A Comparison of Front-Ends for Robust Speech Recognition

*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, the performance of the ZCPA model is further improved by incorporating conventional speech processing techniques into the model output. Spectral and cepstral representations of the ZCPA model output are compared, and the incorporation of dynamic features with several different lengths of time-derivative window are evaluated. Also, comparative evaluations with other front-ends in real-world noisy environments are performed, and result in the superiority of the ZCPA model.

I. Introduction

Automatic speech recognition (ASR) is one of the leading technologies serving as a man-machine interface for real-world applications. In general, the performance of an ASR system is usually degraded when there exist environmental mismatches between training and test phases. One type of mismatch in real environments is the various kinds of background noises which affect the feature extraction stage in an ASR system. In this sense, the front-end for robust speech recognition requires to reduce redundancy and variablity as well as the ability to capture important cues of speech signals, even in noisy environments. One of the most widely used feature representations is cepstral coefficients derived from linear predictive coding (LPC) in which the speech signal is assumed to be the output of the all-pole linear filter simulating the vocal tract of a human being. The ASR systems with LPC-derived cepstrum work well in clean environments, but speech recognition performance is severely degraded in noisy environments.

On the other hand, modeling of the speech perception processes may be more natural for ASR than that of the speech production processes, and there have been many researches devoted to the modeling functional roles of the peripheral auditory systems [1], [2], [3], [4]. Seneff [2] suggested a generalized synchrony detector (GSD) to identify formant peaks and periodicities of the speech signal. Hunt and Lefebvre [5] performed recognition experiments on noisy speech using a dynamic time warping (DTW) recognizer, and showed noise-robustness of the GSD. Perceptual linear prediction (PLP) analysis method [6], [7] is a perception-based technique in which the speech spectrum is transformed to the auditory spectrum by several perceptually motivated relationships before performing conventional linear prediction (LP) analysis. The robustness of the PLP analysis to additive noise was reported in [8]. Subband-Autocorrelation (SBCOR) analysis technique [9] was suggested to extract periodicities present in speech signals by computing autocorrelation coeffcients of subband signals at specifc time-lags, and was shown to outperform the smoothed group delay spectrum for speech recognition tasks under noisy environments.

The superiority of the auditory modeling is more prominent for nonstationary real-world noisy environments where conventional techniques such as spectral subtraction [10] or short-term Wiener fltering [11] may not operate well since they are based on the estimation of noise spectrum and the noise spectrum may change severely over time. Although computational auditory models have been shown to outerform conventional signal processing techniques, especially in noisy environments, modeling peripheral auditory systems is still a diffcult problem. First, studying an auditory model requires interdisciplinary research, including physiology, psychoacoustics, physics, and electrical engineeing. Second, little is known about the exact mechanism of the auditory periphery for detailed construction of the model. Since the auditory model usually involves multistage nonlinear transformations, analytical treatments are intractable, and most auditory models rely heavily on experiments, even though there have been some efforts to analyze auditory models [12], [13], [14], [15]. Furthermore, auditory models require careful determination of many free parameters and much computation time, which make it duffcult for them to be widely used in speech recognition systems.

Zero-crossings with peak amplitudes (ZCPA) model was proposed as a robust front-end for ASR in noisy

* HCI Lab., SAIT
** Department of Electrical Engineering, KAIST
*** Division of Basic Science, KAIST

Manuscript Received : October 9, 1996.

environments [16]. The ZCPA is very simplifed 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) [17] when speech is corrupted by white Gaussian noise.

In this paper, the performance of the ZCPA nodel is further improved by incorporating conventional speech processing techniques into the model. Also, comparative evaluations of several front-ends in real-world noisy environments are presented. This paper is organized as follows. Section II presents a brief review of the ZCPA model for robust feature extraction. In section III the data base, noise material, and speech recognizer used in this papet are described. Performance improvements of the ZCPA model are provided in section IV, followed by comparisons with other front-ends in section V and conclusions in section VI.

