Sound Source Localization for Robot Auditory Systems

Youngkyu Cho, Dongsuk Yook, *Member*, IEEE, Sukmoon Chang, *Member*, IEEE, and Hyunsoo Kim

**Abstract — ***Sound source localization (SSL) is a major function of robot auditory systems for intelligent home robots. The steered response power-phase transform (SRP-PHAT) is a widely used method for robust SSL. However, it is too slow to run in real time, since SRP-PHAT searches a large number of candidate sound source locations. This paper proposes a search space clustering method designed to speed up the SRP-PHAT based sound source localization algorithm for intelligent home robots equipped with small scale microphone arrays. The proposed method reduces the number of candidate sound source locations by 30.6% and achieves 46.7% error reduction compared to conventional methods.*[1]

Index Terms — Sound source localization, steered response power (SRP), search space clustering, small scale microphone array, robot auditory system, intelligent home robot.

I. Introduction

Following on from recent advances in humanoid robot technology, intelligent service robots are expected to work in the living environment in the near future. They will support human activities, such as housekeeping and assistance for elderly people. While much of the previous effort in the development of robot technology focused on robot locomotion and vision systems, establishing an effective communication method between humans and robots is an imperative. Speech recognition is one of the most promising communication tools for human-robot interaction, for both expert and non-expert users, since it offers bidirectional interaction and diverse levels of control. Thus, the development of the robot auditory system plays a potentially important role in developing intelligent home robots working seamlessly with human users [1].

A core component of the robot auditory system for human-robot interaction in home robot environments is sound source localization (SSL) [2]. When a user interacts with a humanoid robot using spoken language, the robot must be able to automatically find the location of the user, i.e., the location of the voice source. Fig. 1 describes a sound source localization scenario for the small service robot developed at Samsung Electronics. For example, if the user says "*Come here*!" to the robot from a distance, the robot must be able to identify the user's location, to respond appropriately. Moreover, the accurate estimation of the sound source location enhances speech quality by beamforming multichannel sound signals. This is useful for robots to recognize distant speech. In addition, sound source localization is one way to find the location of the speaker even in the dark.

Many sound source localization methods have been proposed. For example, methods based on the time difference of arrival (TDOA) use generalized cross correlation (GCC) [3] to estimate the TDOAs and relate them to the location of the sound source. Methods based on high resolution spectral analysis [4] use spatial spectra derived from the signals to locate sound sources. Steered response power (SRP) methods [5] electronically steer the microphone array to locate the sound source with the highest power. The steered response power with the phase transform filter (SRP-PHAT) is a robust method for sound source localization when room reverberation is present [6]. However, SRP-PHAT methods usually employ a grid search scheme that examines a large number of the candidate sound source locations. Therefore, SRP-PHAT using the computationally intensive grid search method cannot be used in real-time systems, such as those of small size service robots with limited computational power.

Several search methods have been proposed for real-time SRP-PHAT [7]-[9]. A hierarchical search method was proposed in [7] that gradually prunes the candidate sound source locations in a coarse-to-fine search. A drawback of this method is that it may prematurely prune the sound source with the highest power, before it reaches the final decision. A hybrid method was proposed to speed up SRP-PHAT in [8]. This method first generates a small set of candidate sound source locations using a TDOA-based search. It then performs a SRP-PHAT-based grid search on this small set of candidate locations. If TDOA estimation is unsuccessful in the first step, then SRP-PHAT will fail in the final decision. This method

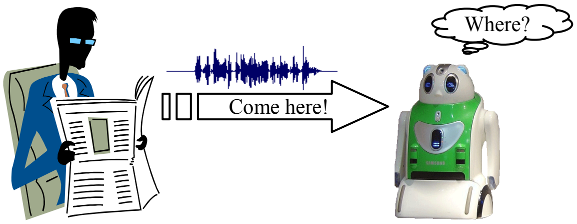


Fig. 1. Sound source localization of a small size home robot to find the user's location.

[1] This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2006-311-D00822). It was also supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute for Information Technology Advancement) (IITA-2008-C1090-0803-0006).

