9-12 AUGUST 2015
inter.noise 2015
SAN FRANCISCO, CALIFORNIA USA

Localization of incoherent multiple sources using three-dimensional sound intensity array

Filip Franek[a)], Jeong-Guon Ih[b)]
Center for Noise & Vibration Control, Department of Mechanical Engineering
Korea Advanced Institute of Science and Technology, Daejeon 305-761, Korea

Efren Fernandez Grande[c)]
Department of Electrical Engineering
Technical University of Denmark, Lyngby, Denmark

Localization of multiple acoustic sources is important in various industrial works. This study describes an approach to localize the incoherent multiple acoustic sources using three-dimensional sound intensity array. Because a standard three-dimensional sound intensity array module using the pressure-to-pressure probes can estimate only a single vector at a time instance, one can detect only a single dominant source or ghost source due to spatial averaging of the signals from multiple sources. By combining the blind source separation technique with double three-dimensional sound intensity array, we attempted the localization of multiple, incoherent sound sources. The enhanced precision of the source detection comes from the double intensity array, of which shape is a coupled tetrahedron sharing. Four sources composed of three speech signals and one Gaussian random noise are adopted in the experiment conducted in an anechoic chamber. The precision of bearing angle detection is found to be within units of degrees that depends on the factors as the spatial proximity between sources, the statistical dependency properties, and the level of signal-to-noise ratio. Performance evaluation measures are suggested over statistical quantities and possible error sources are discussed.

1 INTRODUCTION

Information about a direction and a strength of a propagating sound wave can be obtained through a measurement of a vector quantity such as the particle velocity, or the sound intensity[1]. A practical sound field is usually, however, composed of multiple simultaneous sound events, which are extremely difficult to be distinguished by standard single vector sound intensity

[a)] email: filip@kaist.ac.kr
[b)] email: j.g.ih@kaist.ac.kr
[c)] email: efg@elektro.dtu.dk

method, implemented by scanning the sound field manually[2] or combining the intensity computation with various source separation methods. This study is to tackle such a sound scenario with a multiple sound intensity array modules, 50 mm in diameter, approximately, in order to identify directions of multiple sound sources with a relatively high angular precision within a few degrees. A sound source separation technique is accommodated to the sound intensity measurement, which is, in particular, a combination of a blind audio source separation (BASS) using an independent component analysis method (ICA). The multiple 3D sound intensity modules has been produced in the previous research works[3,4]. An example suggestion is to use a double twisted configuration of 3D sound intensity array that tends to cancel out an angle detection error. The BASS algorithm comes from an extensive research of Koldovsky et al.[5] and it was chosen on the basis of a satisfying separation performance and the provided algorithm. A velocity-to-velocity probe was adopted to implement the Multiple Signal Classification (MUSIC) algorithm, and it was proved on the commercial device[6]. Uncorrelated sound sources could be separately identified although the angular precision was questionable with an error of $\pm 20°$, approximately, and the number of identifiable sound sources was limited by the number of sensors. Another approach introduced by Williams et al.[7,8] employed many microphones to measure the pressure around an open or hard sphere and the spherical harmonic expansion are then used to obtain a vector intensity field to detect multiple dependent as well as independent sound sources. In spite of advantages, this approach requires a number of microphones forming a spherical array with a radius of approximately 0.1 m.

2 THEORETICAL BACKGROUND

The idea of any sound source separation lies in the filtering of single sound sources out of a mixture of sound sources. 4 mostly used techniques in acoustic signal processing are: beamforming, computational auditory scene analysis, multiple signal classification, and blind audio source separation (BASS). Considering the size of microphone array in this work and the known performance information on the foregoing separation techniques, BASS is selected to use, specifically employing the independent component analysis (ICA). The BASS does not assume almost any prior knowledge about the signals to be separated. It uses a spatial diversity of an acoustic sound field, i.e. phase and amplitude differences between sensors, and a statistical independence of sound signals.