II. ZCPA Analysis

The ZCPA model consists of a bank of bandpass cochlear filters and nonlinear stages at the output of each cochlear filter. 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 the adaptive feedbacks [18]. 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 [19], [20], [21]. In the ZCPA model, a synchronous neural firing is simulated as the upwardgoing zero-crossing event of the signal at the output of each bandpass filter, 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. The operation of the ZCPA is significantly different from conventional signal processing techniques. The temporal frequency information of one period of the signal is obtained by zero-crossing intervals, and the temporal intensity information is also incorporated by a peak detector following a saturating nonlinearity. These temporal frequency and intensity information are then accumulated to obtain the final output.

III. Data Base and Recognition Systems

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 [22]. 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 [23], were added to the test data sets at various SNRs for test evaluations in real situations.

The integrated speech recognition system under develpment adopts the neutal network classifier preceded by trace-segmentation algorithm [24] for improved recogiton performance. However, the most widely used recogzer is based on hiddn Markov Markov morel (HMM). Thus, both discrete HMM speech recognizer and multilayer perceptron (MLP) recognizer preceded by trace-segmentation are used to investigate the recognizer independent reliability of features. There have been a lot of schemes proposed to apply neural networks to speech recognition, and static approach utilzing an MLP showed better performance than dynamic approach at least for isolated word recognition tasks [25]. MLP is trained by using error back propagation algorithm [26] 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 same as the normalized time frames, N, which is 64, multiplied by the number of components of a feature vector at one time frame. For HMM recognizer, word-level discrete density HMM construction is performed and 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 oneyllable word or eight for multi-syllable word. Each HMM is iterativcly trained with Baum-Welch algorithm based on maximum likelihood estimation (MLE). The codebook is trained with training data in iterative

manner [27], and the size of codebook is set to be 256.

IV. Toward Performance Improvements

A. Spectral versus Cepstral Representations of the ZCPA

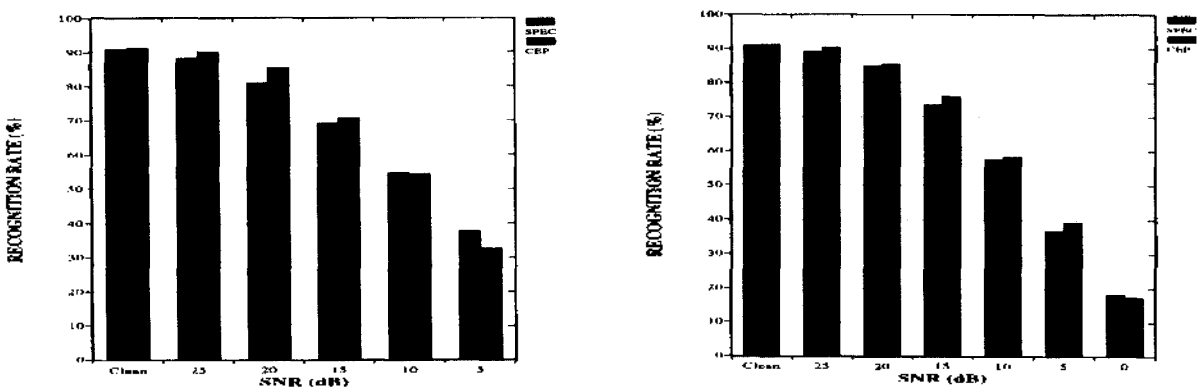
Human voice consists of spectral fine structures caused by pseudo-periodic glottal source and spectral envelope which represents resonance characteristics of the vocal tract. What is required for speech recognition is the latter, since the positions, and sizes of the vocal apparatus in vocal tract are changed accord-ing to the utterance. Spectral envelope and spectral fine structure of speech can be separated in cepstral domain because they are additive in log-spectral domain, and cepstrum is defined as the inverse Fourier transform of the logarithm of magnitude spectrum. This may be one of the reasons of prevailing of cepstral representations in the area of speech recognition. Also, it is known that the input feature vector is made somewhat uncorrelated by the inverse cosine transform - inverse Fourier transform is reduced to the inverse cosine transform in this case since the magnitude spectrum is symmetric about zero.

