[19]M. Morimoto and H. Aokata, "Localization cues of sound sources in the upper hemisphere," J. Acoust. Soc. Jpn (E) **5**, 165–173 (1984).

[20]F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661 (1992).

[21]D. Cabrera and M. Morimoto, "Influence of fundamental frequency and source elevation on the vertical localization of complex tones and complex tone pairs," J. Acoust. Soc. Am. **122**, 478–488 (2007).

[22]E. A. Macpherson and A. T. Sabin, "Binaural weighting of monaural spectral cues for sound localization," J. Acoust. Soc. Am. **121**, 3677–3688 (2007).

[23]J. Qian and D. A. Eddins, "The role of spectral modulation cues in virtual sound localization," J. Acoust. Soc. Am. **123**, 302–314 (2008).

[24]C. Neti, E. D. Young, and M. H. Schneider, "Neural network models of sound localization based on directional filtering by the pinna," J. Acoust. Soc. Am. **92**, 3140–3156 (1992).

[25]P. Zakarauskas and M. S. Cynader, "A computational theory of spectral cue localization," J. Acoust. Soc. Am. **94**, 1323–1331 (1993).

[26]J. Nix and V. Hohmann, "Sound source localization in real sound fields based on empirical statistics of interaural parameters," J. Acoust. Soc. Am. **119**, 463–479 (2006).

[27]M. Raspaud, H. Viste, and G. Evangelista, "Binaural source localization by joint estimation of ILD and ITD," IEEE Trans. Audio, Speech, Lang. Process. **18**, 68–77 (2010).

[28]J. A. MacDonald, "A localization algorithm based on head-related transfer functions," J. Acoust. Soc. Am. **123**, 4290–4296 (2008).

[29]X. Wan and J. Liang, "Robust and low complexity localization algorithm based on head-related impulse responses and interaural time difference," J. Acoust. Soc. Am. **133**, EL40–EL46 (2013).

[30]D. S. Talagala, W. Zhang, and T. D. Abhayapala, "Broadband DOA estimation using sensor arrays on complex-shaped rigid bodies," IEEE Trans. Audio, Speech, Lang. Process. **21**, 1573–1585 (2013).

[31]D. S. Talagala and T. D. Abhayapala, "HRTF aided broadband DOA estimation using two microphones," in *Proc. 12th International Symposium on Communications and Information Technologies (ISCIT)*, Gold Coast, Australia (2012), pp. 1133–1138.

[32]R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **34**, 276–280 (1986).

[33]M. Zhang, W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "HRTF measurement on KEMAR manikin," in *Proc. Australian Acoustical Society Conference (Acoustics 2009)*, Adelaide, Australia (2009) pp. 1–8.

[34]S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**, 179–196 (1997).

[35]S. Teki, M. Chait, S. Kumar, S. Shamma, and T. D. Griffiths, "Segregation of complex acoustic scenes based on temporal coherence," eLife 2 (2013), http://elife.elifesciences.org/content/2/e00699 (Last viewed 11/01/13).

[36]J. Blauert, "Sound localization in the median plane," Acustica **22**, 205–213 (1969/70).