NOVEL GCC-PHAT MODEL IN DIFFUSE SOUND FIELD FOR MICROPHONE ARRAY PAIRWISE DISTANCE BASED CALIBRATION

Jose Velasco[1], Mohammad J. Taghizadeh[2,3,4], Afsaneh Asaei[2], Hervé Bourlard[2,3],
Carlos J. Martín-Arguedas[1], Javier Macias-Guarasa[1], Daniel Pizarro[1]
[1]Department of Electronics, University of Alcalá, Alcalá de Henares, Spain
[2]Idiap Research Institute, Martigny, Switzerland
[3]École Polytechnique Fédérale de Lausanne, Switzerland
[4]Huawei European Research Center, Munich, Germany
{jose.velasco, cj.martin, macias, pizarro}@depeca.uah.es, mohammad.taghizadeh@huawei.com,{afsaneh.asaei, herve.bourlard}@idiap.ch

ABSTRACT

We propose a novel formulation of the generalized cross correlation with phase transform (GCC-PHAT) for a pair of microphones in diffuse sound field. This formulation elucidates the links between the microphone distances and the GCC-PHAT output. Hence, it leads to a new model that enables estimation of the pairwise distances by optimizing over the distances best matching the GCC-PHAT observations. Furthermore, the relation of this model to the coherence function is elaborated along with the dependency on the signal bandwidth. The experiments conducted on real data recordings demonstrate the theories and support the effectiveness of the proposed method.

Index Terms— Generalized cross correlation, Phase transform, Diffuse sound field, Pairwise distance estimation, Microphone array calibration

1. INTRODUCTION

Microphone arrays are widely used to enable high-quality distant audio acquisition. They are an essential part of a plethora of distant technologies ranging from source localization and separation to distant speech recognition [1, 2, 3] and from sound field analysis and monitoring to virtual reality and surveillance [4, 5]. A fundamental pre-processing step to enable the array of microphones to function in synergy consists of the gain, clock and position calibration. In this paper we address the problem of microphone array position calibration or extracting the relative geometry or the shape of the microphone array.

The prior art often rely on activation of known signals to estimate the pairwise microphone distances. This approach is referred to as self-calibration. Sachar et al. [6] presented an experimental setup using a pulsed acoustic excitation generated by five domed tweeters. The transmit times between speakers and microphones were used to find the relative geometry. Raykar et al. [7] used a maximum length sequence or chirp signal in a distributed computing platform. The time difference of arrival of the microphone signals were then computed by cross-correlation and used for estimating the microphone locations. Since the original signal is known, these techniques are robust to noise and reverberation.

In an alternative approach to alleviate the requirement for a known signal, Chen et al. [8] introduced an energy-based method for joint microphone calibration and source localization. The energy of the signal is computed and a nonlinear optimization problem is formulated to perform maximum likelihood estimation of the source-sensor positions. This method requires several active sources for accurate localization and calibration. Pollefeys and Nistre proposed a method for direct joint source and microphone localization which requires matrix factorization and solving linear equations [9]. In a different approach, McCowan et al. [10] proposed a calibration method which does not require activation of a particular signal. This approach relies on the characteristics of a diffuse sound field. A diffuse field can be roughly described as an acoustic field where the signals propagate with equal probability in all directions with the same power. The diffuse field is verified for meeting rooms and car environments [11, 12] and it enables application of well-defined mathematical models for analysis of the acoustic field recordings. A particular property related to diffuse field recordings is the coherence function between pairwise microphone signals which is defined by a sinc function of the distance between the two microphones. Thereby, we can estimate the pairwise distances by least-squares fitting the computed coherence with the sinc function.

In this paper, we derive a new model based on generalized cross correlation with phase transform (GCC-PHAT) for a diffuse sound field. This model elucidates the links between the output of GCC-PHAT and the distance between the microphone pairs. The relation between GCC-PHAT and the coherence has been previously discussed in [13, 14] where PHAT filtering is used as an estimator of the coherence between two signals. The global coherence field described in [15], has a virtually identical formulation to the steered response power with phase transform [16], which can be expressed in terms of GCC-PHAT [17]. Both rely on using the classical beamforming techniques in order to build an acoustic power map of the room, which has been reported in [18] to coincide with the maximum likelihood estimation of the position of the source under low noise and high reverberation conditions. In [19], a novel GCC-PHAT model is established for a point source, being validated with both synthetic and real data. Based on the statistical analysis model of a diffuse sound field, we derive an extension of the GCC-PHAT model for a diffuse field. We present the procedure for estimating the pairwise distance from the GCC-PHAT function of the microphone recordings and elaborate its relation to the coherence-based approach [10].