The output of the ZCPA can be considered as pseudo-log spectrum if the logarithmic function is used for the saturating nonlinearity. Thus cepstral representations can be obtained from the ZCPA spectrum. Let us denote $y(m, i)$ by the ZCPA output at time m, for $i=1, \ldots, N$, where N denotes the number of frequency bins. If the frequency bins of the ZCPA are composed according to the baek or mel scale, cepstral coefficients at time m in the warped frequency scale, $\tilde{c}(m, l)$, can be obtained using

$$\tilde{c}(m, l) = \sum_{i=1}^{N} y(m, i) \cos\left[l(i-\frac{1}{2})\frac{\pi}{N}\right], \quad 1 \le l \le L, \quad (1)$$

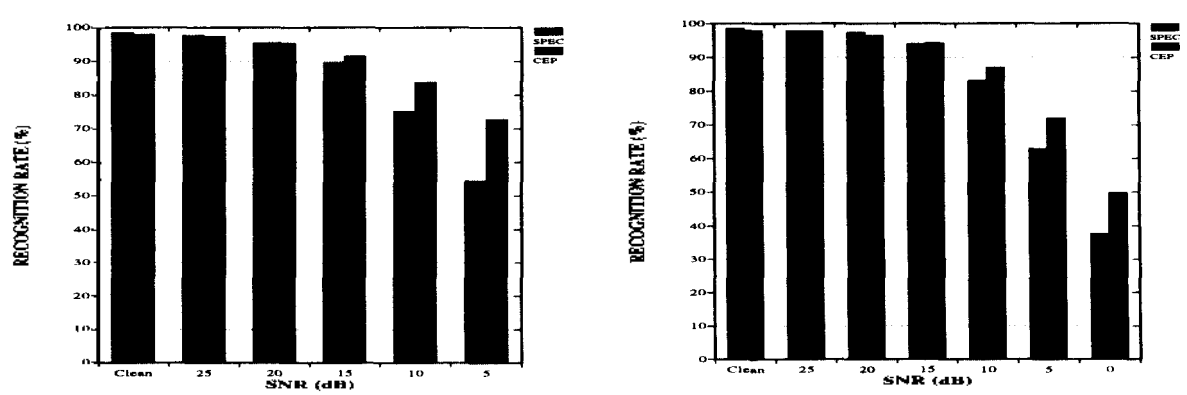
where L is the desired length of the cepstrum [28].

Recognition rates of ZCPA spectrum and ZCPA cepstrum obtained by the HMM recognizer are shown in



(a) White Gaussian Noise (b) Factory Noise

Figure 1. Comparison of ZCPA spectrum and ZCPA cepstrum under various types of noisy conditions. HMM recognizer is used.



(a) White Gaussian Noise (b) Factory Noise

Figure 2. Comparison of ZCPA spectrum and ZCPA cepstrum under various types of noisy conditions. MLP recognizer is used.

