SOUND SOURCE LOCALIZATION WITH BINAURAL HEARING AIDS USING ADAPTIVE BLIND CHANNEL IDENTIFICATION

Ivo Merks, Gerald Enzner[1], Tao Zhang

Starkey Hearing Technologies, Eden Prairie, MN, 55344, USA
E-mail: {ivo_merks, tao_zhang}@starkey.com, gerald.enzner@rub.de

ABSTRACT

This paper applies blind channel identification (BCI) to estimate direction of arrival (DOA) of sound sources with a pair of binaural hearing aids. It compares the Adaptive Eigenvalue Decomposition Algorithm (AEDA) with the Adaptive Principal Component Algorithm (APCA) for blindly estimating the impulse responses from target to hearing aids and these impulse responses are used to estimate DOA. This paper investigates how both the time difference and the level difference of the impulse responses can be used to estimate DOA, and the performance of both algorithms is evaluated for scenarios with different reverberation times, different SNR, and different source positions. The paper also evaluates the tracking behavior in a dual-talker scenario. The results show that AEDA's DOA performance suffers in the presence of noise and reverberation. APCA is fairly insensitive to noise, but it can only handle moderate levels of reverberation.

Index Terms— ***direction of arrival, binaural hearing aids.***

1. INTRODUCTION AND RELATION TO PRIOR WORK

Understanding speech in noise is a difficult task for many hearing aid wearers [1]. For a long time, beamforming algorithms have been shown to help in noisy scenarios [2]. With the advent of wireless technology [3], binaural beamforming algorithms have been applied to further improve speech intelligibility. However, the beamforming algorithms often need to know the direction of arrival (DOA) of the target sound source.

Prior work has investigated a wide variety of algorithms to estimate DOA of sources for different applications [4]. The blind channel identification (BCI) algorithms are especially interesting because they can identify the (relative) impulse response from target to microphone [5, 6] and these impulse responses can be used to enhance the signal [8, 9] as well as to estimate direction of arrival [5, 7, 8, 9]. DOA can be estimated from the geometry of the microphone array and the time delays of arrival (TDOA). The TDOA can be estimated as the difference in the abscissae of the largest values of the estimated impulse responses [4, 5] or the abscissa of the largest value of the cross-correlation of the estimated impulse responses [7, 8, 9]. Using the estimated impulse responses instead of the microphone signals makes the DOA estimate less sensitive to signal properties [8]. In [10], the authors pointed out the Adaptive Eigenvalue Decomposition Algorithm (AEDA) [5] and the Adaptive Principal Component Algorithm (APCA) [6] as algorithms that might be useful for DOA estimation of sound sources in binaural hearing aids. Hearing aids are positioned on the head and that can result in a significant difference between the signal levels at the two hearing aids. This level difference is also present in the estimated impulse responses and can therefore be used as cue for DOA together with the TDOA [11].

The goal of this paper is to evaluate the performance of a DOA method for binaural hearing aids that uses the time difference as well as the level difference of the impulse responses that are estimated blindly with AEDA or APCA. To that end, simulations are done on measured data for scenarios with different SNR's, reverberation times, and source positions. Moreover, a dual talker scenario is investigated. For all scenarios, the performance of the DOA estimation of AEDA an APCA is reported and compared.

This paper is organized as follows. Section 2 describes the signal model, adaptive algorithms, and the DOA estimation methods. Section 3 describes the results of the different simulations and Section 4 draws the main conclusions from this research.

2. THEORY

2.1. Signal Model

Fig. 1 shows the signal model. A sound source at angle ϕ from the

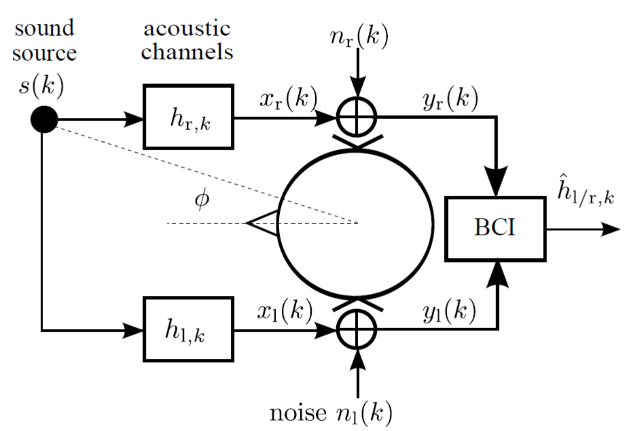

Fig. 1. Signal model of BCI for binaural hearing aids [10].

