# Practical Considerations for Hardware Implementations of the Auditory Model and Evaluations in Real World Noisy Environments

*Doh-Suk Kim, *Jae-Hoon Jeong, *Soo-Young Lee, and **Rhee M. Kil

Abstract

Zero-Crossings with Peak Amplitudes (ZCPA) model motivated by human auditory periphery was proposed to extract reliable features from speech signals even in noisy environments for robust speech recognition. In this paper, some practical considerations for digital hardware implementations of the ZCPA model are addressed and evaluated for recognition of speech corrupted by several real world noises as well as white Gaussian noise. Infinite impulse response (IIR) filters which constitute the cochlear filterbank of the ZCPA are replaced by hamming bandpass filters of which frequency responses are less similar to biological neural tuning curves. Experimental results demonstrate that the detailed frequency response of the cochlear filters are not critical to the performance. Also, the sensitivity of the model output to the variations in microphone gain is investigated, and results in good reliability of the ZCPA model.

## Ⅰ. Introduction

Human auditory system is robust to background noise, and there have been many researches devoted to modeling functional roles of the peripheral auditory systems for robust front-ends of speech recognition systems in noisy environments [1, 2, 3, 4, 5, 6]. Although computational auditory models have been shown to outperform conventional signal processing techniques, modeling peripheral auditory systems is still a difficult problem since it requires an interdisciplinary research including physiology, psychoacoustics, physics, electrical engineering, etc., and since little is known about the exact mechanism of the auditory periphery for mathematical construction of the model. Also, most auditory modeling researches heavily rely on experiments to make the output of computational model coincide with the biological observations, and analytic treatments are intractable since they usually involve multistage nonlinear transformations. Further, auditory models require careful determination of a lot of free parameters which should be determined by trial-and-error methods, and require much computation time, which make it difficult to be used widely in speech recognition systems.

A simple and efficient auditory model, Zero-Crossings with Peak Amplitudes (ZCPA) model, was proposed as a robust front-end for speech recognition systems in noisy environments [7, 8]. The ZCPA is a simplified auditory model and the computational complexity is much less severe than other auditory models, and was shown to outperform both linear predictive coding (LPC) cepstrum and the ensemble interval histogram (EIH) model when speech is corrupted by white Gaussian noise. In this paper some practical considerations for digital hardware implementations of the ZCPA model are addressed, and evaluated for recognition of speech data corrupted by several real world noises as well as white Gaussian noise.

## Ⅱ. ZCPA Analysis

The ZCPA model consists of a bank of bandpass cochlear filters and nonlinear stages at the output of each cochlear filter. Fig. 1 represents the block diagram of the ZCPA analysis.

The cochlear filterbank represents frequency selectivity at various locations along a basilar membrane in the cochlea, and was implemented with Kates' traveling wave filters without adaptive feedback mechanism [7, 8, 9]. Period histogram and interval histogram of firing patterns of auditory nerve fibers reveal that there is a high degree of phase locking in auditory nerve fibers, that is, auditory nerve fibers tend to fire in synchrony with the stimulus [10, 11, 12]. In the ZCPA model, a synchronous neural firing is simulated as the upward-going zero-crossing event of the signal at the output of each bandpass filter,

*Department of Electrical Engineering Korea Advanced Institute of Science and Technology
** Division of Basic Sciences Korea Advanced Institute of Science and Technology

Manuscript Received: December 23, 1996.

and the inverse of time interval between adjacent neural firings is represented as a frequency histogram. Further, each peak amplitude between successive zero-crossings is detected, and this peak amplitude is used as a nonlinear weighting factor to a frequency bin to simulate the relationship between the stimulus intensity and the degree of phase-locking of auditory nerve fibers. The histograms across all filter channels are combined to represent output of the auditory model. Thus frequency information of the signal is obtained by zero-crossing intervals, and intensity information is also incorporated by a peak detector followed by a saturating nonlinearity.

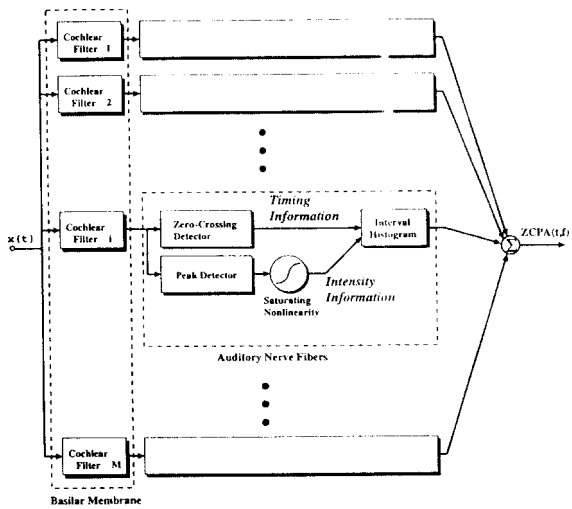



Figure 1. Block diagram of the zero-crossings with peak amplitudes (ZCPA) model.