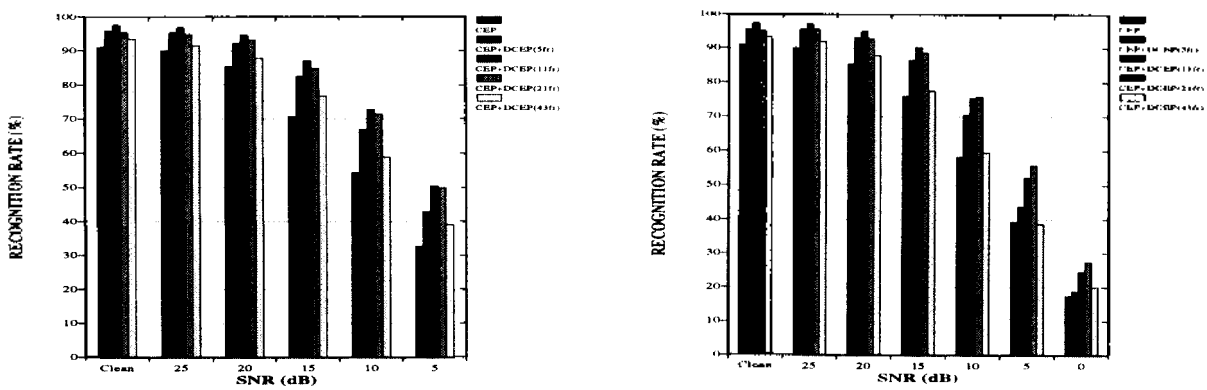
Fig. 1. ZCPA cepstrum is extracted from ZCPA spectrum via Eq.(1). The number of coefficients is 16 for ZCPA spectrum, and 12 for ZCPA cepstrum, respectively. Although less number of coefficients are used in cepstrum, performance of the ZCPA cepstrum is similar to or higher than the ZCPA spectrum, except for some lower SNR conditions of the military operationsroom noise case. Fig. 2 shows the results of the same experiments as Fig. 1 when the MLP speech recognizer is used. Recognition rates of the MLP recognizer are higher than those of HMM recognizer for most cases. Further, the degree of superiority of ZCPA cepstrum to ZCPA spectrum is more clear in the MLP recognition system. Thus it can be concluded that the cepstral representation of the ZCPA model is more useful than spectral representation.

B. Incorporation of Dynamic Features

It is well known that the transition of spectral contents through time plays an important role in human perception of speech [29], and it is common to incorporate dynamic properties of speech into speech recognition systems by augmenting dynamics such as delta and delta-delta features to static features for improved recognition accuracy, not only in clean but also in noisy conditions. Computing delta features is equivalent to an FIR filtering, which rejects lower modulation frequency variations of the speech parameters. If speech and non-speech components occupy different ranges in the parameter domain, they can be separated by filtering in the parameter domain. Actually, the channel characteristics occupy the lower range of the modulation frequency in the logarithmic domain, and lots of techniques, such as cepstral mean normalization (CMN) [30], and RASTA processing and its several variants [31],[32], have been suggested to separate channel effects from speech parameters.

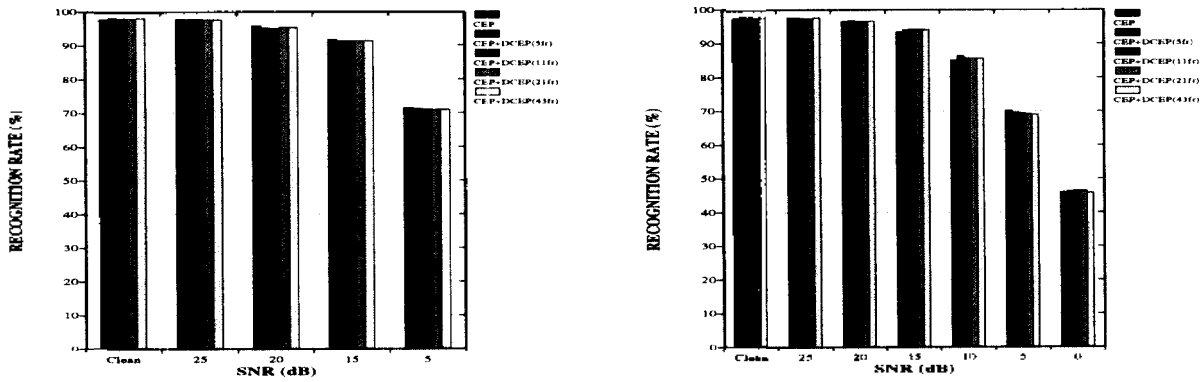
However, it was reported that contribution of dynamic features of the EIH to the performance improvements is much smaller than that of mel-frequency cepstral coefficients (MFCC) [33]. This may be due to the fact that the length of the time-window is channel dependent in the EIH, i.e., it varies inversely with the characteristic frequency of the channel. For example, the length of the time-window at the channel with the low-est characteristic frequency spans up to 50 msec, which is much longer length when compared with the frame rate of about 10 msec. Thus, appropriate dynamic features cannot be obtained with the derivative window of 50 msec duration, which is used in [33]. Even though variable length of the derivative window [34], [35] may be applied to the computation of the delta features of the EIH and ZCPA, it is beyond the scope of this study. Instead, we tried several fixed derivative window lengths: 50.8 msec (5 frames), 111.7 msec (11 frames), 213.3 msec (21 frames), and 436.8 msec (43 frames). As the length of the window is increased, the higher cutoff frequency is decreased to reject higher modulation frequency components.

