# Development of an Optimized Feature Extraction Algorithm for Throat Signal Analysis

Young-Giu Jung, Mun-Sung Han, and Sang Jo Lee

**In this paper, we present a speech recognition system using a throat microphone. The use of this kind of microphone minimizes the impact of environmental noise. Due to the absence of high frequencies and the partial loss of formant frequencies, previous systems using throat microphones have shown a lower recognition rate than systems which use standard microphones. To develop a high performance automatic speech recognition (ASR) system using only a throat microphone, we propose two methods. First, based on Korean phonological feature theory and a detailed throat signal analysis, we show that it is possible to develop an ASR system using only a throat microphone, and propose conditions of the feature extraction algorithm. Second, we optimize the zero-crossing with peak amplitude (ZCPA) algorithm to guarantee the high performance of the ASR system using only a throat microphone. For ZCPA optimization, we propose an intensification of the formant frequencies and a selection of cochlear filters. Experimental results show that this system yields a performance improvement of about 4% and a reduction in time complexity of 25% when compared to the performance of a standard ZCPA algorithm on throat microphone signals.**

**Keywords: Throat microphone signal, Korean phonological feature, zero-crossing with peak amplitude (ZCPA), cochlear filter selection.**

Manuscript received Oct. 20, 2006; revised Feb. 28, 2007.

Young-Giu Jung (phone: + 82 42 860 6711, email: reraj@etri.re.kr) and Mun-Sung Han (email: msh@etri.re.kr) are with Digital Home Research Division, ETRI, Daejeon, Korea.

Sang Jo Lee (email: sjlee@knu.ac.kr) is with Department of Computer Engineering, Kyungpook National University, Daegu, Korea.

## I. Introduction

The development of today's computer environment is being driven by applications which provide user-centric service. User interface technologies are clearly important for emerging domains such as ubiquitous computing environments. In such environments, speech recognition technology is one of the most useful user interface technologies. However, because current noise-canceling technology remains immature, we are currently unable to realize a speech recognition system in a noisy environment such as in a factory or on a street.

Many solutions to this problem have been proposed: spectral subtraction [1], a soft-decision noise suppression filter [2], and a minimum mean-square error short-time spectral amplitude estimator [3], to name but a few. However, these developments have provided insufficient performance improvement. An alternative approach has been to develop speech recognition systems using noise-free microphones such as those using stethoscopic, in-ear, bone, or throat microphones.

Currently, research into recognition technology using noise-free microphones is being conducted in two areas. The first considers only using noise-free microphones and the second uses noise-free microphones as a complementary sensor aiding standard microphones. In this latter case, the noise-free microphone data is typically used for tasks such as end-point detection or speech enhancement.

Research using stand-alone noise-free microphones, has tended to focus on the development of devices which can sense a signal which can be understood by a human listener. There is no literature dealing with how best to construct a speech recognizer that reflects the signal characteristics of these devices. This lack has resulted in the non-optimal performance of these systems and is a significant obstacle to the commercialization of automatic speech recognition (ASR).

The development of recognition technology using noise-free devices has been partially achieved in Japan and the United States. Nakajima and others [4] proposed a stethoscopic microphone for non-audible murmur (NAM) recognition. Their microphone comprised a suction-disk and a polyester plate. As a result of using NAM sampling, consonant speech sounds were extremely difficult to recognize while vowel sounds were extremely clear. In order to solve this problem, they proposed exploring the optimum sensing location to provide a balance in the power ratio between vowels and consonants to support optimal recognition of every phoneme.

Jou and others [5] described various adaptation methods applied to recognize soft whispers recorded with a throat microphone. The adaptation methods include: maximum likelihood linear regression, feature-space adaptation, re-training with downsampling, sigmoid low-pass filtering, and linear multivariate regression. Zhang and others [6] showed novel hardware prototypes which integrate several heterogeneous sensors such as bone microphones, throat microphones, and in-ear microphones into a single headset. They described techniques for using this platform for robust speech detection, enhancement, and recognition in highly noisy, non-stationary environments.

