

values of γ . As shown, a choice of $\gamma = 0.8$ greatly increases the output SNR, however in subjective terms the quality found was not as good as choosing $\gamma = 0.5$. Therefore, the choice of γ should be based on both subjective and objective perspectives. As can be seen from the Table, the proposed method shows significant improvement in SNR over the reported results under different noisy conditions. Also note that the proposed method provides impressive results at higher SNRs, e.g. 25 and 30 dB, where the other methods fail even to hold on to the input SNR value. The spectrograms shown in Fig. 3 demonstrate the quality of the denoised speech.

Conclusion: A new soft thresholding strategy has been proposed from the observation of the amplitude distribution and behaviour of the DCT coefficients in signal-dominant and noise-dominant regions. Unlike conventional approaches, a linear function has been suggested for amplitude thresholding. Significant improvement in the results in terms of output SNR and quality justify the effectiveness of the thresholding devised in this Letter. Research is in progress for determining the optimum value of γ dynamically for each subframe.

© IEE 2002
 Electronics Letters Online No: 20020990
 DOI: 10.1049/el:20020990

S. Salahuddin, S.Z. Al Islam, Md.K. Hasan and M.R. Khan
 (Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh)
 E-mail: khasan@eee.buet.ac.bd

References

- 1 SOON, I.Y., KOH, S.N., and YEO, C.K.: 'Noisy speech enhancement using discrete cosine transform', *Speech Commun.*, 1998, **24**, pp. 249-257
- 2 DONOHO, D.L.: 'De-noising by soft-thresholding', *IEEE Trans. Inf. Theory*, 1995, **41**, pp. 613-627
- 3 BAHOURA, M., and ROUAT, J.: 'Wavelet speech enhancement based on the teager energy operator', *IEEE Signal Process. Lett.*, 2001, **8**, (1), pp. 10-12
- 4 HASAN, M.K., ZILANY, M.S.A., and KHAN, M.R.: 'DCT speech enhancement with new hard and soft thresholding criteria', *Electron. Lett.*, 2002, **38**, (13), pp. 669-670

Transformation of narrowband speech into wideband speech with aid of zero crossings rate

I.Y. Soon, S.N. Koh, C.K. Yeo and W.H. Ngo

An innovative technique, for narrowband to wideband transformation of speech signals, is proposed. The zero crossings rate is used to adaptively control the gain of the synthesised upper band speech leading to significant performance improvement over an existing technique. Results are in fact comparable to more complex techniques. The technique can be implemented at the receiving end alone as it does not require any side information to be transmitted and can be easily implemented using finite impulse response digital filters.

Introduction: It is well known that wideband speech is preferred to narrowband speech by human listeners as shown in an ITU study [1]. Wideband speech normally has a bandwidth of 7 kHz while narrowband speech has a bandwidth of only less than 3.8 kHz. However, in most speech transmission systems, e.g. telephone networks, narrowband speech is transmitted to save on bandwidth. Narrowband speech is good enough for voiced sounds in general as their energy is concentrated in the lower bands but is not as good for unvoiced sounds which spread to the higher frequencies. These higher frequencies are lost when narrowband speech is used, which is why telephone speech sounds muffled. Unvoiced speech signals are important as they contribute significantly to the intelligibility of the speech. Hence it is desirable to reconstruct the high frequency portion of the unvoiced signal from the narrowband signal at the receiving end without requiring any changes to the transmission systems. As the degradation of wideband speech to narrowband speech is a many-to-one mapping process, it is mathematically non-reversible. The only property that can

be exploited to perform the enhancement is the characteristics of the human speech.

There have been many techniques proposed recently to synthesise or estimate the missing high frequencies components. Techniques proposed include using linear predictive (LPC) analysis and vector quantisation [2-6], neural networks [7], statistical recovery method [8] and nonlinear filtering method [9] by Yasukawa. Most of the techniques proposed, with the exception of [9], are computationally quite intensive and the improvement may not justify the computation requirements. The method proposed by Yasukawa is simple and robust as it does not involve any vector quantisation, training issues, linear predictive analysis and synthesis, and its performance is relatively good. However it does suffer from the common problem of over extending the voiced signals which results in minor artifacts or distortion. This problem is also reported in [3, 6].

In the proposed enhancement system, the zero crossing rate (ZCR) is used to adaptively adjust the gain of the synthesised upper band signal. A fixed gain proposed in [9] tends to over enhance the voiced portion of the speech and under enhance the unvoiced portion. In [4], a voiced/unvoiced decision is used to reduce this problem, but due to its binary nature, it does not perform as well as the ZCR. Note that the unvoiced portion normally has high ZCR and the voiced portion has lower ZCR.

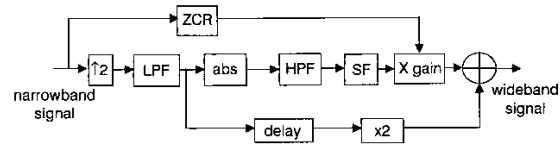


Fig. 1 Block diagram of proposed system

Wideband synthesis: The block diagram of the proposed narrowband to wideband transformation is shown in Fig. 1. To recover the wideband signal from the narrowband signal, the narrowband signal is first interpolated using upsampling-and-lowpass filtering to increase the sampling frequency to 16 kHz. Next, the bandwidth of the upsampled signal is doubled by using a full wave rectifier (absolute function) to obtain an initial estimate of the upper band signal. This is a nonlinear process. The initial estimate, however, overlaps with the lower band signal and hence must be highpass filtered. The 3 dB points of the highpass and lowpass filters are at 3.8 kHz. An additional shaping filter (SF) is used to shape the upper band signal so that it corresponds more to the human speech spectral envelope. The shaping filter is a fourth order finite impulse response (FIR) lowpass filter with a 3 dB attenuation at 5 kHz.