Let us denote the output signal of the $k$-th bandpass filter by $x_k(n)$ and the frame of $x_k(n)$ at time $m$ by $x_k(n;m)$, which is obtained as

$$x_k(n;m) = x_k(n)w_k(m-n),\ k = 1, \ldots, N_{ch} \tag{1}$$

where $w_k(n)$ is a window function of finite length, and $N_{ch}$ is the number of channels, i.e., the number of cochlear filters. Further, let us denote $Z_k$ by the number of upward-going zero-crossings of $x_k(n;m)$, and $P_{kl}$ by the peak amplitude between the $l$-th and $(l+1)$-th zero-crossings of $x_k(n;m)$, respectively. Then the output of the ZCPA at time $m$ is represented as

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

where $N$ is the number of frequency bins, $j_l$ is the index of the frequency bin computed using the $l$-th and $(l+1)$-th zero crossings, and $\delta_{ij}$ is the Kronecker delta. $g(\,)$ is a monotonic function which implements the relation between the stimulus intensity and the degree of phase-locking of auditory nerve fibers. The length of window function, $L_k$, is determined as $10/F_k$ to capture about 10 periods of the signal when the signal is pure sinusoid of frequency, $F_k$, where $F_k$ is the characteristic frequency of the $k$-th channel [5].

Thus, the window lengths become large for low frequencies, and small for high frequencies. As a result, frequency resolutions are finer while time resolutions are poorer at lower frequencies, and vice versa at higher frequencies. This property is consistent with psychoacoustic observations.

The EIH model utilizes level-crossings for frequency information [4]. However, unlike the ZCPA model, multiple level-crossing detectors with different level values are utilized both for frequency and intensity information in the EIH model. In implementing the EIH, one has to determine several parameters such as the number of levels and level values, which are extremely critical for reliable performance. However, there is no elegant method to determine these values, except by trial-and-error. The utilization of zero-crossings in frequency estimation makes the ZCPA model free from unknown parameters associated with the level, more efficient for calculations, and more robust to noise than the EIH model. Let us consider the following signal

$$x(t) = A_s \cos(\omega_s t + \theta) + A_n v(t) \tag{3}$$

where $v(t)$ is a bandlimited white Gaussian noise with a rectangular power spectrum of bandwidth $W$ [rad/sec] and has zero mean and unit variance. Let us suppose that $x(t)$ is filtered by an ideal bandpass filter of bandwidth $B$, and the output of the bandpass filter contains a sinusoidal signal plus bandpass filtered noise. If $r_n$ denotes the perturbation in the level-crossing positions introduced by the noise, the variance of the interval perturbations is obtained as

$$\sigma^2 = E\{|r_n - r_{n+1}|^2\}$$

$$= \frac{2B}{W}\left(\frac{A_n}{\omega_s A_s}\right)^2 \frac{1}{1-(A_l/A_s)^2} \tag{4}$$

where $A_l$ denotes a crossing level value [7, 8]. The variance of the time interval perturbations between two adjacent level-crossings has a minimum value for $A_l = 0$. This implies that higher level values result in higher sensi-

tivity in the estimated intervals and frequencies. The estimated spectra based on zero-crossings have a tendency to enhance the dominant signal component and also to suppress adjacent noise components. This property can be explained by the dominant frequency principle [13] and contributes to the noise-robustness of the ZCPA model. The operation of the ZCPA is significantly different from conventional signal processing techniques such as FFT in that the local frequency and intensity information of one period of the signal is measured and then accumulated to obtain the output.

## Ⅲ. Data Base and Recognition Systems

### 3.1 Data Base and Noise Material

In consideration of practical applications of automatic speech recognition, 50 Korean words which seem to be necessary for control of electric home appliances including TV and VCR were chosen. The utterances from 16 male speakers were sampled at 11. 025 kHz sampling rate with 12 bit precision via SONY ECM-220T condenser microphone. The data base has relatively low quality in consideration of the cost and speed of hardware, which is under development [14]. 900 tokens of 9 speakers were used as training of recognizers, and 1050 tokens of the other speakers as test evaluations.

There are many kinds of noises in real environments which are not stationary in general, and performance evaluation in real situations may be very important for practical applications of ASR. Factory noise, military operations room noise, and car noise, contained in NOISEX-92 CD ROMS [15], were added to the test data sets at various SNRs for test evaluations in real situations. The NOISEX-92 database is produced by the NATO research study group on speech processing in liaison with the ESPRIT SAM (Speech Assessment Methodology) project laboratories, and the noises are from the NATO-RSG-10 noise database [16].

