Zero-Crossing Based Time-Frequency Masking for Sound Segregation

Young-Ik Kim, Sung Jun An, and Rhee Man Kil
Division of Applied Mathematics
Korea Advanced Institute of Science and Technology
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
E-mail: \{youngik, cutaincall, rmkil\}@kaist.ac.kr

Hyung-Min Park
Brain Science Research Center
Korea Advanced Institute of Science and Technology
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
E-mail: hmpark@cs.cmu.edu

(Submitted on December 15, 2005; Accepted on March 20, 2006)

Abstract — This paper presents a new method of zero-crossing based binaural mask estimation for sound segregation under the condition that multiple sound sources are present simultaneously. The masking is determined by the estimated sound source directions using the spatial cues such as inter-aural time differences (ITDs) and inter-aural intensity differences (IIDs). In the suggested method, the estimation of ITDs is utilizing the statistical properties of zero-crossings detected from binaural filter-bank outputs. We also consider the estimation of ITDs with the aid of IID samples to cope with the phase ambiguities of ITD samples in high frequencies. As a result, the proposed method is able to provide an accurate estimate of sound source directions and a good masking scheme for sound segregation while offering significantly less computational complexity compared to cross-correlation based methods.

Keywords — Sound source localization, zero-crossings, noise-robustness, sound segregation, time-frequency masking

1. Introduction

In the human auditory system, a sound source is localized by the differences of signals obtained from both ears. The main cues are inter-aural time differences (ITDs) and inter-aural intensity differences (IIDs). It has been known that the human auditory system is able to select a specific sound source among multiple sound sources even in noisy environments using various cues including these localization cues, for instance, the capability of handling the cocktail party problem. This concept of sound source localization can be applied to sound source segregation using two sensors. In this application, we consider the case that multiple sound sources are present in noisy environments. Here, we can consider the methods using independent component analysis (ICA) [1] for sound segregation. However, ICA based methods are not suitable for this problem because we usually don’t know the number of sound sources and the number of sound sources can be larger than the number of sensors (in our problem, two sensors). Furthermore, ICA based methods are not robust to background noises. From this point of view, we consider the sound segregation method based on the sound source localization using the spatial cues such as ITDs and IIDs. In our approach, a target sound source mixed with multiple sound sources is segregated using the masking scheme in which sound segments in the time-frequency domain originated from a target sound source are selected and other sound segments originated from interfering sound sources are blocked using the estimated sound source direction for each segment. This approach is feasible under the assumption that the signals have non-overlapping time-frequency representation supports. This assumption has been shown to be approximately
true for speech signals[2]. For this approach, Roman et al.[3] suggested the masking method in which the masks in time-frequency domain are trained using a priori knowledge of the target and interferences. However, this method requires the training procedure for every spatial configuration, that is, if the number of sound sources and/or the directions of sound sources are changed, these masks should be retrained. In this sense, it is not favorable for implementing this method in real applications. From this context, we consider the method of identifying the directions of sound sources and segregating the selected sound without any requirement of training the masks in time-frequency domain.

The sound source localization plays an important role for segregating the selected sound. For the sound source localization, we need to obtain the spatial cues such as ITDs and IIDs. It is widely known that ITDs are the main cues used at low frequencies less than 1.5 KHz while IIDs are used in the high frequency range[4]. Estimations of sound source directions using ITDs have smaller variations but could be ambiguous at higher frequencies due to smaller zero-crossing intervals which could be less than ITDs while estimations using IIDs have larger variations. As for the effects of noisy environments, ITDs can be easily affected by background noise while IIDs can be easily biased especially in reverberant conditions[5]. In this sense, a reliable and consistent mapping from the spatial cues to the source directions in noisy environments is required. One of the methods to improve the performance of sound source localization is to combine both ITDs and IIDs[3, 6]. Faller and Merimaa[7] also showed that sound source directions can be estimated even in a reverberant hall by selecting reliable ITD and IID cues which coincide with the inter-aural coherence cues. In this paper, we will concentrate on the reliable estimation of ITDs in noisy multi-source environments.

For computing ITDs, Jeffress’s model, cross-correlation (CC) based sound source localization methods[9, 10, 11, 12, 13, 14] were suggested; however, these methods require high computational complexity involved in the computation of cross-correlation, and they suffer from inaccuracies in estimating the ITDs, especially in noisy multi-source environments since some artificial peaks in addition to the major peaks corresponding to sound source directions are usually generated from the computation of cross-correlation. In this context, we propose a method based on zero-crossings for an accurate and efficient estimation of ITDs in noisy multi-source environments.