Graciarena and others [7] presented a method to combine standard and throat microphone signals for robust speech recognition in noisy environments. The throat microphone, a skin vibration transducer, was used as a complementary sensor. They proposed extending the probabilistic optimum filter (POF) formulation to map the temporal sequence of noisy mel-cepstral features gathered from the standard microphone and the throat microphone, juxtaposed as an extended feature vector, to the clean standard microphone mel-cepstral features. In this previous work, the noise-free devices were used to complement the standard microphone input. However, in the present work, we use only a throat microphone to develop a speaker independent isolated-word recognizer which shows high performance in noisy environments.

There are many differences between standard and noise-free microphones. For instance, noise-free microphones typically sense signals in a limited frequency band and there is a partial loss of formant frequencies. Therefore, it is unlikely that a high performance system can be built if conventional feature extraction algorithms like the mel-frequency cepstral coefficient (MFCC) are applied. In order to solve this problem, it is imperative to select a feature extraction algorithm which matches the signal characteristics of the noise-free device. In this paper, we propose two methods to develop a speaker-independent isolated-word recognizer which guarantees high performance using only a throat microphone. The methods are as follows. Based on the Korean phonological feature theory and a throat microphone signal analysis using MFCC, we ascertain the possibility of developing an ASR system using only a throat microphone. We propose the conditions of the feature extraction algorithm to optimally extract features from a throat microphone. Finally, we propose the following two methods to optimize the feature extraction algorithm: an intensification of the formant frequencies, and the selection of cochlear filters.

The remainder of this paper is organized as follows. In section II, we present the conditions of a feature extraction algorithm which is suitable for throat signal analysis using the Korean phonological feature theory and a throat signal analysis using MFCC. We then verify the validity of the proposed conditions by conducting a performance comparison between a zero-crossing with peak amplitude (ZCPA) algorithm incorporating the proposed conditions and MFCC. In section III, we propose two methods to optimize the ZCPA algorithm [8] for the modeling of throat signals. Using a filter bank selection method, we decrease the time complexity of the ZCPA algorithm considerably. Furthermore, we demonstrate improved performance of the throat signal recognizer using formant analysis intensification.

## II. Characteristics of Throat Microphone Signals

The vocal apparatus consists of several organs: lips, tongue, teeth, hard and soft palates, uvula, larynx, lungs, and so on. These organs are integrated, and together form the complicated cavity which connects the lips to the lungs. If a human produces a meaningful word using the vocal apparatus then he or she has had to train on the phonological features of that sound for a sustained period of time. In this way, phonological features can suggest guide-lines for speech analysis. In this paper, we investigate the characteristics of Korean phonological features (KPF) relating to throat signals and describe the conditions of a feature extraction algorithm which would be best suited for throat microphone signal analysis.

### 1. Korean Phonological Features of Throat Signals

The Korean alphabet, like the English alphabet, consists of consonants and vowels. These are assembled into syllable units and described by letters. There are 21 voiced vowels and 19 consonants but these latter are voiced according to the form and location of surrounding sounds. Essentially, Korean syllables take one of the following forms: consonant + vowel + consonant, consonant + vowel, vowel + consonant, or vowel. Each syllable has a set of phonological features when uttered [9].

Phonological features are the intrinsic properties that

Table 1. Korean phonological feature table having sonorant feature [10].

| IPA | p | p* | pʰ | t | t* | tʰ | K | k* | kʰ | s | s* | tç | tç* | tçʰ | M | n | ŋ | l | h |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hangul | ㅂ | ㅃ | ㅍ | ㄷ | ㄸ | ㅌ | ㄱ | ㄲ | ㅋ | ㅅ | ㅆ | ㅈ | ㅉ | ㅊ | ㅁ | ㄴ | ㅇ | ㄹ | ㅎ |
| Sonorant | - | - | - | - | - | - | - | - | - | - | - | - | - | - | + | + | + | + | - |
| Constricted glottis | - | + | + | - | + | + | - | + | + | - | + | - | + | + | - | - | - | - | - |
| Aspirated | - | - | + | - | - | + | - | - | + | - | - | - | - | + | - | - | - | - | + |

distinguish one phoneme from another. There is a total of 14 phonological features in Korean. We focus on the throat vibration feature in connection with the production of consonants and vowels.