The NATO-RSG-10 database, which is aimed at the evaluation of automatic speech recognition systems and speech communication channels in military situations, contains some examples of representative noise sources such as jet-plane, helicopter, wheel carrier, tank, and command room. Properties of real-world noises used in this paper are described in Table 1, and Fig. 2 shows spectrogram of each noise material. There are periodic sounds of impingement of machinery in both factory noise and military operations room noise. Also, speech noise is contained in the military operations room noise, which makes the problem more difficult. Most of energy of car noise is concentrated at low frequencies due to mechanical characteristics, as shown in Fig. 2 (c). In a car, noise comes from many factors such as the engine, the fans, transmission, tire-surface interaction, and the aerodynamic effects. What makes the actual problem more complicated is the situation of a car. That is, noise can be generated by the passenger and audio equipment besides car itself, and whether the window is opened or not may play an important factor. It was found that the SNR of speech signals recorded in a passenger car with a microphone mounted on the dashboard in front of speaker could drop below -5 dB while the car was in movement with closed windows and without fan [17].

Table 1. Description of real-world noise used in this paper.

| Source | Description |
|---|---|
| Factory Noise | Car floor production, electrical welding |
| Military Operations Room Noise | Operations room of destroyer |
| Car Noise | Volvo-340, 120 km/h, 4th gear, asphalt road |

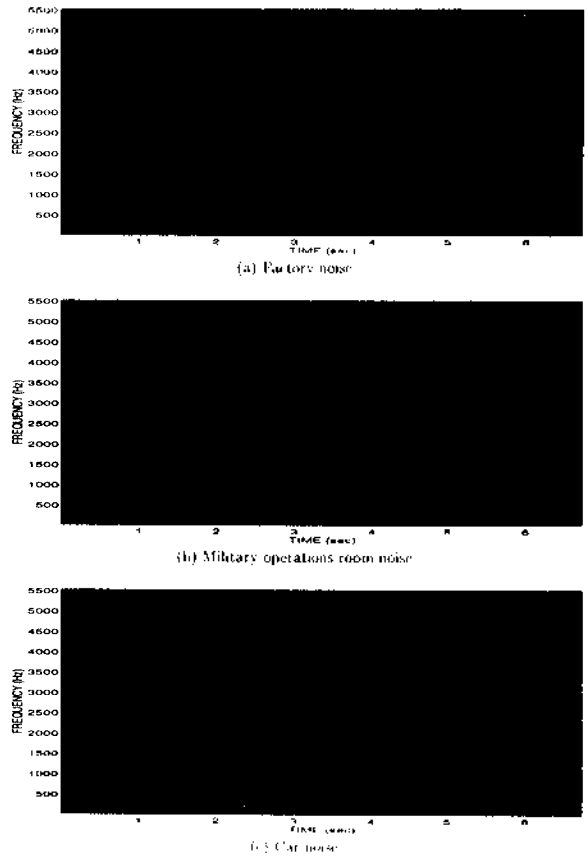



Figure 2. Spectrogram of (a) factory noise, (b) military operations room noise, and (c) car noise.

### 3.2 Speech Recognition Systems

In this paper, both discrete hidden Markov model (HMM) speech recognizer and multilayer perceptron (MLP) recognizer preceded by trace-segmentation [18] are used to investigate the recognizer-independent reliability of features extractions.

Word-level HMM construction is performed since the task is isolated word recognition. Each HMM models a particular word with the left-to-right model. In the left-to-right model, each state has only two transitions, one is going back to its own state and the other is going to the next state. The number of states of the HMM is set to be either five for one-syllable word or eight for multi-syllable word. Each HMM is iteratively trained with Baum-Welch algorithm based on maximum likelihood estimation (MLE). The codebook is trained with training data in iterative manner [19], and the size of codebook is set to be 256.

There have been a lot of schemes proposed to apply neural networks to speech recognition, and static approach utilizing an MLP showed better performance than dynamic approach at least for isolated word recognition tasks [20, 21]. However, the problem of time-variation of speech should be handled before classification by static neural network, since the number of input neurons is fixed whereas the length of speech signal varies at each pronunciation. Trace-segmentation algorithm [18] is a good candidate for normalization of time scale without serious computation time. For each isolated word cumulative distances of input features are calculated at each frame, and an overall trace of the feature is then divided into $(N-1)$, representing equivalent amounts of feature changes between each normalized time interval. New input features may be formed by interpolation to provide the equivalent amount of change between adjacent time frames. This simple time normalization procedure reduces redundancies of speech period, especially for steady long-pronounced vowels. MLP is trained by using error back propagation algorithm [22] with new input features passed through trace-segmentation, where each output neuron indicates a particular word. Thus, the number of output neurons is same as the number of vocabulary words. The number of hidden neurons is twice that of output neurons, and the number of input neuron is the normalized time frames, $N$, which is 64, multiplied by the number of components of a feature vector at one time frame.