2.1 ICA

In order to achieve sufficient separation, the ICA assumes that at least one audio source to be normally distributed and, at the same time, all sources in the sound mixture are statistically independent. In a real room environment, assuming that sound waves are generated by d sound sources ($s_{j,j=1,\ldots,d}$), while m microphones ($p_{i,i=1,\ldots,m}$) record a direct path contribution from each sound source, as well as its delayed and filtered version using a transfer functions $a_{i,j}$of a given environment between j^{th} source and i^{th} microphone. Such system is described mathematically by convolutive mixture as follows:

$$p_{i(n)} = \sum_{j=1}^{d}\left(a_{i,j} * s_j\right) = \sum_{j=1}^{d}\sum_{\tau=0}^{M_{i,j}-1}\left(a_{(i,j)}(\tau)s_j(n-\tau)\right). \quad (1)$$

This convolution is expressed as the original sources being modified by filters $a_{i,j}$ of the length $M_{i,j}$. The BASS algorithm can be used either in time or frequency-time domain; we adopted a time-domain algorithm[5] to preserve the temporal information that is important for the sound intensity calculation. In order to find the separated sound source estimates, deconvolution of Eq. (1) is needed. In this work, T-ABCD[5] algorithm is used to this end.

2.2 3D Sound intensity

A 3D sound intensity array configuration needs to be arranged from the minimum of 4 microphones to detect 3 orthogonal vectors leading to finding a 3D spatial vector. A computation of the used configuration is based on a single tetrahedron configuration as depicted in Fig. 2, where the derivation of the sound intensity comes from a Taylor approximation as was described by Pascal[9]. The sound intensity in the frequency domain is thereafter given as follows[1]:

$$I_i(\omega) = \frac{1}{2}\, Re\left\{P(\omega)\left(\frac{j}{\rho\omega}\frac{\Delta P_i(\omega)}{\Delta r}\right)\right\}\ ,\ (i = x, y, z). \tag{3}$$

The measured pressures P and their gradients in each Cartesian coordinates in frequency domain are sufficient for obtaining the desired spatial intensity vector $I(\omega)$.

3 EXPERIMENTAL STUDY

The procedure of the BASS method combined with the 3D sound intensity technique is symbolically summarized in Fig. 1. Recorded pressures p_i of mixed sound sources are passed to the BASS algorithm, which estimates original sound sources waveforms at those microphone positions in estimated vectors v_j. These vectors contain separated sound sources for each microphone, therefore preserving the time delay information among the microphones. The estimated signals are then passed by a group of 4 vectors to the 3D sound intensity algorithm that processes single estimated sound sources separately and indicates bearing angles. The proposed method was subjected to 4 sound sources composed of 3 speech signals and a noise with Gaussian distribution.

A twisted double module tetrahedron is illustrated in the left-hand side of Fig. 1. Each of 2 tetrahedron array has got spacing of 50 mm among each microphone and they share one microphone at the apex[4]. Compression drivers are used as sound sources and the experiments are conducted in an anechoic condition with reliable low frequency limit of 150 Hz set by a size of the anechoic chamber. Acoustic drivers are placed between 0.9 m and 1.5 m away from the array, and therefore they are considered to be in the direct sound field only[10]. The angles related to source positions served as a reference for separated sources.

The resulting convolutive mixture of all sound sources playing simultaneously can be seen at the top of Fig. 2. The BASS method is applied to the 4 microphone recordings as described in Fig. 1 and separated estimated signals are depicted below the mixture together with single original sound signals for a comparison. The Efficient ICA (EFICA) algorithm has been chosen for the separation. The BASS T-ABCD algorithm then allows to choose more parameters to tailor the deconvolution performance of the mixed signal. We took 1 second of the signals and applied a generalize cross-correlation (GCC) similarity measure, a hierarchical clustering with Toeplitz weightings, and the filter length $L_{i,j}$ of 20. All of other parameters were unchanged from their default state. The chosen parameters were chosen on the basis of having better performance compared to other options. In our case of the BASS method, the number of separated sound sources depends on the number of microphones.

The angle detection of separated sources plotted in Fig. 3 shows the reference sound source angle by the grey dashed lines and the separated source angles by the black solid lines with corresponding marker. Numerical details about the single and separated sound sources are given in Table 1 for the azimuth and elevation angles. The single number of angles of reference as well as separated signals are found over arithmetic average, and their difference results in an error $\Delta\theta$ for the azimuth angle, and error $\Delta\phi$ for the elevation angle. The standard deviation σ is obtained from the fluctuations around the averaged angles of separated signals. The angle distribution can also be express by a 3D histogram as shown in Fig. 4, that allows to see a different picture of the angle error spread. The vertical axis corresponds to the azimuth angle, the horizontal axis to the elevation angle and the depth is an occurrence of angles in the frequency domain. Such a picture would be seen, for example, by a camera from a position of the microphone array.