Acoustically, from the point of view of articulation, phonological features can be classified as sonorant, consonantal, or syllabic. Sonorant features involve vibrations in the throat. All vowels and some consonants have sonorant features in Korean. Sonorant consonants include "ㅁ(m)", "ㄴ(n)", and "ㅇ(ŋ)", and "ㄹ(l)".

Other consonants do not have sonorant features. However, in utterance patterns, the remaining consonants can be classified by two features: constricted glottis and aspirated features. The constricted glottis feature separates plain consonants (ㅂ,ㄷ,ㄱ,ㅅ) from tense (ㅃ,ㄸ,ㅉ,ㄲ) and aspirate consonants (ㅍ,ㅌ,ㅊ,ㅋ). These two features are accompanied by tensing of the throat. Tense and aspirate consonants also have the constricted glottis feature. However, tense and aspirate consonants can be differentiated by the presence or absence of the aspirated feature. It involves making a vibration in the throat and aspirate consonants possess it.

Another important characteristic is phoneme change, in particular nasalization and intersonorant obstruent voicing. A nasalization refers to the fact that consonants with a [-sonorant] feature (ㅂ,ㄷ,ㄱ,ㅍ,ㅌ,ㅋ,ㄲ,ㅅ,ㅆ,ㅈ,ㅊ,ㅎ) are changed to a nasal with a [+sonorant] feature when combined with a nasal consonant (ㅁ,ㄴ,ㅇ). Intersonorant obstruent voicing is the phenomenon which occurs when an obstruent is changed to a voiced sound when sandwiched between phonemes with [+sonorant] features. In the Korean consonant set, the plain phonemes (ㄱ,ㄷ,ㅂ,ㅊ,ㅎ) show intersonorant obstruent voicing between voiced sounds with the [+sonorant] feature.

In Korean, 9 consonants out of 19 have the [+sonorant] feature. The [-sonorant] feature of the remaining 10 consonants is frequently changed to the [+sonorant] feature in a word because of nasalization and intersonorant obstruent voicing. Due to these Korean phonological features, we believe that a high-performance isolated-word recognizer could be developed using only a throat microphone. Furthermore, this analysis reveals that the feature vectors (such as zero-crossing, peak, and period pitch) used for voiced/unvoiced classification will be useful in throat signal modeling.

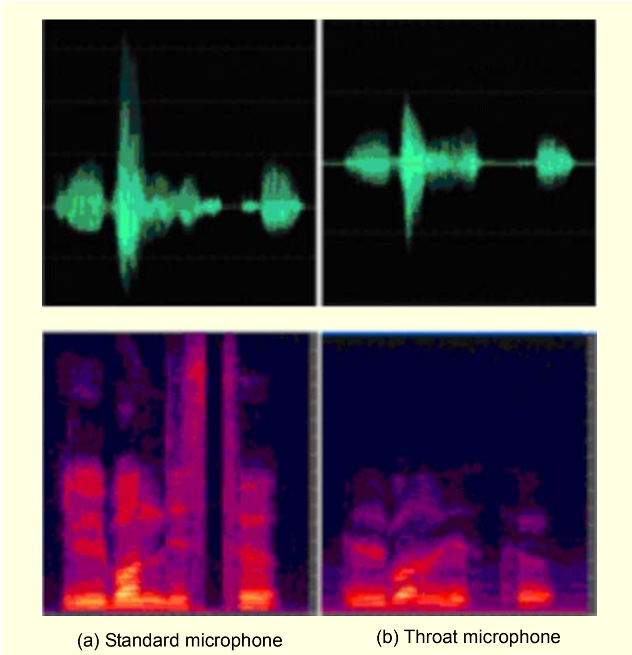


Fig. 1. Outputs of the standard and throat microphones.

## 2. Throat Microphone Signal Analysis and Modeling

Speech recorded with a throat microphone is not incomprehensible to the human ear. Figure 1 shows the output signal of standard and throat microphones. The upper two charts show that an utterance is different in the magnitude of the amplitude profile, but the shape of the two signals seems similar. However, the lower part of the figure demonstrates that there are significant differences in the spectrum.