Fig. 3 summarizes recognition results of HMM recognition system as the derivative window length is varied. ZCPA cepstrum and delta-cepstrum are used as feature vector, and two independent codebooks are constructed for cepstrum and delta-cepstrum, respectively, under the assumption that the static and dynamic features are independent each other. Recognition rate obtained by using static feature (cepstrum) only is depicted as the leftmost bar (CEP) at each plot to indicate the improvements incurred by combination of static and dynamic features. It is clear that contribution of dynamic features is poor if the derivative window length is too short (5 frames) or too long (43 frames). And derivative window length of 11 frames shows the best performance on average for the



(a) White Gaussian Noise (b) Factory Noise

Figure 3. Recognition rate (%) obtained with ZCPA cepstrum and ZCPA cepsrum augmented by delta-cepstrum with various derivative window lengths under various types of noisy conditions. HMM recognizer is used.



(a) White Gaussian Noise

(b) Factory Noise

Figure 4. Recognition rate (%) obtained with ZCPA cepstrum and ZCPA cepstrum augmented by delta-cepstrum with various derivative window lengths under various types of noisy conditions. MLP recognizer is used.

ZCPA. However, the trend of MLP recognizer is significantly different from that of HMM recognizer. There is no performance improvement by utilizing dynamic features for MLP recognizer. Further, different time-derivative window lengths do not make any differences in recognition rates. This is because the delta features obtained by a linear combination of static features can be represented internally in the hidden representations of the MLP recognizer with static features only. Thus, it is sufficient to use only static features for MLP recognition systems.

Even though somewhat longer window lengths are preferable for ZCPA with HMM recognizer, it is not sufficient to conclude that this is the best. As mentioned before, different lengths of bandpass signals are considered in computing the ZCPA output according to the characteristic frequency of the channel while the frame rate is fixed. Thus it may be possible to apply different lengths of time-derivative window according to the characteristic frequency of the channeln [34], [35], and it remains as future works.

V. Summary of Results and Comparison with Other Front-ends

In this section, the improved performance of the ZCPA cepstrum (ZCPAC) is compared with other frontends including LPC cepstrum (LPCC), mel-frequency cpstral coefficients (MFCC), subband autocorrelation (SBCOR), perceptual linear prediction (PLP), and EIH cepstrum (EIHC) in various types of noisy environments. Table 1 summarizes comparison of several features concatenated by time-derivative versions of them. Recognition rates are obtained by HMM recognizer. The window length of time-derivative features is 11 frames, i.e., 111.7 msec for all front-ends. For LPCC, speech signal is first multiplied by hamming window of 20.3 msec duration every 10.15 msec, and 8 LPC coefficients and 12 cepstral coefficients are obtained successively. For MFCC, 16 mel-scale triangular bandpass filters are used in frequency domain to obtain 12 coefficients. To calculate SBCOR, 16 hamming bandpass filters, which are also used in both the ZCPA and the EIH, are used in frequency domain. In PLP processing, 16 critical-band filters are used and LPC order is set to 8. Performance of the several EIH cepstrum were evaluated by varying the number of levels and level values, and only the best case among them is shown. (7 level crossing detectors are used.)

