Localization of sound sources by means of unidirectional microphones

Piervincenzo Rizzo (corresponding author),

Department of Civil and Environmental Engineering, University of Pittsburgh, 949 Benedum Hall, 3700 O'Hara Street, Pittsburgh, PA 15261, USA, E-mail: pir3@pitt.edu; Phone: +1-412-6249575; Fax: +1-412-6240135.

Giacomo Bordoni

former Visiting Scholar, now at Italcementi Group, Via Camozzi 124, 24121 Bergamo (Italy).

Alessandro Marzani

DISTART, University of Bologna, Italy, Viale del Risorgimento 2, 40136 Bologna, Italy.

Jeffrey Vipperman

Department of Mechanical Engineering and Materials Science, University of Pittsburgh, 531 Benedum Hall, 3700 O'Hara Street, Pittsburgh, PA 15261, USA.

ABSTRACT

This paper describes the results of a new approach devoted to the localization of ground borne acoustic sources. It is demonstrated that an array made of at least three unidirectional microphones can be exploited to detect the position of the source. Pulse features extracted either in the time domain or in the frequency domain are used to identify the direction of the incoming sound. This information is then fed into a semi-analytical algorithm aimed at identifying the source location. The novelty of the method presented here consists on the use of unidirectional microphones rather than omnidirectional microphones and on the ability to extract the sound direction by considering features like sound amplitude rather than the time of arrival. Experimental tests have been undertaken in a closed environment and have demonstrated the feasibility of the proposed approach. It is believed that this method may pave the road toward a new generation of reduced size sound detectors and localizers, and future work is described in the conclusions.

Keywords: sound source localization; unidirectional microphones, polar directivity pattern, feature extraction.

I. INTRODUCTION

Locating acoustic point sources using passive sensor arrays is of considerable interest in several fields of civil, mechanical, industrial, and biomedical engineering. Sound source localization (SSL), where there is no a priori knowledge of the sound source position, suits a variety of applications that span from teleconferencing tracking of speakers (Huang *et al* 2000) and human robot interaction (Kim *et al* 2004), to surveillance systems and air/ground objects recognition (Wilson *et al* 2002). The latter, for instance, is pivotal in military applications to enhance the patrol capability of troops and to identify and locate snipers. In conferencing scenarios, SSL can be used for directing a pan-tilt-zoom camera toward the speaker or the audience such that the viewing experience is more interesting and/or network bandwidth is used more efficiently (Rui *et al* 2005).

The passive detection and tracking of targets that emit acoustical energy present several advantages over active sounding and radar systems. First, passive systems do not radiate a signal which can be used to indicate the presence and identity of the tracking station. Second, active detection systems are subjected to the reception of "false echoes" by virtue of multiple reflections. Finally, active tracking systems are more prone to counter measures.

In the last two decades several methods for the passive localization of acoustic sources or the determination of the direction of a sound source were proposed (Olson 1946, Knapp and Carter 1976, Wang and Kaveh 1985, Brandstein and Silvermant 1997, Benesty 2000, Bourennane and Bendjama 2002, Merimaa 2002, Reid and Milios 2003, Liu and Milios 2005). These methods are based on the use of single or multiple omnidirectional microphone arrays, linearly (Wang and Chu 1997, Kleban 2000), triangularly (Nakadai 2003), circularly (Birchfield and Gillmor 2001, Rui and Florencio 2003, Rui *et al* 2005), or sparsely (Bian *et al* 2005, Kobayashi *et al* 2008)

arranged. These approaches can be divided into three categories (Merimaa 2002, Bian *et al* 2005, Kwon *et al* 2007, Atmoko *et al* 2008): beamformer based locators; high-resolution spectral estimation based locators; and time-of-delay based locators. In the first category the source is localized by maximizing the output of a steerable beamformer (Elko *et al* 1994, Ikeda *et al* 2007). The approach combines delay-and-sum beamforming with statistical analysis to trace the position of the acoustic source. The method is effective when the source is emitting continuously but it may suffer when concurrent sound sources are present. In the second category, beamforming-based techniques are combined with high resolution spectral analysis. The methods were mainly developed for narrowband signals (Brandstein and Silverman 1997). Extensions to wideband sounds were introduced by analyzing signal components at several subbands (Wang and Kaveh 1985). One of the limits of these approaches is the capability to distinguish concurrent sound sources. Bourennane and Bendjama (Bourennane and Bendjama 2002) introduced an analysis based on higher order statistics to overcome this problem. The first two categories discussed above require a search over the potential directions of sound sources. The third category, instead, is based on time delay estimation (TDE) (Valin *et al* 2003, Liu and Milios 2005, Kwon *et al* 2007, Jin *et al* 2008, Kobayashi *et al* 2008, Lv and Zhang 2008). In the TDE the arrival time delay between signals detected by each pair of the microphones array is estimated. The delays are used to determine the direction of arrival of sound. Finally, the source location is obtained by intersecting the direction lines towards the source out of the multiple microphone arrays given their geometry and the relative directions of arrival. In order to measure the time delays techniques based on cross-correlation (Olson 1946, Atmoko *et al* 2008, Brutti *et al* 2008, Hu *et al* 2008), eigenvalue decomposition (Benesty 2000), least mean square (LMS) adaptive filters (Olson 1946) and crosspower-spectrum phase (Omologo and Svaizer 1994, Kim

et al 2008) were proposed. Knapp and Carter (Knapp and Carter 1976) pointed out that locators based on the arrival time difference present significant computational advantages compared to steered beamformers or high resolution spectral estimation methods.

Researchers (Rui *et al* 2005, Ikeda *et al* 2007, Kim *et al* 2008) have proposed hybrid methods that combine two of the three aforementioned categories. One of the common limits associated with the arrangements proposed is the need of large spacing between the microphone elements and/or the requirement that acoustic sources need to stay farther from the microphones in order to retain accuracy.

This paper presents the results of a passive scheme based on the use of unidirectional microphones for the identification of the position of the acoustic source with respect to the microphones' center of mass. Sound direction is projected by exploiting the unidirectionality of certain features associated with the microphones output level. The incidence value estimated by each microphone is then fed into a semi-analytical algorithm aimed at identifying the source location. The objective of the paper is to demonstrate the feasibility of the proposed method to evaluate the position of a sound source with an array of at least three unidirectional microphones. The study is part of an ongoing effort to pave the road toward a new generation of sound detectors and localizers that can minimize the spacing between the microphone elements. The main elements of novelty introduced with this study are the exploitation of unidirectional microphones and the use of features that are not related to the time of arrival of the sound.

The paper is structured as follows. The semi-analytical algorithm to extract the location of the sound source from an array of unidirectional microphones is presented in section 1. Section 2 illustrates the signal processing algorithm that is combined with the semi-analytical approach. The experimental results of the initial tests are presented in section 3. Finally, conclusions and

directions for future research are presented in section 4.

II. ANALYTICAL MODEL

A. Unidirectional microphones

Unidirectional microphones are sound transducers that possess a strong dependence on the incidence angle of the acoustic source. Directionality is achieved through a compact internal array microphone in a perforated housing. Such microphones are used in all those applications or circumstances that require rejection of noise, filtering of unwanted acoustic signals, or elimination of reverberation phenomena. The variation of the detected sound wave amplitude (microphone output level) with the direction of the wave propagation (angle of sound incidence) is described by the polar directivity pattern (PDP). Generally, the PDPs detect a plurality of amplitude values of a propagating wave approaching at different angles (Olson 1946). When the microphone's sensitivity (output level) is not function of the incidence angle the microphone is called omnidirectional. Mathematically a plane PDP of an unidirectional microphone may be expressed as (Sessler and West 1975)

$$M(\vartheta) = K \times abs\left[(a + \cos\vartheta)(b + \cos\vartheta)\right] \qquad (1)$$

where M is the microphone output level, a and b are frequency dependent constants based on the microphone's design and the sound speed, ϑ is the sound's angle of incidence relative to the microphone axis. The constant K depends on the sensitivity of each sensor composing the array microphone, frequency and speed of the incoming sound as well as microphone design parameters. Particularly Eq. (1) describes a microphone of order two, in which the incidence angle ϑ is the argument of the second power of the cosine term. Figure 1 shows an example of a PDP expressed by Eq. (1) in polar coordinates for a case where $a = 0.69$ and $b = -0.29$. As

shown by Sessler and West (Sessler and West 1975), for low-frequency sounds these values provide the highest directivity to second order gradient unidirectional microphones. The PDP has been normalized with respect to the maximum value expected at along the microphone axis $\vartheta = 0°$, by considering instead of K a constant equal to $\hat{K} = K / M(0°)$.

B. Source location: semi-analytical formulation

To describe the background hypothesis, let us consider an array of three unidirectional microphones A, B and C, of coordinates (x_i, y_i) (i = A, B, C), as in Fig. 2a. The origin O of the x-y coordinate system corresponds to the center of mass of the triangle ABC. The angle δ_i, between the reference y-axis and the i-th microphone axis, is introduced to univocally orient each microphone. For instance, in Fig. 2a $\delta_A = 0°$, $\delta_B = -120°$ and $\delta_C = 120°$. Let M_A, M_B, and M_C be the normalized pulse features recorded by microphones A, B, and C, respectively, as a result of an emission from the acoustic source. These values were normalized with respect to their maximum values expected at $\vartheta_i = 0$. By assuming that each microphone's PDP satisfies Eq. (1) we obtain:

$$\hat{K} \times abs\{[a + \cos(\vartheta_A)][b + \cos(\vartheta_A)]\} = M_A \tag{2a}$$

$$\hat{K} \times abs\{[a + \cos(\vartheta_B)][b + \cos(\vartheta_B)]\} = M_B \tag{2b}$$

$$\hat{K} \times abs\{[a + \cos(\vartheta_C)][b + \cos(\vartheta_C)]\} = M_C \tag{2c}$$

which can be generalized as:

$$\hat{K} \times abs\{[a + \cos(\vartheta_i)][b + \cos(\vartheta_i)]\} = M_i \qquad (i = \text{A, B, C}) \tag{3}.$$

In polar coordinates, the value M_i can be geometrically described as a circle of radius M_i. By solving numerically Eqs (2a), (2b), and (2c) for the unknown values of ϑ_A, ϑ_B, ϑ_C, the

intersection between each of the microphone lobes and the relative circle are determined. These intersection points identify the possible directions of the incoming sound directions. However, these directions are not univocally determined. In fact, as the microphone i-th possesses four lobes, the normalized value M_i may identify up to eight incidence angles, i.e. eight intersections between the PDP and the circle.

For each angle, the algorithm calculates the equation of the semi-straight line having origin in the microphone coordinates (x_i, y_i) and orientation $\vartheta_{i,j} + \delta_i$. In general, for the i-th microphone ($i = A, B, C$) the equation will be:

$$y_{i,j} = \tan(\vartheta_{i,j} + \delta_i)\left[x - x_i\right] + y_i \tag{4}$$

where $j = 2, 4$ or 8.