Zero-crossings have been used to find noise-robust speech features[15, 16, 17]. One example of such use is the zero-crossing peak-amplitudes (ZCPAs) proposed by Kim et al.[17]. They considered a model of the neural transduction of acoustic signals based on two parallel mechanisms of auditory nerve fibers: rate and temporal representations. They demonstrated that the auditory model based on zero-crossing features is more robust in noisy environments than other popularly used feature extraction methods, such as linear predictive coding derived cepstrum (LPCC) or mel-frequency cepstral coefficients (MFCC). This is mainly due to the dominant frequency principle[18, 19] which states that the number of zero-crossings per unit time is close to two times the frequency of the dominant signal when one exists. From this observation, we propose a method of estimating ITDs using the zero-crossing time differences (ZCTDs) detected from the filter-bank outputs of the left and right sensors. This approach is in accordance with Jeffress’s hypothesis in which the time difference is actually measured using delay components and coincidence detectors. In this approach, one of the notable properties in the statistics of ITD estimates is that their variances are closely related to the signal-to-noise ratios (SNRs) of filtered signals, enabling us to identify reliable samples according to the variances of ITD estimates. This direction of ITD estimation has a similar flavor to methods which use the variance of estimates. Bradstein[20] used the variance of phase deviation to obtain the minimum variance unbiased estimator of the delay of incoming signals from sensor arrays. Colburn and Isabelle[21] also used the variance of the maximum likelihood estimate of interaural phase difference to predict the performance in interaural time discrimination and binaural detection. In our work, the variances of ITD estimates are used to calculate the SNRs of filtered signals from two sensors, and used for the reliable estimation of ITDs. As a result, the suggested method is able to provide an accurate estimate of multiple sound source directions and a good masking scheme for sound segregation while offering significantly less computational complexity compared to the conventional cross-correlation based methods.

2. Estimation of ITDs using Zero-Crossings

In the estimation of ITDs, it is important to obtain reliable samples, for instance, samples with high SNR, as much as possible. This direction of finding the reliable ITD samples is possible since zero-crossings corresponding to the dominant source signal are mainly detected according to the dominant frequency principle. First, as done
in the ZCPA coding, a series of bandpass filter is applied to each left and right sensor signals. Here, let us denote \( x_i(t) \) as the output signal of the \( i \)th channel of the filter-bank. The estimation of ITDs is performed separately for each channel. Suppose there are \( N \) (upward) zero-crossings, and zero-crossing times are represented by \( t_n \), \( n = 1, 2, \cdots, N \) satisfying \( x_i(t_n) = 0 \). To distinguish the signals generated from the left and right sensors, we use \( x_i^L(t) \) and \( x_i^R(t) \) as the signal at the \( i \)th channel of the left and right sensors, respectively. We now describe the principle of determining the ITD using zero-crossings. Let us define zero-crossing time in the left channel as \( t_n^L \) for \( n = 1, 2, \cdots, N \), and in the right channel as \( t_n^R \) for \( n = 1, 2, \cdots, M \), where \( N \) and \( M \) represent the number of zero-crossings detected from the left and right channel signals, respectively.

In the auditory processing, the ITDs and IIDs convey same information of source directions. From this observation, we consider the method of selecting valid ITD-IID sample pairs. First, for \( t_n^L \) we consider the following candidates of ITD samples:

\[
\Delta t_i(n, m) = t_n^L - t_m^R
\]

where \( t_n^R \) is selected within a window in which the time interval is given by the range of \( t_n^L - T \) and \( t_n^L + T \) for the time span \( T \). The time span \( T \) is determined as \( 1 \text{ ms} \). Here, the problem is to determine the proper \( t_n^R \) for the given \( t_n^L \). To solve this problem, we consider the following IID samples:

\[
\Delta p_i(n, m) = 10 \log_{10} \frac{p_n^L}{p_m^R}
\]

where \( p_n^L \) and \( p_m^R \) represent the left power at time \( n \) and the right power at time \( m \) respectively. They are defined by

\[
p_n^L = \frac{1}{2W} \sum_{t=t_n^L-W}^{t_n^L+W} (x_i^L(t))^2 \quad \text{and} \quad p_m^R = \frac{1}{2W} \sum_{t=t_m^R-W}^{t_m^R+W} (x_i^R(t))^2
\]

where \( W \) is set to \( 5/\text{center frequency of the } i \text{th channel according to the Ghitza’s perceptual window}[22] \).