## Ⅳ. Practical Consi lerations for Digital Hardware Implementations

There exists a lot of stand-alone applications of automatic speech recognition technology in which the whole platforms such as workstations and personal computers cannot be used. Therefore it is necessary to develop stand-alone hardware such as ASICs (Application Specific Integrated Circuits), and several factors of the auditory model should be modified and optimized. Also, even though the ZCPA is a simplified auditory model and the computational complexity is much less severe than other auditory models, the required computation time is still greater than conventional feature extraction algorithms.

Thus, several factors of the developed auditory model should be considered for efficient digital hardware implementations, which is under development [14].

### 4.1 Choice of Cochlear Filters

Both the number of bandpass filters and that of frequency bins are set to 16, since it is more effective to use powers of two as the number of parameters for digital hardware implementations. Frequency range between 1.5 bark and 17.5 bark is divided into 16 frequency bins equally spaced by one bark according to the critical-band rate [23].

For cochlear filters, it is recommended to use finite impulse response (FIR) filters than infinite impulse response (IIR) filters for digital hardware implementations because roundoff noise and coefficients quantization errors are much less severe in FIR filters than in IIR filters, and stability of IIR filters should be carefully considered. In [7], cochlear filters of the ZCPA were implemented with Kates' traveling wave (TW) filters. Fig. 3 shows frequency response of cochlear filterbank implemented with Kates' TW filters [9]. Kates' traveling wave filter sections are actually IIR filters, and these IIR filters are cascaded by the number of frequency bands. Thus it is not profitable for digital hardware if only fixed-point calculations are available, and it is necessary to design the cochlear filterbank with FIR filters.

Frequency response of filterbank consists of 16 hamming bandpass filters (FIR filters), which are designed by window method, is shown in Fig. 4. Even though the desired filter shape is not aimed to follow biological neural tuning curve in detail, the center frequencies of filterbank are determined between 200 Hz and 4000 Hz by the frequency-position relationship on the basilar membrane [24], which is represented as

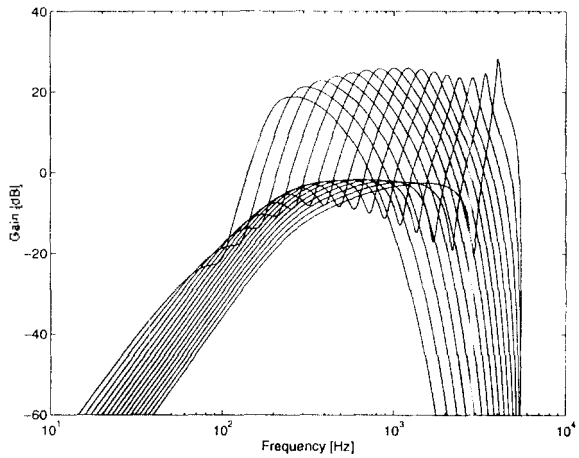



Figure 3. Frequency response of cochear filterbank implemented with TW filters.

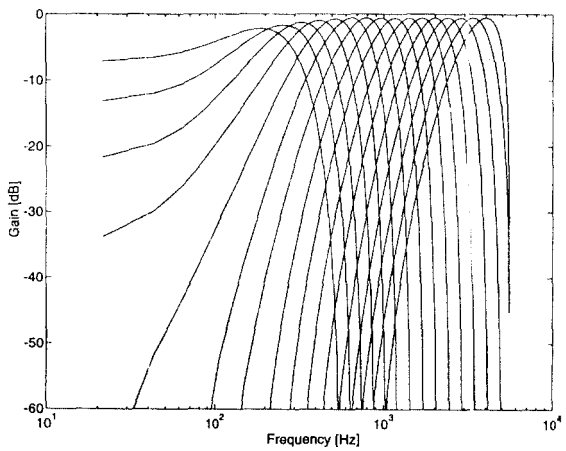



Figure 4. Frequency response of cochlear filterbank implemented with FIR filters.

Table 2. Comparison of recognition rate (%) of the ZCPA obtained using the TW filters and FIR filters.