[1] Gerald Enzner is affiliated with Ruhr-Universität Bochum. His part of the work was performed on leave for technology consultancy to Starkey Hearing Technologies in fall 2011.

head transmits a signal $s(k)$ that arrives at the left and right ear through the acoustic channels $h_{l,k}$ and $h_{r,k}$ .There is also noise present at each ear, $n_l(k)$ and $n_r(k)$. The sum of the acoustically filtered signals and the noise results in the microphone signal of the left hearing aid, $y_l(k) = x_l(k) + n_l(k)$, and the right hearing aid, $y_r(k) = x_r(k) + n_r(k)$. In BCI, the algorithms estimate the left and right acoustic paths in the vectors

$$\hat{\mathbf{h}}_{l/r} = [h_{l/r,0}\ h_{l/r,1} \dots h_{l/r,L-1}]^T, \quad (1)$$

based on only the most recent observations contained in vectors

$$\mathbf{y}_{l/r} = [y_{l/r}(k)\ y_{l/r}(k-1) \dots y_{l/r}(k-L+1)]^T. \quad (2)$$

2.2. Adaptive Algorithms

In [10], two BCI algorithms were identified as candidates for on-line Head-Related Impulse Response (HRIR) estimation with hearing aids: Adaptive Eigenvalue Decomposition Algorithm (AEDA) and Adaptive Principal Component Algorithm (APCA). This section briefly summarizes the two algorithms.

AEDA [5] poses the BCI task as a minimum eigenvalue estimation problem. The error signal is defined as:

$$e(k) = \hat{\mathbf{h}}_r^T(k)\mathbf{y}_l(k) - \hat{\mathbf{h}}_l^T(k)\mathbf{y}_r(k). \quad (3)$$

If the adaptive filters match the HRIRs and no noise is present, this cross-relation processing would obviously yield an error signal $e(k) = 0$. AEDA defines the update step of the iterative minimization as follows:

$$\begin{aligned} \hat{\mathbf{h}}_l(k+1) &= \hat{\mathbf{h}}_l(k) + \mu e(k)(\mathbf{y}_r(k) + e(k)\hat{\mathbf{h}}_l(k)) \\ \hat{\mathbf{h}}_r(k+1) &= \hat{\mathbf{h}}_r(k) + \mu e(k)(\mathbf{y}_l(k) - e(k)\hat{\mathbf{h}}_r(k)) \end{aligned} \quad (4)$$

The algorithm normalizes for input level by using a normalized step factor:

$$\mu = \mu_0 \big/ (\mathbf{y}^T_l(k)\mathbf{y}_l(k) + \mathbf{y}^T_r(k)\mathbf{y}_r(k) + \delta), \quad (5)$$

where μ_0 is the adaptation constant and δ is an offset to avoid divergence at low input levels. After each iteration, a unit-norm constraint is enforced via normalization on the estimated filter:

$$\hat{\mathbf{h}}_{l/r}(k) = \hat{\mathbf{h}}_{l/r}(k) / \sqrt{\hat{\mathbf{h}}^T_l(k)\hat{\mathbf{h}}_l(k) + \hat{\mathbf{h}}^T_r(k)\hat{\mathbf{h}}_r(k)} \quad (6)$$

APCA [6] solves the BCI problem as iterative channel identification and equalization tasks. First a two-channel matched filter array is applied to get an estimate of the source signal:

$$\hat{s}(k) = \hat{\mathbf{h}}_l^{T\lrcorner}(k)\mathbf{y}_l(k) + \hat{\mathbf{h}}_r^{T\lrcorner}(k)\mathbf{y}_r(k), \quad (7)$$

where $\hat{\mathbf{h}}_{l/r}^{T\lrcorner}(k)$ are the time-reversed acoustic channels. This m\atched filter maximizes the SNR of the source signal and this operation is also known as focusing in acoustic and seismic imaging [9]. Subsequently, the individual error signals are calculated:

$$e_{l/r}(k) = y_{l/r}(k-L+1) - \hat{\mathbf{h}}_{l/r}(k)\hat{\mathbf{s}}(k), \quad (8)$$

where $\hat{\mathbf{s}}(k) = [\hat{s}(k)\ \hat{s}(k-1) \dots \hat{s}(k-L+1)]^T$ is a vector with the most recent equalizer output samples. The estimated source signal is used as input signal in an LMS update:

$$\hat{\mathbf{h}}_{l/r}(k+1) = \hat{\mathbf{h}}_{l/r}(k) + \mu e_{l/r}(k)\hat{\mathbf{s}}(k). \quad (9)$$

Similar to AEDA, the normalized step factor (5) as well as the enforcement of the unit-norm constraint (6) is applied after each iteration.