The difference in the amount of information in the spectrum manifests as a significant performance difference in an ASR system. To observe the performance difference, we evaluated the performance of our system on both standard and throat microphone data. In our system, we adopted the MFCC with 13 coefficients for feature extraction, and the time delay neural network (TDNN) as a recognizer. The experimental data

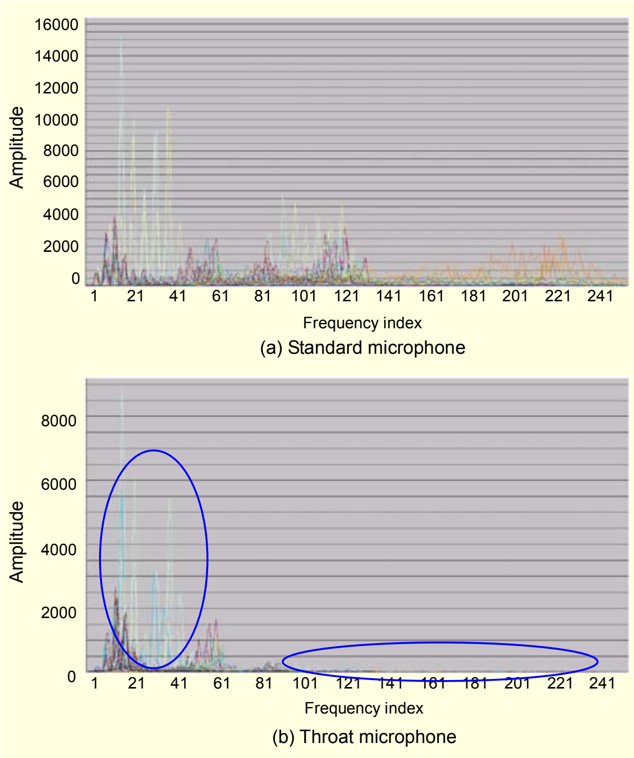


Fig. 2. Energy variation of each frame in frequency domain.

consisted of 50 words spoken by 150 males. Each speaker uttered the words once in a quiet office environment and the sounds were recorded by both a standard and a throat microphone. The utterances were sampled at 16 kHz with 16-bit resolution. One hundred sets were used for training, and the other 50 sets were used as test patterns. The experimental results demonstrate that a recognition system using data from a throat microphone shows poor performance (about 30% less) when compared to a system using a standard microphone. Table 2 shows the performance of the recognizer according to each type of microphone.

To ascertain the underlying cause of this performance difference, we analyzed a throat microphone signal using MFCC.

Figure 2 shows the frequency information of each frame after the application of pre-emphasis, Hamming window, and fast Fourier transform (FFT) processes. The horizontal axis represents the frequency domain and the vertical axis represents the amplitude in the frequency band. Different colors indicate different frames.

As shown in Fig. 2, the frequency distribution of the two signals is similar below 2 kHz. Above this point, the data captured from the throat microphone is reduced. Above 4 kHz it is negligible.

This signal analysis clarifies why the isolated-word recognizer using the throat microphone data shows degradation in performance: much of the required data (that above 2 kHz) is simply not present. In fact, the most significant problem is the loss of frequency information in the band between 2 kHz and 4 kHz. The second formant frequency lies within this band, and its loss exerts a significant negative effect on the recognition of both vowels and voiced consonants, which appear there weakly.

To solve this problem, we suggest that the following conditions must be considered by a feature extraction algorithm in order to reliably analyze the signal from a throat microphone. First, it should have a sensitive band-pass filter which is similar to that found in the human auditory system. Second, it is more useful to use feature vectors for voice/non-voice classification, such as zero-crossing rate, peak, and pitch period.

In this study, we use a finite impulse response (FIR) filter to satisfy the first condition. The FIR filter consists of 16 Hamming band-pass filters, and it outputs sample information on each frequency band for every sample. Therefore, we can apply various feature vectors to the frequency bands. Regarding the second condition, we adopt an established feature vector based on signal processing utilizing acoustic features [11].

To verify the feasibility of our solution, we developed an ASR system using a ZCPA algorithm which incorporates the above mentioned modifications. We experimentally demonstrate that using this ZCPA algorithm improves the recognizer's performance by approximately 16% over that achieved by MFCC. The experimental environment and results are presented in section III.