On clean speech, the recognition of all front-ends are similar to each other. As the noise level increases, the recognition rate of the ZCPA becomes higher than that of the other front-ends for all kinds of noises. The usefulness of the ZCPA in noisy environments is maximum when speech data is corrupted by white Gaussian noise. However, for speech data corrupted by real-world noises, the differences in recognition rate between the ZCPA and the others are reduced compared with the other kinds of noisy environments. Also, the performance of SBCOR is higher than that of PLP below 20dB SNR when speech data is corrupted by white Gaussian noise. Howe- ver, PLP outerforms SBCOR under real-world noisy environments on the contrary.

Summary of recognition rates of MLP recognizer is shown in Table 2. Since the incorporation of time-derivative features does not improve recognition accuracy in MLP system, results obtained with only static features are summarized. ZCPA cepstrum outperforms all the other features, too. Further, MLP recognizer outperforms HMM recognition systems especially at noisy conditions.

Table 1. Comparison of recognition rates (%) obtaied using several features augmented by time-derivative features under various types of noisy environments where HMM recognition system is used.

(a) White Gaussian noise

SNR(dB)	LPCC	MFCC	SBCOR	PLP	EIHC	ZCPAC
Clean	94.4	97.5	96.4	98.2	97.4	97.6
25	74.4	92.1	94.1	96.0	97.0	96.9
20	38.5	74.0	90.0	85.9	93.5	94.6
15	12.0	38.3	72.7	55.5	84.3	87.0
10	4.2	12.1	43.2	25.7	66.3	72.7
5	2.6	4.9	16.9	7.7	45.3	50.5

(b) Factory noise

SNR(dB)	LPCC	MFCC	SBCOR	PLP	EIHC	ZCPAC
Clean	94.4	97.5	96.4	98.2	97.4	97.6
25	91.2	95.9	94.3	97.1	96.6	97.1
20	79.0	90.5	91.3	95.0	94.4	95.1
15	52.6	67.3	77.8	81.6	86.7	90.3
10	20.7	33.7	50.6	52.4	70.0	75.4
5	7.9	10.5	22.2	25.7	46.3	52.3

(c) Military operations room noises

SNR(dB)	LPCC	MFCC	SBCOP	PLP	EIHC	ZCPAC
Clean	94.4	97.5	96.4	98.2	97.4	97.6
25	91.5	96.1	95.0	96.9	96.2	96.7
20	81.4	89.2	91.3	94.5	92.2	94.0
15	53.5	70.9	76.9	82.4	79.8	85.8
10	23.8	39.4	47.6	55.3	61.8	68.0
5	7.7	16.0	23.1	29.0	30.6	37.3

(d) Car noise

SNR(dB)	LPCC	MFCC	SBCOR	PLP	EIHC	ZCPAC
Clean	94.4	97.5	96.4	98.2	97.4	97.6
10	93.6	95.7	95.5	98.1	97.2	97.6
5	93.2	95.5	95.1	97.3	97.5	97.2
0	91.5	94.6	94.5	96.3	96.1	96.6
-5	84.8	90.7	89.6	92.6	93.5	93.6
-10	68.2	78.2	75.8	72.9	84.2	86.9

Table 2. Comparison of recognition rates (%) obtained using several features under various types of noisy environments where MLP recognition system is used. Only the results obtained with static features are shown.

(a) White Gaussian noise

SNR(dB)	LPCC	MFCC	SBCOR	PLP	EIHC	ZCPAC
Clean	95.0	97.0	96.0	98.4	98.3	97.8
25	79.0	93.0	93.8	96.0	97.4	97.6
20	50.7	73.2	85.4	85.6	95.3	95.7
15	24.9	42.8	69.8	61.7	88.1	91.7
10	10.2	21.7	46.6	32.9	77.0	83.1
5	5.1	13.9	26.3	15.8	62.7	71.5

(b) Factorky noise

SNR(dB)	LPCC	MFCC	SBCOR	PLP	EIHC	ZCPAC
Clean	95.0	97.0	96.0	98.4	98.3	97.8
25	91.0	95.5	93.7	97.1	97.7	97.6
20	80.3	86.8	88.8	92.1	95.6	96.6
15	56.3	64.8	74.8	78.7	90.5	93.4
10	29.9	32.1	49.8	51.6	79.0	85.1
5	10.6	12.0	25.0	27.5	63.0	70.2