Since the ITD and IID samples represent the same information of sound source direction, we consider to make a map of sound source directions represented by the angles measured from the frontal axis versus ITD values and also a map of sound source angles versus IID values. These maps can be made by investigating the ITD or IID values for the corresponding sound source angles. Here, we can identify the corresponding angle for the ITD value \( \theta_{ITD} \) from the map of angles versus ITD values and also the corresponding angle for the IID value \( \theta_{IID} \) from the map of angles versus IID values. Then, we can search the best matching ITD-IID sample pair by searching the minimum angle difference between \( \theta_{ITD} \) and \( \theta_{IID} \), that is, we can find the time index \( k \) for the proper ITD value as

\[
k = \arg \min_m |\theta_{ITD}(\Delta t_i(n, m)) - \theta_{IID}(\Delta p_i(n, m))|\]

where \( \theta_{ITD}(\Delta t_i(n, m)) \) and \( \theta_{IID}(\Delta p_i(n, m)) \) represent the angle values for the measured values of \( \Delta t_i(n, m) \) and \( \Delta p_i(n, m) \) respectively. For the illustration of searching the best matching ITD-IID sample pair, refer to Figure 1. As a result, we get the ITD sample \( \Delta t_i(n) \) at the \( i \)th channel as

\[
\Delta t_i(n) = t_n^L - t_k^R
\]

where \( k \) satisfies (5).

Although we select the best matching ITD-IID sample pair, there are perturbations when we measure the value of (6) due to the environmental noise and/or measurement error. This can be described by the following equations:

\[
t_n^L = t_n^L + r_n^L \quad \text{and} \quad t_k^R = t_k^R + r_k^R
\]

where \( t_n^L \) and \( t_k^R \) represent the zero-crossing points without noise in the left and right sensor signals respectively, and \( r_n^L \) and \( r_k^R \) represent the perturbation of zero-crossing points due to noise in the left and right sensor signals
Figure 1: Selection of a pair of ITD-IID samples: an example of combined azimuth angle estimate with the channel center frequency of 3.4 KHz where three speech sources are located at azimuth angles of -30, 0, 30 degrees.

respectively. Here, we assume that both $r_L^n$ and $r_R^k$ are identically and independently distributed with mean zero. Suppose there is a true time delay $\Delta$ between two sensors, that is,

$$\bar{t}_L^n = \bar{t}_R^k + \Delta.$$  

Then, the mean and the variance of the time difference $\Delta t_i(n)$ are given by

$$E[\Delta t_i(n)] = E[\Delta + r_L^n - r_R^k] = \Delta \quad \text{and}$$

$$Var(\Delta t_i(n)) = E[(\Delta t_i(n) - \Delta)^2] = Var(r_L^n) + Var(r_R^k).$$

since we assume that $r_L^n$ and $r_R^k$ are independent of each other and have zero mean. Here, we need to analyze the variances of perturbation, $Var(r_L^n)$ and $Var(r_R^k)$. By investigating these variances, we can show that the variance of $\Delta t_i(n)$ is determined by

$$Var(\Delta t_i(n)) \approx \frac{1}{2w^2} \left( \frac{1}{10^{SNR_L/10}} + \frac{1}{10^{SNR_R/10}} \right).$$

(11)

where $SNR_L$ and $SNR_R$ represent the SNRs of the left and right channels respectively. If there is no intensity difference between two sensors, the SNR of the filtered signal can be approximated as

$$SNR \approx 10 \log_{10} \left( \frac{1}{w^2 Var(\Delta t_i(n))} \right).$$

(12)