Youngkyu Cho and Dongsuk Yook (corresponding author) are with the Speech Information Processing Laboratory, Department of Computer and Communication Engineering, Korea University, Seoul, 136-701, Republic of Korea (e-mail: ccameo@voice.korea.ac.kr and yook@voice.korea.ac.kr). They thank Samsung Electronics for their support. Sukmoon Chang is with Pennsylvania State University, Middletown, PA 17057, USA (e-mail: sukmoon@psu.edu). Hyunsoo Kim is with Samsung Electronics, Korea (e-mail: hyunsoo72.kim@samsung.com).

Contributed Paper
Manuscript received July 15, 2009

0098 3063/09/$20.00 © 2009 IEEE

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on February 25,2010 at 11:35:57 EST from IEEE Xplore. Restrictions apply.

decreases sound source localization performance in low signal-to-noise (SNR) environments due to its dependency on TDOA estimation to generate the small set of candidate locations. In [9], the cross-correlation functions are used to find the time delays that may correspond to the sound source location. An inverse mapping function relates a relative time delay to a set of candidate locations. Only the output powers of the locations that are inversely mapped by the time delays are considered to find the maximum power location. However, this method may fail to find the maximum power location, because it searches only a few locations that are inversely mapped by the relative time delay. In summary, none of these methods are admissible; their search results may differ from those of a full grid search, especially in noisy reverberant environments.

We are interested in an efficient search method that always locates the sound source with the highest power in real-time using SRP-PHAT for small size robots with small scale microphone arrays. The resolution of the sound source localization result obtained by the SRP-PHAT method depends on the sampling rate and the microphone array geometry. If we can divide the whole search space in advance into sub-regions, each containing only the points with the same TDOAs of sound signals, and examine only the representative point from each sub-region to locate the sound source with the highest power, we can dramatically reduce the computational cost of the conventional grid search method. In this paper, we propose the search space clustering (SSC) method that efficiently divides the search space into sub-regions, each of which contains only the points with the same TDOAs. The proposed method constructs a look-up table that contains the complete set of the unique regions having different TDOA values for the sound signals. By searching only one representative point per region, the computational cost can be greatly reduced without diminishing the accuracy of sound source localization compared to a full grid search. We demonstrate that the proposed method can be effectively applied to the auditory system of a small size service robot with a four-sensor microphone array.

This remainder of this paper is organized as follows. In sections II and III, we review SRP-PHAT theory and propose the SSC method for real-time SRP-PHAT using small scale microphone arrays. Sections IV and V evaluate the performance of SRP-PHAT using the proposed SSC method in simulation and real room environments, respectively. Finally, we draw conclusions in section VI.

II. REVIEW OF SRP-PHAT

Consider an array of N microphones. Given the source signal, $x_n(t)$, received by the n-th microphone at time t, the output, $y(t,q)$, of the delay-and-sum beamformer is defined as follows:

$$y(t,q) = \sum_{n=1}^{N} x_n(t + \tau_{n,q}) \tag{1}$$

where, $\tau_{n,q}$ is the direct time of travel from location q to the n-th microphone. Filter-and-sum beamformers using a weighting function may be used to deal with complex noises, such as reverberation. In the frequency domain, the filter-and-sum version of (1) can be written as:

$$Y(\omega,q) = \sum_{n=1}^{N} G_n(\omega) X_n(\omega) e^{j\omega\tau_{n,q}} \tag{2}$$

where, $X_n(\omega)$ and $G_n(\omega)$ are the Fourier transforms of the n-th microphone signal and its associated filter, respectively. In (2), the microphone signals are phase-aligned by the steering delays and summed after the filter is applied.

The sound source localization algorithm based on the steered response power steers the microphone array to focus on each spatial point, q, and calculates the output power, $P(q)$, of the microphone array for the focused point q as follows:

$$\begin{aligned}
P(q) &= \int_{-\infty}^{\infty} |\, Y(\omega)\,|^2 \, d\omega \\
&= \int_{-\infty}^{\infty} \left(\sum_{l=1}^{N} G_l(\omega) X_l(\omega) e^{j\omega\tau_{l,q}} \right) \left(\sum_{k=1}^{N} G_k^*(\omega) X_k^*(\omega) e^{-j\omega\tau_{k,q}} \right) d\omega \\
&= \sum_{l=1}^{N} \sum_{k=1}^{N} \int_{-\infty}^{\infty} \left(G_l(\omega) X_l(\omega) e^{j\omega\tau_{l,q}} \right) \left(G_k^*(\omega) X_k^*(\omega) e^{-j\omega\tau_{k,q}} \right) d\omega \\
&= \sum_{l=1}^{N} \sum_{k=1}^{N} \int_{-\infty}^{\infty} \Psi_{lk}(\omega) X_l(\omega) X_k^*(\omega) e^{j\omega(\tau_{l,q} - \tau_{k,q})} d\omega
\end{aligned} \tag{3}$$

where, $\Psi_{lk}(\omega) = G_l(\omega) G_k^*(\omega)$. In SRP-PHAT, the filter to reduce the effect of reverberation is defined as:

$$\Psi_{lk}(\omega) = \frac{1}{\left| X_l(\omega) X_k^*(\omega) \right|} \quad . \tag{4}$$

After calculating the steered response power, $P(q)$, for each candidate location, the point $\hat{q}$ that has the highest output power is selected as the location of the sound source, i.e.,

$$\hat{q} = \arg\max_{q} P(q) \quad . \tag{5}$$

The SRP-PHAT method has a very high computational cost, since it uses a grid search method to find the maximum power location, $\hat{q}$, in (5) by computing the output power at every point in the grid defined in three dimensional space. In the next section, we propose a search space clustering method that effectively reduces the search space of SRP-PHAT without decreasing its accuracy.

III. SEARCH SPACE CLUSTERING

Although SRP-PHAT calculates the output power at every point in the grid to find the sound source location with the highest output power, the output powers of the candidate locations are not all unique due to the finite sampling rate of

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on February 25,2010 at 11:35:57 EST from IEEE Xplore. Restrictions apply.

TABLE I
SEARCH SPACE CLUSTERING ALGORITHM

```
Given the initial block $b$ representing the entire search space:
$B \leftarrow \{b\}$
$C \leftarrow \phi$
while $B$ is not empty
    Calculate the TDOAs of each microphone pair at every
    vertex of $b$.
    if the TDOAs at all vertices of $b$ are the same
        $B \leftarrow B - \{b\}$
        $C \leftarrow C \cup \{b\}$
    else
        The block $b$ is divided into a set of smaller size blocks,
        $b_1, b_2, ..., b_8$, as shown in Fig. 2.
end
Any two blocks with the same TDOAs in $C$ are merged.
The centroids of the blocks in $C$ are stored in a look-up table.
```

the analog to digital conversion. As shown in (3), the output power of a candidate location depends on the microphone signals and the phase differences caused by the steering delay required to focus the microphone array on the location. The phase difference is also in the discrete time domain, since the signals are converted into the digital domain. If any two candidate points are closely located, such that the phase differences, i.e., TDOAs, in the discrete time domain are the same, the output powers for the two locations are indistinguishable by (3). Therefore, if the candidate locations that have the same TDOAs are clustered together and the representative coordinate of each clustered group is stored in a look-up table in advance, then the computational cost can be greatly reduced, because only the pre-computed coordinates of the points in the look-up table will be searched using (3) in (5) without going through every point in the grid.

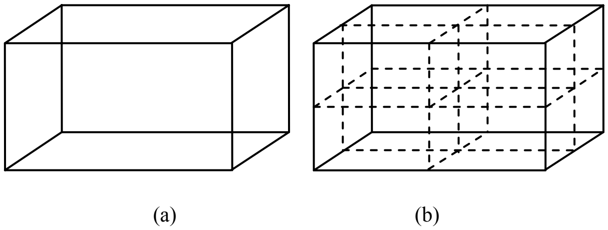
The look-up table is constructed in advance using a top-down clustering algorithm. First, the sample difference, $s_{k,l,q}$ that corresponds to TDOA for each microphone pair (microphones k and l) at the location q is obtained as:

$$s_{k,l,q} = \mathrm{round}\left(r \times (\tau_{l,q} - \tau_{k,q})\right) \tag{6}$$

where, r is the sampling rate. Initially, the entire search space is clustered as a single block. If the block is not sufficiently small and contains at least two points with different TDOAs, it is divided into eight smaller-sized blocks. This division process is repeated until every block in the search space becomes sufficiently small to contain only the points with the same TDOA. For instance, consider a microphone array system of four sensors. We can compute six TDOAs for a location q and collect them into a TDOA vector, since there are six microphone pairs in the system. That is,

$$s_q = [s_{1,2,q}, s_{1,3,q}, s_{1,4,q}, s_{2,3,q}, s_{2,4,q}, s_{3,4,q}]^T . \tag{7}$$