The rest of the paper is organized as follows: The definition of GCC-PHAT and its model for the point sources is stated in Section 2, showing its behavior with respect to the source direction of arrival and the model extension for a diffuse sound field. In Section 3, the procedure for pairwise distance estimation is presented and contrasted with the alternative technique based on coherence fitting. The experimental evaluation on real data recordings is conducted in Sec-

tion 4, and the conclusions are drawn in Section 5.

2. GCC-PHAT IN DIFFUSE SOUND FIELD

In this section, we explain the new GCC-PHAT model for a point-source that establishes the links between the microphone array geometry and the GCC-PHAT output. We derive its extension for a diffuse sound field.

2.1. Generalized Cross-Correlation

The generalized cross-correlation (GCC) has been widely used for time-difference-of-arrival estimation and it is the basis for many acoustic source localization algorithms. The GCC of the signals recorded by two microphones is defined as:

$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi_{ij}(\omega)\, X_i(\omega)\, X_j^*(\omega)\, \mathrm{e}^{\mathrm{j}\omega\tau}\, \mathrm{d}\omega, \qquad (1)$$

where $X_i(\omega)$ and $X_j(\omega)$ denote the signals recorded by microphones i and j in Fourier domain; ω is the angular frequency, $[.]^*$ stands for the conjugate transpose operation, and $\mathrm{j} = \sqrt{-1}$. The weighting function $\Psi_{ij}(\omega)$ is designed to optimize a given performance criteria. Many different functions have been proposed in the literature depending on the context, and among all of them, the phase transform (PHAT), defined as:

$$\Psi_{ij}(\omega) = \frac{1}{|X_i(\omega)\, X_j^*(\omega)|} = \frac{1}{|X_i(\omega)|\,|X_j(\omega)|}\,, \qquad (2)$$

has been found to perform very well for acoustic localization in reverberant environments, leading to the GCC-PHAT method [20] (also known as the crosspower-spectrum phase [15]). The PHAT can be seen as a filter which discards the amplitude and preserves the phase of the signal. The advantage of using it is that no assumptions are made about the signal or room conditions, which are typically unknown. This procedure has received considerable attention due to its simplicity and robustness in real world scenarios [18].

2.2. Analytic Model for a Point Source

The authors of [19] derive an analytical model for accurately predicting the behavior of the SRP-PHAT power maps for wideband signals, taking into account both the room geometry and the microphone array topology. They also show that the model is independent of the spectral content of the recorded signals, for both anechoic and reverberant conditions.

We consider a scenario where a single source is present and generates a baseband signal with bandwidth ω_0, thus $X_i(\omega) = 0$, $\forall\omega > \omega_0$. Assuming a free-space propagation model and discarding the distance dependent attenuation which is not relevant to our purposes, the signal at microphone j can be represented as a time-shifted version of $X_i(\omega)$, i.e. $X_j(\omega) = X_i(\omega)\mathrm{e}^{-\mathrm{j}\omega\tau_p}$ where τ_p is the time-difference of arrival between the two microphones.

From the model proposed in [19], and considering the anechoic propagation case, it is easy to show that when GCC-PHAT is applied to the signals captured by the microphone array, the resulting correlation can be approximated as a sinc function ($\mathrm{sinc}(x) = \frac{\sin(x)}{x}$), through

$$\begin{aligned} R_{\text{PHAT}}^{\text{point-source}}(\tau,\tau_p) &\approx \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} \mathrm{e}^{\mathrm{j}\omega(\tau-\tau_p)}\,\mathrm{d}\omega \\ &= \frac{\omega_0}{\pi}\,\mathrm{sinc}\left(\omega_0\,(\tau-\tau_p)\right)\,. \end{aligned} \qquad (3)$$

It may be noted that τ_p depends on the position of the source signal and it is limited by the distance d between two microphones such that $\tau_p \in \left[\frac{-d}{c}, \frac{d}{c}\right]$ with c being the speed of sound.