(a) Results of HMM recognizer

| Noise | | WGN | | FAC | | MOP | | CAR | |
|---|---|---|---|---|---|---|---|---|---|
| Filterbank | | TW | FIR | TW | FIR | TW | FIR | TW | FIR |
| | Clean | 88.2 | 90.8 | 88.2 | 90.8 | 88.2 | 90.8 | 88.2 | 90.8 |
| | 25 | 85.8 | 88.1 | 87.5 | 89.1 | 86.4 | 87.5 | 87.8 | 90.0 |
| | 20 | 76.7 | 80.9 | 82.4 | 84.9 | 79.3 | 80.6 | 87.9 | 90.1 |
| S | 15 | 63.9 | 69.2 | 66.8 | 73.5 | 64.4 | 65.3 | 88.5 | 90.3 |
| N | 10 | 43.4 | 54.7 | 45.8 | 57.6 | 44.7 | 48.2 | 87.9 | 90.7 |
| R | 5 | 24.2 | 37.7 | 27.0 | 36.9 | 20.9 | 26.0 | 87.5 | 90.3 |
| (dB) | 0 | - | - | 12.7 | 18.3 | 8.9 | 10.4 | 84.8 | 88.1 |
| | -5 | - | - | - | - | - | - | 76.9 | 81.1 |
| | 10 | - | - | - | - | - | - | 63.7 | 64.8 |

(b) Results of MLP recognizer

| Noise | | WGN | | FAC | | MOP | | CAR | |
|---|---|---|---|---|---|---|---|---|---|
| Filterbank | | TW | FIR | TW | FIR | TW | FIR | TW | FIR |
| | Clean | 97.8 | 98.5 | 97.8 | 98.5 | 97.8 | 98.5 | 97.8 | 98.5 |
| | 25 | 96.2 | 97.2 | 97.0 | 97.3 | 97.2 | 97.6 | 97.7 | 98.5 |
| | 20 | 93.3 | 93.8 | 95.3 | 95.8 | 95.9 | 95.9 | 97.4 | 98.4 |
| S | 15 | 82.6 | 87.0 | 90.8 | 92.4 | 89.4 | 90.2 | 97.6 | 98.5 |
| N | 10 | 60.2 | 73.3 | 77.2 | 81.0 | 70.8 | 74.3 | 97.2 | 98.1 |
| R | 5 | 34.4 | 53.9 | 54.2 | 61.8 | 44.3 | 48.7 | 97.3 | 97.8 |
| (dB) | 0 | | | 29.0 | 35.5 | 21.1 | 22.2 | 96.5 | 97.1 |
| | -5 | | | | | | | 95.4 | 95.7 |
| | -10 | | | | | | | 89.9 | 90.5 |

$$F = A(10^{ax} - 1) \tag{5}$$

where $F$ is frequency in Hz, $x$ is the normalized distance along the basilar membrane with value from 0 to 1. The appropriate constants for the human cochlea, $A = 165.4$ and $a = 2.1$, are used in this paper. And the bandwidths are set to be proportional to the equivalent rectangular bandwidth (ERB) [25]. ERB is the bandwidth of an hypothetical rectangular filter, and is represented as the quadratic fit as a function of the center frequency of the auditory filter

$$\mathrm{ERB} = 6.23F^2 + 93.39F + 28.52 \tag{6}$$

where $F$ is frequency in kHz [25]. Further, the maximum number of tabs is limited to 100 for appropriate level of hardware implementations, and the characteristics of several lower frequency channels are sacrificed by the limitation as shown in Fig. 4.

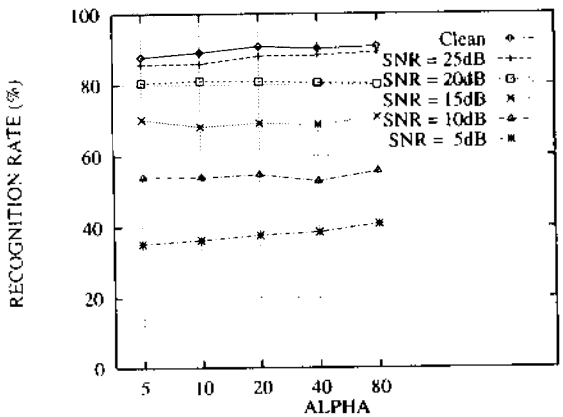
Table 2 summarizes recognition rates of the ZCPA obtained using the TW filters and FIR filters. Results of HMM recognizer are shown in (a), and those of MLP recognizer in (b). WGN, FAC, MOP, and CAR denote white Gaussian noise, factory noise, military operations room noise, and car noise, respectively. Even though TW filters are designed to mimic neural tuning curve shapes in detail, recognition rate obtained by hamming filters is higher than that obtained by TW filters regardless of the types of noise and SNR, on the contrary. As a result, the shape of the filter does not seem to be critical for recognition performance, which is in agreement with the result of [4]. And the critical part of the auditory model is the neural transduction stage. Thus it is sufficient to use FIR filterbank if one considers digital hardware implementations of the ZCPA. Moreover, recognition rate of MLP system is much higher than that of HMM system. The reason may be as follows. HMM is trained with maximum likelihood estimation (MLE) by which expectation value for samples of its own class given the model is maximized. If the topology and assumptions associated with the model are correct, the resulting recognition system is optimal classifier. However, there is no guarantee that the model is correct. Further, the first-order Markov assumption may not fit for the real situations. On the contrary, the MLP recognizer is trained with all samples of all classes, so that the decision boundaries between classes are formed as hyperplanes.