Given a block, as in Fig. 2 (a), we calculate eight TDOA vectors, i.e., one at each of the eight vertices of the block. If these eight TDOA vectors are not identical, the block is divided into eight smaller blocks, as shown in Fig. 2 (b). After the division process is completed, any two blocks with identical TDOA vectors are merged. Finally, the centroid of each block is stored in a look-up table. The proposed search space clustering method is very effective for small scale microphone array systems, where the number of unique points (i.e., centroids) with different TDOAs is much less than the total number of candidate locations in the grid. The search space clustering algorithm is summarized in Table I.



(a) (b)

Fig. 2. If the TDOAs at the eight vertices of a given block in (a) are not the same, the block is divided into eight smaller blocks, as in (b).

IV. EVALUATION IN SIMULATION ENVIRONMENTS

We evaluated SRP-PHAT with the proposed SSC method using the sound data simulated by the image method described in [10] to analyze performance under various SNR and reverberation conditions. Sound data were generated in a 5 x 6 x 2 m^3 room. We used a square microphone array of 0.17 x 0.17 m^2 with four sensors. The center of the microphone array was located at (2.5, 2, 0.3) meters in the room. The sound sources were placed in various positions, as shown in Fig. 3. The heights of the sound sources were 1.2 and 1.6 meters. We evaluated the performance of SRP-PHAT with SSC under three different SNR conditions (10, 20, and 30 dB) and three different reverberation times (20, 200, and 300 milliseconds). The source signal was contaminated with white noise to create different SNR signals. The source sound duration was 30 seconds. The frame length was 128 milliseconds and the sampling rate was 16 kHz.

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on February 25,2010 at 11:35:57 EST from IEEE Xplore. Restrictions apply.

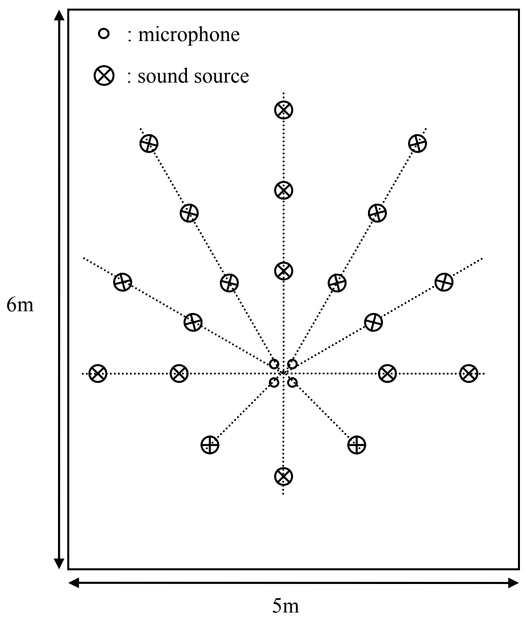


Fig. 3. Locations of the microphones and the sound sources for the simulated data experiments.

We compared the performance of SRP-PHAT with SSC to the conventional SRP-PHAT with the grid search scheme to analyze accuracy. We used various grid sizes ranging from 1 to 20 centimeters for conventional SRP-PHAT. Conventional SRP-PHAT had to search 60,000,000 candidate locations, for a grid size of 1 centimeter. Conversely, there were only 5,203 representative locations in the look-up table generated by SSC. That is, a reduction in the search space by a factor of 11,531.

