Pollster j. acad.res. 04(01) 106-118, 2017
© Pollster Journal of Academic Research, Pollster Publications
ISSN: 2411-2259, 2017, Vol (04), Issue (01)
www.pollsterpub.com

A Novel Method for Sound Localization Based on Improved Correlation algorithms & Spatial Gradients Approach

Muhammad Mubashir[1], Md Rishad Ahmed[2], Muhammad Ahmad Shabir[3]

[1,2,3] Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, 250022, P.R. China

Email address: mubi_aquarius@yahoo.com

ABSTRACT

Sound localization has been essential in the overall scheme of simulating the placement of an auditory cue. In our work, we have proposed improved correlation algorithms, spatial gradients approach, and inverse algorithm over cross channel algorithms based on time queues comparing them with biological inspired methods. The correlation algorithms were tested with sources in the plane in front of them and were able to determine the position of a source within less than a degree of error. The spatial gradients approach was also tested in the same experiment but in a spherical/3d environment and had the same accuracy, moreover but the spread of the estimate for a given source position was much more reliable. A related technique, transfer function based ones were found to have cross channel algorithms with an error of about 2 degrees having an edge over ordinary human hearing with a chance of around 5 degrees of error in the best scenario; whereas our proposed methodology has given the more accurate results with even less than 1 degree of error. Finally, we gave a comparison based on complexity limitations of each of these and accuracy of our practiced work.

Keyword: Sound localization, auditory, time queues, correlation, cross-correlation, cross-correlative derivative, spatial gradients, inverse, cross channel, spherical

INTRODUCTION

Sound localization (SL) is the process of determining the position of a sound source by using the strings of recorded audio. It is a key necessity for all sound-related frameworks and has persuaded much research. Distance, vertical angle, and the azimuth angle are the three-dimensional position to describe localization. The main purpose of the work, but not limited to, is to combine cross correlation

algorithms (CCA) and spatial gradients approach (SGA) for the better accuracy while localizing a sound source. In real-life, SL situations needs to consider that more than one sound source could be animated in the nature's domain. Therefore it is also important to gauge those positions of various synchronous Sound sources.

- Time difference of arrival (time delays)
- Distance estimation (angle with the sound source)

SL done in real-life situations needs with consider that more than person sound source could be animated in the domain. SL need been considerably pushed ahead keeping by refining traditional techniques for example, absolute distance of-arrival (DOA) estimation, learning-based methodologies, shaft forming-based approaches, subspace methods, source clustering through duration of the time and following strategies like Kalman filters. Same time actualizing these techniques, a few facets applicable to SL need been constructed apparent including: type and number for microphones used, versatility and number of claiming sound sources, vigorousness against commotion and resonance, kind of array geometry to be occupied, sort from claiming platforms to raise upon, and so.

As is illustrated in this article, SL is quite mature. To instance, (Nakadai, 2015) elucidates a percentage SL meets expectations in light of binaural strategies also multiple-microphone arrays, (Argentieri, 2013) and (Argentieri, 2015) displays an overview on binaural robot audition, (Xiaofei, 2012) adorn a general study from claiming SL to Chinese, and (Okuno, 2015) displays a review of the audition field overall. An approaches for ascertaining time difference of arrival (TDOA), also inter-aural time difference (ITD) is measuring the time difference between the moments of zero level crossings of the signals (Huang, 1999). To analyze the time delay between of two signals, cross correlation is a much more supportive as to figure out the time delays for propagation of acoustic signals over microphone array (Rhudy, 2009). In our proposed study, correlation algorithms and spatial gradients approach were proposed to localize the sound in a plane and spherical/3d environment respectively. Moreover, in our study, we proposed inverse algorithm rather than of cross channel algorithms and the results shows that it had better accuracy than other related methods.