Since the training scheme of MLP system is based on decision boundary between all classes, the classification capability of MLP is much higher than HMM. Further, all contextual cues are included in the input pattern in MLP recognition system, since the whole word is considered as one pattern.

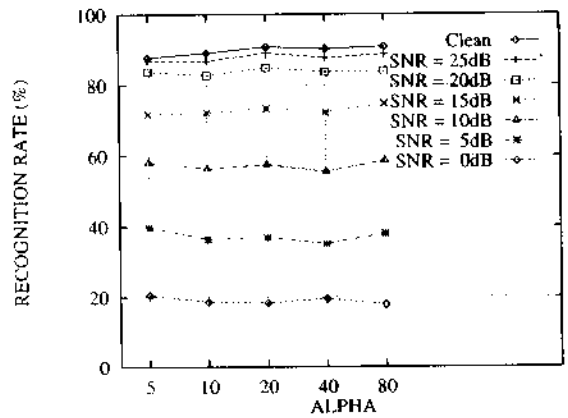
### 4.2 Sensitivity to Input Gain

Empirically determined saturating nonlinearity of the ZCPA, $g(\,)$ in Eq.(2), is $\log(1+\alpha x)$, where $x$ denote the peak amplitude and $\alpha$, which is set to 20, determines the slope of the nonlinear function. Unlike conventional feature extraction algorithms such as LPC cepstrum or MFCC, the effect of changes in microphone input gain to the output of the ZCPA is not additive. Actually, the input gain term is appeared to be additive at $c_0$ of MFCC. However, the effect of variations in microphone input gain to the output of the ZCPA depends on the quantity, $\alpha x$. If $\alpha x$ is sufficiently large, then $\log(1+\alpha x) \approx \log(\alpha)+\log(x)$ and the gain term in the input signal is separated as an additive term at model output. However, if $x$ is small such that the above approximation does not hold, the gain term is not additive any more. Since final value of each frequency bin is computed by considering signals of several neighboring channels which are associated with the frequency of that bin, it can not be said that the effect of gain term is additive. Actual investigations of the output of the model for speech signals reveal the fact that the gain term is not additive at the output of the model. Thus, there is no guarantee that the model is not sensitive to the input gain. If so, the performance of the model may not be independent of the volume of the microphone or distance from microphone to speakers, and the reliability of the model may not be maintained. One solution is to normalize amplitude of the input signal by the maximum amplitude of the signal, for example. However, the whole utterance should be stored in memory for normalization, and it is almost impossible for practical applications where real-time processing is required.

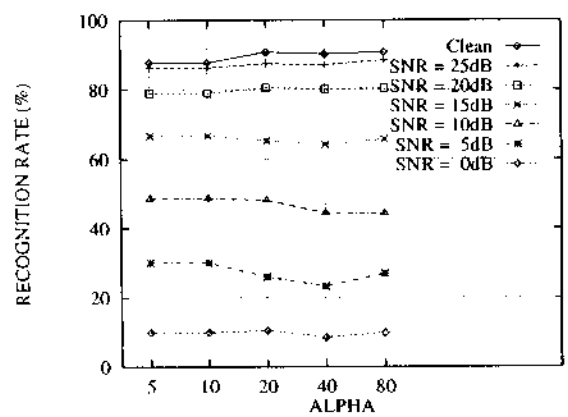
Thus it is necessary to investigate the performance of the model as the microphone input gain is varied. On the other hand, this problem is concerned with the slope of saturating nonlinearity of the ZCPA model since the change in input gain can be represented as the slope of saturating nonlinearity. Fig. 5 summarizes recognition rates of the ZCPA as $\alpha$ is varied under several types of noisy conditions, where HMM recognition system is used. If one suppose that $\alpha$ is set to 20, then the condition when $\alpha$ is 5 is equal to the reduced microphone input gain by a quarter. And the increased microphone input gain by four times can be equal to the condition of $\alpha = 80$. As shown in Fig. 5, it can be hardly said that the performance of the ZCPA is sensitive to the slope of saturating nonlinearity, $\alpha$ Thus, the output of the ZCPA is not sensitive to input gain variations.