For instance, Fig. 2b presents the case where $M_A = 0.25$. In this scenario the circle intersects two lobes identifying four potential directions $\vartheta_{A,j}$ ($j=1,2,3,4$). Replacing these angles in Eq. (4) yields to the lines $y_{A,1}, \ldots, y_{A,4}$ shown in Fig. 2b. Applying the same approach to microphones B and C, lines $y_{B,j}$, and $y_{C,j}$ are determined. By intersecting all pairs ($y_{k,j}$ and $y_{h,j}$ $k \neq h$, $k,h = A, B, C$) of direction lines, the position of the sound source is univocally identified by the common coordinate resulted from these intersections.

III. EXPERIMENTAL SETUP

A. Hardware/software

The hardware/software system utilized to conduct the experimental program consisted of a portable National Instruments PXI unit running under LabView, a commercial speaker, and a unidirectional microphone model *Audio-Technica AT815b* (Audio-Technica 2008). The LabView program was designed to perform signal generation, acquisition, and storage. A 5-

cycles Gaussian-windowed sinusoidal toneburst was sent to the speaker. This choice provides relatively short duration and small bandwidth. A frequency sweep between 500 Hz to 10 kHz at 500 Hz steps was conducted. Thus, a total of 20 frequencies were considered. The sound detected by the microphone was digitized at 100 kHz sampling frequency.

One microphone was used. In order to simulate the presence of three microphones, three tests were conducted using the same test protocol. Figure 3 shows the spatial and the in-plane location of the microphone in position A, B, and C during tests 1, 2, and 3, respectively. During each test the microphone was manually rotated by 10º. As such the propagating sound was detected from 36 different incidence angles, prior to completing one revolution. Finally, in order to verify the repeatability of the setup, the procedure has been repeated ten times. Thus, a total of two hundreds (20x10) sound waveforms were recorded at a given microphone's orientation.

Table 1 summarizes the experimental protocol. The amount of data recorded by such protocol was next used to simulate the acquisition of a single toneburst by deploying three microphones simultaneously.

Typical pulse waveforms acquired by microphone A (test #1) when the excited toneburst was equal to 500 Hz are shown in Figure 4. In particular, the pulses acquired during one half microphone revolution at 30º increment are presented. The expected decrease of the signal amplitude as the microphone is rotated away from the speaker is evident.

B. Signal processing

Prior to applying the semi-analytical formulation described in section 2.2, the digitized waveforms were pre-processed. The signal processing algorithm is schematized in the flow chart of Figure 5. Each signal was initially windowed in order to reject unwanted reflections or sound

components extraneous to the direct sound path. In order to retain the appropriate time window the expected time of arrival of the signal was considered. The sound's time of flight was determined by cross-correlating the recorded signal with the excited toneburst. To add redundancy to the approach a second cross-correlation was carried out between the signal recorded for a certain angle ϑ and the time waveform recorded at $\vartheta = 0$ degrees. In a practical application where neither the sound source at the origin nor the time waveform at $\vartheta = 0$ is available, other methods may be adapted. For instance, one of these methods may be related to the detection of the peak amplitude (max) and then consider the time of arrival as the moment that the signal amplitude passed a threshold set at the 10% of the max value.

A typical time history and the corresponding fast Fourier transform of the 500 Hz signal recorded by microphone A at $\vartheta = 0°$ before windowing are shown in Figs. 5a and 5b, respectively. It is possible to observe the partial distortion due to the reflections of the sound from the laboratory boundaries. Figure 5c shows the signal after windowing. The duration of the windowed time history was chosen in order to retain the same number (five) of generated cycles. As such the duration of the processed data was around 10% longer than the duration of the excited toneburst. By comparing the frequency domains in Fig. 5d and 5b it is clearly visible that the time window does not alter the overall frequency content of the signal. As such, the important information of the signals is retained.

Once the recorded signal was windowed appropriately, six features from the time domain and three features from the frequency domain were considered. This step was conducted to investigate which signal's features maximize the microphone's unidirectionality. In the time domain the features of maximum absolute amplitude (max), peak to peak amplitude (ppk), variance (var), root mean square (RMS), K-factor (KF), and crest factor (CF) were selected. The

expression of the last four features is:

$$\text{var} = \sum_{l=1}^{N} \frac{(x_l - \mu)^2}{N-1}, \qquad RMS = \sqrt{\frac{\sum_{l=1}^{N} x_l^2}{N}}, \quad KF = \max(x_l) \cdot RMS, \qquad CF = \frac{\max(x_l)}{RMS} \quad (4)$$

where N is the number of points of the windowed time history $x(t)$ and μ is mean of the signal. The maximum amplitude (max FFT), the root mean square (RMS FFT), and the area of the frequency domain were selected as well. As the frequency spectrum spans from zero to the Nyquist frequency, the interval for the computation of the frequency-domain based features include 1 kHz bandwidth centered at the toneburst frequency.

C. Polar directivity patterns

The selected features were calculated at each acquisition angle. This step resulted in the creation of the PDP relative to feature. Figure 7 shows the PDP associated with the time domain based features extracted from the 500 Hz sound. One of the main outcomes from this part of the study is that the shape of the polar pattern is dependent on the selected features. The features (max, ppk, RMS) directly related to the microphone output level are in good agreement to the PDP provided by the microphone's manufacturer (Audio-Technica 2008). The differences are associated with the experimental setup that was not conducted in an anechoic chamber and with the test protocol in which the microphone was rotated manually at discrete steps of 10 degrees. Since the microphone was spun manually, small errors in the orientation (angle ϑ) might have occurred.

The features of the variance and the K-factor show stronger unidirectionality. Not surprisingly the crest factor (CF) denotes instead the behavior typical of omnidirectional microphones. This is a mathematical artifact due to the definition of CF. As the polar pattern

associated with the max and the RMS are very similar, the ratio between such features is close to 1 over the entire angle range. On the opposite side their product "penalized" the values closer to zero. As such the K-factor is strongly unidirectional.

The differences observed among the selected features in the time domain for a 500 Hz pulse were also observed at the higher frequencies. As the method presented here is based on the microphone directionality, the CF was not further considered. In addition, because some features were qualitatively identical, only the max and the K-factor were retained.

Figure 8 shows the PDP extracted from the frequency domain. As the shape of all patterns is very similar only the feature of the maximum value of the frequency spectrum (max FFT) is considered in this study.

Figure 9a shows the experimental PDP associated with the maximum absolute amplitude of the time history (max) of the detected pulse at higher frequencies. As expected the shape of the PDP changes with the frequency and a higher number of lobes were expected. At frequencies above 2 kHz the agreement between the experimental data and the manufacturer's specification degrades. This is mainly associated to the step-wise angle acquisition procedure described in Table 1. As the microphone's normalized output level is more sensitive to small incidence angle variations, the acquisition taken every 10º results in lower resolution. The PDPs associated with the K-factor at frequencies 500 Hz, 1 kHz, and 5 kHz are shown in Fig. 9b.

In order to apply the semi-analytical approach described in section 2, the equation of the experimentally obtained PDP needed to be derived. The experimental data of the extracted feature as a function of the incidence angle in Cartesian coordinate was plot. For instance, Fig. 10a shows the experimental data associated with the maximum absolute of the time history (max) of the 500 Hz pulses. The data are the average of the ten acquisition made at each angle

step. In order to satisfy the symmetry along the 0º-180º direction, the experimental data in the range 0º-180º were also used in lieu of the experimental data acquired from 180º-360º. By using the MATLAB function *fit* and the Fourier polynomial type, the equation that best fit the experimental data was found and it is superimposed as continuous line in Fig. 10a. By comparing the value of the residual R^2, the Fourier equation of order 7

$$M = f(\vartheta, a, b) = a_0 + \sum_{n=1}^{7} a_n \cos(n\vartheta) + \sum_{n=1}^{7} b_n \sin(n\vartheta) \tag{5}$$

was chosen. The empirical coefficients a_n and b_n are summarized in Table 2.

The experimental data and the fitting curve for the 5 kHz pulse processed in the time domain and for the 500 Hz and 5 kHz pulse processed with the maximum of the frequency domain (max FFT) are shown in Fig. 10b – Fig. 10d, respectively. As the frequency of the toneburst increased the interpolating function was less accurate. This is once more associated with the low resolution of the experimental protocol that resulted at the high frequencies.

Figure 11 shows the experimental PDP and the "best fit" polar pattern for the feature of the max amplitude, max FFT, and K-factor for 500 Hz toneburst.

IV. EXPERIMENTAL RESULTS

In the semi-analytical approach developed to localize the sound source position, Eq. (1) is now replaced by the equation of the fitting curve extracted from the experimental patterns. The formulation described in section II.B is used here to identify the location of the speaker. The approach is tested using the normalized features' values of the absolute maximum and the K-factor of the time history, and the maximum amplitude value of the Fourier transform (max FFT). The results from recording pulses at 500 Hz, 1000 Hz and 5000 Hz are presented. The microphone arrangements relative to three microphones array systems are discussed. Following

the angle notation introduced in Figs. 2 and 3, the arrangements are: 30º/0º/-30º, 60º/0º/-60º, and 120º/0º/-120º, where the first angle is relative to microphone C, the second angle to microphone A, and the last angle to microphone B.

Figures 12 shows the results for the 500 Hz toneburst detected by using the array 30º/0º/-30º and the max amplitude as a signal feature. The normalized values of the microphone output levels averaged over the ten acquisitions from detectors C, A, and B were equal to 0.711, 0.999, and 0.669, respectively. Such values are represented by the circles superimposed in the corresponding PDP of Fig. 12a. As each circle intersects the main lobe only, the normalized amplitudes identified two potential incidence angles ($\vartheta_{i,j} + \delta_i$) per microphone, namely + 72º and -12° for microphone C, ± 2º for microphone A, and + 15º and – 75° for microphone B. The algorithm then traced the lines departing from the microphone coordinate along the direction of these potential incidence angles. The lines are indicated in Fig. 12b as $y_{A,1}$ and $y_{A,2}$, $y_{B,1}$ and $y_{B,2}$, and $y_{C,1}$ and $y_{C,2}$. The intersection points among the lines $y_{A,j}$ and $y_{B,j}$, $y_{A,j}$ and $y_{C,j}$, and $y_{B,j}$ and $y_{C,j}$ (with $j=1, 2$) were determined. These points are clustered into groups that are associated to the microphone pairs, namely A-B, A-C, and B-C. The numerical framework created for each point a circular area centered in the point of one array and radius equal to a chosen tolerance. If at least one point from each of the other groups lies within the circle, then the algorithm has found the sound source position with an approximation provided by the tolerance. The procedure is repeated for all the points found in the intersections. If more than one area (three-point solution) is found the algorithm selects the smallest area. The tolerance is necessary to take into account measurement errors due to the test protocol and experimental errors. The tolerance has to be large enough to include one intersection per microphone pair, but it has to be sufficiently small to avoid that false positive, i.e. mathematical solutions of the problem, are considered as physical

solutions. To improve the algorithm performance any solution lying within the area of the triangle ABC is rejected.