4 DISCUSSION

From Fig. 2, one can notice that the sources have been relatively well separated with little artefacts. One can see the performance of the angle detection of separated sources in Fig. 4. The best result is achieved with the Gaussian noise which has got a flatter angle distribution along the frequency for both, azimuth and elevation angle. A good separation is achieved by the distinctive probability distribution of the original noise signal as compared to the Laplacian distribution of speech signals. These speech signals are prone to higher errors caused by their similar statistical properties and also by containing a background noise of a Gaussian character from recording. Depending on how close are the sources to each other, the error bias also changes. One can observe that the signal s_3 is closest to the signal s_2 and, therefore, it seems that the separation has been affected by this proximity and it can be seen that how s_3 is approaching the angle estimation of s_2 for both angles. It is seen that signal s_4 bears the most fluctuations from the separated signals. This is due to the lower SNR of the signal. The speech signal s_1, s_3 and the noise signal s_2 were well separated and the angle error is within a few degrees.

The error depends on the separation performance, but it is within $\pm 4°$ for the well separated signals and $\pm 6°$ for the other signals. One cannot beforehand know what signals will be separated better, but this can be evaluated on the basis of standard deviation, that is low for well-separated, and thus more accurate signals. In this manner, ranking of separated signal can be performed and only well-separated signals can be taken as reliable results. This also rise a question what happens in case more than 4 sources are present. It depends on a chosen setting of the algorithm, because the similarity measure, and weightings of clustering determine how many signals are distinguished from the unmixing process. However, in this case, 3 sources will be well separated with small standard deviation and the 4th signal will contain the mixture of the remaining sound sources with a noise reflecting large standard deviation.

The 3D histogram representation reflects the artefact of separation, in which one can be clearly observe a smearing effect. In all the separated sound sources, there is an artefact from the neighbouring sound sources. The larger the standard deviation, the more prominent the smearing is, and there are more local maxima caused by the components of other sound sources that are not completely separated from the mixture. These faulty peaks are not at a position of their original sources, but they are pointing at their direction.

A question of the best averaging of the separated angles could arise. Currently, an arithmetic mean is utilized, but other statistical measures are also examined. As depicted in Fig. 4, 3D histogram shows local maxima off the centre due to an error of the separation. For that reason, other approaches for finding maxima of skewed distribution of separated angles reflect worsened precision. A phase calibration of microphones can be also addressed to reduce the angle error, but it does not show significant improvement in the low frequency end, so it is omitted.

5 CONCLUSIONS

A suggested demonstration is given for a 3D sound intensity angle detection of multiple incoherent sound sources when combined with the BASS returns reasonably good results. Although the resultant source localization is less accurate compared to that using single sound source, it is encouraging because they are within an error of a few degrees. The separation performance depends on statistical, spatial and energy properties of sound sources. The most appropriate separated algorithm for good separation performance and correct number of separated sources is not easily defined yet, which can be regarded as a disadvantage of the current implementation, with an additional burden in computational. It is observed that the measurements in reverberant rooms worsen the separation due to added background noise, and it would need longer filters to decompose the longer impulse response of a room. Further experiments should be conducted to figure out limits of the method.

6 ACKOWLEDGEMENTS

This work has been partially supported by BK21 Plus Project and NRF (2012-0000977 and 2012R1A2A2A01009874).

7 REFERENCES

1. Fahy, F., “Sound Intensity”, Taylor & Francis, (2002)
2. Janssens, K.; De Weer, D.; Bianciardi, F. & Sondergaard, T. “On-line sound brush measurement technique for 3D noise emission studies SAE”, Technical Paper, (2013)
3. Cho, S.-K. “Acoustic source localization by using double-module intensity array”, Master Thesis, KAIST (2012)
4. Ih, J.-G.; Woo, J.-H. & Cho, S.-K. “Acoustic source localization by using twisted double-module 3D intensity array” INTER-NOISE and NOISE-CON Proceedings, 247, 2779-2784, (2013)
5. Koldovsky, Z. & Tichavsky, P. “Time-domain blind separation of audio sources on the basis of a complete ICA decomposition of an observation space” IEEE Transactions on, 19, 406-416, (2011)
6. Basten, T.; de Bree, H.; Druyvesteyn, W. & Wind, J. “Multiple incoherent sound source localization using a single vector sensor” ICSV16, (2009)
7. Williams, E. G.; Valdivia, N.; Herdic, P. C. & Klos, J. “Volumetric acoustic vector intensity imager” The Journal of the Acoustical Society of America, Acoustical Society of America, 120, 1887-1897, (2006)
8. Williams, E. G. & Takashima, K. “Vector intensity reconstructions in a volume surrounding a rigid spherical microphone array”, The Journal of the Acoustical Society of America, Acoustical Society of America, 127, 773-783, (2010)
9. Pascal, J.-C. & Li, J.-F. “A systematic method to obtain 3D finite-difference formulations for acoustic intensity and other energy quantities” Journal of Sound and Vibration, Elsevier, 310, 1093-1111, (2008)
10. Jacobsen, F. “*Sound intensity and its measurement and applications”* in “Current Topics in Acoustical Research”, 3, 87-91, (2003).