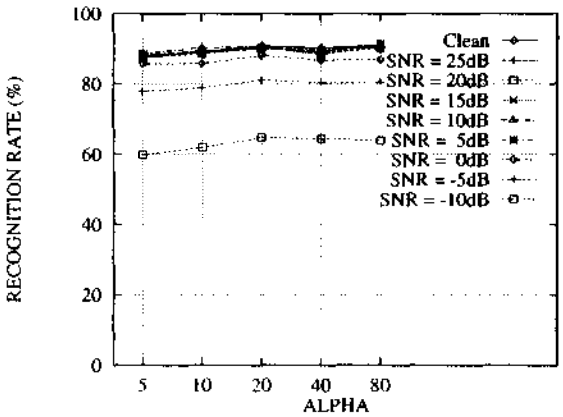

(a) White Gaussian noise




(b) Factory noise




(c) Military operations room noise




(d) Car noise

Figure 5. The effect of slope of nonlinear function in the ZCPA under various types of noisy conditions. HMM recognition system is used, and nonlinear function of the ZCPA used in this experiment is $\log(1+\mathrm{ALPHA}\cdot x)$.

## V. Conclusions

The ZCPA model based on human auditory periphery was proposed as a robust front-end for speech recognition systems in noisy environments. Although the ZCPA model is computationally efficient compared with other auditory models, it still requires much computation time than conventional speech processing techniques. Thus it is recommended to implement the developed auditory model with digital hardwares such as ASICs or digital signal processors for practical applications of ASR in noisy environments, and two factors of the ZCPA model are addressed in this paper. First, IIR filters which constitute the cochlear filterbank are replaced by FIR filters, which have less similarities to biological observations, to make it easy to implement with digital hardwares. Experimental results demonstrate that the detailed frequency response of the cochlear filters are not critical to the performance. Second, the sensitivity of the model output to the variations in microphone input gain is investigated in consideration of real-time processing of speech signals, and results in good reliability of the model even when the input gain is increased by 16 times as large. In most of evaluations, several real-world noises as well as white Gaussian noise are considered, and two recognizers are used to investigate the recognizer-independent reliability of the features. The MLP classifier shows much better recognition rates than the discrete HMM classifier in all cases.

## References

1. J. Allen, "Cochlear modeling," *IEEE ASSP Magazine*, pp. 3-29, 1985.
2. S. Seneff, "Pitch and spectral estimation of speech based on auditory synchrony model," *Proc. ICASSP*, pp. 36.2.1-36.2.4, 1984.
3. J. R. Cohen, "Application of an auditory model to speech recognition," *J. Acoust. Soc. America*, vol. 85, pp. 2623-2629, 1989.
4. O. Ghitza, "Auditory nerve representation as a basis for speech processing," *Advances in Speech Signal Processing* (S. Furui and M. M. Sondhi, eds.), pp. 453-485, New York: Marcel Dekker, 1992.
5. O. Ghitza, "Auditory models human performances in tasks related to speech coding and speech recognition," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 1, part II, pp. 115-132, 1994.
6. K. Wang and S. A. Shamma, "Self-normalization and noise-robustness in early auditory representations," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 3, pp. 421-435, 1994.
7. D. S. Kim, J. H. Jeong, J. W. Kim, and S. Y. Lee, "Feature extraction based on zero-crossings with peak amplitudes for robust speech recognition in noisy environments," *Proc. ICASSP*, (Atlanta, USA), pp. 61-64, May 1996.
8. D. S. Kim, S. Y. Lee, and R. M. Kil, "Auditory representations for robust speech recognition in noisy environments," *J. Acoust. Soc. Korea*, vol. 15, no. 5, pp. 90-98, 1996. (in Korean).
9. J. M. Kates, "A time-domain domain digital cochlear model," *IEEE Trans. Signal Processing*, vol. 39, no. 12, pp. 2573-2592, 1991.
10. M. B. Sachs and E. D. Young, "Encoding of steady state vowels in the auditory-nerve: representation in terms of discharge rate," *J. Acoust. Soc. America*, vol. 66, pp. 470-479, 1979.
11. E. D. Young and M. B. Sachs, "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers," *J. Acoust. Soc. America*, vol. 66, no. 5, pp. 1381-1403, 1979.
12. B. Delgutte and N. Y. S. Kiang, "Speech coding in the auditory nerve: I.," *J. Acoust. Soc. America*, vol. 75, no. 3, pp. 866-878, 1984.
13. B. Kedem, "Spectral analysis and discrimination by zero-crossings," *Proc. IEEE*, vol. 74, pp. 1477-1493, November 1986.
14. S. Y. Lee, K. H. Ahn, D. S. Kim, J. W. Cho, J. H. Jeong, J. W. Kim, S. O. Kwon, and R. M. Kil, "Voice command: A digital neuro-chip for robust speech recognition in real-world noisy enviromments (Invited talk)," *Proc. ICONIP*, (Hong Kong), pp. 283-287, Sep. 1996.
15. A. Varga and H. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an

experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247-251, 1993.