The solution found by the algorithm for the case of 500 Hz with microphone arrangement equal to 30º/0º/-30º are represented by the three filled circles in Fig. 12b. The planar coordinates of these dots are x = -0.04 m y = 3.18 m for microphone pair A-B, x = -0.06 m, y = 3.74 m for microphone pair A-C, and x = -0.11 m y = 3.45 m for microphone pair B-C. The locations of these points are in good agreement with the true position of the speaker (x = 0.00 m, y = 3.15 m).

Same approach and signal feature were used to examine the 500 Hz pulses recorded by the array configuration 120º/0º/-120º. The results are presented in Figure 13. As the feature's values associated with the microphone at 120º, 0º and -120º were equal to 0.211, 0.999, and 0.258, respectively, each microphone identified two possible direction lines identified in Fig. 13a as $y_{j,i}$ (j=A, B, C; i=1, 2). Only the intersection between $y_{B,2}$ and $y_{C,2}$ was found. However in order to identify the position of the speaker the algorithm needs two other intersections: one between $y_{A,i}$ and $y_{B,i}$ and one between $y_{A,i}$ and $y_{C,i}$. For detector B, for instance, the "true" direction of the propagation sound was 135º (please see Fig. 3) with respect to the microphone's zero axis. As such this solution would have been possible only if the circle of the feature value had intersected the secondary lobe of the polar pattern.

The outcomes from Fig. 13 suggest that the equation of the PDP needs to be evaluated more accurately especially at those directions where small angle variations induce significant changes of the microphone output level.

Figure 14 shows the location provided by the algorithm when the features of max amplitude, K-factor, and max FFT were applied to examine the pulses at 500 Hz, 1 kHz, 5 kHz recorded by microphone array at 30º/0º/-30º and 60º/0º/-60º.

Figure 14a compares the results of the max amplitude feature applied to the three different frequency pulses recorded by the microphone array 30º/0º/-30º. The triangles are obtained by connecting the points of intersection of the semi-lines $y_{j,i}$. Ideally, such points should be identical and the triangle collapse into a single point. This point shall represent the location of the sound source. Figure 14a shows that as the pulse frequency increases the accuracy of the detecting system degrades. This is due to the precision of the experimental PDP. For instance, by observing the microphone directivity discussed in Fig. 9a, it is evident that the experimental pattern associated with max amplitude lacks needed to be determined with resolution higher than 10º.

Figure 14b compares the results obtained from the three different features applied to the 500 Hz recorded by the microphone array 30º/0º/-30º. Although the max FFT provided a very small triangle the two features extracted from the time domain provided more accurate results when compared to the real position of the speaker represented with the filled circle.

The same three features and pulse frequency were used to evaluate the performance of the microphone array 60º/0º/-60º. The results are presented in Fig. 14c. By comparing Fig. 14c with Fig. 14b it is evident that the identification of the source position is less accurate when angular position of the microphones increases. This may suggest that a larger number of acoustic detectors should be employed if a planar 360º detection strategy is desired.

V. DISCUSSION AND CONCLUSIONS

This paper presents the results of a novel scheme based on the use of unidirectional microphones to determine the planar position of an acoustic pulse source. Few statistical features extracted from the time and the frequency domains are exploited to calculate the experimental

16

polar directivity pattern of the microphone. Multiple arrangement scenarios where three microphones are oriented at a certain angle from each other are simulated. For each arrangement the location of the sound source is located by intersecting all possible sound directions identified through the polar pattern.

Overall the results shown in Fig. 14 demonstrate that the array system 30º/0º/-30º provided the more accurate results, using either time or frequency domain-based features. The performance of the algorithm degrades when higher frequencies are considered. Moreover the outcomes of this study make obvious that the accuracy of the localization algorithm is related closely to the accuracy of the experimental polar pattern.

The results presented in this paper also suggest that in order to accomplish the identification of the sound source location over a full 360º planar direction, an array of six microphones should be deployed.

The study, that has demonstrated the feasibility of the proposed method as an efficient, inexpensive, and compact alternative to the current methods based on omnidirectional microphones and the estimation of the sound arrival time, is part of an ongoing effort to pave the road toward a new generation of sound detectors and localizers that can minimize the spacing between the microphone elements. The method shows computational advantages over the existing techniques and it may offer a suitable tool for further developments and field deployment.

One of the potential limitations of the proposed method is related with the source's signal strength. As the method is based on normalized values of signal features, in field applications it is impossible to establish whether the detected signal is intrinsically weak or it has been generated far away from the sound detectors. A feasible approach to overcome this problem is

normalizing the signal captured by the array of unidirectional microphones with the signal captured by one omnidirectional microphone. As such the three-point solution of features values is always comprised between 0 and 1.

Future efforts will evaluate the capability of the proposed methodology to detect a sound source location in 3-D problems.

ACKNOWLEDGEMENTS

REFERENCES

Atmoko H, Tan D C, Tian G Y and Fazenda B 2008 Accurate sound source localization in a reverberant environment using multiple acoustic sensors *Meas. Sci. Technol.* **19** 024003 (10pp), doi: 10.1088/0957-0233/19/2/024003.

Audio-Technica 2008, http://www.audio-technica.com (last viewed September 1, 2008).

Benesty J 2000 Adaptive eigenvalue decomposition algorithm for passive acoustic source localization *The Journal of the Acoustical Society of America* **107** 384-391, doi: 10.1121/1.428310.

Bian X, Abowd G D and Rehg J M 2005 Using sound source localization in a home environment *Pervasive Computing* **3468** 19–36, doi: 10.1007/b136550.

Birchfield S and Gillmor D 2001 Acoustic source direction by hemisphere sampling *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP-01*) **5** 3053-3056, doi: 10.1109/ICASSP.2001.940302.

Bourennane S and Bendjama A 2002 Locating wide band acoustic sources using higher order statistics *Applied Acousics* **63** 235-251, doi: 10.1016/S0003-682X(01)00039-1.

Brandstein M S and Silverman H F 1997 A practical methodology for speech source localization with microphone arrays *Computer Speech & Language* **11** 91-126.

Brutti A, Omologo M and Svaizer P 2008 Comparison between different sound source localization techniques based on a real data collection *Proceedings of the Hands-Free Speech Communication and Microphone Arrays (HSCMA-08)* 69-72, doi: 10.1109/HSCMA.2008.4538690.

Elko G W, Chou T C, Lustberg R J and Goodwin M M 1994 A constant-directivity beamforming microphone array (A) *The Journal of the Acoustical Society of America* **96** 3244, doi:

10.1121/1.411096.

Hu J-S, Yang C-H and Wang C-K 2008 Sound source localization by microphone array on a mobile robot using eigen-structure based generalized cross correlation *Proceedings of the IEEE ARSO 2008, Workshop on Advanced Robotics and its Social Impacts* 1-6, doi: 10.1109/ARSO.2008.4653625.

Huang Y, Benesty J and Elko G W 2000 Passive acoustic source localization for video camera steering *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-00)* **2** II909-II912, doi: 10.1109/ICASSP.2000.859108.

Ikeda A, Mizoguchi H, Sasaki Y, Enomoto T and Kagami S. 2007 2D Sound source localization in azimuth & elevation from microphone array by using a directional pattern of element *Proceedings of IEEE Sensors 2007* 1213-1216, doi: 10.1109/ICSENS.2007.4388627.

Jin S, Kim D, Kim H S, Lee C H, Choi J S and Jeon J W 2008 Real-time sound source localization system based on FPGA *Proccedings of the 6th IEEE International Conference on Industrial Informatics (INDIN-08)* 673-677, doi: 10.1109/INDIN.2008.4618186.

Kim H-D, Choi J-S, Kim M and Lee C-H 2004 Reliable detection of sound's direction for human robot interaction *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-04)* **3** 2411- 2416, doi: 10.1109/IROS.2004.1389769.

Kim H-D, Komatani K, Ogata T and Okuno H G 2008 Design and evaluation of two-channel-based sound source localization over entire azimuth range for moving talkers *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-08)* 2197-2203, doi: 10.1109/IROS.2008.4650947.

Kleban J T 2000 Combined acoustic and visual processing for video conferencing systems *CAIP Center for Advanced Information Processing Technical Reports TR-251* (MS Thesis, Rutgers

University).

Knapp C H and Carter G C 1976 The generalized correlation method for estimation of time delay *Proceedings of IEEE **ASSP** Trans. Acoust. Speech Signal Processing* **24** 320–327, ISSN: 0096-3518.

Kobayashi T, Kameda Y and Ohta Y 2008 Sound source localization with non-calibrated microphones *Advances in Multimedia Modeling* **4903** 134-143, doi: 10.1007/978-3-540-77409-9.

Kwon B, Park Y and Park Y-S 2007 Sound source localization using the compensation method in robot platform *Proceedings of the IEEE International Conference on Control, Automation and Systems (ICASSP-07)* 1911-1914, doi: 10.1109/ICCAS.2007.4406659.

Liu H and Milios E 2005 Acoustic positioning using multiple microphone arrays *The Journal of the Acoustical Society of America* **117** 2772-2782, doi: 10.1121/1.1895005.

Lv X and Zhang M 2008 Sound source localization based on robot hearing and vision *International Conference on Computer Science and Information Technology* 942-946, doi: 10.1109/ICCSIT.2008.26.

Merimaa J 2002 Application of a 3-D microphone array *Proceedings of the 112th Convention of the Audio Engineering Society*, 1-10, Paper 5501.

Nakadai K, Matsuura D, Okuno H G, and Kitano H 2003 Applying Scattering Theory to Robot Audition System: Robust Sound Source Localization and Extraction *Proceedings of the 2003 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS-03)* **2** 1147-1152.

Olson H F 1946 Gradient Microphones *The Journal of the Acoustical Society of America* **17**, 192-198, doi: 10.1121/1.1916315.

Omologo M and Svaizer P 1994 Acoustic event localization using a crosspower-spectrum phase based technique *Proceedings of the IEEE International Conference on Acoustics, Speech, and*

Signal Processing (ICASSP-94) **2** II273-II276, doi: 10.1109/ICASSP.1994.389667.

Reid G L and Milios E 2003 Active stereo sound localization *The Journal of the Acoustical Society of America* **113**, 185-193, doi: 10.1121/1.1518469.

Rui Y and Florencio D 2003 New direct approaches to robust sound source localization *Proceedings of the ICME International Conference on Multimedia and Expo* **1** 737-740, doi: 10.1109/ICME.2003.1221023.

Rui Y, Florencio D, Lam W and Su J 2005 Sound source localization for circular arrays of directional microphones *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-05)* **3** 93-96, doi:10.1109/ICASSP.2005.1415654.

Sessler G M and West J E 1975 Second Order Gradient Uni-Directional Microphones Utilizing an Electret Transducer *The Journal of the Acoustical Society of America* **58** 273-278, doi: 10.1121/1.380657.