2.3 Direction of Arrival Estimate

The impulse responses from AEDA or APCA can be used to estimate the direction of arrival [7, 8, 9]. Using the impulse responses instead of the microphone signals has the advantage that the impulse responses have been adapted to the sound source and are therefore less noisy and variable than the microphone signals themselves. Given the estimated impulse responses $\hat{h}_l(k)$ and $\hat{h}_r(k)$, the cross-correlation between the impulse response is calculated.

$$\rho(k) = \sum_{n=0}^{N} \hat{h}_l(k)\hat{h}_r(n-k) \qquad k = 0,1,\dots,N-1 \quad (10)$$

The criterion for localization is the absolute value of the cross-correlation

$$J_{\Delta T}(\Delta T) = \|\rho(k)\|, \quad (11)$$

and $\Delta T = k/f_s$, where f_s is the sample frequency. The lag that corresponds to the maximum is the time difference between the microphones:

$$\max_{\Delta T}(J_{\Delta T}(\Delta T)). \quad (12)$$

This estimated delay can be converted to a direction of arrival estimate using the geometry of the binaural microphone array [8]. Since hearing aids are worn on the head, the conversion should take into account the propagation of the sound around the head which can be modeled as [11]:

$$\Delta T(\phi) = d/c\,(\sin\phi + \phi), \quad (13)$$

where d is the diameter of the head and c is the speed of sound. Besides the time difference, the level difference (due to shadowing of the head) between the microphones can also be used as a cue for DOA. This level difference is frequency dependent and irregular and cannot be easily modeled because of the head scattering effect. The most straightforward way is to use a table look-up that has been created using transfer function measurements $H_{l/r}(\phi,\omega)$ at 5 degrees interval in an anechoic chamber.

$$\Delta L(\phi,\omega) = 20\log_{10}\|H_r(\phi,\omega)/H_l(\phi,\omega)\|. \quad (14)$$

From the estimated impulse responses, the level difference can be calculated as

$$\Delta\hat{L}(\omega) = 20\log_{10}\left\|\hat{H}_r(\omega)\big/\hat{H}_l(\omega)\right\|,$$

where $\hat{H}_{l/r}(\omega)$ is the frequency-domain representation of $\hat{\mathbf{h}}_{l/r}(k)$. The following criterion calculates the match between the estimated level and the table in a least-squares sense over all frequencies ω.

$$J_{\Delta L}(\phi) = \left(-\alpha\left\|\Delta L(\phi,\omega) - \Delta\hat{L}(\omega)\right\|^2\right), \quad (15)$$

where α is chosen to normalize $\left\|\Delta L(\phi,\omega)-\Delta\hat{L}(\omega)\right\|^2$ to $\rho(\phi)$. A combined DOA estimation can be done by combining the time difference and level criteria into one criterion [11]:

$$J_{\Delta T+\Delta L}(\phi)=\left(-\alpha\left\|\Delta L(\phi,\omega)-\Delta\hat{L}(\omega)\right\|^2+\left\|\rho(\phi)\right\|\right). \quad (16)$$

3. RESULTS

This describes the simulations to evaluate the algorithms.

3.1 Experimental Set-up

The simulations were done using recordings with two BTE hearing aids mounted on a Knowles Electronics Manikin (KEMAR). The recordings were made in four different locations: an anechoic chamber (T60 = 0 s), a sound-treated lab (T60 = 0.2 s), an office (T60 = 0.4 s), and a reverberant room (T60 = 1.0 s). The target sound source was 1 m from KEMAR. The level of the target sound from the front was set to be 65 dB SPL at the ear. The babble noise (used as background noise) was recorded in a large food court of a mall during lunch hours. The sample frequency was 16 kHz.