2.3. Extension to the Diffuse Sound Field

A diffuse field is defined as an acoustic field consisting of a superposition of an infinite number of sound waves traveling with random phases and amplitudes such that the energy density is equivalent at all points. More precisely, all points in the field radiate equal power and random phase sound waves, with the same probability for all directions, and the field is homogeneous and isotropic [21]. The analytic studies to model the diffuse sound field often rely on the statistical approach by considering an infinite number of free propagation plane waves, referred to as the plane wave model. In the plane wave model, a diffuse field is characterized as the superposition of a large set of plane waves impinging from all directions.

The spatial uniformity in a diffuse field can be expressed through integration of waves arriving from all directions [22, 23]. For two microphones, integrating over all directions is equivalent to integrating over all possible time-differences of arrival $\tau_p \in \left[\frac{-d}{c}, \frac{d}{c}\right]$ [22]. Therefore, the GCC-PHAT obtained in a diffuse field can be approximated by the GCC-PHAT model for a single source through the integration of uncorrelated sources arriving uniformly at all possible time-differences of arrival:

$$\begin{aligned} R_{\text{PHAT}}^{\text{diffuse}}(\tau,d) &\approx \int_{-\frac{d}{c}}^{\frac{d}{c}} R_{\text{PHAT}}^{\text{point-source}}(\tau,\tau_p)\;\frac{c}{2d}\,\mathrm{d}\tau_p \\ &= \frac{c}{2\pi d}\left(\mathrm{Si}\left(\omega_0\,(\tau+d/c)\right) - \mathrm{Si}\left(\omega_0\,(\tau-d/c)\right)\right), \end{aligned} \qquad (4)$$

where $\mathrm{Si}(x) = \int_0^x \mathrm{sinc}(t)\,\mathrm{d}t$ is the *sine integral*. The model expressed by (4) only depends on the distance between microphones d, and the signal bandwith ω_0. Furthermore, for large enough ω_0, the model can be approximated by a scaled version of the rectangular function:

$$\Pi\left(\frac{c\tau}{2d}\right) = \begin{cases} 0 & : |\tau| > \frac{d}{c} \\ \frac{1}{2} & : |\tau| = \frac{d}{c} \\ 1 & : |\tau| < \frac{d}{c} \end{cases} \qquad (5)$$

Fig. 1 demonstrates an example of the model and the real data measurements, for two different bandwidth values. Note that the values of $|\tau| > \frac{d}{c}$ do not provide relevant information about the distance between the two microphones while they nevertheless introduce some noise. Hence, it is easy to increase the signal-to-noise ratio by discarding those τs which do not have physical meaning based on the prior knowledge on the dimensions of the room or the physical setup.

3. MICROPHONE ARRAY CALIBRATION

In this section, we explain how the model of GCC-PHAT in diffuse sound field can be exploited to estimate the pairwise distance between two microphones for microphone array geometry calibration.

3.1. Distance Estimation Based on GCC-PHAT Model

The GCC-PHAT function for the signals of two microphones is obtained from (1)–(2) thus $R_{\text{PHAT}}(\tau)$ denotes the output based on the real data recordings. From the GCC-PHAT model expressed in (4), the distance between microphones can be estimated by fitting the model as:

$$\hat{d} = \underset{d,K}{\arg\min} \sum_{\tau=-\tau_{\max}}^{\tau_{\max}} \left(K\, R_{\text{PHAT}}(\tau) - R_{\text{PHAT}}^{\text{diffuse}}(\tau,d)\right)^2\,, \qquad (6)$$