Valin J M, Michaud F, Rouat J and Letoumeau D 2003 Robust sound source localization using a microphone array on a mobile robot *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-03)* **2** 1228- 1233, ISBN: 0-7803-7860-1.

Wang H and Chu P 1997 Voice source localization for automatic camera pointing system in videoconferencing *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)* **1** 187–190, doi: 10.1109/ICASSP.1997.599595.

Wang H and Kaveh M 1985 Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources *IEEE Transaction on Acoustics, Speech and Signal Processing* **33** 823-831, ISSN: 0096-3518.

Wilson D K, Sadler B M and Pham T 2002 Simulation of detection and beamforming with acoustical ground sensors *Proceedings of SPIE, Seismic and Acoustic Unattended Ground*

Sensors I **4743** 50-61, doi: [10.1117/12.443525](10.1117/12.443525).

.

FIGURE CAPTIONS

Figure 1. - Polar directivity pattern of a unidirectional second order gradient microphone.

Figure 2 – (a) Directivity Patterns of the microphones' setting and localization of the source. (b) Angles found and corresponding lines of possible direction.

Figure 3 – Schematic arrangement of the pilot study. (a) 3-D microphone placement; (b) planar view of the test setup with microphones oriented at 120º from each other. Dimensions are expressed in millimeters. Drawing not in scale.

Figure 4 – Time waveforms detected at various angles of incidence between the microphone and the speaker.

Figure 5 – Flowchart adopted for the experimental characterization of the microphone PDP.

Figure 6 (a) Time domain recorded. (b) Recorded signal FFT. (c) Time domain windowed signal. (d) Windowed signal FFT.

Figure 7 – Polar directivity pattern associated with the selected features from the time domain at 500Hz.

Figure 8 – Polar directivity pattern associated with the selected features from the frequency domain at 500Hz.

Figure 9 – Polar directivity pattern associated with (a) the feature of the maximum amplitude of the signals and (b) the feature of the K-factor.

Figure 10 – Experimental directivity pattern in Cartesian coordinates and curve fit (a) 500 Hz max amplitude; (b) 500 Hz max FFT; (c) 5000 Hz max amplitude; (d) 5000 Hz max FFT.

Figure 11: Experimental polar directivity pattern associated with the features of max amplitude, max FFT, and K-factor of the 500 Hz propagating toneburst.

Figure 12 – Source localization by using the feature of the maximum amplitude of the 500 Hz tonebursts. (a) Microphone configuration (30°, -0°, -30°) and experimental polar directivity patterns. (b) Localization of the sound source. The coordinates of the axes are expressed in meters.

Figure 13 – Source localization by using the feature of the maximum amplitude of the 500 Hz tonebursts. (a) Microphone configuration (120°, -0°, -120°) and experimental polar directivity patterns. (b) Localization of the sound source. The coordinates of the axes are expressed in meters.

Figure 14 – Location of the sound source obtained by using: (a) the feature of the maximum amplitude of the time waveform and microphones oriented at 30º/0º/-30º; (b) various features from the 500 Hz signal recorded by the microphones oriented at 30º/0º/-30º; (c) various features from the 500 Hz signal recorded by the microphones oriented at 60º/0º/-60º. The solid circle identifies the location of the sound source.

Figure 1

(a)

Acoustic source

Microphone A

(x_A, y_A)

Microphone C

(x_C, y_C)

O

Microphone

(x_B, y_B)

(b)

$y_{A,2}$

$y_{A,1}$

0.5

1

0.5

0.5

1

$M_A = 0.25$

0.5

$y_{A,3}$

$y_{A,4}$

1

Figure 2

(a)

1180

1180

A

2000

2000

C

B

1732

(b)

15°

2000

A

1732

$\delta_C=120$ $\delta_B=-120°$

C O B

2000
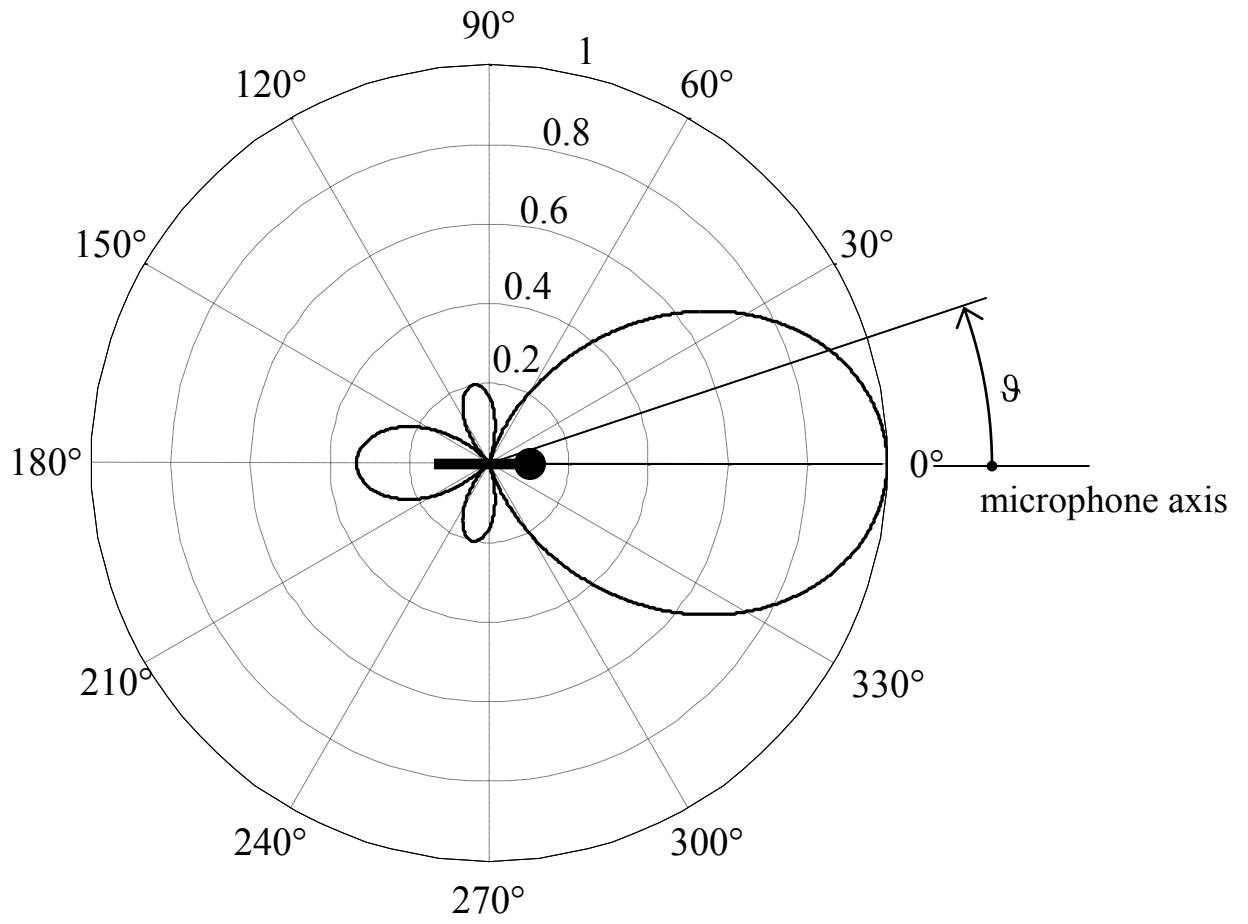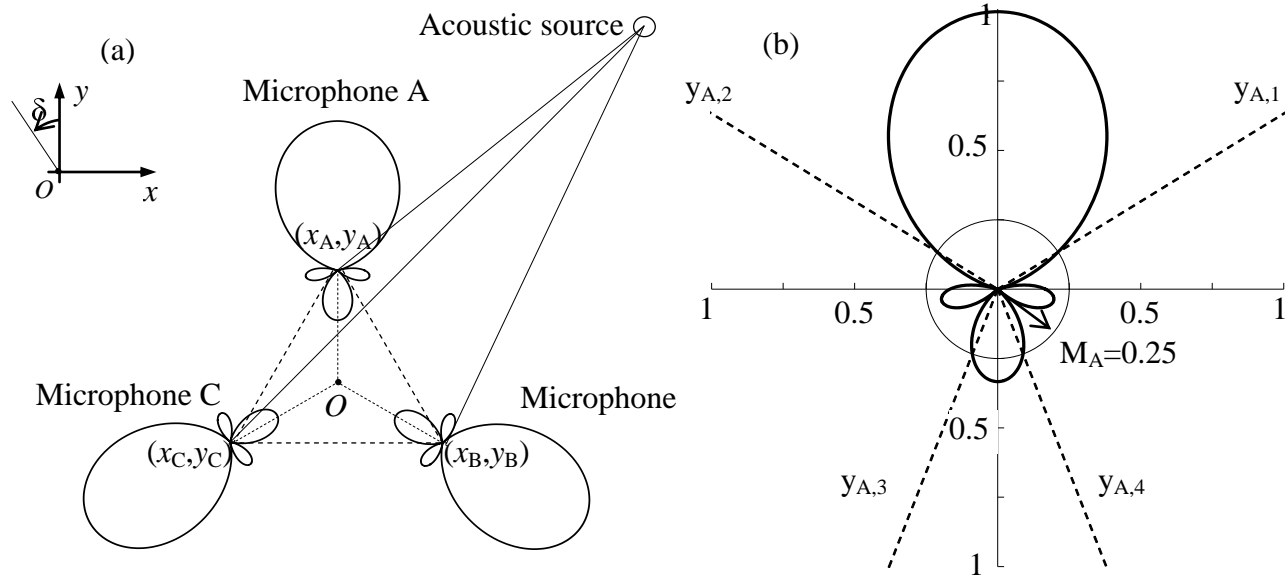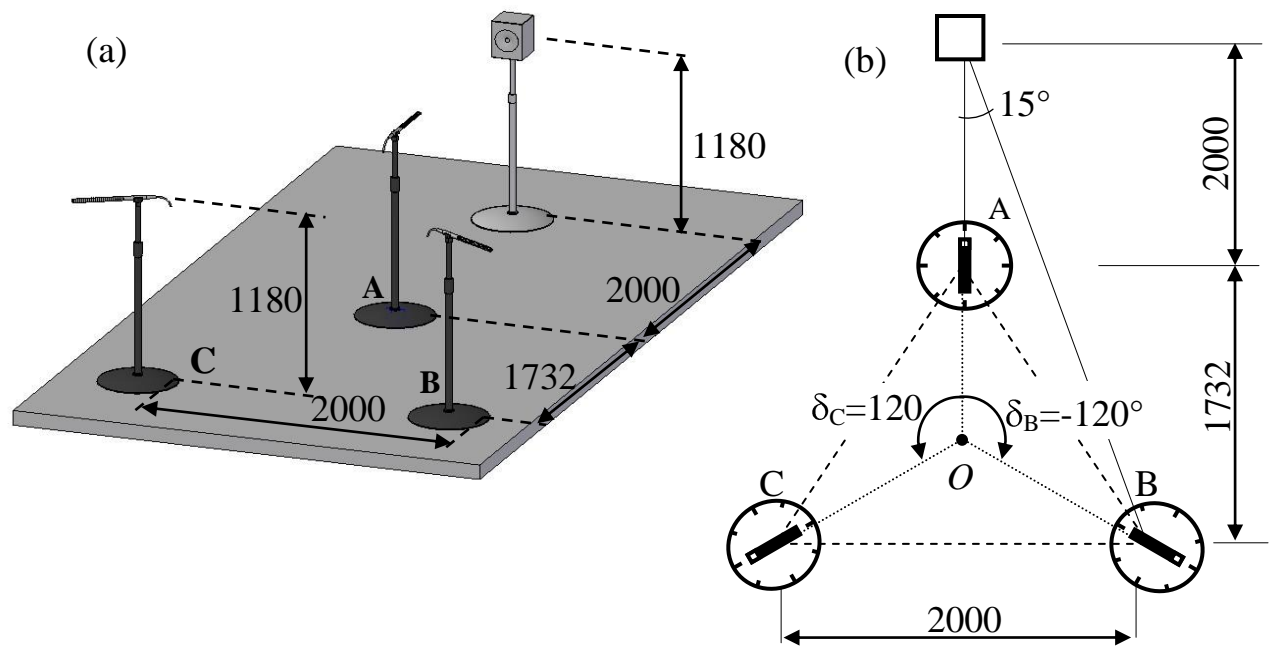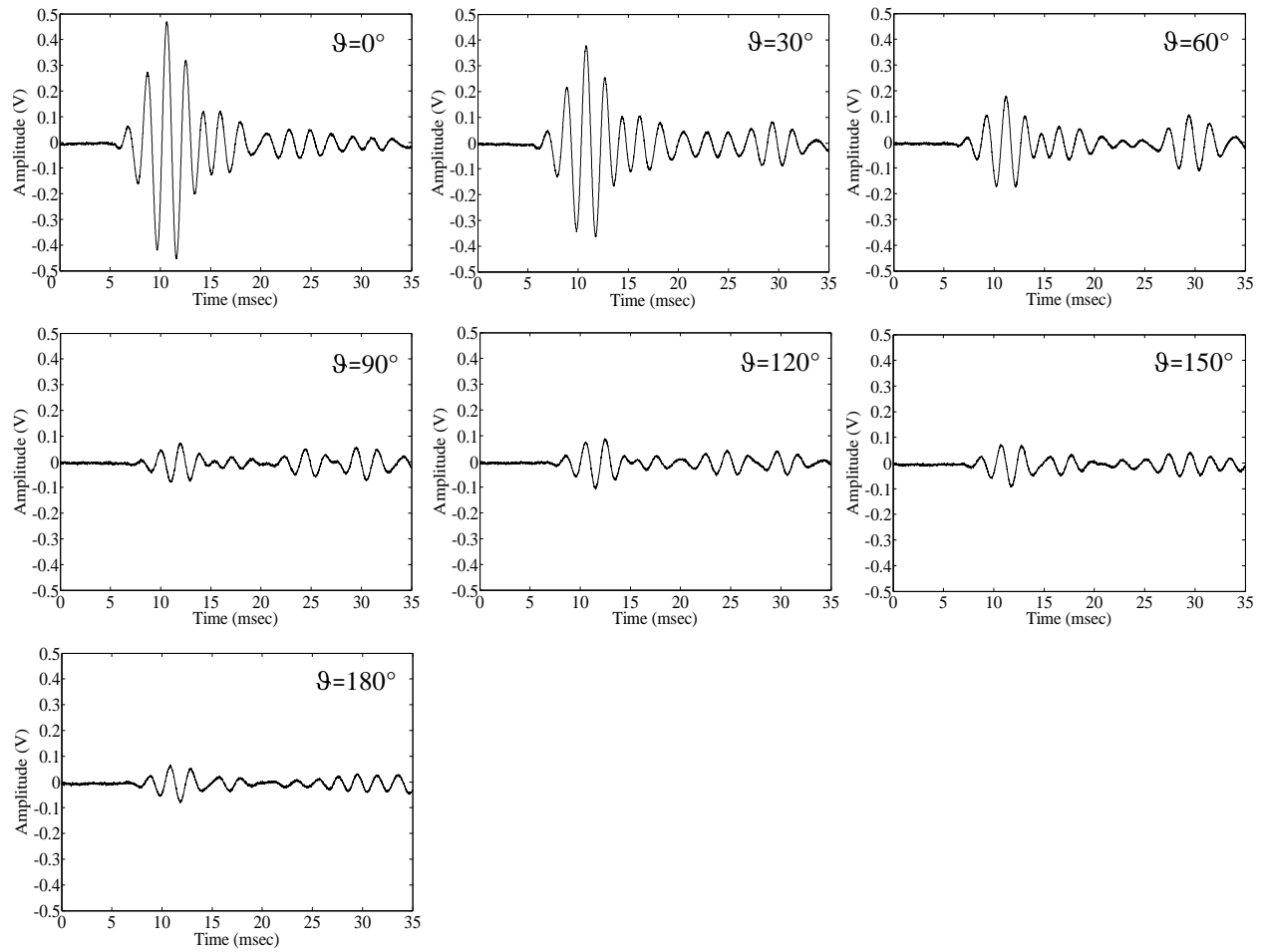
Figure 3