The aim of this work is to localize the sound in a plane and spherical/3d environment proposing correlation algorithms and spatial gradients approach respectively. Moreover, in our study, we proposed inverse algorithm rather than of cross channel algorithms which worked better than of that as proved in the later sections. As the result of a given experiment previously, the transfer function based ones were found to have cross channel algorithms with an error of about 2 degrees having an edge over ordinary human hearing with a chance of around 5 degrees of error in the best scenario; whereas our proposed methodology has given the more accurate results with even less than 1 degree of error.

The main contributions of this study are as follows:

a) To localize the sound in a plane and spherical/3d environment proposing improved correlation algorithms and spatial gradients approach respectively.
b) Proving better accuracy in locating the sound source as of other related methods i.e. cross channel algorithms and head related transfer functions.

The rest of the paper is organized as follows. In Section 2, we epitomized the proposed architecture for our work while other sub-sections include the details working of correlation algorithms (CCA), spatial gradients approach (SGA), inverse and cross channel algorithms and heart related transfer functions (HRTF). A brief comparison of CCA, SGA, and HRTF is typified in sub-section K which proves the better accuracy of our proposed study over others. Section 3 elucidates discussion & results of our proposed work following with the devoted Section 4 for conclusion of proposed study.

MATHODOLOGY

A. *Time Delays*

Firstly, we used the set options related to correlation algorithms as presented in fig. 2. We have two microphones, one at position $M_1$and the other at $M_2$and there will be a single source of sound at an unknown position P, relative to the microphones such that the sound approaches them at an angle alpha. We'll go here to identify this angle and using the inter-aural time delay between the two microphones we would find specifically the time between sound hitting M_1 and it hitting M_2, for example here we have seen that it hits M_2 first.

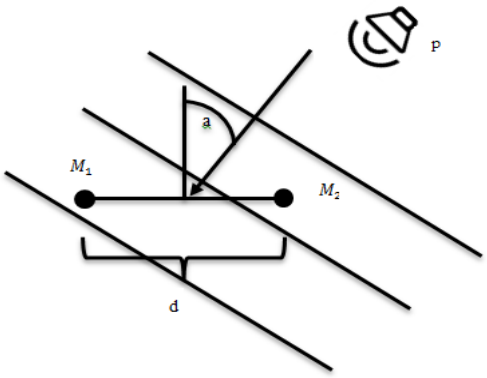

Fig. 2 identifying azimuthal angle "alpha" using T_d

For these algorithms we'll be assuming that the difference between microphones and the sound source will be significantly more than the actual of the apparatus. This is so that we can model that the sound approaches as a plain wave and thus the angle which produces with the microphones is constant. We'll be using the inter-aural time delay to calculate alpha using eq. 1 here and once we find the delay value

we can rearrange and solve. $T_{d,max}$ is the maximum possible time delay quite clearly occurs when our sound source is in line with those two microphones.

$$T_{d=}\ T_{d,max}Sin(a)\ \ldots eq.1$$

$T_{d,max} = d/c$ is maximum possible delay, when $a = -90\ deg$.

B. *Correlation Algorithms*

Correlation is just a straight-forward property about two continuous functions. Here we'll initiate it to $X_1$and X_2 which would be the signals from two microphones respectively utilizing a portion counterbalance parameter 'ᴛ' which will be our estimate of time difference. Eventually observing the composition of the signals from the two microphones, we consider that they comprise about true signals 'S' which may be the thing that we really need to watch from the proper source. Some noise signal is assumed negligible and whatever which does come up they would autonomous about each other conclusion of unique indicator such that their correspondence turns out to be zero. Accordingly, our equation for correlation turns into as eq. 2, from which it will be exactly clear to perceive that this will bring a maximum value when 'ᴛ' equals the actual true delay.

$$R_{x1,x2}(\tau) = \int_{-\infty}^{\infty} X_1(t)X_2(t+\tau)dt$$

Also $X_1(t) = s(t) + n_1(t)\ and\ X_2(t) = s(t - T_d) + n_2(t)$ giving

$$R_{x1,x2}(\tau) = \int_{-\infty}^{\infty} s(t)s(t - T_d + \tau)dt \cdots eq.2$$