The parameters of the algorithms were optimized to get a fast initial convergence and to have as little error as possible after convergence for clean speech in the anechoic environment. The length of the impulse response is 21 samples, so that it is long enough to model sound arriving from +90 or -90 degrees, which has the largest time difference. The impulse responses were initialized with random values to avoid having any a priori information.

3.2 Anechoic

To assess the upper boundary on performance, the DOA estimation performance of the algorithms, using the time difference criterion, the level difference criterion and the combined criterion, were assessed in the anechoic chamber scenario with steady-state speech shaped noise as target signal.

Fig. 2 shows the DOA estimation performance of both algorithms in steady state after convergence. DOA estimation performance is fairly similar for both algorithms and it is better for sources from the front and it is worse for sources from the side.

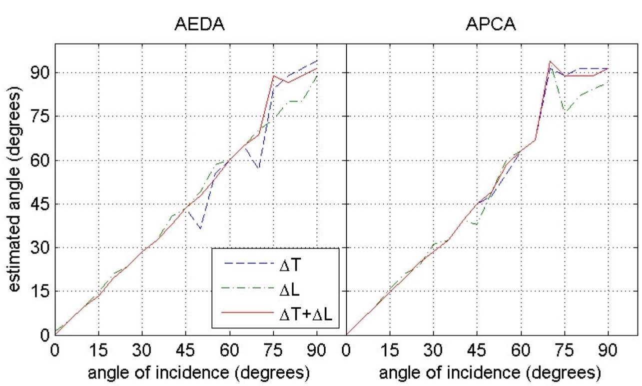

Fig. 2. Estimated angle of arrival versus actual angle for speech shaped noise in an anechoic chamber for the AEDA and APCA.

Fig. 3 analyzes this in detail by plotting the time-difference (12), the level difference (15), and the combined criterion (16) versus the angle for a source at 0 degrees and a source at 45 degrees.

The top plots in Fig. 3 show the criteria for the source at 0 degrees. The maximum of the criteria are at 0 degrees indicating that this is the most likely direction of the source. The time-difference criterion (ΔT) has a sharper peak for APCA than for AEDA which means that APCA's DOA estimation (based on time difference) is more reliable than AEDA's DOA estimation. The level-difference criterion (ΔL) is similar for AEDA and APCA and the combined criterion improves the reliability of DOA estimation for APCA and AEDA. The bottom plots in Fig. 3 show the criteria for the source at 45 degrees. The maximum of the criteria match again the actual direction of the source (45 degrees). However, the maximum of the criteria is in general not as pronounced for this source direction.

For AEDA, the time difference criterion is quite flat and the second largest peak is almost as large as the largest peak. Although the level-dependent criterion is also fairly flat for source directions around the actual source direction, the overall criterion shows a more distinct peak than the individual criteria.

For APCA, the time-difference criterion has a more distinct peak resulting in a more reliable DOA estimate. The level-difference criterion only slightly improves on the time-difference criterion.

These results have shown that APCA localization is more reliable in the anechoic scenario and the next sections will investigate how the results change in more adverse scenarios where speech is the source signal and noise or reverberation are present.

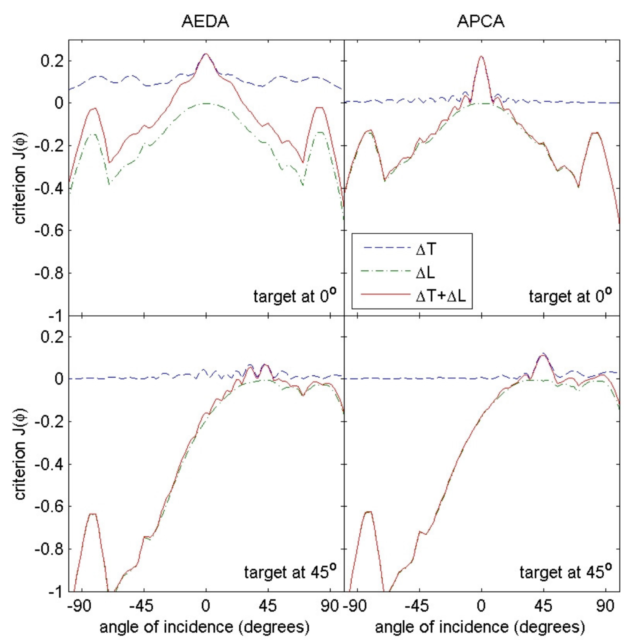

Fig. 3. Criteria as function of angle for source at 0 degrees (top) and at 45 degrees (bottom) for AEDA (left) and APCA (right).