(c) military operations room noise

SNR(dB)	LPCC	MFCC	SBCOR	PLP	EIHC	ZCPAC
Clean	95.0	97.0	96.0	98.4	98.3	97.8
25	93.0	96.3	94.9	98.2	97.8	97.9
20	82.5	91.0	89.6	93.2	96.2	96.7
15	61.3	73.3	71.3	79.8	90.8	93.2
10	39.6	41.0	41.0	55.0	77.1	83.4
5	17.0	20.3	20.7	32.4	55.5	60.4

(d) Car noise

SNR(dB)	LPCC	MFCC	SBCOR	PLP	EIHC	ZCPAC
Clean	95.0	97.0	96.0	98.4	98.3	97.8
10	95.0	97.2	95.0	98.1	98.4	98.1
5	94.0	94.9	94.1	95.5	97.9	97.7
0	92.8	97.0	91.2	91.8	97.2	97.1
-5	86.8	85.9	84.4	83.4	96.6	96.6
-10	66.4	62.6	66.3	69.4	91.0	93.0

VI. Conclusions

The ZCPA model based on human auditory periphery was proposed as a robust front-end for speech recognition systems in noisy environments, and shown to be robust to additive white Gaussian noise than both LPCC and the EIH in our previous work. In this paper, performance of the ZCPA model is further improved and evaluated in several real-world noisy environments. For further improvements in recognition performance, several conventional speech processing techniques are also incorporated into the ZCPA model. Spectral representation of the features are extended into cepstral representations, which demonstrates better recognition rates in general with less number of coefficients. Also, several different lengths of time have been tried to obtain good time-derivative features of the developed auditory model. Relatively longer length in the time-derivative window results in better recognition accuracy with the HMM classifier. However, it does not make much differences with the MLP classifier. The MLP classifier shows much better recognition rates than the discrete HMM classifier in all cases. Also, comparative evaluations of the ZCPA model with several feature extraction methods demonstrate the robustness of the ZCPA model in noisy environments.

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," in *proc. Int.Conf. on Acoust., Speech,signal Processing*, 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 models and 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.
5. M. Hunt and C. Lefebvre, "Speech recognition using a cochlear nodel," in *proc. Int. Conf. on Acoust., Speech, Signal Processing*, pp. 37.7.1-37.7.4, 1986.
6. H. Hermansky, "Perdptual linear predictive(PLP) analysis of speech," *J. Acoust. Soc. America*, vol. 87, no. 4, pp. 1738-1752, 1990.
7. J. C. Junqua, H. Wakita, and H. Hermansky, "Evaluation and optimization of perceptually-based ASR front-end," *IEEE Trans. speech and Audio Processing*, vol. 1, no. 1, pp. 39-48, 1993.
8. J. C. Junqua, *Toward robustness in isolated-word automatic Speech recognition*. PhD thesis, University of Nancy 1, STL Monograph, 1989.
9. S. Kajita and F. Itakura, "Robust feature extraction using SBCOR analysis," in *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, pp. 421-424, 1995.
10. S. F. Boll, "Suppression of noise in speech using the SABER method," in *Proc. Int. on Acoust., Speech, Signal Processing*, pp. 606-609, 1978.
11. J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 197-210, 1978.
12. M. Hunt and C. Lefebvre, "Speech recognition using an auditory model with pitch-synchronous analysis," in *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, pp. 20.5.1-20.5.4, 1987.
13. X. Yang, K. Wang, and S. A. Shamma, "Auditory representations of acoustic signals," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 824-839, 1992.
14. 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.
15. A. Morris, J.L. Schwartz, and P. Escudier, "An information theoretic investigation into the distribution of phonetic information across the auditory spectrogram," *Computer Speech and Language*, vol. 2, pp. 121-136, 1993.
16. D.-S. Kim, J.-H. J.-W. Kim, and S.-Y. Lee, "Feature extraction based on zero-crossings with peak amplitudes for robust speech recognition in noisy environments," in *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, (Atlanta, USA), pp. 61-64, May 1996.
17. O. Ghitza, "Auditory nerve representation as a basis for speech processing," in *Advances in Speech Signal Processing* (S. Furui and M. M. Sondhi, eds.), pp. 453-485, New York : Marcel Dekker, 1992.
18. J. M. Kates, "A time-domain digital cochlear model," *IEEE Trans. Signal Processing*, Vol. 39, no. 12, pp. 2573-2592, 1991.
19. 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.
20. E. D. Yonug 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.
21. 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.
22. 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 environments (Invited talk)," in *proc. Int. Conf. on Neural Information Processing*, (Hong Kong), pp. 283-287, Sep. 1996.
23. 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.