Figure 4

Figure 5

Figure 6

max value

peak to peak

variance

RMS

k factor

crest factor

Figure 7

max FFT

AREA FFT

RMS FFT

Figure 8

(a)

500 Hz   1 kHz   3 kHz

5 kHz   8 kHz   10 kHz

(b)

500 Hz   1 kHz   5 kHz
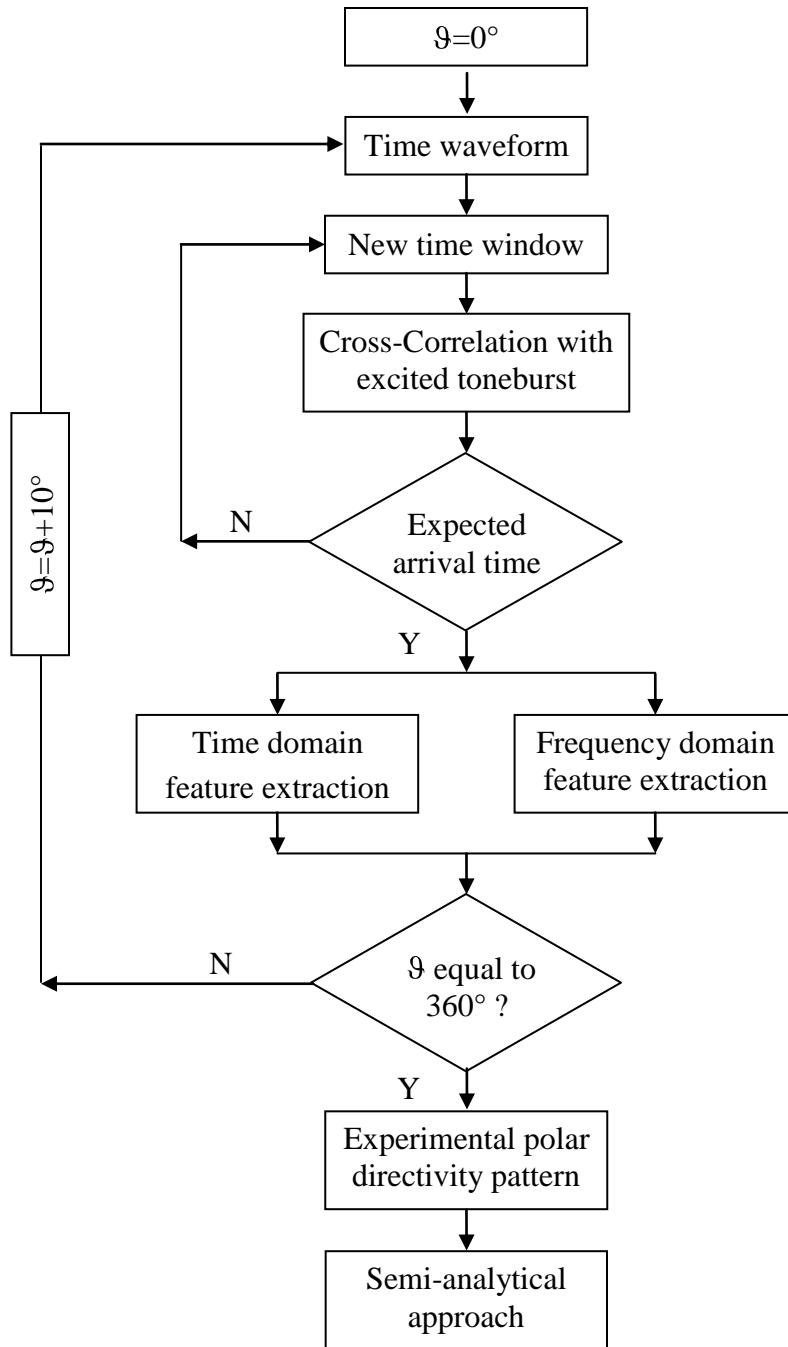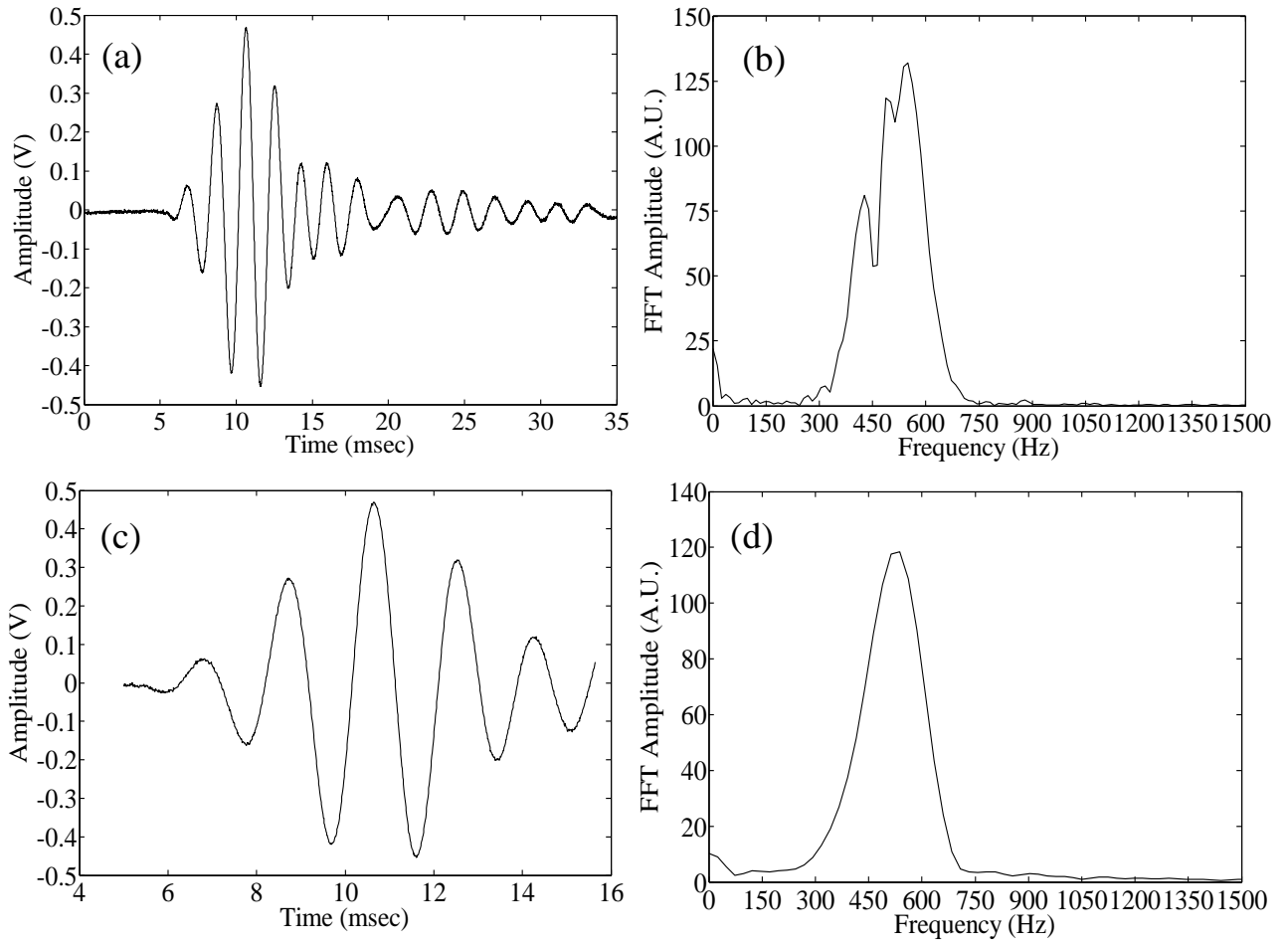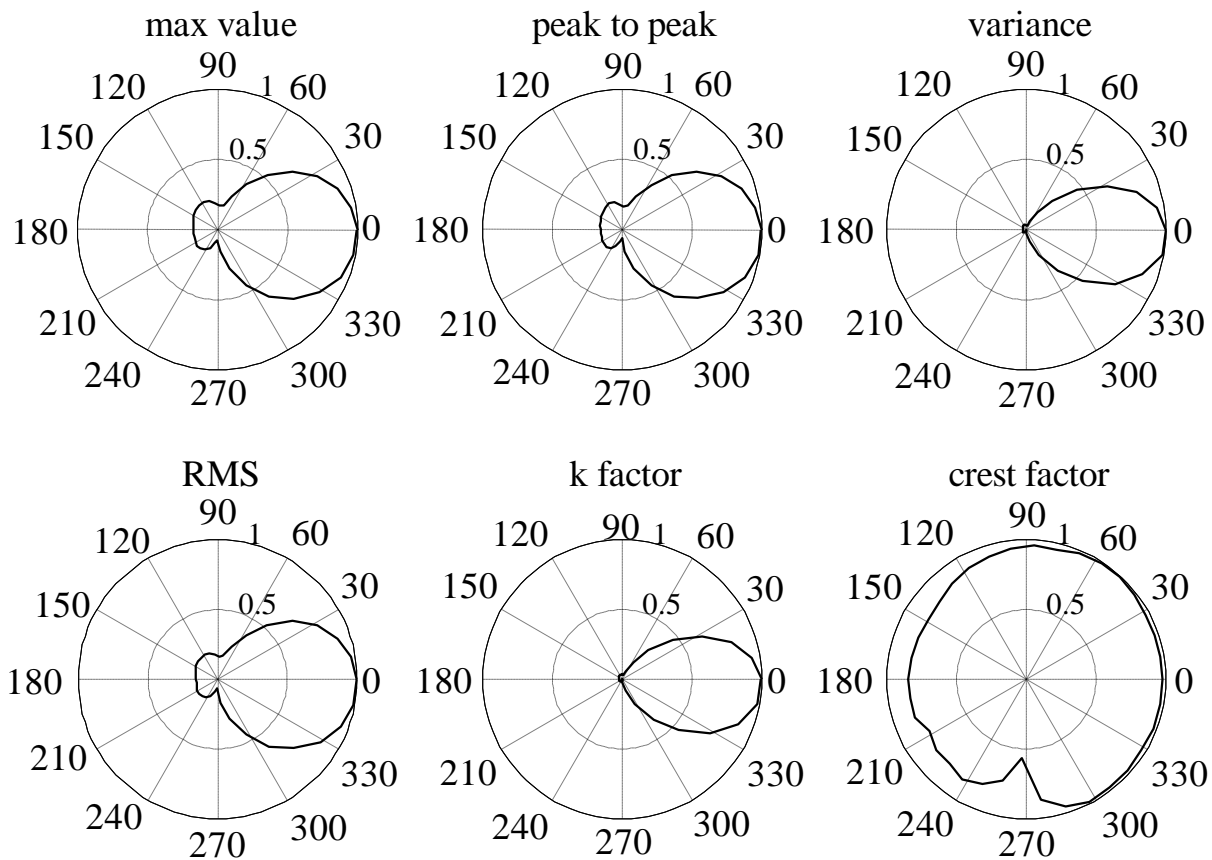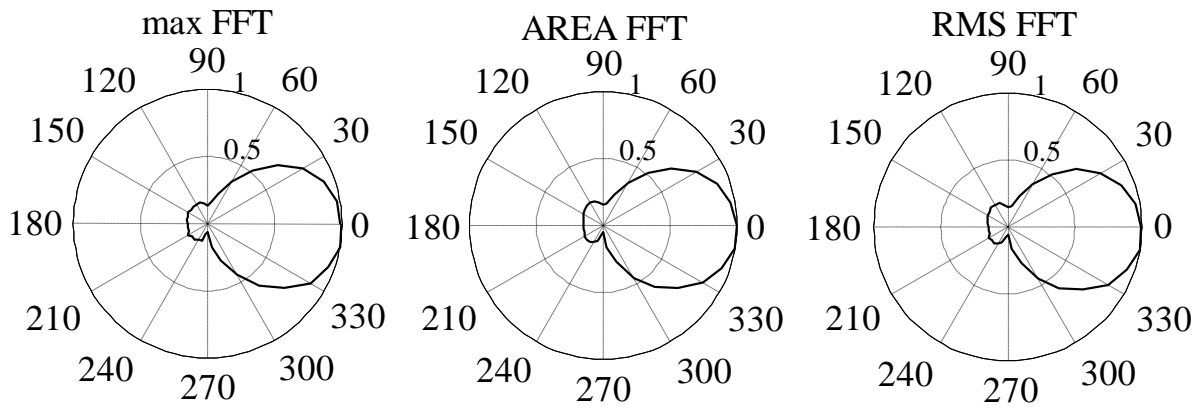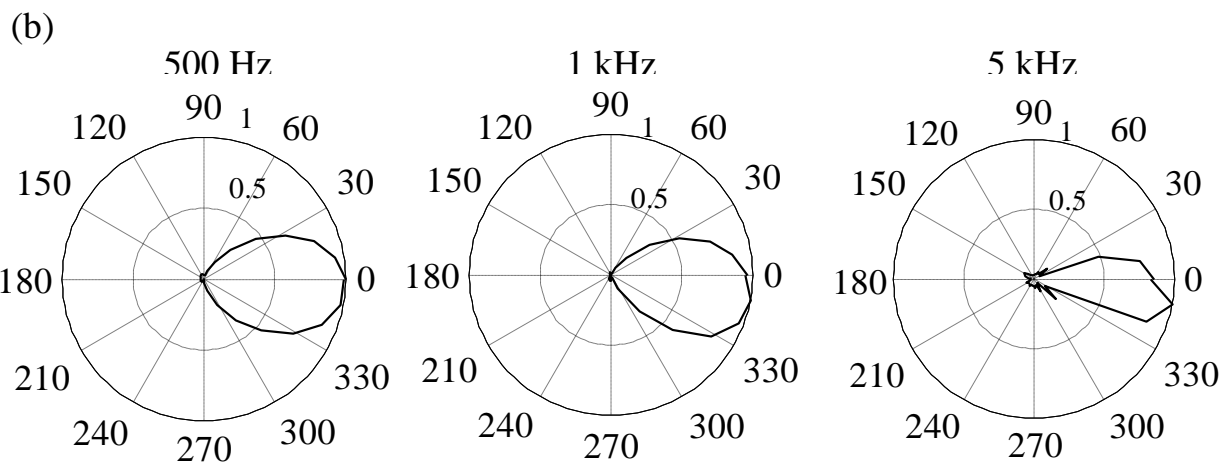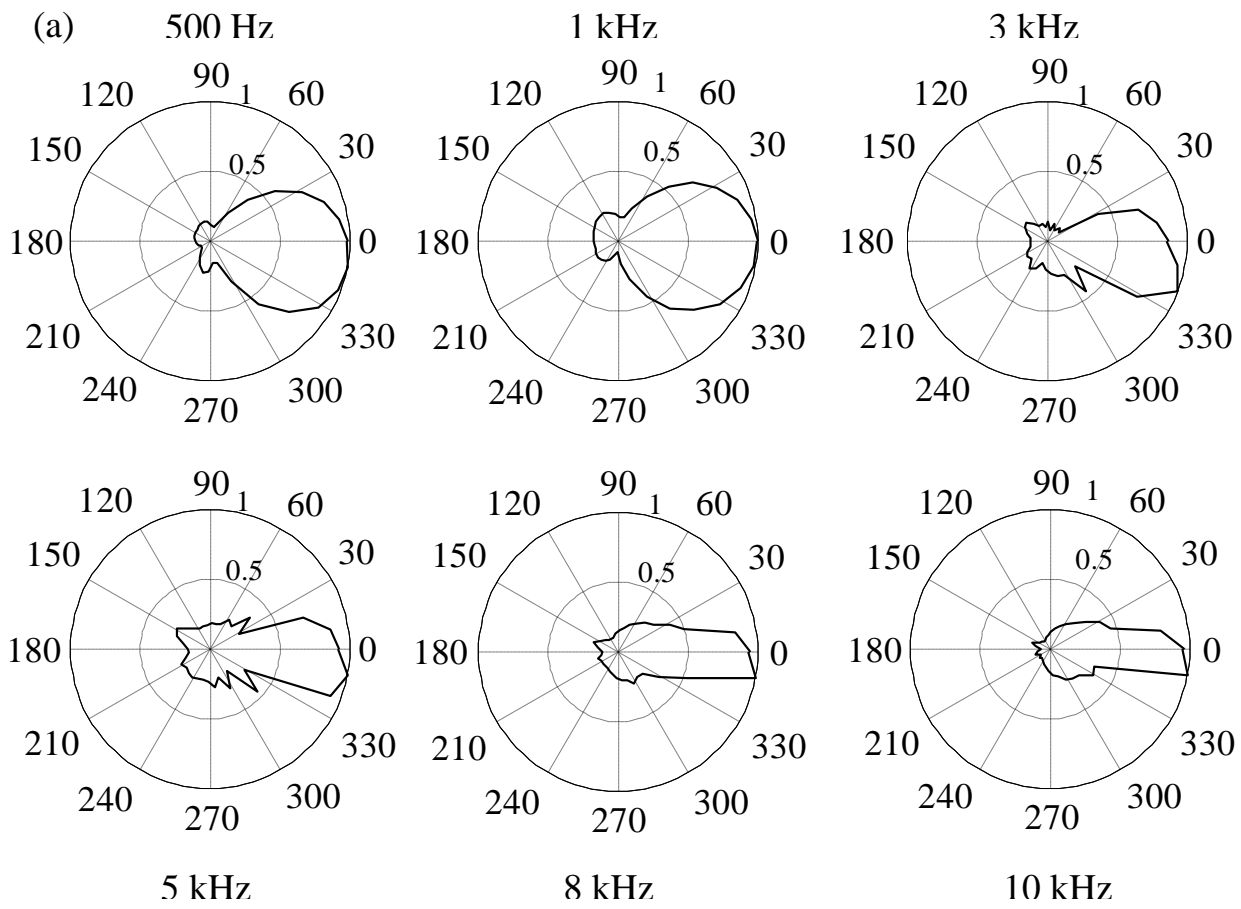
Figure 9