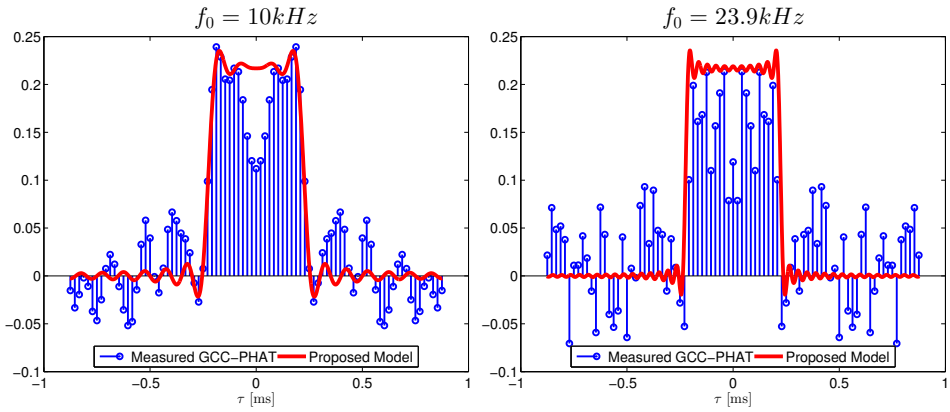

Fig. 1. The proposed GCC-PHAT model (4) contrasted with the measured GCC-PHAT on real data recordings in a diffuse sound field recorded at the room described in Section 4.1. The dependency on the signal bandwidth is demonstrated: the left graphic uses $f_0 = 10$ kHz and the right one uses $f_0 = 23.9$ kHz. We can see that for larger f_0 the model gets closer to the ideal case expressed in Eq. (8). Moreover we can see that the model fitting is better for smaller f_0 which is related to the fundamental limitations of a diffuse sound field for pairwise distance estimation [24].

where τ is discretized according to the sampling frequency and $\tau_{\max} = \frac{d_{\max}}{c}$. $d_{\max}$ indicates the expected maximum pairwise distance between any two microphones in the array, and can be estimated using geometrical considerations regarding the maximum room dimensions and the expected array geometry and locations. The additional parameter $K > 0$ is necessary since, in real scenarios, the model overestimates the amplitude of the correlation, which is lower due to the noise. Stacking the components for all values of τ, we obtain:

$$\mathbf{R}_{\text{PHAT}} \triangleq [R_{\text{PHAT}}(-\tau_{\max}), \dots, R_{\text{PHAT}}(\tau_{\max})]\,,$$
$$\mathbf{R}^{\text{diffuse}}_{\text{PHAT}}(d) \triangleq \left[R^{\text{diffuse}}_{\text{PHAT}}(-\tau_{\max}, d), \dots, R^{\text{diffuse}}_{\text{PHAT}}(\tau_{\max}, d)\right],$$

and after being Euclidean normalized, we obtain $\hat{\mathbf{R}}_{\text{PHAT}}$ and $\hat{\mathbf{R}}^{\text{diffuse}}_{\text{PHAT}}$. It it straightforward to show that, for discrete τ, minimizing the quadratic error $\left(K\, R_{\text{PHAT}}(\tau) - R^{\text{diffuse}}_{\text{PHAT}}(\tau, d)\right)^2$ is equivalent to minimizing the angle between the normalized vectors. Hence, denoting the inner product between two unit vectors by $\langle ., . \rangle$, we can rewrite Eq. (6) as:

$$\hat{d} = \arg\max_{d} \left\langle \hat{\mathbf{R}}_{\text{PHAT}}, \hat{\mathbf{R}}^{\text{diffuse}}_{\text{PHAT}}(d) \right\rangle \tag{7}$$

Given all the (offline-calculated) unitary vectors $\hat{\mathbf{R}}^{\text{diffuse}}_{\text{PHAT}}(d)$, the one that is better aligned with the $\hat{\mathbf{R}}_{\text{PHAT}}$ computed from the data can be found efficiently, indicating an estimate of the pairwise distance d.

3.2. Relation to the Coherence

The GCC-PHAT and coherence are two terms which are closely interconnected [13, 14]. The coherence of two signals is defined as the cross spectrum normalized by the square roots of the auto spectra. It has been shown that the real-part of the coherence of the signals at each frequency in a diffuse sound field is a $\text{sinc}\left(\frac{\omega d}{c}\right)$ function of the microphone distances [25]. This property is exploited by McCowan et al. to estimate the microphone pairwise distances [26].

In this section we show that the model introduced in equation (4) is, in fact, a low-pass filtered version of the inverse Fourier transform of the coherence-based approach [26]. Based on Eqs. (3) and (4), the GCC-PHAT model for the diffuse sound field can be written as:

$$\begin{aligned} R^{\text{diffuse}}_{\text{PHAT}}(\tau, d) &\approx \int_{-\omega_0}^{\omega_0} \int_{-\frac{d}{c}}^{\frac{d}{c}} \mathrm{e}^{\mathrm{j}\omega(\tau-\tau_p)}\, \frac{c}{2d}\, \mathrm{d}\tau_p\, \mathrm{d}\omega \\ &= \int_{-\omega_0}^{\omega_0} \frac{c}{2d\omega\mathrm{j}} \left(\mathrm{e}^{\mathrm{j}\omega(t+\frac{d}{c})} - \mathrm{e}^{\mathrm{j}\omega(t-\frac{d}{c})}\right) \mathrm{d}\omega \\ &= \int_{-\omega_0}^{\omega_0} \text{sinc}\left(\frac{\omega d}{c}\right)\, \mathrm{e}^{\mathrm{j}\omega}\, \mathrm{d}\omega \end{aligned} \tag{8}$$