Table 1 *- Summary of azimuth and elevation angle detection for multiple separated sources in anechoic condition. The top line corresponds to Signal-to-Noise ratio, reference angle, separated angle, error angle, and standard deviation.*

Sound source	SNR (dB)	Ref. $\theta_{ref}(°)$	Sep. $\theta_{sep}(°)$	Error $\Delta\theta_{sep}(°)$	Std. dev. $\sigma_\theta(°)$	Ref. $\theta_{ref}(°)$	Sep. $\theta_{sep}(°)$	Error $\Delta\theta_{sep}(°)$	Std. dev. $\sigma_\theta(°)$
s1	24.1	-12.2	-13.9	-1.7	4.0	11.8	13.7	1.9	3.1
s2	35.7	54.4	55.9	1.5	4.3	-15.3	-15.0	0.3	1.5
s3	26.3	29.2	31.1	1.9	4.7	-10.7	-11.8	-1.1	3.1
s4	22.7	-34.8	-31.4	3.4	10.3	-45.5	-47.7	-2.2	4.5

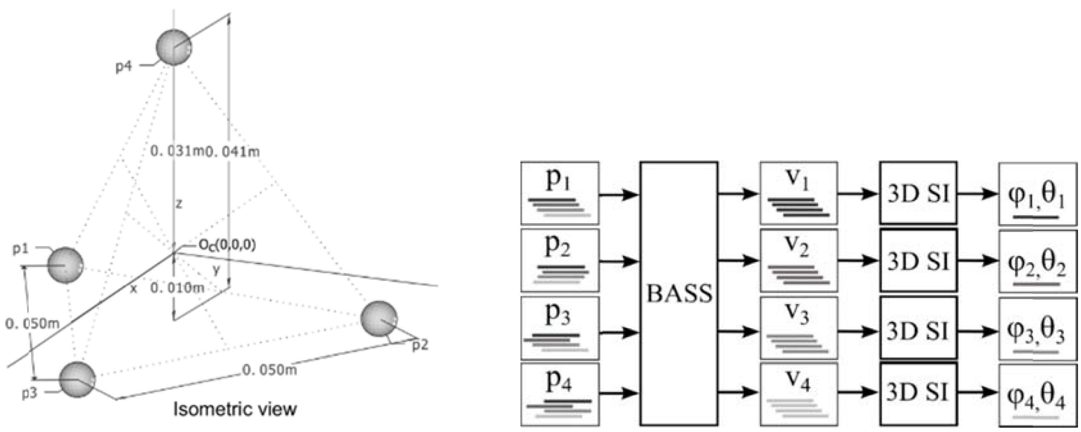

Fig. 1 *- A symbolic algorithm flow of the incoherent sound sources detection over the 3D tetrahedron sound intensity array.*