C. *Cross Correlation Algorithms*

It just takes the discrete time approach of the conversion that we finished from eq. 2 to eq. 3 which provides us the value $y$comparing of the correspondence at the amount about tests of the estimate delay i also by maximizing y we might figure out the value i which does this and use that to calculate our estimate of time delay and, consequently, the alpha.

$$y(i) = \tilde{R}_{x1,x2}(iT_s) = \sum_{k=0}^{K} X_1(kT_s)X_2((k-i)T_s) \cdots eq.3$$

The value of i that maximizes y leads to approximation $\tilde{T}_d$.

D. *Proposed CCA*

This method is a good start and we can obtain sufficiently accurate results by using a single bit of quantization from our microphones which means rather than extended floating point values between -1 and 1, each sample of our impulse stream is either 0 or 1. We can do this because the approach of our correlation relies on time based queues rather than of actual relative amplitudes of the two signals. Using this 1 bit quantization makes our instructions lots simpler and to the point where it can officially be generated within hardware.

In order to maneuver this effectively in high position of this method we need to choose high sampling rate for audio for example if we are to use the standard 44.1kHz sampling rate and microphones are about 20cm apart, this means the maximum delay between them will only about 21 samples which means we can only identify 21 different regions of space. By increasing the sampling rate we reduce the size these regions and hence we have more and our position improves.

This algorithm relies on finding a maximum point for that we can use any maximum algorithm finding we like, but by making a small adjustment to the equation we can make that lot easier.

E. *Proposed Cross Correlative Derivative Algorithm*

The cross correlative derivative algorithm is slight change around past toward noticing that maximum in y will be zero point in its first derivative with negative esteem and, secondly, we might only find an estimation for the first derivative and figure the place that dives in from positive value to negative value. Accordingly, we can still use single bit quantization and actually with that the summation has becomes effectively simpler up down counter. Consequently, we don't have exorbitant multiplications though we bring really low circuit activity that obviously consumes less power. Hence, it’s a useful thing to utilize for a mobile application.

Solving $\frac{dy}{di} = 0\ with\ \frac{d^2y}{di^2} < 0.$

$$\Delta y(i) = y(i) - y(i-1) = \sum_{k=0}^{K} X_1(k)[X_2(k-1) - X_2(k-i+1)]$$

F. *Throughput Analysis*

The simulation scenario for correlation algorithms is as follows:
We commence by generating some tones which came from our microphones, conjecturing, we won’t have any noise. Subsequently, we simulate it coming from some other position. Fig. 3 is the top down view of apparatus, with the thing in middle being where our microphones are, and the blue thing where we have our source. Whilst putting out some distance i.e. an angle of about 60 degrees and applying simulation, we can quite clearly see that the right audio is somewhat ahead

of the left audio as in fig. 4 because of where the source now is in the relationship of microphones.

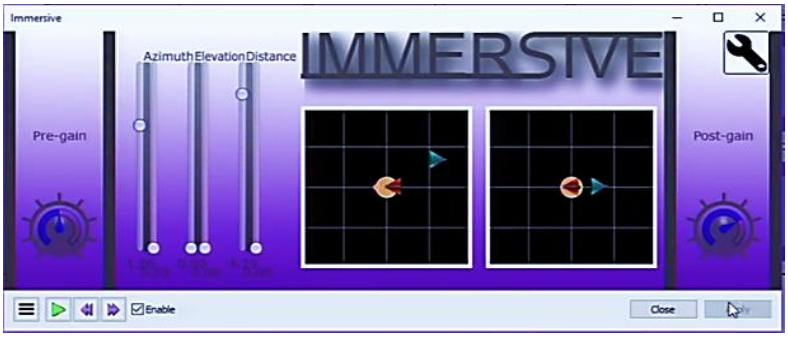

Fig. 3 The top down view of apparatus with the microphone and the sound source.