Hence, we can see that the GCC-PHAT model for a diffuse sound field is the Fourier transform of the sinc (real-part of the coherence) ideally filtered at ω_0. Since the proposed model is the inverse Fourier transform of the coherence-based model, removing high values of τ in the GCC-PHAT calculation, implies removing fast changes in the coherence and lead to denoising the coherence; Fig. 2 demonstrates an example of the denoising effect achieved via suppressing the time coefficients corresponding to $\tau > \tau_{\max}$. As we will see during the experimental evaluation presented in Section 4, the GCC-PHAT model in a diffuse sound field outperforms the coherence-based approach [26], while improving the computational cost.

4. EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of the proposed technique for pairwise distance estimation using real data recordings collected at the Idiap smart meeting room.

4.1. Acoustic Recording Setup

We use the geometrical setup of the MONC corpus to record the sound field in a meeting room [27]. The enclosure is a $8 \times 5.5 \times 3.5\,\text{m}^3$ rectangular room and it is moderately reverberant. It contains a centrally located $4.8 \times 1.2\,\text{m}^2$ rectangular table. Nine microphones are located on a planar area parallel to the floor at a height of 1.15 m: Eight of them are located on a circle with diameter 20cm and one microphone is at the origin. The microphones are Sennheiser MKE-2-5-C omnidirectional miniature lapel microphones. The floor of the room is covered with carpet and surrounded with plaster walls and two big windows; the room is mildly reverberant with a reverberation time less than 200 ms. The room is almost silent and no

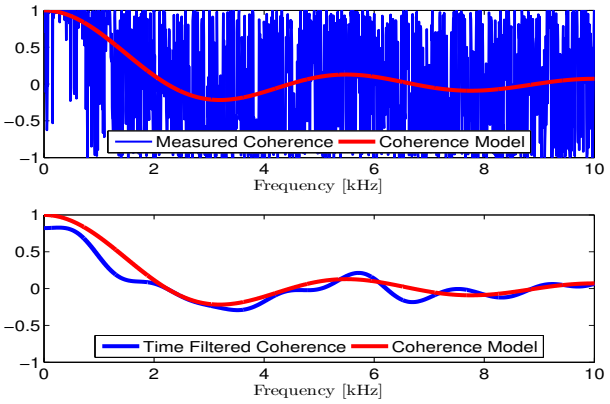

Fig. 2. The frame-based coherence measured using the real data and the theoretical sinc model in the original form (top) and after time filtering (bottom) based on suppression of the GCC-PHAT output at the large time intervals that do not correspond to the physical setup.

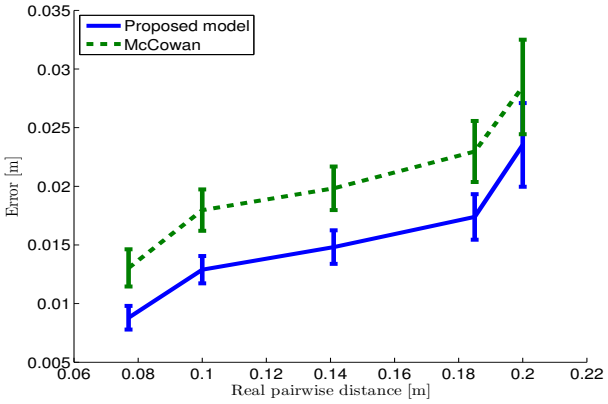

Fig. 3. Comparison between the average error in distance estimation using the proposed GCC-PHAT model (7) and the McCowan's coherence-based method [10].

sound source is generated; there is ambient noise due to the street and computer fans. The sampling rate is 48 kHz. The experiments are conducted using $c = 343$ m/s that corresponds to 20° Celsius temperature of the room.

4.2. Analysis Parameters

The recordings are processed frame by frame in frames of 4096 samples (85.3 ms) after applying the Tukey window. The FFT is calculated using 8192 samples (after zero-padding). The maximum distance between microphones was restricted to $1.5m$, so that all τ in GCC-PHAT corresponding to longer distances were not considered. The set of possible distances are discretized within the range of $[0.05, 1.5]$ m with one millimeter resolution.