Fig. 4 and Fig. 5 show the sound source localization performance for varying SNR levels and reverberation times. Accuracy was measured as the percentage of correctly estimated directions of arrival (DOA) in the azimuth. The estimated DOA that lies within ±10°of the true azimuth was counted as correct. Fig. 4 compares the sound source localization accuracy of SRP-PHAT with SSC to conventional SRP-PHAT, for various grid sizes (1, 10, 15, and 20 centimeters), under three different SNR conditions (10, 20, and 30 dB). The reverberation time in Fig. 4 was 20 milliseconds. Similarly, Fig. 5 compares sound source localization accuracy for three different reverberation times (20, 200, and 300 milliseconds). The average SNR in Fig. 5 was 30 dB. SRP-PHAT with the SSC outperformed conventional SRP-PHAT with the grid sizes of 10, 15, and 20 centimeters, regardless of the SNR levels in Fig. 4 and the reverberation times in Fig. 5. Fig. 4 and 5 illustrate that the SRP-PHAT with SSC performed with the same accuracy as conventional SRP-PHAT with the finest grid size of 1 centimeter. This demonstrates that SRP-PHAT with SSC always locates the sound source with the highest output power that conventional SRP-PHAT with the finest grid size can find.

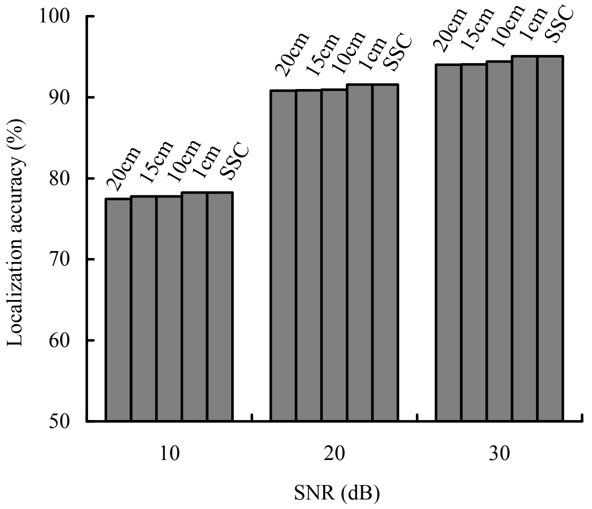


Fig. 4. Sound source localization accuracies of SRP-PHAT with the proposed SSC method, compared to conventional SRP-PHAT, with varying grid sizes (1, 10, 15, and 20 centimeters), and three different SNR conditions (10, 20, and 30 dB).

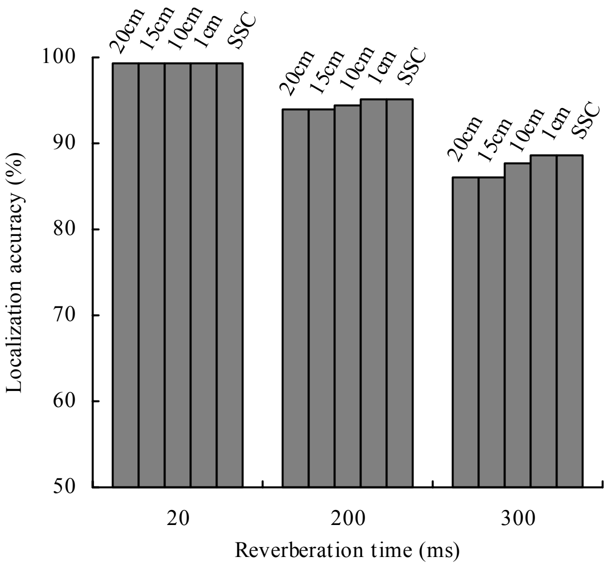


Fig. 5. Sound source localization accuracies of SRP-PHAT with the proposed SSC method, compared to conventional SRP-PHAT, with various grid sizes (1, 10, 15, and 20 centimeters), under three different reverberation times (20, 200, and 300 milliseconds).

V. Evaluation In Real Environments

We recorded sound data in a classroom to evaluate the performance of SRP-PHAT with the SSC in real environments. A square microphone array of 0.17 x 0.17 m^2 with four sensors was attached on the shoulder of a plaster cast of the small size home robot shown in Fig. 6. The location of the microphones and the sound sources, the room size, and the source speech for the recording in the real environment, were the same as the conditions used for the simulated data generated in the previous section. The room reverberation time

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on February 25,2010 at 11:35:57 EST from IEEE Xplore. Restrictions apply.