## III. Optimization of the ZCPA Algorithm and Experimental Results

In the previous section, we described the conditions required for an effective feature extraction algorithm for throat signals. In this section, we propose an optimization method for a ZCPA algorithm [8] for a throat recognition system. For the experiment, we use the same data as in the previous experiment shown in Table 2.

To verify the performance of our recognizer according to each feature extraction algorithm, we tested the recognition rate of the proposed recognizer on standard microphone data. Table 3 shows this recognition rate. The results indicate that

Table 2. Comparison of recognizer performance.

| | Standard microphone | Throat microphone |
|---|---|---|
| Recognition rate (%) | 97.14 | 67.8 |

Table 3. The recognition rate with standard microphone.

| | Standard microphone | |
|---|---|---|
| Feature extraction | MFCC | ZCPA |
| Recognition rate (%) | 97.14 | 98.86 |

Table 4. Comparison of recognition rates of our experimental results with related work.

| | Feature extraction | Recognition rate (%) |
|---|---|---|
| Lee's system [12] | PLP Cepstrem | 52.3 |
| | MFCC+CMS | 75.0 |
| Our system | MFCC (8 kHz) | 60.5 |
| | MFCC+CMS | 74.8 |
| | ZCPA | 83.63 |

both algorithms achieve high, and nearly identical, levels of performance.

Table 4 compares the performance of our initial system with that achieved in previous work dealing with throat microphones reported by Lee [12]. Lee developed a speaker-dependent isolated-word recognizer using a throat microphone. When the MFCC algorithm was used, the performance of Lee's system was similar to the performance of our system. This result shows the limits of MFCC in throat signal analysis. Also, by examining the results from the down-sampled data (8 kHz), we conclude that the limited frequency information is not the only reason for the low performance of the recognizer using MFCC. The relatively high performance of the ZCPA algorithm suggests that our proposed throat signal modeling method is feasible. In Table 4, the cepstrum mean subtraction (CMS) algorithm was used as the channel normalization (CN) algorithm of our system.

Next, we optimized the ZCPA algorithm according to the specific characteristics of the analyzed throat signal. Through this optimization, we will be able to reduce the time complexity of the algorithm and improve its performance.

### 1. Optimization of the ZCPA Algorithm

The significant information contained in the signal from a throat microphone is of a relatively low frequency. Of all the information, the voiced-sound-related component is critical; however, the formant frequencies of F1 and F2 appear weakly in the spectrum of a throat microphone signal. To solve this problem, we propose modifying the window length according to the average pitch length. The ZCPA algorithm analyzes a signal by using a different window size for each channel [13].

Table 5. Recognition rate of recognizer at each parameter C (Feature extraction algorithm : ZCPA + RASTA).

| Parameter C | Window length (ms) | Recognition rate (%) |
|---|---|---|
| C = 10 | 3–50 | 85.16 |
| C = 20 | 5–100 | 87.47 |
| C = 30 | 9–150 | 87.3 |
| C = 40 | 12–200 | 86.81 |

Table 6. Center frequency at each histogram bin.

| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Hz | 200 | 264 | 340 | 429 | 533 | 656 | 801 | 972 |
| Index | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Hz | 1172 | 1408 | 1685 | 2001 | 2394 | 2845 | 3376 | 3999 |

Table 7. Recognition rates of the throat signal isolated-word recognizer by the cochlear filter selection.

| Removed bank number | Recognition rate (%) | Speedup rate (%) |
|---|---|---|
| None | 89.5 | 0 |
| Bank 16 | 89.1 | 6.25 |
| Bank 15,16 | 88.9 | 12.5 |
| Bank 14, 15, 16 | 88.6 | 18.75 |
| Bank 13,14,15,16 | 89.4 | 25 |

The window lengths become long for low frequencies, and short for high frequencies. Equation (1) is the function that calculates the length of windows in ZCPA algorithms:

$$L_k = \frac{C}{F_k}, \qquad (1)$$

where $F_k$ is the center frequency of the corresponding band-pass filter, and C is a constant, determining the window length. We tested four different values of parameter C: 10, 20, 30, and 40. We measured the recognition rate achieved with each value of C and report the optimal one for throat signal analysis. Table 5 shows the recognition rate and window length for each value of C. In Table 5, RelAtive SpecTrAl (RASTA) [14] was used as the CN of our system.