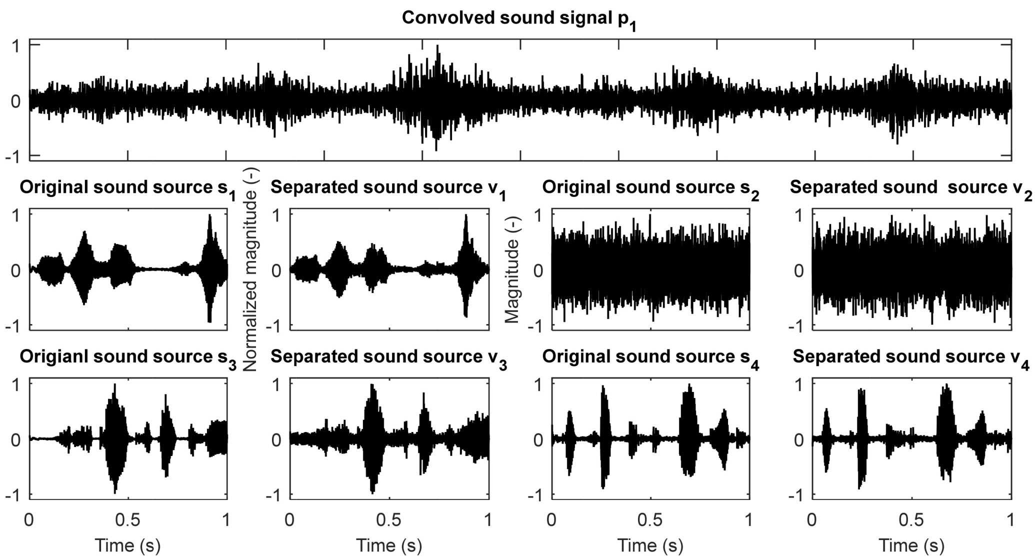

Fig. 2 *- Time domain representation of a convolved signal obtained from microphone* p_1*, when all original signals are playing simultaneously, is depicted in the uppermost graph The original sound signals are in the left column, depicted while playing alone. The estimates of separated sound sources are in the right column.*

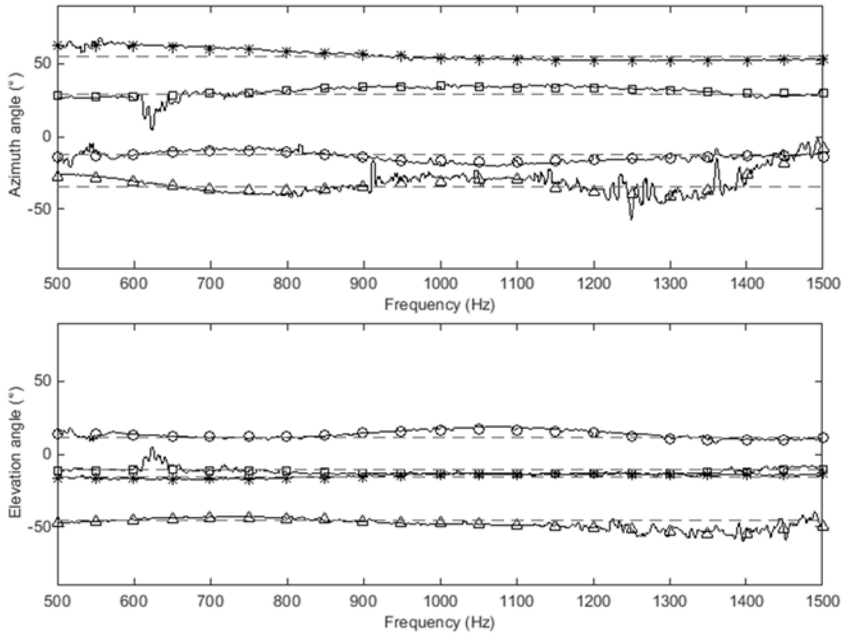

Fig. 3 - *Angle estimation of multiple separated signals. Estimated separated source 1 is represented as circle ○, source 2 as star *, source 3 as square □, and source 4 as triangle Δ.*

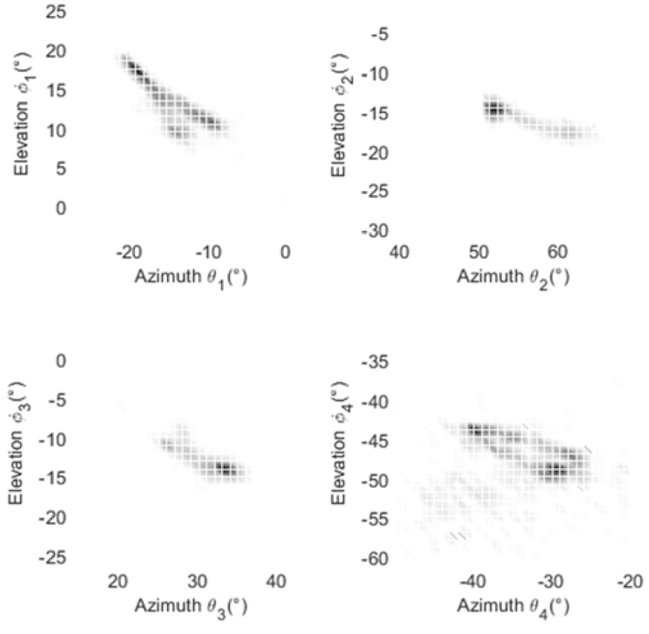

Fig. 4 - *3D histogram of azimuth and elevation angle detection for separated sound sources. The reference angle is located in the centre of each histogram. From top left to bottom right are displayed angles of separated sources from 1 to 4. Vertical axis correspond to azimuth angle, horizontal axis to elevation angle and depth in grey scale is occurrence of angles in frequency domain.*