USING A DIFFERENTIAL MICROPHONE ARRAY TO ESTIMATE THE DIRECTION OF ARRIVAL OF TWO ACOUSTIC SOURCES

Fotios Talantzis[1], *Anthony G. Constantinides*[2] *and Lazaros C. Polymenakos*[1]

[1]Autonomic & Grid Computing Group, Athens Information Technology, Markopoulo Av., 19002, Athens, Greece
[2]Dept. of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, UK

ABSTRACT

We present a system that estimates the direction of arrival of two competing acoustic sources using two closely spaced receivers that form a differential microphone array. The main advantage of the proposed array topology is that null steering can be essentially performed by adapting a set of two scalars. The direction of arrival estimation relies on the successful estimation of the relative delays between the microphone signals using the decorrelation constraint. Processing is performed in real-time by operating on blocks of recorded data. We examine the performance of the system for different block sizes and investigate its robustness in environments of strong multipath reflections where algorithms often fail to distinguish between the true direction of arrival and that of a dominant reflection. The overall performance of the system is compared to the simple omni-directional array topology. The results indicate that the examined framework can track the two directions of arrival adequately.

Index Terms: Direction Of Arrival, Differential Arrays

1. INTRODUCTION

Estimating the Direction Of Arrival (DOA) for camera steering in automated video-conferencing systems is typically approached by employing microphone arrays for the collection of acoustic data in frames. For the single source to multiple microphones scenario, the problem is typically dealt by estimating the time delay between microphone signal pairs [1, 2]. The delay can then be converted to the corresponding DOA by simple geometrical calculations.

When dealing with multiple DOAs from distinct acoustic sources the simple one source model has to be altered to a more complicated structure that requires estimation of more parameters. Additionally, if the system is used in reverberant environments, the estimation system often fails to distinguish between the true DOA and that of a dominant reflection. Algorithms for multi source tracking typically extend the one source model to systems that estimate multiple delays according to some independence criterion [3, 4].

Another ubiquitous signal processing problem is that of background noise cancelation. The utilization of directional microphone arrays can provide a limited solution to the problem since it can only cope with fixed noise sources. Using differential microphone arrays has shown to be a better approach to the problem [5, 6] since they can provide means for adaptive tracking of sources as they move.

For the purposes of the present context we make use of the properties of a two sensor differential array to establish a DOA estimation system for two sources. By extending the system of Elko et al. [6] to deal with two acoustic sources we present a system that scans the broadside of the array by pointing nulls at different directions. The aim is to find the ones that minimize the correlation of the recorded signals at the two sensors. The scanning process is performed by simple tuning of two scalars that control the directivity response of two back-to-back cardioid microphones. Even though the mathematical framework we propose is based on the anechoic case it remains robust for enclosures with reverberant characteristics. In fact, simulations show that it can resolve adequately the DOA estimation problem and subsequently generate consistent estimations under high reverberation times.

The paper is organized as follows. In Section 2 we formulate the DOA estimation problem and present the mathematical foundations of the system. Section 3 exhibits the performance of the system under different criteria such as reverberation level and architectural constraints of real-time systems. Section 4 summarizes the outcomes of the study.

2. SIGNAL MODEL

We concentrate on the case involving two sources and two microphones. Let us then consider a two element omni-directional microphone array positioned arbitrarily in an acoustical enclosure with the microphones being d metres apart. The sound sources are assumed to be in the far-field of the array and therefore we can approximate the spherical wavefront emanating from any of the sources as a plane wavefront of sound waves arriving at the microphones in a parallel manner. For the case in which the environment is non-reverberant the following Fourier domain signal is being recorded at the m^{th} microphone (where $m = 1, 2$) if the n^{th} (where $n = 1, 2$) source was impinging from angle θ_n with respect to the mid-point of the array:

$$X_m(\omega) = S_n(\omega) e^{(-1)^m j\omega \frac{d\cos\theta_n}{2c}} \quad (1)$$

where $X_m(\omega), S_n(\omega)$ are the discrete L-point Fourier transforms of the microphone and source signals at frequency ω respectively. Also, c denotes the speed of sound (typically defined as $343m/s$). Angle θ_n is the DOA of the n^{th} source and thus one of the parameters we are attempting to estimate. Forming the output of a first-order differential array involves the generation of the following signal:

$$Y_1(\omega) = X_1(\omega) - e^{-j\omega T} X_2(\omega) \quad (2)$$

where T is some time delay in seconds. If alternatively we delay the signal at the 2^{nd} microphone we obtain:

$$Y_2(\omega) = X_2(\omega) - e^{-j\omega T} X_1(\omega) \quad (3)$$

After some lengthy but straightforward manipulations, substitution of (1) into (2) and (3) leads to the following relationships for the output of the k^{th} (where $k = 1, 2$) differential array:

$$\frac{Y_k(\omega)}{S_n(\omega)} = e^{-j\omega T/2} 2j \sin\left[\omega\left(T/2 - (-1)^k \frac{d\cos\theta_n}{2c}\right)\right] \quad (4)$$

