Feature Extracting in the Presence of Environmental Noise, using Subband Adaptive Filtering

Salina Abdul Samad , Aini Hussain, Khairul Anuar Ishak and Ali O. Abid Noor
Department of Electrical, Electronic and System Engineering,
Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia, UKM,
43600 Bangi ,Selangor ,Malaysia
E-mail: salina@vlsi.eng.ukm.my, aini@vlsi.eng.ukm.my,nuarscc@vlsi.eng.ukm.my,ali511@vlsi.eng.ukm.my

Abstract

In this work, a new feature extracting method in noisy environments is proposed. The approach is based on subband decomposition of speech signals followed by adaptive filtering in the noisiest subbbands of speech. The speech decomposition is obtained using low complexity octave filter bank, while adaptive filtering is performed using the normalized least mean square algorithm. The performance of the new feature was evaluated for isolated word speech recognition in the presence of a car noise. The proposed method showed higher recognition accuracy than conventional methods in noisy environments.

1. Introduction

Speech feature extraction in the presence of noise has been a key focus in recent speech recognition researches [1]-[3].Environmental noise often effect feature extraction in speech recognition. Noisy environments like cars have an adverse effect on the performance of the Mel scale derived features [4]. The line spectral frequency (LSF) representation and cepstral coefficient representation have comparable performances for a general speech recognition system. Environmental noise such as car noise has low-pass characteristics which may degrade the performance of LSF or mel scaled cepstral coefficient (MELCEP) representations. It is observed that significant amount of spectral power of car noise is localized under 500 Hz [5]. Based on this fact the linear prediction analysis LP is performed in low and high frequency subbands [5]. This kind of frequency domain decomposition can be generalized to cases in which the noise is frequency localized. In this paper we propose a subband decomposition on a noisy signal, performing subband adaptive filtering to remove the low-pass noise. The feature extracting in subbands is based on octave filterbank followed by adaptive noise cancellation in the lower subbands. In many applications of noise cancellation, the changes in signal characteristics could be quite fast. This requires the utilization of adaptive algorithms, which converge rapidly. From this perspective, the normalized least mean square NLMS algorithm can be used to enhance the lower part of the speech spectrum. The method called octave cepestral with noise cancellation OCNC. Robustness of Parameters extracted using the new method is compared with some well known techniques such as LPCC, MFCC and MELCEP.

2. Methodology

Consider the arrangement shown in Figure1, the noisy signal *s(n)* is decomposed into sub-signals with the aid of analysis filterbank *H(z)*.This filterbank is an octave implementation which closely matches the frequency response of the human perceptual. In each level of the decomposition, a perfect reconstruction quadrature mirror QMF bank is used for signal splitting [6]. The lowpass prototype filter is given by

$$H_0(z) = \sum_{n=0}^{L-1} z^{-n} h_0(n) \tag{1}$$

According to the QMF bank, the highpass filter is given by

$$H_1(z) = H_0(-z) \tag{2}$$

The computational complexity of the filter bank can be reduced to a half of the direct implementation by the use of polyphase decomposition in each stage. The polyphase representation of the filter given in (1) is expressed as

$$H_0(z) = \sum_{k=0}^{1} z^{-k} F_k(z) \tag{3}$$

where $F_k(z)$ is the k[th] polyphase component of the prototype filter and given by

$$F_k(z) = \sum_{n=0}^{L-1} z^{-n} f_k(2n+k) \tag{4}$$

Another cost reduction can be achieved by shifting the down-samplers to the inputs of the filters utilizing the noble identities of multirate systems [6].

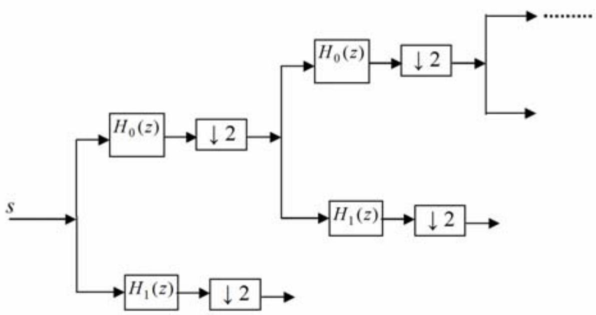

Figure 1**.** The analysis stage of the feature extracting algorithm

The noise cancellation model shown in Figure 2 is used in each of the three lower subbands. Adaptive process is performed using the normalized least mean square algorithm NLMS [7]. Thias algorithm is used to control a low order FIR filter. The adaptive process can be expressed by the following set of equations

$$\mathbf{w}_i(n) = \mathbf{w}_i(n-1) + \mu \mathbf{x}_i e_i(n) \tag{5}$$

$$e_i = \hat{s}_i - y_i \tag{6}$$

$$y_i = \mathbf{x}_i \mathbf{w}_i^T \tag{7}$$

Where $\mathbf{w}$ is weight coefficient vector, $\mathbf{x}$ is the input vector at time *n*, *y* is the output of the adaptive filter, *e* is the error signal, *i* is an index to represent the lower subbands *i=1,2,3* and μ is the adaptation step size. For the NLMS case μ is divided by the power of the input signal. This way, the algorithm will have a better tracking capability for non stationary signals, such as ambient noise. In addition to that, the subband splitting of the input signal will have a positive impact on the convergence speed of the adaptive filter , since the input spectrum will much whiter than the original signal.