When there are low values for parameter C, there are unreliable frequency estimates because the window lengths are shorter than the average pitch period. When there are high values for the parameter, the windows are too long for the low-frequency sub-bands which can obstruct the stationarity assumption. This experiment suggests that the optimal value of parameter C is 20.

2. Optimization According to the Cochlear Filter Selection

The ZCPA algorithm used for the experiment was coded for speech signal analysis; therefore, its time complexity is reduced and its performance is improved if we apply the characteristics of the proposed throat signal [14]. In this section, we describe the use of a cochlear filter selection method to reduce time complexity without degrading the recognition rate. The key idea of the cochlear filter selection method is to remove the output of unnecessary cochlear filters. We can explain the cochlear filter selection method using the equation of the ZCPA algorithm. The output of the ZCPA algorithm at time $m$ as given in [8] is

$$y(m,i)=\sum_{k=1}^{N_{ch}}\sum_{l=1}^{Z_{k-1}}\delta_{ij_l}\,g(p_{kl}),\qquad 1\le i\le N, \tag{2}$$

where $N$ is the number of frequency bins, and $\delta_{ij}$ the Kronecker delta. For each channel, the index frequency bins, $j_l$, are computed by taking the inverse of the time interval between the $l$-th and $(l+1)$th zero-crossings, for $l=1,\cdots, Z_k-1$. The value of the frequency histogram at the frequency bin, $j_l$, is then increased by $g(P_{kl})$, and g() is a monotonic function which implements this saturation nonlinearly (a log function is currently used).

Before the experiment, we analyzed the output of the cochlear filter. In the ZCPA algorithm, the FIR filter was used as the cochlear filter. The frequency response of the filter bank consisted of 16 Hamming band-pass windows, which were designed using the window method, are shown in Fig. 3. Table 6 shows the 16 frequency bins and the frequency domain of the ZCPA algorithm.

We first analyzed the output of the FIR filter for each channel to determine which could be removed without affecting performance. Figure 4 shows the signal shape of each channel after FIR filtering.

In Fig. 4, we can see that a similar waveform appears in bank 13 and above. This characteristic of the throat signal can be used to reduce the time complexity without degrading the recognition rate by removing these portions of the signal. Table 7 shows the recognition rate of the throat signal isolated-word

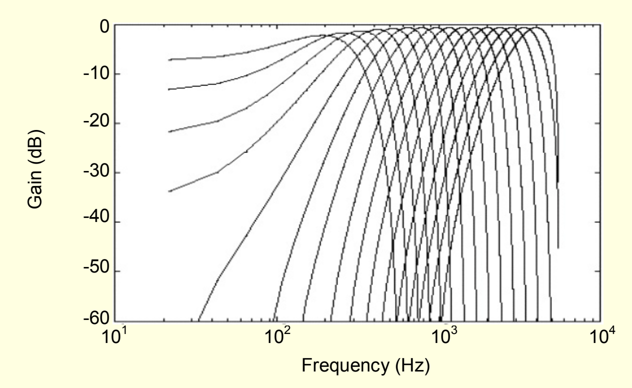


Fig. 3. Frequency response of cochlear filter bank implemented with FIR filter [11].