Since the diffuse noise is expected to be broadband and with equal power in all frequencies, $\omega_0 = 2\pi f_0$ has been in fact determined by the antialiasing filter ($f_0 = 23.9$ KHz). A more restrictive filtering allows a better fitting, as demonstrated in Fig. 1.

4.3. Pairwise Distance Estimation Performance

Fig. 3 shows the estimation error of pairwise distances for the two models; the bars represent the 99% confidence interval, assuming a normal distribution. The improvement of the proposed model in terms of pairwise distance estimation is statically significant, but it does not lead to better results in the calibration of the position of the microphones based on multidimensional scaling method [28].

4.4. Numerical Approximation for the Proposed Model

The mathematical approximation is suitable as Matlab® provides a symbolic implementation for the sine integral which can be sometimes quite slow. We suggest using the numerical approximation described in [29, p. 231]:

$$\mathrm{Si}\,(x) \approx \begin{cases} \sum_{n=0}^{N-1} \frac{(-1)^n x^{2n+1}}{(2n+1)(2n+1)!} & : |x| \leq 1 \\ \frac{\pi}{2} - f(x)\cos(x) - g(x)\sin(x) & : x > 1 \\ f(-x)\cos(x) - g(-x)\sin(x) - \frac{\pi}{2} & : x < 1 \end{cases} \tag{9}$$

which has a low error ($|\epsilon(x)| < \max\{\frac{1}{(2N+1)2!}, 3 \times 10^{-7}\}$), and it can speed up the implementation of the proposed model. Functions $f(x)$ and $g(x)$ are calculated as [1] :

$$f(x) = \frac{1}{x}\left(\frac{x^8 + a_1x^6 + a_2^4 + a_3x^2 + a_4}{x^8 + b_1x^6 + b_2^4 + b_3x^2 + b_4}\right) \tag{10a}$$

$$g(x) = \frac{1}{x^2}\left(\frac{x^8 + c_1x^6 + c_2^4 + c_3x^2 + c_4}{x^8 + d_1x^6 + d_2^4 + d_3x^2 + d_4}\right) \tag{10b}$$

The above approximation speeds up the process more than one million times. The time that it takes to perform pairwise distance estimation using each frame is 40 times faster than real time. The new GCC-PHAT model is also 30 times faster than the alternative coherence-base approach.

5. CONCLUSIONS

In this paper, a new model for GCC-PHAT in diffuse sound field is proposed which establishes the links between GCC-PHAT output and the microphone array geometry. To estimate the pairwise distances, the GCC-PHAT is computed for a pair of microphone signals and the distance that generates the best fitting model is estimated. It was shown that this model is in fact equivalent to an inverse Fourier transform of an ideally filtered coherence of the two signals. The experiments conducted on real data recordings demonstrate the effectiveness of the proposed approach for pairwise distance estimation. Furthermore, it suggests a simple denoising scheme for the coherence function via suppression of the GCC-PHAT activation at the time intervals which do not meet the physical constraints. The model was shown to perform significantly faster than the coherence-based counterpart and it is applicable for real time calibration setups.

6. ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Economy and Competitiveness under project SPACES-UAH (TIN2013-47630-C2-1-R), and by the FPU Grants Program of the University of Alcalá. Afsaneh Asaei acknowledges the SNSF 200021-153507 grant on PHASER project.

[1] $a_1 = 38.027264$, $a_2 = 265.187033$, $a_3 = 335.677320$, $a_4 = 38.102495$, $b_1 = 40.021433$, $b_2 = 322.624911$, $b_3 = 570.236280$, $b_4 = 157.105423$, $c_1 = 42.242855$, $c_2 = 302.757865$, $c_3 = 352.018498$, $c_4 = 21.821899$, $d_1 = 48.196927$, $d_2 = 482.485984$, $d_3 = 1114.978885$ and $d_4 = 449.690326$.

7. REFERENCES

[1] H. T. Do, *Robust cross-correlation-based methods for sound-source localization and separation using a large-aperture microphone array*, Ph.D. thesis, Brown University, 2011.

[2] A. Asaei, M. Golbabaee, H. Bourlard, and V. Cevher, "Structured sparsity models for reverberant speech separation," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 3, pp. 620–633, 2014.