Now if we choose the microphone spacing such that $T = d/c$ we can further simplify to:

$$\frac{Y_k(\omega)}{S_n(\omega)} = e^{-j\omega\frac{d}{2c}} 2j \sin\left[\omega\frac{d}{2c}\left(1 - (-1)^k \cos\theta_n\right)\right] \quad (5)$$

For frequencies where $\left[\omega\frac{d}{2c}\left(1 - (-1)^k \cos\theta_n\right)\right] << \pi$ we may use the small angle approximation: $\sin\phi \approx \phi$. Implicitly this assumption states that the microphone spacing d has to be smaller then the acoustic wavelength over the frequency range of interest. Thus:

$$Y_k(\omega) \approx P(\omega) j\omega \frac{d}{c}(1 - (-1)^k \cos\theta_n) \quad (6)$$

where

$$P(\omega) = S_n(\omega) e^{-j\omega\frac{d}{2c}} \quad (7)$$

The first-order differential arrays formed have a monopole term and a first-order dipole term $\cos\theta_n$ that controls the directional response. Also note that the array has a first order linear differentiator frequency dependence. We can compensate for this by filtering each of the $Y_k(\omega)$ signals by an integrator having a transfer function:

$$H(\omega) = \frac{d}{c}\frac{1}{j\omega} \quad (8)$$

For each of the two arrays the processing has resulted into the following transfer functions:

$$\begin{aligned} Y_1(\omega) &= H(\omega)\left[X_1(\omega) - e^{-j\omega T} X_2(\omega)\right] \approx P(\omega)(1 + \cos\theta_n) \\ Y_2(\omega) &= H(\omega)\left[X_2(\omega) - e^{-j\omega T} X_1(\omega)\right] \approx P(\omega)(1 - \cos\theta_n) \end{aligned} \quad (9)$$

An obvious implementation to realize a steerable differential sensor would be to vary T in either $Y_1(\omega)$ or $Y_2(\omega)$ and thus control the gain of the array towards a specific direction. Rather than dealing with exponential terms though we can combine $Y_1(\omega)$ and $Y_2(\omega)$ to provide a conceptually simpler realization as follows:

$$Z_1(\omega) = Y_1(\omega) - \beta Y_2(\omega) = P(\omega)\left[(1 + \cos\theta_1) - \beta(1 - \cos\theta_1)\right] \quad (10)$$

We have now generated a system of back-to-back cardioid microphones that have the same phase centre. The microphones we assumed for the analysis were omni-directional ones. Alternatively, we could directly use cardioid microphones. The attractive property of this topology is that we can choose β such that $Z_1(\omega)$ has a null at an angle of:

$$\theta_2 = \arccos\left[\frac{\beta - 1}{\beta + 1}\right] \quad (11)$$

Note that for $0 < \beta < 1 \leftrightarrow 90^o > \theta_2 > 180^o$ i.e. null is always in rear half plane. Authors in [5, 6] used a similar configuration to adapt β and point a null towards a *single* noise source in that half-plane. By extending the configuration we can introduce the final step in estimating two DOAs:

$$Z_2(\omega) = Y_2(\omega) - \alpha Y_1(\omega) = P(\omega)\left[(1 - \cos\theta_2) - \alpha(1 - \cos\theta_2)\right] \quad (12)$$

Using the same logic, in Eq. (12) we can choose α such that $Z_2(\omega)$ has a null at an angle of:

$$\theta_1 = \arccos\left[\frac{1 - \alpha}{1 + \alpha}\right] \quad (13)$$

where for $0 < \alpha < 1 \leftrightarrow 0^o < \theta_2 < 90^o$ i.e. null is always in front half plane. Thus, we now have a system that can point two nulls in different half-planes. The overall system for two sources impinging upon the array at angles θ_1 and θ_2 can be seen in Fig. 1.

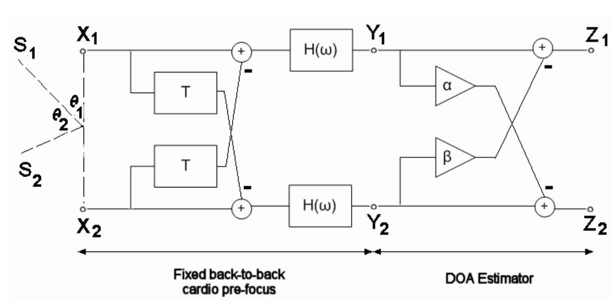

Fig. 1. Block diagram of the DOA estimation system.