16. H. Steeneken and F. Geurtsen, "Description of the RSG-10 noise database," Tech. Rep., TNO Institute for Perception, 1990.
17. I. Lecomte, M. Lever, J. Boudy, and A. Tassy, "Car noise processing for speech input," *Proc. ICASSP*, pp. 512-515, 1989.
18. H. F. Silverman and N. R. Dixon, "State constrained dynamic programming (SCDP) for discrete utterance recognition," *Proc. ICASSP*, pp. 169-172, 1980.
19. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communications*, vol. COM-28, pp. 84-95, January 1980.
20. D. S. Kim and S. Y. Lee, "Intelligent judge neural network for speech recognition," *Neural Processing Letters*, vol. 1, no. 1, pp. 17-29, 1994.
21. D. S. Kim, K. W. Hwang, and S. Y. Lee, "Intelligent judge neural network for speaker independent isolated word recognition," *Proc. WCNN*, (San Diego, USA), pp. IV 577-582, Jun. 1994.
22. D. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Internal Representation by Error Propagation*, vol. 1. MIT Press, 1986.
23. E. Zwicker and E. Terhart, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," *J. Acoust. Soc. America*, vol. 68, pp. 1523-1525, 1980.
24. D. Greenwood, "A cochlear frequency-position function for several species-29 years later," *J. Acoust. Soc. America*, vol. 87, no. 6, pp. 2592-2650, 1990.
25. B. C. J. Moore and B. R. Glasberg, "Suggested formula for calculating auditory-filter bandwidth and excitation patterns," *J. Acoust. Soc. America*, vol. 74, pp. 750-753, 1983.

▲Doh-Suk Kim

Doh-Suk Kim received the B.S. degree in electronics engineering from Hanyang University, Seoul, Korea, in 1991, and the M.S., and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology(KAIST) in 1993 and 1997, respectively.

He is currently a post-doctoral fellow in the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology(KAIST), Taejon, Korea. His research interests include auditory modeling, speech signal processing, robust speech recognition, and neural networks. He is a member of IEEE(Institute of Electrical and Electronics Engineering) and Acoustical Society of Korea.

▲Jae-Hoon Jeong

Jae-Hoon Jeong received the B.S. and M.S. degrees in electronics engineering from Korea Advanced Institute of Science and Technology(KAIST) in 1993 and 1995, respectively.

He is currently in the Ph.D course at the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology(KAIST), Taejon, Korea. His research interests include speech recognition, musical timber recognition, and neural networks.

▲Soo-Young Lee

Soo-Young Lee received the B.S. degree in Electronics from the Seoul National University, Seoul, Korea, in 1975, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science, in 1977, and the Ph.D. degree in Electrophysics from the Polytechnic Institute of New York(PINY) in 1984. From 1977 to 1980, he was a project engineer with the Taihan Engineering Co., Seoul, Korea. From 1980 to 1983, he was a Senior Research Fellow at the Microwave Research Institute, PINY, N. Y., USA. From 1983 to 1985, he served as a Staff/Senior Scientist at the General Physics Corp., MD, USA. After a short stay at the Argonne National Lab., USA, he joined the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Seoul, Korea. His current research interests include optical computing and neural networks. He has published or presented more than 80 papers in optical implementation of neural networks, neural network architectures and applications, and numerical simulation techniques for electromagnetics. He was the Guest Editor of the Special Issue on Neural Networks of the Proceedings of the KIEE, February 1989. He organized the Korea-USA Joint Workshop on Optical Neural Networks in 1990.

▲Rhee M. Kil

Rhee M. Kil received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea in 1979 and the M.S. and Ph.D. degrees in computer engineering from the University of Southern California, Los Angeles, U.S.A. in 1985 and 1991, respectively.

From 1979 to 1983 he was with Agency for Defense Development, Taejon, Korea where he was involved in the development of Infrared Imaging System. From 1987 to 1991 his research topic was concentrated on the theories and applications of connectionist models. His dissertation was on the learning algorithms of connectionist models and their applications to the nonlinear system

control. From 1991 to 1994, he was with Research Department in Electronics and Telecommunications Research Institute, Taejon, Korea. In 1994, he joined the Division of Basic Sciences in Korea Advanced Institute of Science and Technology, Taejon, Korea, as an assistant professor. His general research interests lie in the areas of pattern recognition, system identification, data coding, and nonlinear system control. His current interests focus on learning based on evolutionary computation and information representation.