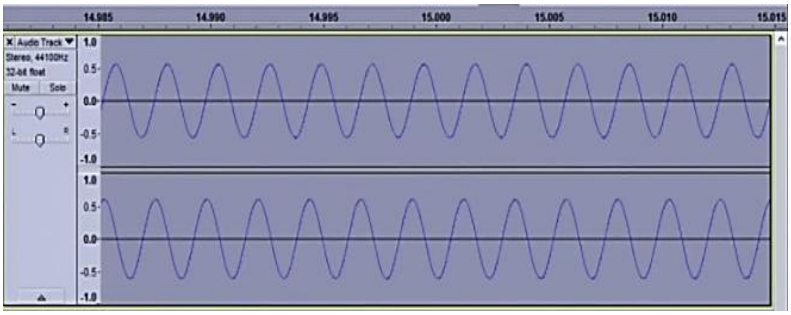

Fig. 4 Right audio signal ahead of the left audio signal.

Consequently, we got the reasonable approximation to this.

G. *Proposed Spatial Gradients Approach*

The alternative to the correlation algorithms, sticking with the time based ones, is generally believed a spatial gradient approach.

For this motivation we utilized four microphones orchestrated toward the corners of a square. We have been looking at time delay just between two but for this method we have to generalize a time delay of 'τ_r' which we define now to be the time delay between the center of the apparatus and some relative point r. By expanding the signal term in a tailor expansion and taking this to first order in our delay for each of our signals we get the eq. 4 and eq. 5. We note that because of the shape of our apparatus we have the same delay between the microphone 1 at the center and the central microphone 2 and similarly between the microphones 3 and 4 in figure 5 ,so we can just re-label these as constant parameters 'τ_1' 'τ_2' as in eq. 4 and eq. 5.

$$\text{s(t+}\tau(\boldsymbol{r})) = s(t) + \tau(\boldsymbol{r})\dot{s}(t) + \frac{1}{2}\tau(\boldsymbol{r})^2\ddot{s}(t) + O(\tau(\boldsymbol{r})^3)$$

$$X_1 \approx s + \tau_2\dot{s}$$

$$X_2 \approx s - \tau_2\dot{s}$$

$$X_3 \approx s - \tau_1\dot{s}$$

$$X_4 \approx s + \tau_1\dot{s}$$

$$\tau_1 = \frac{1}{2}\frac{d}{c}\cos a = \frac{1}{2\dot{s}}(X_4 - X_3) \cdots eq.4$$

$$\tau_2 = \frac{1}{2}\frac{d}{c}\sin a = \frac{1}{2\dot{s}}(X_1 - X_2) \cdots eq.5$$

Using $\dot{s}(k) \approx \frac{s(k)-s(k-1)}{T_s}$

$$\boldsymbol{a} = \begin{bmatrix} \dot{s}(k) \\ \vdots \\ \dot{s}(k+N) \end{bmatrix}$$

$$\boldsymbol{y_1} = \begin{bmatrix} X_4(k) - X_3(k) \\ \vdots \\ X_4(k+N) - X_3(k+N) \end{bmatrix}$$

$$\boldsymbol{y_2} = \begin{bmatrix} X_1(k) - X_2(k) \\ \vdots \\ X_1(k+N) - X_2(k+N) \end{bmatrix}$$

$$\tau_1 = \frac{\boldsymbol{a} \cdot \boldsymbol{y_1}}{\boldsymbol{a} \cdot \boldsymbol{a}}$$

$$\tau_2 = \frac{\boldsymbol{a} \cdot \boldsymbol{y_2}}{\boldsymbol{a} \cdot \boldsymbol{a}}$$

Eventually, by taking combinations of these equations we can isolate these different delay values, put them in terms of our original signals and the derivative of S, which we calculate using the approximation. This approximation does amplify one of the effects of high frequency noise but we assume that the noise is negligible.