3.3 Reverberation

This section assesses the performance of the algorithm in environments with different amounts of reverberations. The performance of DOA estimation for a single speaker at 0, 45 and

90 degrees is assessed with both algorithms. Since the signal is now speech instead of steady-state speech shaped noise, the localization performance is quantified as the percentage of localizations that are within 5 degrees of the actual position. Algorithm performance is investigated after initial convergence (first 5 s).
Fig. 4 shows that the performance of the DOA estimation strongly depends on the amount of reverberation and the angle of incidence. The performance decreases for increasing reverberation time and the performance is poor for high reverberation. APCA is especially for high reverberation much better than AEDA.

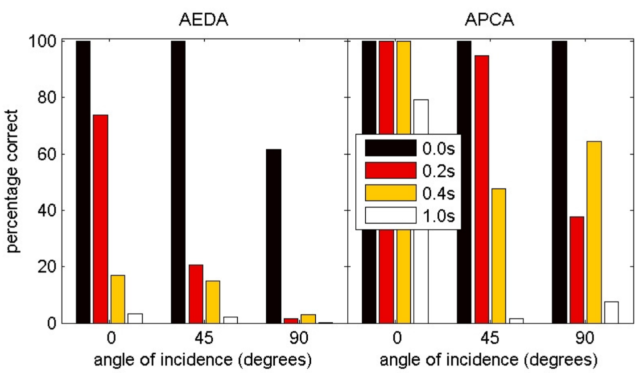

Fig. 4. Percentage correct for DOA estimation as function of reverberation time in seconds (see legend) and angle of target (on x-axis) for AEDA and APCA.

3.4 Noise

Because beamforming can be used to improve speech intelligibility in noise, the performance of DOA estimation in noise is also important. The performance of DOA estimation of a single speaker at 0, 45, and 90 degrees is assessed when recorded babble noise is present at SNR decreasing from 20 dB to 0 dB in steps of 5 dB.
Fig. 5 shows the performance, which has been calculated as in the previous section. The difference between the algorithms is quite large. APCA perform well for even the lowest SNR, while AEDA performs poorly for SNR smaller than 15 dB.

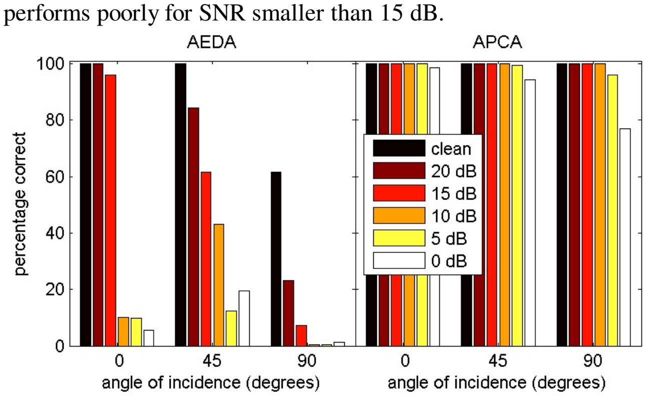

Fig. 5. Percentage correct for DOA estimation as function of SNR (see legend) and angle of target (on x-axis) for AEDA and APCA

3.5 Dual Talker Scenario

This section evaluates the ability of the algorithms to follow two talkers who speak one sentence (3s) in turn. One talker is at +45 degrees and the other talker is at -45 degrees. There is no background noise and no reverberation present. Fig. 6 shows that APCA is able to follow the active talker quite well and has found the correct location halfway through the sentence. There is slight overshoot during re-convergence. An analysis (similar to Fig. 3) shows that the time difference criterion is the most important criterion in the localization.