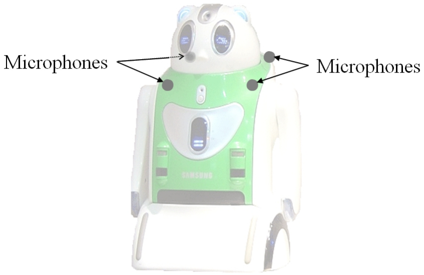


Fig. 6. Location of microphones for a small size home robot.

was about 530 milliseconds, and the average SNR of the recorded data was about 23 dB. Only those frames with SNR greater than 20 dB were processed for sound source localization.

Table II compares SRP-PHAT with SSC to conventional SRP-PHAT for various grid sizes (1, 10, 15, and 20 centimeters), in terms of their location accuracy, the number of candidate locations, and processing time (real-time factor). Accuracy was measured as the percentages of correctly estimated DOAs in azimuth as well as in elevation. An estimated DOA that lies within $\pm 10^\circ$ of the true azimuth and elevation was considered correct. As shown in Table II, when compared to conventional SRP-PHAT with a grid size of 20 centimeters, SRP-PHAT with SSC reduced the number of candidate locations by 30.6%, whilst achieving error reductions of 41.9% in azimuth (from 10.5% to 6.1%) and 50.7% in elevation (from 14.6% to 7.2%). SRP-PHAT with SSC dramatically reduced the number of the candidate locations by a factor of 11,531 compared to the conventional SRP-PHAT method with its finest grid size of 1 centimeter, without any loss of localization accuracy. SRP-PHAT with the proposed SSC method can run in 0.56 times of real-time.

Fig. 7 shows the performance of SRP-PHAT with SSC relative to the performance of conventional SRP-PHAT. The figure shows the proposed method yields performance improvements of 4.7% for azimuth accuracy and 8.1% for elevation accuracy relative to conventional SRP-PHAT for a grid size of 20 centimeters, which corresponds to an error reduction of 41.9% in azimuth and 50.7% in elevation, as discussed previously.

TABLE II
LOCALIZATION ACCURACY, NUMBER OF SEARCH POINTS AND REAL-TIME FACTOR

	Grid Search				Search Space Clustering
	20cm	15cm	10cm	1cm	
Azimuth (%)	89.5	89.9	93.6	93.9	93.9
Elevation (%)	85.4	88.8	89.2	92.8	92.8
Number of Search Points	7,500	19,000	60,000	$6\text{x}10^7$	5,203
Real-Time Factor	0.63	1.0	2.8	220	0.56

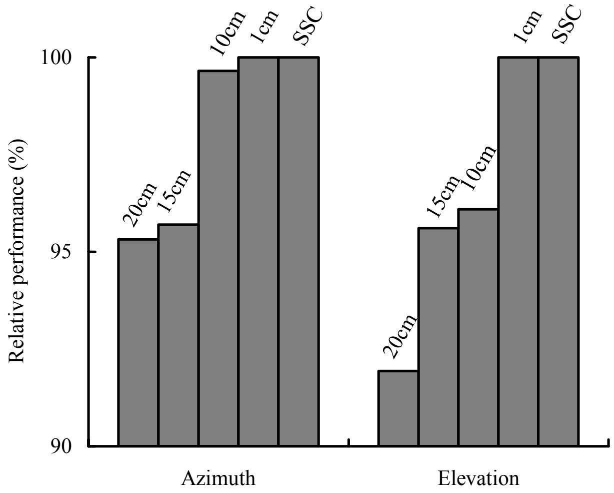


Fig. 7. Sound source localization performance of the SRP-PHAT with the proposed SSC method compared to the conventional SRP-PHAT with various grid sizes (1, 10, 15, and 20 centimeters) in real environment.

VI. CONCLUSION

In this paper, we proposed a novel search space clustering method for SRP-PHAT that significantly reduces the computation time of conventional SRP-PHAT without performance degradation. The proposed method achieves this performance gain by the dramatic reduction in the search space, using the TDOA estimations in a way that guarantees the reduced search space always includes the location with the highest power. As a result, SRP-PHAT with SSC will always find the highest output power locations, unlike conventional methods in [7]-[9]. Therefore, SRP-PHAT with the proposed SSC method is admissible for sound source localization.