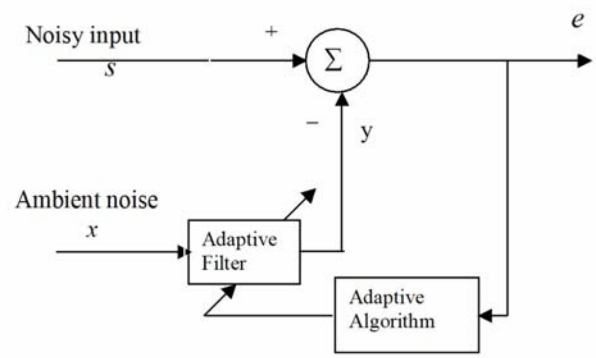

Figure 2. Noise cancellation model

Let the final subband signals are be denoted by s_k *for k=1,2,3,….M,*where *M* is the total number of subbands. We define the normalized power of the subband signals as

$$P(k) = \frac{1}{N_k} \sum_{n=1}^{N_k} \left| s_k(n) \right|^2 \tag{8}$$

N_k is the number of samples in the *k-th* band. The OCNC parameters, $OC(m)$, which form the feature vector is defined in a similar way to the Mel frequency cepestral MELCEP coefficients[8]

$$OC(m) = \sum_{k=1}^{M} \log(P(k)) \cos\left(\frac{m(k-0.5)}{M}\pi\right) \tag{9}$$

for $m=1,2\ldots,12$

3. Simulation results

A speech recognition system based on vector quantization VQ (template matching) is used with 8 bit codebook. The speech signal is sampled at 16 kHz and the so called car noise is upsampled to 16 kHz. The noisy speech is obtained with the car noise recording, for four levels of signal to noise ratios 0,-10,-20,-30,-40 dB, assuming that the noise is additive. Simulation studies are performed on isolated Malay digits from 0-9. The utterances of 60 speakers are used. The isolated word recognition system is trained with 36 speakers male and the performance evaluation is done with the remaining 24 speakers.

The filter bank structure of Figure 1 is applied to the speech signal (up to 4 levels) to achieve the subband decomposition. This filter bank is based on using 32-tap FIR filters after Johnston (type 32-D) [9]. The adaptive filter stage is performed using the NLMS algorithm, controlling a 16-tap FIR filter, with step size set to 0.01, Figure 3 shows initial adaptive filter test.

The window size is chosen as 20 *ms* (320samples) with an overlap of 10 *ms* (160 samples). The OCNC parameters are derived as in Equation (9) and the feature vector is constructed from these OCNC parameters. The performance of the OCNC is compared to three well known techniques, LPCC, MFCC and MELCEP. The results are depicted in Figures 4,5 and 6. The OCNC representation exhibits robust performance in the isolated word recognition application and it outperforms other techniques as these results proved.