Hence, provided geometry is such that there is one source either side of the mid-point of the differential array we can use the system for DOA estimation. The aim would actually be to scan the broadside of the array for the pair of values of α and β that point nulls towards the correct DOAs of the two sources and convert those values to the correct angles θ_n. In order to do this we look for the pair of values that minimize the average value of the off-diagonal terms in the correlation matrix of $Z(\omega) = [Z_1(\omega), Z_2(\omega)]^T$ over all frequencies ω. The criterion used to estimate α and β can be formalized as:

$$\{\alpha, \beta\} = \arg\min_{\alpha,\beta} \left\{ \frac{1}{L} \sum_{\omega} R(\omega) \right\} \tag{14}$$

where $R(\omega)$ is estimated as:

$$R(\omega) = Z_1(\omega) Z_2^*(\omega) = Z_2(\omega) Z_1^*(\omega) \tag{15}$$

3. SIMULATIONS AND DISCUSSION

The performance of the system is examined for a series of simulated experiments. The results are compared against those of a simple omnidirectional two element array. The configuration of this reference system is identical to the one presented in [3]. For the two source to two sensor scenario the system attempts to estimate the delays τ_1 and τ_2 corresponding to angles θ_1 and θ_2. It then compensates for them by shifting the recorded signals of Eq. (1). The delays are thus estimated using the decorrelation criterion (as in Eq. (15)) between the outputs of the following network:

$$\begin{aligned} \widehat{Y}_1(\omega) &= X_1(\omega) e^{j\omega\tau_2} - X_2(\omega) e^{-j\omega\tau_2} \\ \widehat{Y}_2(\omega) &= -X_1(\omega) e^{j\omega\tau_1} + X_2(\omega) e^{-j\omega\tau_1} \end{aligned} \tag{16}$$

To keep the comparison fair, the search range was kept identical to the one used for the differential array. Two speech signals of duration 5 sec sampled at $f_s = 44.1$ kHz were used. The value of T was chosen to be an integer number of samples (choose d accordingly) in order to make delaying by T trivial.

Even though the model assumes a non-reverberant environment, we test the robustness of the system in reverberant conditions. For this experiments where performed for three different enclosures distinguished by their reverberation times T_{60}. For the used sampling rate f_s these result in impulse responses h of different lengths. The impulse responses are generated using the image model [7]. The simulated room dimensions are $[5, 4, 3]$ m. These where then convolved with the speech signals to create the microphone signals. Moreover, 15dB of additive noise was also introduced to the signals. The process was repeated for ten random displacements and rotations of the relative geometry between the sources and the receivers inside the room. The simulation variables are given below:

$Variable$	$Value$
Distance between receivers (m)	0.0385
Distance between $\mathbf{s}_1$ and mid-point of receivers (m)	1.077
Distance between $\mathbf{s}_2$ and mid-point of receivers (m)	1.077
Value of θ_1 (degrees)	111.80
Value of θ_2 (degrees)	68.2
T_{60} (sec)	0.15, 0.30, 0.50
Length of h (samples)	6615, 13230, 22050

Most of the DOA estimation techniques are required to operate in real time. We must therefore assume that data at each sensor m are collected over t frames of L samples. For this reason we first brake the created time-domain microphone signals in frames of L samples. These are then converted into the Fourier domain using an L-point short time discrete Fourier transform and fed into the algorithm.

The system does not perform any type of adaptive search in order to minimize the correlation in Eq. (15). Instead the system searches over all possible combinations of α and β in real-time before deciding on the set of constants that results in minimization. Of course any adaptive version of the search process, like gradient descent variants, can help towards a more efficient method. Nevertheless for the given geometry, all simulations showed that the system can operate on-line even when calculating all possible combinations. The set of values used for α and β is derived by calculating the resolution of the array. This is actually a sequence of discrete values in degrees:

$$\delta\theta(\tau) = \arcsin\left[\frac{\tau c}{f_s d}\right] \tag{17}$$

with τ being the discrete sequence $[-\tau_{max}, -\tau_{max} + 1, ..., \tau_{max} - 1, \tau_{max}]$ and $\tau_{max} = round(df_s/c)$. The $round(.)$ operator denotes rounding to the closest integer.

For each frame of data processed, the system uses Eq. (14) to return two estimates $\widehat{\theta}_1$ and $\widehat{\theta}_2$ corresponding to the two sources. The squared error of the estimation for the n^{th} source at any frame is then computed as:

$$\sigma_n = (\theta_n - \widehat{\theta}_n)^2 \tag{18}$$

The Root Mean Squared Error (RMSE) metric is the performance measure used to evaluate the system. For a given source-sensor setup this is defined to be the square root of the average value of σ_n over all frames. A total of 10 simulations is performed with the overall source-sensor setup being randomly rotated and displaced inside the room. The relative geometry between the sources and the sensors is not altered though. In the figures to follow we present the average RMSE over all ten simulations. Thus, the lower the average RMSE value, the better the performance.