For more detailed description of (11) and (12), refer to [23]. The above equation implies that the SNR of the filtered signal can be described by the variance of ITD samples and also the frequency of the filtered signal. Using the formula of (12), we can estimate the SNR approximately from the variance of ITD samples and the center frequency of bandpass filter used in the filter-bank. Then, the estimated SNR can be effectively used to construct the histogram of ITD samples: the measured ITD samples are weighted by the estimated SNR and accumulated in the histogram of ITD samples. For the sound source localization, the peak values of the histogram are identified and the corresponding ITD values are determined. Then, the sound source direction can be determined from the map of angles versus ITD values. Using this approach, the reliable estimation of ITDs is possible under noisy multi-source environments. As an example, the histograms of the measured angles using the CC method and the suggested
Figure 2: Estimation of multiple sound source directions using (a) the CC method and (b) the ZCTD method with SNR weighting: histograms of measured angles for four sound sources (two male and two female speech samples from the TI-DIGIT database) located at azimuth angles of -10, 0, 10, and 40 degrees under white Gaussian noise with a SNR of 5dB each.

ZCTD method with SNR weighting are illustrated in Figure 2 when four sound sources (two male and two female speech samples from the TI-DIGIT database) located at azimuth angles of -10, 0, 10, and 40 degrees mixed with white Gaussian noise with a small SNR (= 5 dB) each, were present simultaneously. These results showed that the histogram using the ZCTD method with SNR weighting presented four major peaks corresponding to the target source directions more clearly and with fewer artificial peaks than the CC method. Furthermore, the suggested method offers significantly less computational complexity compared to cross-correlation based methods[23].

3. Sound Segregation Algorithm

We will describe the algorithm for sound segregation based on the sound source localization using zero-crossings. Here, we assume that the sound source is captured by two sensors, L and R. Sensor L is used as the reference sensor, that is, $\Delta t_i(n)$ is estimated from sensor L. In this estimation, the identification of the reliable ITD samples with high SNR is necessary to determine the direction of the sound source accurately. For this purpose, we can consider the identification of reliable ITD samples using (12). Here, we assume that the sound source localization using zero-crossings[23] is done so that we have sound source ITD values $ITD(j)$, $j = 1, \cdots, K$ corresponding to K sound sources. Afterwards, the sound segregation algorithm based on the sound source localization using the binaural zero-crossing time differences (ZCTDs) is described as follows:

Step 1. (Selection of valid ITD-IID sample pairs) Select valid ITD-IID sample pairs $(\Delta t_i(n, k), \Delta p_i(n, k))$ in which $k$ satisfies the condition of (5).

Step 2. (Estimation of the powers of sound sources) For each time frame $\tau$ and frequency $i$, the following procedure is applied:

- For the ITD samples $\Delta t_i(n)$ of (6), $n = 1, \cdots, M$ associated with $M$ zero-crossing points in a time frame, select the nearest sound source ITD value. Then, the zero-crossing point is assigned to the selected sound source corresponding to the nearest sound source ITD value, that is, one of $ITD(j)$, $j = 1, \cdots, K$.
- For each zero-crossing point in a frame, the signal powers $p_{ni}$, $n = 1, \cdots, M$ between the current and previous zero-crossings are calculated.
- For each sound source, the signal powers associated with the assigned zero-crossings are accumulated: for $j = 1, \cdots, K$. 

129
Suppose the target sound source is the $i$th sound source. Then, $\delta(t,i)$ represents the target signal reconstructed using all-one mask under no interference and $\delta(t,l)$ represents the estimated target signal reconstructed from the binary mask generated by the masking method. The masking is done by comparing the powers associated with zero-crossings. The advantages of the suggested algorithm are 1) the robustness to noise due to the dominant frequency principle of zero-crossings, 2) the computational complexity involved in estimating ITDs, and 3) no need to train the masks for sound segregation.