24. H. F. Silverman and N. R. Dixon, "State constrained dynamic programming (SCDP) for discrete utterance recognition," in *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, pp. 169-172, 1980.
25. D.-S. Kim and S.-Y. Lee, "Intelligent judge neural network for speech recognition," *Neural Processing Letters*, vol. 1, no. 1, pp. 17-20, 1994.
26. D Rumelhart, G.E. Hinton, and R. J. Williams, *Learning Internal Representation by Error Propagation*, vol. 1. MIT Press, 1986.
27. 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.
28. L. R. Rabiner and S. H. Juang, *Fundamentals of speech recognition*. Prentice-Hall Inc., 1993.
29. S. Handel, Listening: *An introduction to the perception of auditory events*. The MIT Press, 1993.
30. J.-C. Junqua, D. Fohr, J.-F. Mari, T. Applebaum, and B. Hanson, "Time derivatives, cepstral normalization, and spectral parameter filtering for continuously spelled names over the telephone," in *Proc. Europ. Conf. on Speech Communication and Technology*, 1995.
31. H. Hermansky, N. Morgan, A. Bayya, and P. Kohn, "Compensation for the effect of the communication channel in auditory-like analysis of speech (RASTA-PLP)," in *Proc. Europ. Conf. on Speech Communication and Technology*, pp. 1367-1370, 1991.
32. H. Hermansky, E. Wan, and C. Avendano, "Speech enhancement baed on temporal processing," in *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, pp. 405-408, 1995.
33. S. Sandhu and O. Ghitza, "A comparative study of mel cepstra and EIH for phone classification under adverse conditions," in *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, (Detroit, USA), pp. 409-412, 1995.
34. J. Smolders and D.V. Compernolle, "In search for the relevant parameters for speaker independent speech recognition," in *Proc. Int. Conf. on Acoust., Speech, Signal Processing*, Vol. II, pp. 684-687, 1993.
35. K. Aikawa, H. Singer, H. Kawahara, snd Y. Tohkura, "A dynammic cepstrum incorporating time-frequency masking and its application to continuous speech recognition," in *Proc. Int.Conf. on Acoust., Speech, Signal Processong*, Vol. II, pp. 668-671, 1993.

▲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 engieering from Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea, in 1993 and 1997, respectively.

From 1993 to 1996, he was a part-time researcher with the Systems Engineering Research Institute, Taejon, Korea. He served as a post-doctoral Fellow at KAIST from March 1997 to September 1997.

From November 1997 to October 1998, he was with the Acoustics and Speech Research Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ, USA, as a Post-doctoral Member of Technical Staff. He is now with the Human & Computer Interaction Lab., Samsung Advanced Institute of Technology (SAIT), Suwon, Korea. His research interests include auditory psychophysics, speech recognition, speech coding, and objective speech quality assessment.

▲Jae-Hoon Jeong

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

He is currently a Ph.D. candidate 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., Columbia, MD, USA. After a short stay at the Argonne National Lab., Argonne, IL, USA, he joined the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Seoul, Korea. His current resesrch 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 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 Reserch 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.