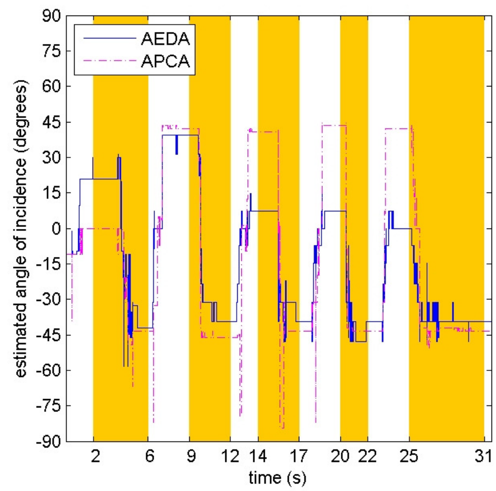

Fig. 6. Estimated angle of incidence versus time for AEDA and APCA. The color of the background indicates the angle of the source: white = 45 degrees. Non-white = -45 degrees.

AEDA is able to detect a change in direction, but it is not able to estimate the correct direction for either speaker. The DOA estimate of AEDA uses only the level difference criterion, since the time-level criterion shows no distinction between the sources.

4. CONCLUSIONS

This paper applied blind channel identification to the problem of Direction of Arrival (DOA) estimation with binaural hearing aids. On the basis of previous research, APCA and AEDA were chosen as algorithm candidates for this investigation. The algorithms estimated the impulse responses from the sound source to the hearing aids and used these impulse responses to estimate DOA of the sound source. The results of the simulations show that AEDA cannot estimate DOA well when even modest levels of noise and reverberation are present. This is because that the time-level criterion of AEDA is flat, so that multiple directions are almost equally likely to be the direction of the source. APCA is fairly insensitive to noise and can handle modest levels of reverberation too. For a practical application, APCA would need to perform better since reverberation times of 0.4s is not extraordinary and desired sources might be further away than 1 meter. Further improvements might be gained by using a frequency–domain implementation or step-size control. In addition, future evaluations should include scenarios with both noise and reverberation. Finally, the computational complexity has to be evaluated to ensure that the algorithms are feasible for hearing aids.

5. REFERENCES

[1] S. Kochkin. MarkeTrak VII: "Obstacles to adult non-user adoption of hearing aids," The Hearing Journal, vol. 60 (4), pp. 27-43, April 2007.

[2] P.M. Peterson, N.I. Durlach, W.M. Rabinowitz, and P.M. Zurek. "Multimicrophone adaptive beamforming for interference reduction in hearing aids," J Rehabil. Res. Dev., vol. 24(4), pp.103-10. Fall 1987

[3] A. Boothroyd, K. Fitz, J. Kindred, S. Kochkin., H. Levitt, B.C. Moore, B.C. and J. Yanz,. "Hearing aids and wireless technology," Hearing Review, vol. 14(6), pp. 44-47 (2007)

[4] N. Madhu, and R. Martin. "Acoustic source localization with microphone arrays," Advances in Digital Speech Transmission R. Martin, U. Heute, C. Antweiler (Eds.), John Wiley & Sons Ltd, Chichester, England, (2008), pp. 135-166.

[5] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," J. Acoust. Soc. Am., vol. 107, no. 1, pp. 384–391, Jan. 2000.

[6] D. Schmid and G. Enzner, "Robust subsystems for iterative multichannel blind system identification and equalization," in Proc. IEEE Pacific Rim Conf. on Commun., Comput. and Signal Process., Victoria, Can., Aug. 2009, pp. 889–893.

[7] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.II/273-II/276 Apr 1994.

[8] H.J.W. Belt and C.P Janse. "Signal Localization Arrangement". US Patent 6,774,934, August 2004.

[9] J. Schmalenstroeer, R. Haeb-Umbach, "Online Speaker Change Detection by Combining BIC with Microphone Array Beamforming", in *Proc. Interspeech 2006*, Pittsburgh, USA, Sept. 2006

[10] G. Enzner, I. Merks, and T. Zhang, "Adaptive filter algorithms and misalignment criteria for blind binaural channel identification in hearing-aids," Proc. 20th European Signal Processing Conference. Bucharest, Romania, Aug. 2012, pp. 315-319.

[11] M. Raspaud, H. Viste, G. Evangelista, "Binaural Source Localization by Joint Estimation of ILD and ITD," Audio, Speech, and Language Processing, IEEE Transactions on , vol.18, no.1, pp.68-77, Jan. 2010

[12] D. de Vries and A. J. Berkhout. "Wave theoretical approach to acoustic focusing," J. Acoust. Soc. Am. Volume 70, Issue 3, pp. 740-748, Sept. 1981.