Figure 10

Figure 11

(a)

2° -2°
0.999
0.5
1

0° -42°
42°
0.5
0.711
1

45° 0°
-45°
0.5
1
0.670

(b)

$y_{A,1}$ $y_{A,2}$
6
$y_{B,1}$ $y_{C,1}$
4
2
(x_A, y_A)
$y_{C,2}$ $y_{B,2}$
4 2 2 4
(x_C, y_C) (x_B, y_B)
2
4

Figure 12

(a)

2° -2° 0.999

0.5

1

-75° 0.211

0.5

1

0°

75°

72° 0.258

0.5

1

0°

-72°

(b)

$y_{A,1}$ $y_{A,2}$

6

4

$y_{C,1}$ $y_{B,1}$

2

(x_A,y_A)

4 2 2 4

(x_C,y_C) (x_B,y_B)

2

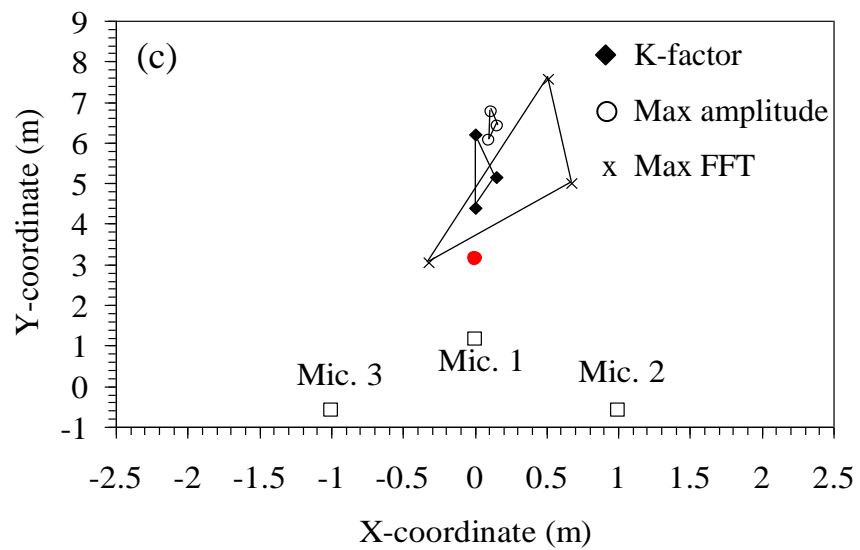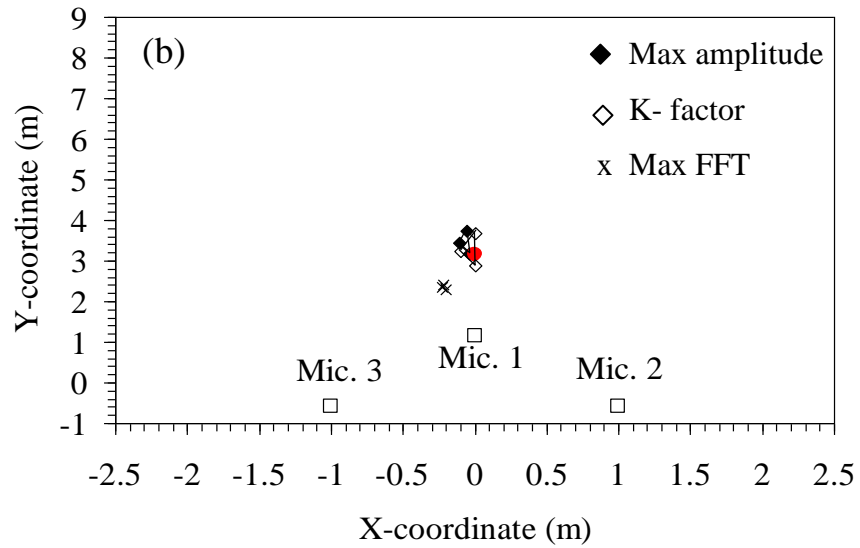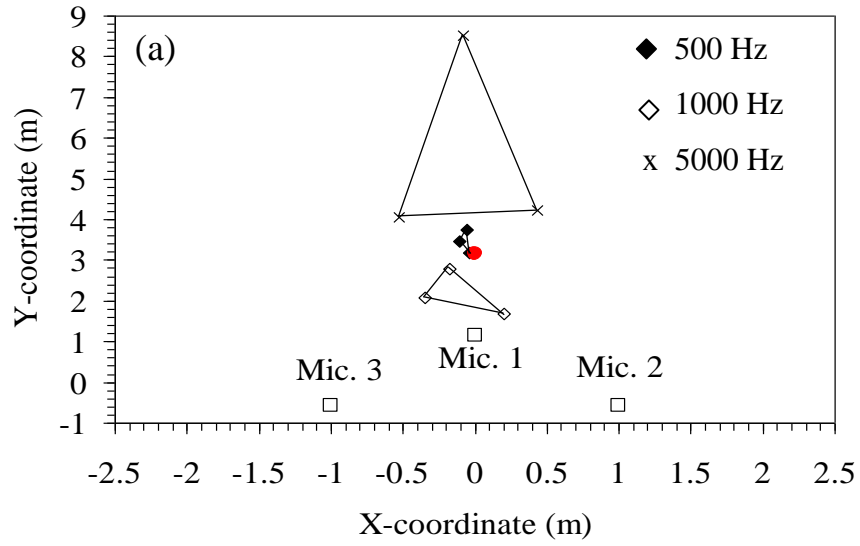$y_{C,2}$ $y_{B,2}$