4. Simulation

For the simulation of sound segregation, the sound sources were transformed by the Head-Related-Transfer-Function (HRTF)[24] and decomposed by a gamma-tone filter-bank composed of 128 channels, in which the center frequency of each channel was between 0.08 and 5.0 KHz. In this simulation, we used the data set collected by Cooke[25] which contains ten voiced speech signals and ten noise intrusions, encompassing a variety of common acoustic interferences such as telephone ring, rock music, and other speech utterances. We assume that the target sound source was located in the frontal axis, that is, 0 degree in all cases. In the case of two sound sources, the interfering sound source was located at an azimuth angle of 5 or 30 degrees. We also consider the spatial configuration that the interfering sound sources were located at azimuth angles of -5 and 5 degrees or located at azimuth angles of -30 and 30 degrees in the case of three sound sources while located at azimuth angles of -5, 5, and 30 degrees or located at azimuth angles of -30, 5, and 30 degrees in the case of four sound sources. The simulation of sound segregation was made using the suggested ZCTD masking (ZCTDM) method and the two types of cross-correlation based methods: 1) the ITD value consistent with the IID value was selected among the ITD values corresponding to the peak values of the histogram constructed from the cross-correlation of binaural signals within a time frame and masking was decided according to the nearest sound source ITD value (CCNM), and 2) the ITD value was estimated using the cross-correlation method and masking was performed on the ITD-IID space using the previously learned masks[3] which were trained for sound samples under the given spatial configuration (CCLM). We also made the simulation for sound segregation using the ideal binary masks which were obtained from a priori knowledge of the target and interferences. To compare the ideal masking, CCNM, CCLM, ZCTDM methods, we made the plots of estimated masks in time-frequency domain when the target sound source was located in the frontal axis, that is, 0 degree in all cases. In the case of two sound sources, the interfering sound source was located at azimuth angles of 5 or 30 degrees.

In principle, the suggested algorithm determines which sound source is more suitable for each zero-crossing using the ITD and IID value pair. The masking is done by comparing the powers associated with zero-crossings. The advantages of the suggested algorithm are 1) the robustness to noise due to the dominant frequency principle of zero-crossings, 2) the computational complexity involved in estimating ITDs, and 3) no need to train the masks for sound segregation.

Step 3. (Selection of segments in time-frequency domain) For each time frame $\tau$ and frequency $i$, the accumulated power for the selected sound source (or target sound source) is compared with the powers for other sound sources. If the power for the target sound source is larger than the sum of powers for other sound sources, the masking value of the segment is 1 (passing). Otherwise, the masking value of the segment is 0 (blocking). Suppose the target sound source is the $j$th sound source. Then,

$$M(\tau,i) = \begin{cases} 1 & \text{if } P_j/\sum_i P_i > 0.5 \\ 0 & \text{otherwise} \end{cases}$$

(13)

In principle, the suggested algorithm determines which sound source is more suitable for each zero-crossing using the ITD and IID value pair. The masking is done by comparing the powers associated with zero-crossings. The advantages of the suggested algorithm are 1) the robustness to noise due to the dominant frequency principle of zero-crossings, 2) the computational complexity involved in estimating ITDs, and 3) no need to train the masks for sound segregation.

4. Simulation

For the simulation of sound segregation, the sound sources were transformed by the Head-Related-Transfer-Function (HRTF) and decomposed by a gamma-tone filter-bank composed of 128 channels, in which the center frequency of each channel was between 0.08 and 5.0 KHz. In this simulation, we used the data set collected by Cooke which contains ten voiced speech signals and ten noise intrusions, encompassing a variety of common acoustic interferences such as telephone ring, rock music, and other speech utterances. We assume that the target sound source was located in the frontal axis, that is, 0 degree in all cases. In the case of two sound sources, the interfering sound source was located at azimuth angles of 5 or 30 degrees. We also consider the spatial configuration that the interfering sound sources were located at azimuth angles of -5 and 5 degrees or located at azimuth angles of -30 and 30 degrees in the case of three sound sources while located at azimuth angles of -5, 5, and 30 degrees or located at azimuth angles of -30, 5, and 30 degrees in the case of four sound sources. The simulation of sound segregation was made using the suggested ZCTD masking (ZCTDM) method and the two types of cross-correlation based methods: 1) the ITD value consistent with the IID value was selected among the ITD values corresponding to the peak values of the histogram constructed from the cross-correlation of binaural signals within a time frame and masking was decided according to the nearest sound source ITD value (CCNM), and 2) the ITD value was estimated using the cross-correlation method and masking was performed on the ITD-IID space using the previously learned masks which were trained for sound samples under the given spatial configuration (CCLM). We also made the simulation for sound segregation using the ideal binary masks which were obtained from a priori knowledge of the target and interferences. To compare the ideal masking, CCNM, CCLM, ZCTDM methods, we made the plots of estimated masks in time-frequency domain when the target sound source was located in the frontal axis and the interference sound sources were located at azimuth angles of -30 and 30 degrees as illustrated in Figure 3. In this simulation, the masks generated by the CCLM and ZCTDM methods showed similar results and closer to the ideal binary mask compared to the mask generated by the CCNM method which presented the noisy patterns especially in high frequencies. To measure the performance of sound segregation quantitatively, a segregated signal is reconstructed from a binary mask and the following SNR, the ratio in decibel of the target signal power to the power of error frequencies. To measure the performance of sound segregation quantitatively, a segregated signal is reconstructed from a binary mask and the following SNR, the ratio in decibel of the target signal power to the power of error frequencies. To measure the performance of sound segregation quantitatively, a segregated signal is reconstructed from a binary mask and the following SNR, the ratio in decibel of the target signal power to the power of error frequencies.