[3] Afsaneh Asaei, *Model-based Sparse Component Analysis for Multiparty Distant Speech Recognition*, Ph.D. thesis, École Polytechnique Fédéral de Lausanne (EPFL), 2013.

[4] M. Mattila V. Veijanen V. Pulkki T. Hiekkanen, T. Lempiainen, "Reproduction of virtual reality with multichannel microphone techniques," in *Proceeding of 122nd AES Convention*, 2007.

[5] Giuseppe Valenzise, Luigi Gerosa, Marco Tagliasacchi, E Antonacci, and Augusto Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, 2007, pp. 21–26.

[6] J. M. Sachar, H. F. Silverman, and W. R. Patterson, "Microphone position and gain calibration for a large-aperture microphone array," *IEEE Transactions on Speech and Audio Processing*, vol. 13(1), 2005.

[7] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudspeakers in distributed computing platforms," *IEEE Transactions on Speech and Audio Processing*, vol. 13(1), 2005.

[8] M. Chen, Z. Liu, L. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad-hoc microphone arrays," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007.

[9] Marc Pollefeys and David Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data.," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008, pp. 2445–2448.

[10] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Transactions on Audio,Speech and Language Processing*, vol. 16(3), 2008.

[11] M. J. Taghizadeh, P. N. Garner, H. Bourlard, H. R. Abutalebi, and A. Asaei, "An integrated framework for multi-channel multi-source localization and voice activity detection," in *IEEE workshop on Hands-free Speech Communication and Microphone Arrays*, 2011.

[12] J. Bitzer, K. U. Simmer, and K. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999.

[13] Maurizio Omologo and Piergiorgio Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*. IEEE, 1994, vol. 2, pp. II–273.

[14] Alessio Brutti, Maurizio Omologo, and Piergiorgio Svaizer, "Speaker localization based on oriented global coherence field," in *Proceedings of Interspeech*, 2006, vol. 7, p. 8.

[15] M. Omologo and P. Svaizer, "Use of the cross-power-spectrum phase in acoustic event location," *IEEE Trans. on Speech and Audio Processing*, vol. 5, pp. 288–292, 1993.

[16] J. DiBiase, H. Silverman, and M. Brandstein, *Microphone Arrays*, chapter Robust Localization in Reverberant Rooms, pp. 157–180, 2001.

[17] J.H. DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*, Ph.D. thesis, Brown University, 2000.

[18] Cha Zhang, D. Florencio, and Zhengyou Zhang, "Why does phat work well in low noise, reverberative environments?," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 31 2008-april 4 2008, pp. 2565 –2568.

[19] Jose Velasco, Carlos J. Martín-Arguedas, Javier Macias-Guarasa, Daniel Pizarro, and Manuel Mazo, "Proposal and validation of an analytical generative model of srp-phat power maps in reverberant scenarios," Tech. Rep. GEINTRA-RR-1-2014, GEINTRA Research Group, Department of Electronics, University of Alcala, Spain, November 2004, http://www.geintra-uah.org/RR-14-01.pdf.

[20] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320 – 327, aug 1976.

[21] T.J. Schultz, "Diffusion in reverberation rooms," *Journal of Sound and Vibration*, vol. 16(1), 1971.

[22] Boaz Rafaely, "Spatial-temporal correlation of a diffuse sound field," *The Journal of the Acoustical Society of America*, vol. 107, no. 6, pp. 3254–3258, 2000.

[23] Allan D Pierce et al., *Acoustics: an introduction to its physical principles and applications*, McGraw-Hill New York, 1981.

[24] Mohammad J. Taghizadeh, Philip N. Garner, and Hervé Bourlard, "Enhanced diffuse field model for ad hoc microphone array calibration," *Signal Processing*, vol. 101, 2014.

[25] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, "Measurement of correlations coefficients in reverberan sound fields," *Journal of the Acoustical Society of America*, vol. 27, 1955.

[26] Iain McCowan, Mike Lincoln, and Ivan Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 16, no. 3, pp. 666–670, 2008.

[27] "The multichannel overlapping numbers corpus (MONC)," Idiap resources available online:, `http://www.cslu.ogi.edu/corpora/monc.pdf`.

[28] T. F. Cox and M. A. A. Cox, "Multidimensional scaling," *Chapman-Hall, 2001*.

[29] Milton Abramowitz, Irene A Stegun, et al., *Handbook of mathematical functions*, vol. 1, Dover New York, 1972.