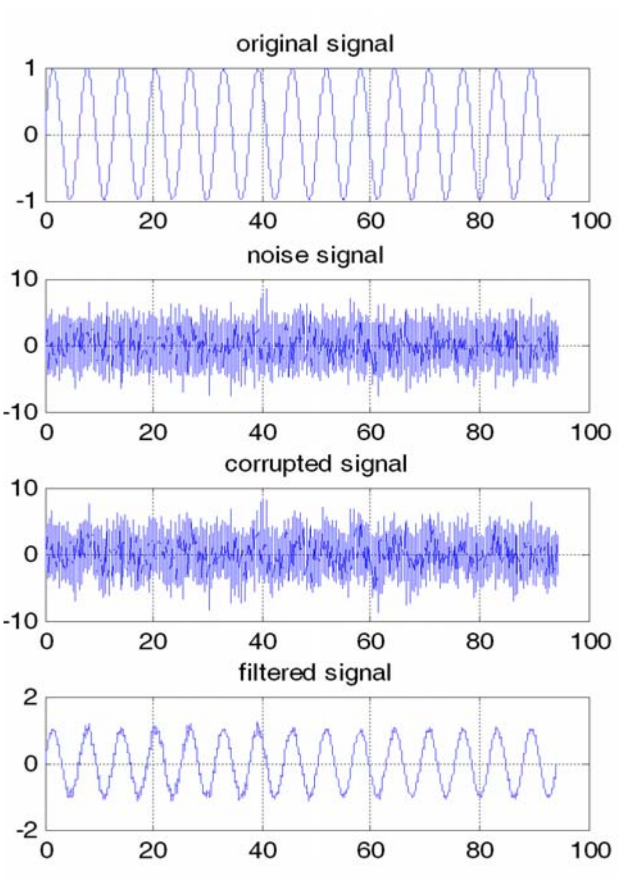

Figure 3. Adaptive filtering test

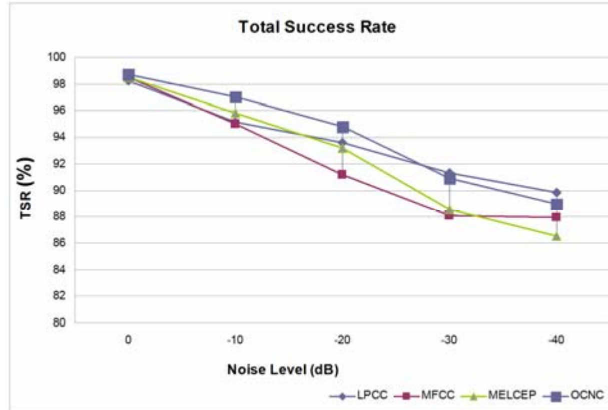

Figure 4. Performance comparisons in terms of total success rate

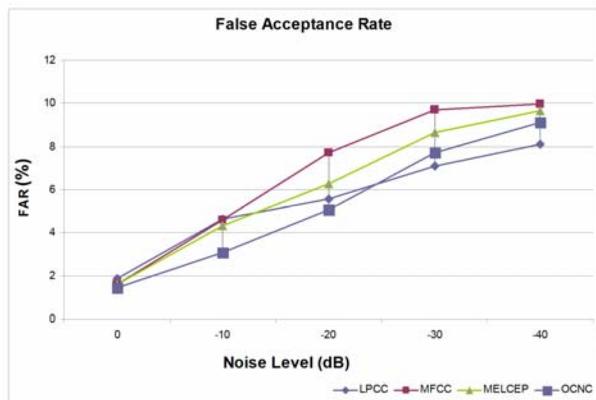

Figure 5. Performance comparisons in terms of false acceptance rate.

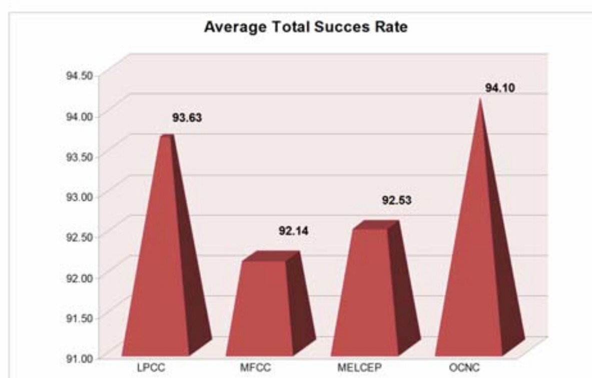

Figure 6. Performance comparisons in terms of average total success rat

4. Conclusion

In this paper, a new set of speech feature parameters based on subband adaptive filtering, OCNC's are introduced. It is experimentally observed that the OCNC's representation provides the highest recognition rate for speaker independent isolated word recognition in the presence of noise.

References

[1] H. Park, T. Takiguchi, and Y. Ariki, "Integrated Phoneme Subspace Method for Speech Feature Extraction", *EURASIP Journal on Audio, Speech, and Music Processing.* Volume 2009, Article ID 690451,6 pages. doi:10.1155/2009/690451.

[2] S.A.R. Al-Haddad, S.A. Samad, A. Hussain, K.A. Ishak and A.O.A. Noor, "Robust Speech Recognition Using Fusion Techniques and Adaptive Filtering" *American Journal of Applied Sciences* 6 (2), 2009, ISSN 1546-9239. pp. 290-295.

[3] S.A.R. Al-Haddad, S.A. Samad, A. Hussain, K.A. Ishak and A.O.A. Noor, "Robust digit recognition with dynamic time warping and recursive least squares", *ASM Science Journal*, Vol. 2(1), 2008.pp.75-81.

[4] K. Hermus, P. Wambacq, and H. Van Hamme, "A review of signal subspace speech enhancement and its application to noise robust speech recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 45821, 15 pages, 2007.

[5] E. Erzin and A.E. Cetin, "Line spectral frequency representation of subbands for speech recognition", *Elesevier Signal Processing* (44)1995. pp. 117-119.

[6] Vaidyanathan, P.P., Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial, *Proceedings of the IEEE*, Vol.78, No.1, January 1990, pp56-93.

[7] Sayed, A. H., *Adaptive Filters*. Wiley, NJ, 2008.

[8] S. Sigurdsson, K. B. P. and T. Lehn-Schiøler, "Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music", *Proceedings of the Seventh International Conference on Music Information Retrieval (ISMIR)* 2006.pp.286-289.

[9] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1983.