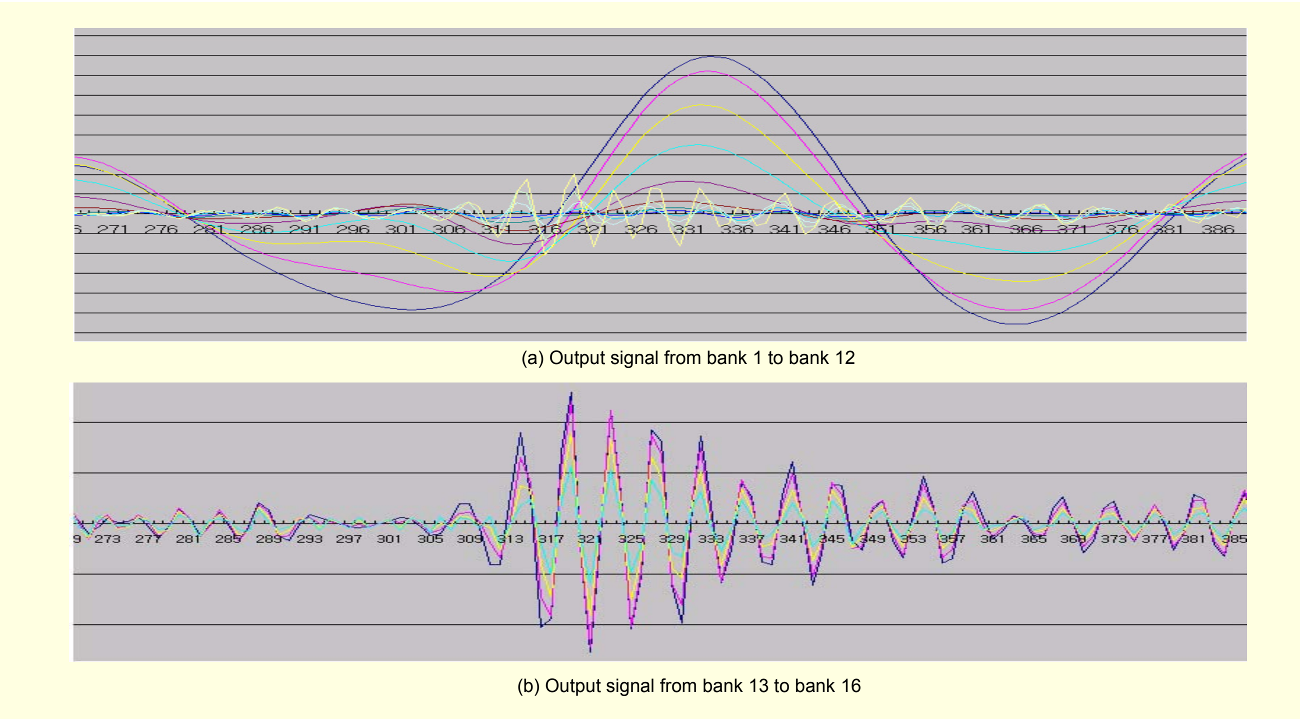


Fig. 4. Output signal of each bank after FIR filtering.

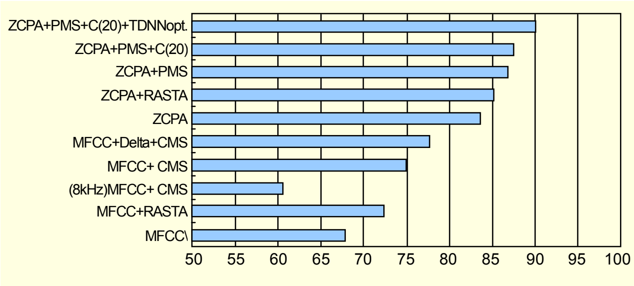


Fig. 5. A performance improvement of the speaker-independent isolated-word recognizer using the throat microphone.

recognition system incorporating cochlear filter selection with a different number of filters.

The results of this experiment demonstrate that we can safely remove frequency bins 13, 14, 15, and 16 without reducing performance. In fact, the time complexity of the ZCPA algorithm is reduced by about 25% using the proposed method. Finally, Fig. 5 shows the performance improvement of our speaker-independent isolated-word recognizer using a throat microphone compared to the developmental stages described earlier in this paper.

## IV. Conclusion

A speech interface is one of the most useful user interfaces in man-machine interaction, but in noisy environments, we have yet to achieve a reliable applied isolated-word speech recognition system. In this paper, we present a speaker-independent isolated-word recognition system which we developed using only a throat microphone.

So far, studies of the throat microphone have focused on using the signals to compliment standard microphone signals because they have a limited frequency band and a partial loss of formant frequencies. However, based on KPF, this study has demonstrated that it is possible to develop an ASR system using only a throat microphone. From KPF and an analysis of throat microphone signal based on MFCC, the conditions of the feature extraction algorithm to robustly extract feature vectors from a throat microphone signal were proposed. Finally, a ZCPA algorithm was optimized to improve the performance of the ASR system using only a throat microphone.