$$\text{SNR} = 10 \log_{10} \frac{\sum_t S_T^2(t)}{\sum_t (S_T(t) - S_E(t))^2}$$

(14)

where $S_T$ represents the target signal reconstructed using all-one mask under no interference and $S_E(t)$ represents the estimated target signal reconstructed from the binary mask generated by the masking method.

The simulation results of average SNRs for 100 trials using the ideal masking, CCNM, CCLM, and ZCTDM methods for sound segregation in the cases of two, three, and four sound sources were illustrated in Figures 4,
Figure 3. Comparison of masking methods: (a) ideal masking, (b) CCNM, (c) CCLM, and (d) ZCTDM methods.

Figure 4: Sound segregation for two sources (a) located at azimuth angles of 0 and 5 degrees, and (b) located at azimuth angles of 0 and 30 degrees.
Figure 5: Sound segregation for three sources (a) located at azimuth angles of -5, 0, and 5 degrees, and (b) located at azimuth angles of -30, 0, and 30 degrees.

Figure 6: Sound segregation for four sources (a) located at azimuth angles of -5, 0, 5, and 30 degrees, and (b) located at azimuth angles of -30, 0, 5, and 30 degrees.

5, and 6 respectively. These results showed that 1) the ideal masking method provided the best performance, 2) the difference of SNRs between the ideal masking and other methods became larger as the number of sources increased, and 3) the ZCTDM method outperformed the CCNM method in all cases, and 4) provided the similar performance with the CCLM method. Overall, the ZCTDM method showed the comparable performance with the CCLM method without the need of training masks for every spatial configuration.

5. Conclusion

We have suggested a method of sound source segregation using the masking method in which ITDs are estimated using zero-crossings detected from binaural filter-bank outputs in order to get more reliable ITD estimates in noisy multi-source environments. We also consider the estimation of ITDs with the aid of IID estimates to cope with the phase ambiguities of ITD samples in high frequencies. As a result, the proposed method is able to provide an accurate estimate of sound source directions and a good masking scheme for sound segregation while offering significantly less computational complexity compared to cross-correlation based methods. Simulation results for sound segregation in various interfering sound sources show that the suggested ZCTD masking method gives us
the better performance than the cross-correlation based masking method and the comparable performance with
the method using previously learned masks which should be adjusted for every spatial configuration. The good
features of the suggested ZCTD masking method are 1) the robustness to noise, 2) the less computational com-
plexity than the cross-correlation based methods, 3) no need to train the masks for sound segregation, and 4) the
comparable performance with the method using learned masks.

Acknowledgement

This work was supported by the Brain Neuroinformatics Research Program of the Korean Ministry of Com-
merce, Industry, and Energy.

References


Young-Ik Kim is currently a Ph.D. candidate of the Division of Applied Mathematics, Korea Advanced Institute of Science and Technology. His research interests include machine learning algorithms, auditory and visual scene analysis, binaural processing for speech segregation, recognition, and dereverberation.

Sung Jun An is currently a Ph.D. candidate of the Division of Applied Mathematics, Korea Advanced Institute of Science and Technology. His research interests are applications of machine learning, graphical modeling, noise robust feature extraction and automatic speech recognition.

Rhee Man Kil received the Ph.D. degree from the university of Southern California in 1991, and currently an associate professor of the Division of Applied Mathematics, Korea Advanced Institute of Science and Technology. His general research interests lie in theories and applications of machine learning and his current interests focus on model selection in regression, noise-robust feature extraction, and binaural speech segregation and recognition.

Hyung-Min Park received the Ph.D. degree from Korea Advanced Institute of Science and Technology in 2003. He is currently working in the Language Technologies Institute, Carnegie Mellon University as a Post-Doc. His current research interests include binaural speech segregation, noise-robust speech recognition, and independent component analysis.