4
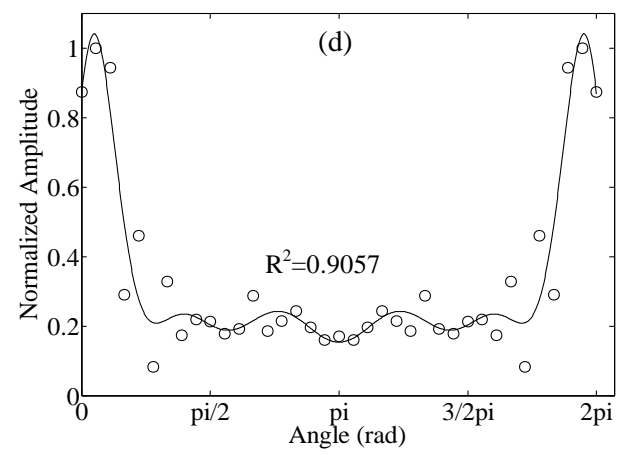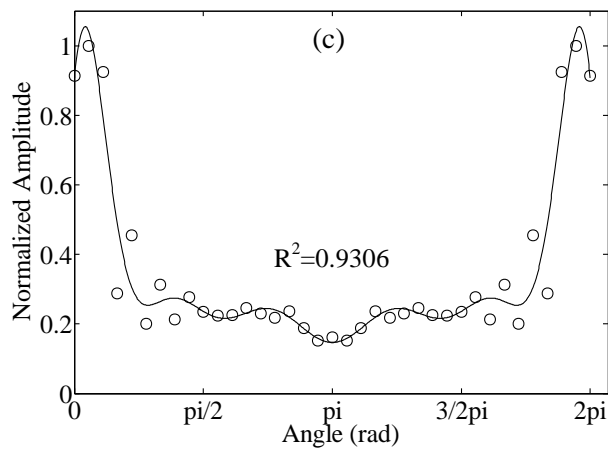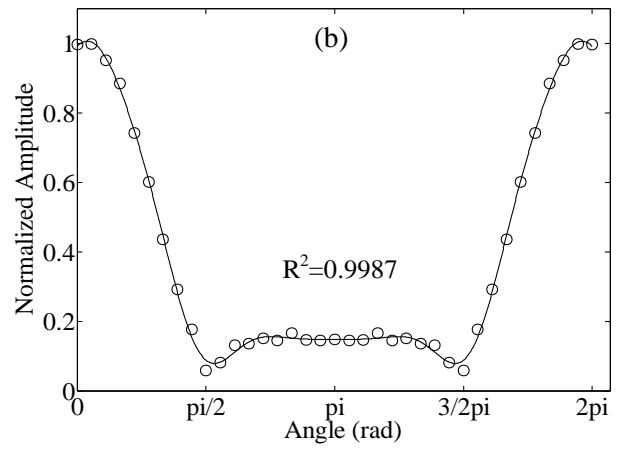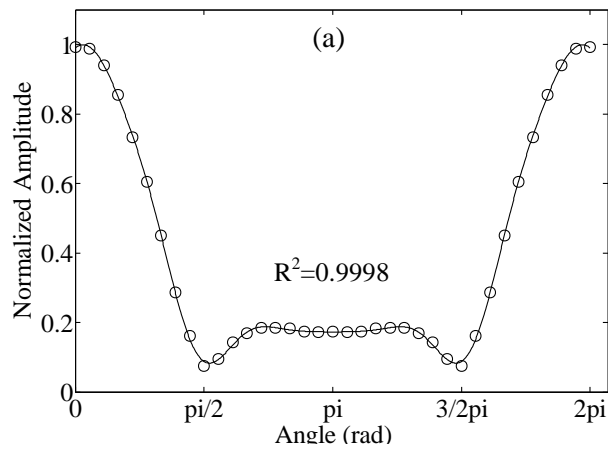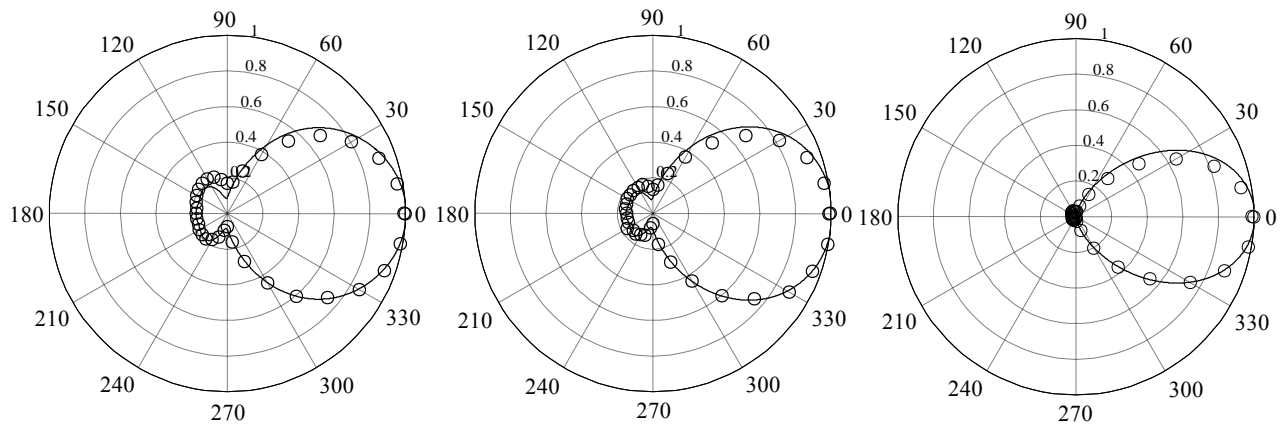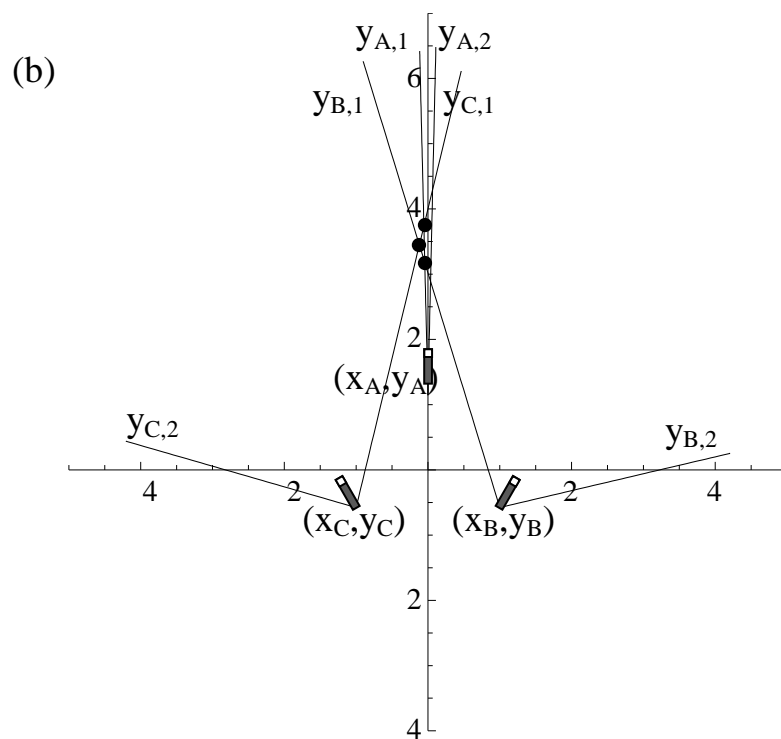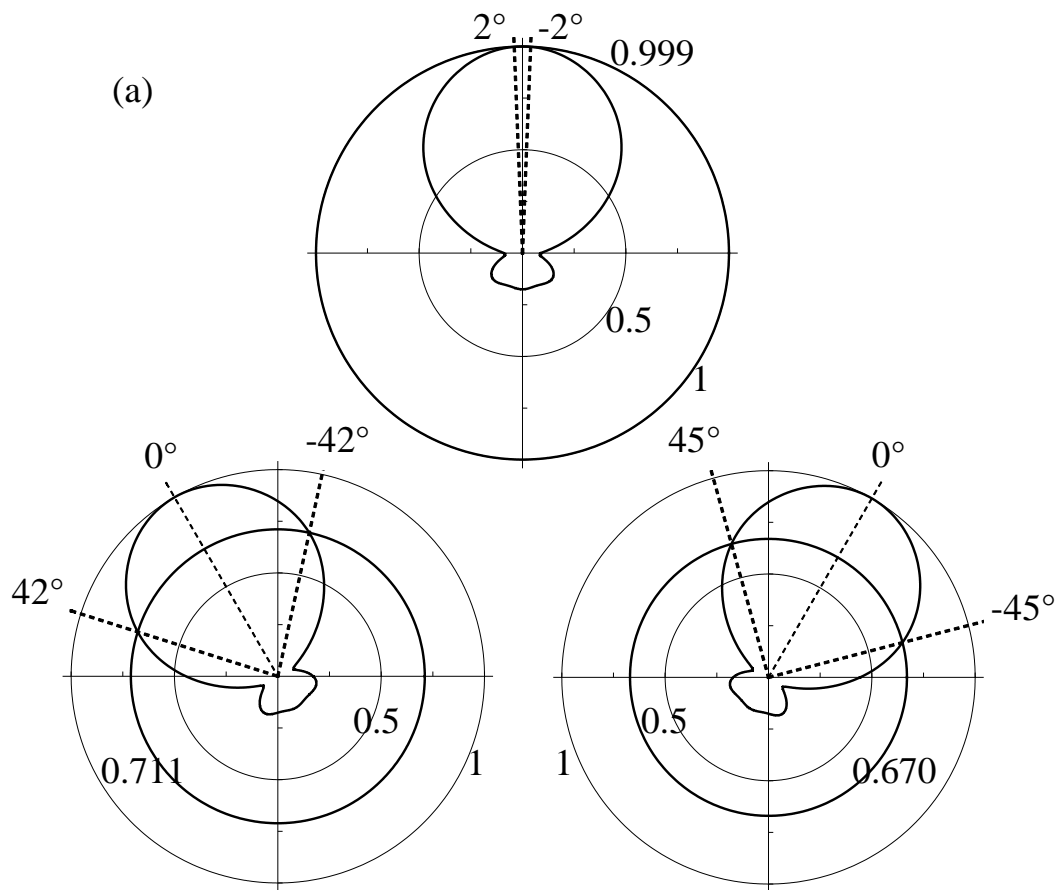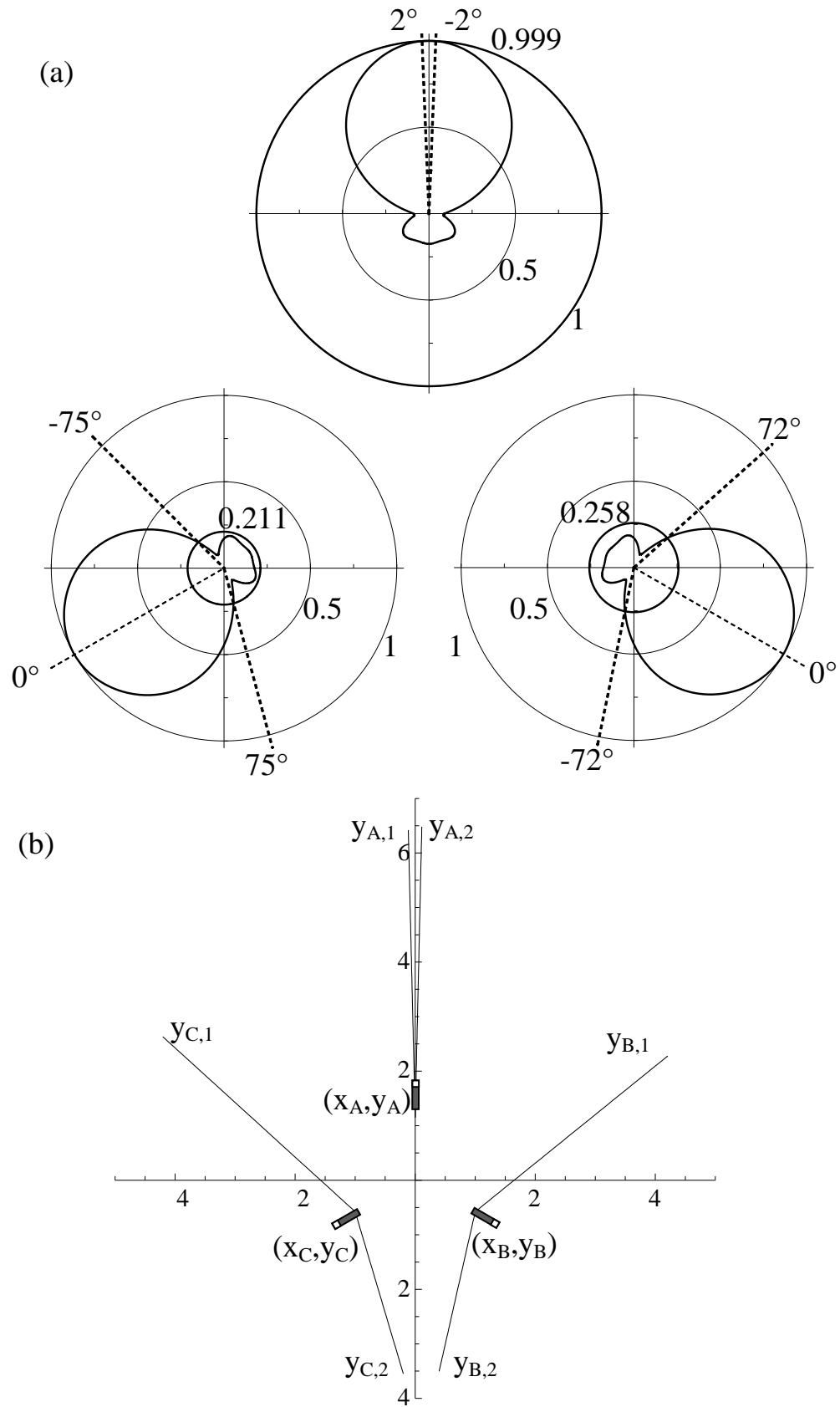
Figure 13

Figure 14

Table 1 - Summary of the tests conducted in this study.

est#	Initial/End frequency (Hz)	Frequency increment (Hz)	Measurements per frequency	Initial/End Angle (º)	Angle increment (º)
A	500 / 10,000	500	10	0 / 350	10
B	500 / 10,000	500	10	0 / 350	10
C	500 / 10,000	500	10	0 / 350	10

Table 2 – Experimental coefficients of the polar directivity pattern of the unidirectional microphone used in this study.

n	a_n	b_n
0	0.3875	
1	0.3787	0.0002219
2	0.2347	0.0002751
3	0.04661	0.00008194
4	-0.04718	-0.0001106
5	-0.01263	-0.000037
6	0.01534	0.00005394
7	0.001065	0.000004368