These conditions involve a sensitive band-pass filter and the use of feature vectors for voiced/unvoiced classification. The optimization method uses window length adjustment and cochlear filter selection. The experimental results show a performance improvement of about 4% and a reduction in time complexity of 25% when compared to a standard ZCPA algorithm on a throat microphone signal. As shown in Fig 5, our system using only the throat microphone achieves improved performance.

## References


[1] S.F. Boll "Suppression of Acoustic Noise Speech Using Spectral Subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-27, Apr. 1979, pp. 113-120.

[2] R.J. McAulay and M.L. Malpass, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, Apr. 1980, pp. 137-145.

[3] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, Apr. 1985, pp. 443-445.

[4] Y. Nakajima, H. Kashioka, K. Shikano, and N. Campbell, "Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin," *ICASSP'03*, vol. 5, 2003, pp. 708-711.

[5] S.C. Jou, T. Schultz, and A. Waibel, "Adaptation for Soft Whisper Recognition Using a Throat Microphone," *Proc. ICSLP*, Jeju Island, Korea, Oct. 2004.

[6] Z. Zhang, Z. Liu, M. Sinclair, A. Acero, L. Deng, J. Droppo, X. Huang, and Y. Zheng, "Multi-Sensory Microphones for Robust Speech Detection, Enhancement, and Recognition," *ICASSP'04*, vol. 3, May 2004, pp. 781-784.

[7] M. Graciarena, H. Franco, K. Sonmez, and H. Bratt, "Combining Standard and Throat Microphones for Robust Speech Recognition," *IEEE Signal Processing Letters*, vol. 10, no. 3, Mar. 2003, pp. 72-74.

[8] D.S. Kim, S.Y. Lee, and Rhee M. Kil, "Auditory Processing of Speech Signals for Robust Speech Recognition in Real-Word Noisy Environments," *IEEE Tran. Speech and Audio Processing*, vol. 7, no. 1, Jan. 1999.

[9] H.O. Gu, *Understanding of Korean Phonology*, The Institute of Language Culture, 1999.

[10] J.Y. Shin and J.E. Cha, *System of Korean Phonology*, Institute of Language Culture, 2003.

[11] C.K. Un and S.C. Yang, "A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF," *IEEE Trans. Acoust., Speech Signal Processing*, ASSP-25, Dec. 1977, pp. 565-572.

[12] Y.C. Lee, S.H. Lee, H.S. Hong, M.S. Han and P.S. Ma, "A Study on Speech Recognition for Neck-Microphone Input Signal," *18th Korea Information Processing Society Conference*, vol. 9, Nov. 2002, pp. 747-750.

[13] B. Gajic and K.K. Paliwal, "Robust Speech Recognition Using Features Based on Zero Crossings with Peak Amplitudes," *ICASSP '03*, vol. 1, Apr. 2003, pp. 64-67.

[14] H. Hermansky and N. Morgan, "RASTA Processing of Speech," *IEEE Trans. Speech Audio*, vol. 2, no. 4, pp. 578-589.

**Young-Giu Jung** received the BS degree in computer science from Silla University, Korea, in 2000. He received the MS degree in computer engineering from Kyungpook National University, Korea, in 2002. Since 2004, he has been with Electronics and Telecommunications Research Institute (ETRI), Korea, as a Member of Research Staff. His research interests include natural language processing, speech recognition, and multimodal recognition.

**Mun-Sung Han** received the BS degree in mathematics from Seoul National University, Korea, in 1977. He received the MS degree and completed the PhD course in computer science in Indiana University, Bloomington, USA, in 1985 and 1988. Since 1989, he has been with Electronics and Telecommunication Research Institute (ETRI), Korea, as Senior /Principle Member of Research Staff. His research interests include speech recognition, multimodal recognition, and brain computing.

**Sang Jo Lee** received the BS degree in mathematics education from Kyungpook National University, Korea, in 1974. He received the MS degree in computer science from KAIST, Korea, in 1976. He received the PhD degree in computer engineering from Seoul National University, Korea, in 1994. He is currently a professor of Kyungpook National University, Korea. His research interests include natural language processing, knowledge processing, information retrieval, machine learning, and semantic web.