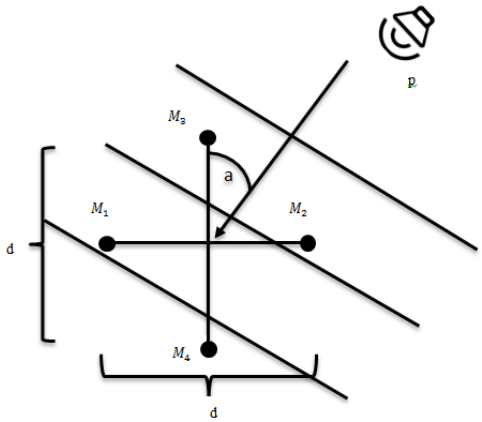

Fig. 5 SGA framework

H. *Head Related Transfer Function*

These have been very good methods coming from understanding of the physics of the situation. However, the nature is being working on this problem a lot longer then we have. Hence, many researchers thought to work on what nature has done

and observe what happens with actual human. Accordingly, when turns out to human, we have very well designed ears and heads such that they actually apply physical filtering to the sound before it reaches in an ear and filtering is highly dependent on the angle on which the sound approaches the body. Consequently, this effect can be modeled by the fine art response filter H, which is the Head Related Transfer Function (HRTF). The signal is received by the microphone 0 is just the original signal sent by the source convolved with this transfer function. The model is used in industry for identifying the head related functions.

Fig. 6 Head Related Transfer Function model

I. *Inverse and Cross Channel Algorithm*

Considering that we have 2 strings of input data with different transfer functions applied, we need to find some way of making sense for both of them. Hence, we apply Inverse filter and try to get back the original signal that will only work if our inverse is for the same angles of the transfer function; so that we can use the method of testing our estimates of the angle and indeed this does work. The eq. 6 is what we actually maneuvered to resolve it. However, the impact is that this is not very good because the inverse is very difficult to calculate. Though the transfer functions are very complex functions, they involve both time queues and frequency based facts so it is very difficult to get an exact inverse for them. However, to steal or reuse the functions makes it three times of the size of the original filter, so that they make our convolution operation a lot more expensive. Hence ideally we won't like to use inverse algorithms. On the other hand, Cross-Channel algorithms don't use inverses. As we can actually measure these transfer functions, we just apply the one for the other ariel of the microphone and, by the summitry of convolution, we have both the signals that should now be identical if we applied it for right angle. We can now optimize this for our estimate angles and hopefully use that to find the correct angle of the approached sound

$$arg\min_{\hat{\theta},\hat{\emptyset}} \sum \left(R_{left} * \left[H_{left}^{(\hat{\theta},\hat{\emptyset})} \right]^{-1} - R_{right} * \left[H_{right}^{(\hat{\theta},\hat{\emptyset})} \right]^{-1} \right)^2 \ldots eq.6$$

J. *Complexity*

Considering the complexity to implement the correlation based algorithms, as demonstrated above, it would be precise simple to implement and can even be done

built in to the hardware. The spatial gradients approach takes slightly more complex equations yet the principle purpose is that they bring no optimizations that there may be no seeking for the turning points, thereabouts, it will be constantly a fixed calculation; thus it ought to kick a preferred estimation from estimated time which is dependably a decent property for our algorithm on need. The transfer function based techniques will actually take similar amount of time for each estimate angle to test as the correlation algorithms but because the functions are very complex, their shapes may not have a distinct maximum or minimum in them, and with optimization of two variables, the optimization steps take a lot more iterations than it actually takes to run the algorithm.

Table 1. Comparing CCA, SGA and HRTF-base methods

Method	Progress
HRTF-based	Similar work for each test angle + Complex functions make optimization harder
SGA	More complex but no turning point searching
CCA & CDA	Can reasonably be done in hardware

K. *Overcoming Spatial Limitations*

Not all of these can actually give us the value for the whole field of solutions that the source get lined, for example using 2 microphones for the correlation algorithms; we might just determine only those things in the plane in front of us. The transfer function based techniques will actually take similar amount of time for each estimate angle to test as the correlation algorithms but because the functions are very complex, their shapes may not have a distinct maximum or minimum in them, and with optimization of two variables, the optimization steps take a lot more iterations than it actually takes to run the algorithm.