SRP-PHAT with the proposed SSC method has many advantages over conventional SRP-PHAT using the grid search method. First, SRP-PHAT with SSC always finds the sound source location with the global maximum output power under varying SNR conditions and reverberation times. Second, we observed that the proposed SSC method dramatically reduced the number of the candidate locations to be searched by SRP-PHAT, radically improving computation time. For example, the reduction in the candidate locations, compared to conventional SRP-PHAT, ranged from 30.6% for the grid size of 20 centimeters to over 99.9% for the finest grid size of 1 centimeter. Finally, despite the dramatic search space reduction, SRP-PHAT with SSC, still achieved an error reduction of 41.9% in azimuth and 50.7% in elevation. In the real environment experiments, the robot was able to distinguish 18 directions for incoming sounds with 93.9% of the azimuth accuracy and 92.8% of the elevation accuracy. Following these promising results, the proposed method can be successfully used for the real-time auditory system of small size home robots with four sensor microphone arrays.

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on February 25,2010 at 11:35:57 EST from IEEE Xplore. Restrictions apply.

The proposed method greatly reduces the size of the search space and thus the computational cost of SRP-PHAT for the small scale microphone arrays. Although we observed a significant reduction in the size of the search space when large scale microphone arrays were used, the reduction rate was not as dramatic as in the case of small scale microphone arrays. We are currently developing a search space clustering algorithm that is scalable to the larger systems.

REFERENCES

[1] Y. Oh, J. Yoon, J. Park, M. Kim, and H. Kim, "A name recognition based call-and-come service for home robots," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 247-253, 2008.

[2] K. Kwak and S. Kim, "Sound source localization with the aid of excitation source information in home robot environments," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 852-856, 2008.

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

[4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 276-280, 1986.

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

[6] J. DiBiase, H. Silverman, and M. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward, pp. 157-180, Springer, 2001.

[7] D. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Transactions on Speech and Audio Processing*, vol. 12, pp. 499-508, 2004.

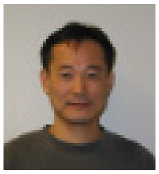
[8] J. Peterson and C. Kyriakakis, "Hybrid algorithm for robust, real-time source localization in reverberant environments," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 1053-1056, 2005.

[9] J. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 2510-2526, 2007.

[10] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943-950, 1979.

Youngkyu Cho was born in Seoul, Korea, in 1975. He received the M.S. degree in computer science and engineering from Korea University, Korea, in 2002. He is pursuing a Ph.D. degree at the Korea University. His current research interests are acoustic modeling, speaker recognition, and sound source localization using a microphone array.

Dongsuk Yook received the B.S. and M.S. degrees in computer science from Korea University, Korea, in 1990 and 1993, respectively, and the Ph.D. degree in computer science from Rutgers University, New Jersey, USA, in 1999. He worked on speech recognition at IBM T.J. Watson Research Center, New York, USA, from 1999 to 2001. Currently, he is a professor of the Department of Computer and Communication Engineering, Korea University, Korea. His research interests include machine learning and speech processing. Dr. Yook is a member of IEEE.



Sukmoon Chang received the M.S. degree in computer science from Indiana University, Indiana, USA, in 1995 and the Ph.D. degree in computer science from Rutgers University, New Jersey, USA, in 2002. He worked on image and signal processing at the Center for Computational Biomedicine Imaging and Modeling, Rutgers University, from 2002 to 2004. He is currently a professor in Computer Science, School of Science, Engineering, and Technology, Pennsylvania State University. His research interests include image and signal processing and machine learning. Dr. Chang is a member of IEEE.

Hyunsoo Kim graduated from the School of Electrical Engineering, Seoul National University. He also completed his postgraduate course and received his Ph.D. degree from the School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, New South Wales, Australia. He has worked for the Motorola Sydney Research Center and the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. Currently, he is working for the Telecommunication Research Center, Samsung Electronics. His research interests include human-computer interface, image/speech processing and multi-modal interface technology.

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on February 25,2010 at 11:35:57 EST from IEEE Xplore. Restrictions apply.