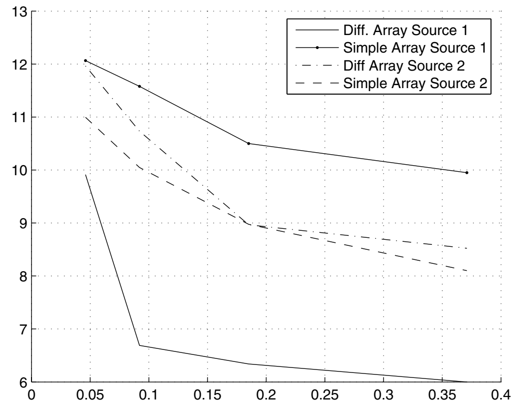

Fig. 2. RMSE for varying block size L. Shown for $T_{60} = 0.15$ sec.

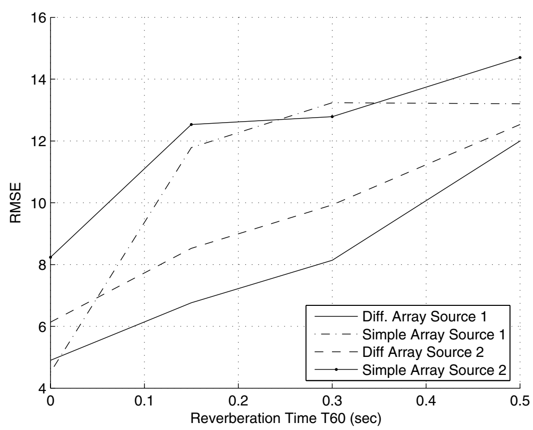

Fig. 3. RMSE for varying $T60$. $L = 0.371\ sec.$

Since DOA estimation systems are designed to operate in real time, a crucial characteristic is the size of L to be selected. In real systems, an accurate estimate of the angles should be given repeatedly after small segments of time in order to remain responsive. At the same time algorithms require enough data to be able include the effect of reverberation in their estimations. We thus examine the effect of the size of L by considering a series of different block sizes. These are $L = [0.046, 0.092, 0.185, 0.371]$ sec. Figure 2 expresses the effect of L on the performance of the proposed system and the simple omnidirectional one. In all cases $T_{60} = 0.15$ sec. The figure shows that as more data are available for estimating the DOAs, performance of both systems increases. The proposed system remains more robust in all cases for the corresponding sources.

Another crucial performance factor is the effect of reverberation. Fig. 3 summarizes the effect for the case when $L = 0.185$ sec. As expected, as the room becomes more reverberant the performance of the estimating systems degrades because reflections get mistaken for the actual DOAs of the sources. Nevertheless, the differential microphone array system estimates the DOA of both sources with greater accuracy in most cases. This is a result of the narrower beams created by the differential array is used. In essence, given the same search range of Eq. (17), the omnidirectional array does not estimate the delays that minimize the cross-correlation as accurate since the wider beams include the signal effect of neighbouring search angles.

4. CONCLUSIONS

A differential microphone array structure was used to reduce the problem of esimating the DOAs of two acoustic sources to a simple identification of a pair of scalars. These are chosen to minimize the cross-correlation of the output of the network. The system is limited in the sense that it requires geometry to be such that there is one source either side of the mid-point of the differential array. Simulations have showed though that under this constraint the system can produce DOA estimations in a fast and accurate manner. For the examined relative geometry, the system remained more robust for almost any combination of block size and T_{60} than the simple omni-directional implementation.

Acknowledgements

Special thanks to Dr. Darren B. Ward for valuable suggestions.

5. REFERENCES

[1] C. H. Knap and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoust. Speech and Sig. Proc.*, vol. 24, no. 4, pp. 320–327, 1976.

[2] P. Aarabi and S. Mavandadi, "Multi-source time delays of arrival estimation using conditional time-frequency histograms," *Information Fusion*, vol. 4, no. 2, pp. 111–122, 2003.

[3] J. P. Rosca, J. Ruanaidh, A. Jourjine, and S. Rickard, "Broadband direction-of-arrival estimation based on second order statistics," *NIPS*, pp. 775–781, 1999.

[4] E. D. Claudio, R. Paris, and G. Orlandi, "Multi-source localization in reverberant environments," *ICASSP*, vol. 2, pp. II921–II924, 2000.

[5] H. Teutsch and G. W. Elko, "First- and second-order adaptive differential microphone arrays," *Proc. 7th International Workshop on Acoustic Echo and Noise Control*, pp. 35–38, 2001.

[6] G. W. Elko and H. Teutsch, "A simple adaptive first-order differential microphone," *Proc. IEEE ASSP Workshop on Applications of Signal Proc. to Audio & Acoustics*, 1995.

[7] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–948, Apr. 1979.