Similarly, spatial gradients approach works well in the plane but it gives no sense of elevation; similarly, the solution is to add more microphones, make them outside the plane, and repeat them with the each plane you have to try and get the best fit of them. This is the area where the transfer function comes to their own because using just 2 microphones you can have an understanding within the whole 3d environment.

However it is not necessarily full 3d in a way as we found out that the effects are different in effect between front and back.

Table 2. Limitations & comparison of CCA, SGA, and HRTF-based methods.

Method	Comparison
HRTF-based	Full 2D range but still has front/back reversal
CCA & CDA	Cone of confusion
SGA	Works in both plane & 3D/Spherical environment

Results & Discussion

Speaking of the accuracy, the correlation algorithms were tested with sources in the plane in front of them and were able to determine the position of a source within less than one degree of error, which makes it quite impressive. The spatial gradients approach was also tested in the same experiment and had the same accuracy but the spread of the estimate for a given source position was much more reliable, accordingly, it actually works better. The transfer function based ones were also tested in a separate experiment and they were found to have specifically the cross channel algorithm with an error of about 2 degrees, which is obviously less effective than one degree but is still far more accurate than the typical human hearing which is found to be at best around 5 degrees of error. However, these aren't technically better ways of doing it than having a genuine person on the scene because none of these can really handle noise as all of our assumptions here are the noise is negligible and noise is uncorrelated to one another into the signal. Noise can be the internal noise within the microphones, can be other sound sources, and it can even be the echo from the original sound source, so literally any surface would give you an echo whether it is wall, floor, sealing, other objects around you, they will cause delayed versions of the original signal which will screw up correlation calculations and generally introduce error in results. Consequently, these are by no means ideal solutions but they do occasionally worth.

Future prospective

In this work, we have proposed a method to localize the sound with the source unknown proposing correlation algorithms and spatial gradients approach and comparing it with other present techniques and we successfully achieved the targeted results as shown in Section 3. Furthermore, we believe that this worth of effort will lead our research clinched alongside two ways:

- To develop/propose a model to localize heart sound from lung sound and other respiratory sounds.
- To work for switching the human robotics to make decisions on sound signals rather than of images for their movement.

Conclusion

In this paper, we proposed a sound localization method improving correlation algorithm, its derivatives, spatial gradients approach, and inverse algorithm over cross channel algorithms; comparing them with biological inspired methods. As result, previously, the transfer function based ones were found to have cross channel algorithms with an error of about 2 degrees having an edge over ordinary human hearing with a chance of around 5 degrees of error in the best scenario; whereas our proposed methodology had accuracy of more accurate results with even less than 1 degree of error.

References

A. Deleforge, F. Forbes, R. Horaud, Acoustic space learning for sound-source separation and localization on binaural manifolds, Int. J. Neural Syst. 25 (1) (2015)1–19.

A. Deleforge, R. Horaud, Learning the direction of a sound source using head motions and spectral features, Tech. rep. Institut National Polytechnique de Grenoble,2011.

A. Portello, P. Danès, S. Argentieri, Acoustic models and kalman filtering strategies for active binaural sound localization, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2011, pp. 137–142.

A. Saeidi, F. Almasganj, Heart sound localization via combinational subspace methods for long-term heart monitoring, in: *Biomedical Signal Processing and Control*, 2017, vol. 31, pp. 434-443.

B. Garcia, M. Bernard, S. Argentieri, B. Gas, S. Argentieri, A. Portello, M. Bernard, P. Danès, B. Gas, M. Bernard, et al., Sensorimotor learning of sound localization for an autonomous robot, in: Proceedings of EAA Congress on Acoustics, Forum Acusticum, Springer, 2012, pp. 188–198.

C. Rascon, I. Meza, Localisation of sound sources in robotics: a review, in: *Robotics & Autonomous Systems*, 2017, vol. 96, no. C, pp. 184-210.

H.G. Okuno, K. Nakadai, Robot audition: Its rise and perspectives, in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, 2015, pp. 5610–5614.

H. Liu, M. Shen, Continuous sound source localization based on microphone array for mobile robots, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2010, pp. 4332–4339.

J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, N. Sugie, “A model based sound localization system and its application to robot navigation,” Robot. Auton. Syst, vol. 27, no. 4, pp. 99–209, 1999.

K.-C. Kwak, An enhanced multimodal sound localization with humanlike auditory system for intelligent service robots, Int. J. Latest Res. Sci. Technol. 2(6)(2013)26–31.

K. Nakadai, K. Nakamura, Sound source localization and separation, Wiley Encyclopedia of Electrical and Electronics Engineering.

K. Youssef, S. Argentieri, J.-L. Zarader, Multimodal sound localization for humanoid robots based on visio-auditive learning, in: Proceeding of IEEE International Conference on Robotics and Biomimetics, ROBIO, 2011, pp. 2517–2522.

L. Chen, Y. Liu, F. Kong, N. He, "Acoustic Source Localization Based on Generalized Cross-correlation Time-delay Estimation," *Procedia Engineering,* vol. 15, no. 1, pp. 4912-4919, 2011.

L. Sinapayen, K. Nakamura, K. Nakadai, H. Takahashi, and T. Kinoshita, “Swarm of micro-quadrocopters for consensus-based sound source localization.” *Advanced Robotics*, 2017, 1-10.

L. Xiaofei, L. Hong, A survey of sound source localization for robot audition, CAAI Trans. Intell. Syst. 7(1) (2012) 9–20.

M. Rhudy, B. Bucci, J. Vipperman, J. Allanach, B. Abraham, "Microphone Array Analysis Methods Using Cross-Correlations," *ASME International Mechanical Engineering Congress and Exposition,* vol. , no. ,pp. 281-288, 2009.

M. Z. S. Ahmed, R. Lobo, C. R. Somaiah, Sound localization used in robotics, in: Proceedings of IRF International Conference, 2015, pp. 18–24.

P. Zahorik, Direct-to-reverberant energy ratio sensitivity, J. Acoust. Soc. Am. 112(5)(2002)2110–2117.

R. Suzuki, T. Takahashi, H.G. Okuno, Development of a robotic pet using sound source localization with the hark robot audition system, J. Robot. Mechatronics 29(1) (2017) 146–153.

S. A. H. Sabzevari, M. Moavenian, Sound localization in an anisotropic plate using electret microphones, in: *Ultrasonics*, 2016, pp. 73–114.

S. Argentieri, A. Portello, M. Bernard, P. Danès, B. Gas, Binaural systems in robotics, in:J. Blauert (Ed.), The Technology of Binaural Listening, Springer BerlinHeidelberg,Berlin,Heidelberg,2013,pp.225–253.

S. Argentieri, P. Danès, P. Souères, A survey on sound source localization in robotics: From binaural to array processing methods, Comput. SpeechLang. 34(1)(2015)87–112.

S. Lana, K.N.K.N.H. Takahashi, T. Kinoshita, Consensus-based sound source localization using a swarm of micro-quadrocopters, in: Proceedings of the Conference of the Robotics Society of Japan, 2015, pp. 1–4.

T. Otsuka, K. Nakadai, T. Ogata, H.G. Okuno, Bayesian extension of music for sound source localization and tracking, in: Proceedings of Annual Conference of the International Speech Communication Association, INTERSPEECH, 2011, pp. 3109–3112.

X. Wan, Z. Wu, Sound source localization based on discrimination of cross-correlation function, in: *Applied Acoustic*, 2013, vol. 74, no. 1, pp. 28-37.

Y. Sasaki, M. Kabasawa, S. Thompson, S. Kagami, K. Oro, Spherical microphone array for spatial sound localization for a mobile robot, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2012,pp.713–718.