The zero crossings, z , at sample, n , are defined mathematically as:

$$z(n) = |\text{sign}(x(n)) - \text{sign}(x(n-1))|$$

where

$$\begin{aligned} \text{sign}(x) &= 1 & x(n) > 0 \\ &= 0 & x(n) = 0 \\ &= -1 & x(n) < 0 \end{aligned}$$

The unvoiced portion of the speech signal has an average of 75 zero crossings in 15 ms while the voiced portion has an average of approximately 20 zero crossings in 15 ms. A 15 ms Hamming window is then applied to the zero crossings signal, z , to obtain the average short time zero crossing rate (ZCR). The upperband gain can be obtained by dividing the ZCR by a factor of ten. The divisor can be adjusted if different degree of enhancement is desired. The synthesised upperband signal has to be multiplied by the gain before adding to the delayed narrowband signal. The delay is to compensate for the time taken to filter the highband signal, so that the correct temporal alignment of the highband and lowband signal can be achieved. All the filters used are implemented using digital FIR filters.

Results: Ten wideband speech files taken from the TIMIT database which has a sampling rate of 16 kHz are first lowpassed at 3.8 kHz and downsampled to obtain the 8 kHz sampled narrowband speech files. The reason behind using downsampled files instead of recorded narrowband speech files is that objective measures can be used to

evaluate the enhancement system. The computation of SNR in the time domain is not useful as the phase of the estimated high frequency components will not match that of the original 16 kHz signal. A better way will be to use the average log spectral distance (LSD) computed in the Fourier domain.

The average log spectral distance, LSD, between the original wideband speech, X , and the reconstructed wideband speech, Y , is defined as:

$$LSD = \frac{1}{K} \sum_{k=0}^{K-1} \left[\frac{1}{\omega_s} \int_{-0.5\omega_s}^{0.5\omega_s} (\log_e |X(e^{j\omega})| - \log_e |Y(e^{j\omega})|)^2 d\omega \right]^{0.5}$$

where K is the number of frames with frame size of 512 samples and ω_s is the sampling frequency.

A similar measure, but with the computation restricted to only the upper band, is also used by some researchers [4, 6]. However some synthesis techniques do not reconstruct the lower band perfectly and such distortions may go unnoticed. Hence it is more appropriate to compute the distance over the entire spectrum.

Table 1: Average log spectral distance

Speech file	ZCR	Yasukawa	Narrowband
F1	1.1037	1.2386	1.6268
F2	1.2975	1.4537	1.7394
F3	1.1853	1.2582	1.9041
F4	1.2022	1.3035	1.9600
F5	1.1628	1.3178	1.4268
M1	1.1286	1.2298	1.8154
M2	1.1217	1.1825	1.8066
M3	1.2199	1.4674	1.5419
M4	1.2063	1.4018	1.6563
M5	1.2633	1.3896	1.7477
Average	1.1891	1.3243	1.7225

The results for 5 sets each of male and female speech are shown in Table 1. The average log spectral distance is computed between the enhanced speech using the proposed ZCR method and the original wideband speech. The same is computed for the Yasukawa [9] method. It can be seen from Table 1 that both methods produce wideband speeches which are superior to the narrowband speech. However, the proposed ZCR method performs significantly better than Yasukawa's. Note that all the speeches are aligned temporally before the distance computation. In essence, the ZCR method produces synthesised wideband speech which is closer to the original speech with no irritating distortion heard. The enhanced speech also sounded natural to all human listeners.

Conclusion: The results of this Letter show that voiced and unvoiced signals should be treated differently in the narrowband to wideband transformation and an effective way to accomplish it is to use the zero crossing rate to adaptively adjust the upper band gain. The proposed technique is still low in computational requirements despite the additional step and most human listeners prefer the reconstructed wideband signal to the original narrowband signal. The proposed technique does not require transmission of any side information and hence it can be applied to all narrowband speeches.

© IEE 2002

4 March 2002

Electronics Letters Online No: 20021040

DOI: 10.1049/el:20021040

I.Y. Soon, S.N. Koh, and W.H. Ngo (School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, S 639798, Singapore)

E-mail: eiysoon@ntu.edu.sg

C.K. Yeo (School of Computer Engineering, Nanyang Technological University, Nanyang Avenue, S 639798, Singapore)

References

- 1 'Paired Comparison test of wide-band and narrow-band telephony' ITU, COM 12-9-E, March 1993

- 2 CHENNOUKH, S., GERRITS, A., MIET, G., and SLUIJTER, R.: 'Speech enhancement via frequency bandwidth extension using line spectral frequencies'. Proc. ICASSP, 2001, Vol. 1, pp. 665-668
- 3 AVENDANO, C., HERMANSKY, H., and WAN, E.A.: 'Beyond Nyquist: Towards the recovery of broad-bandwidth speech from narrow-bandwidth speech'. Proc. Eurospeech 95, 1995, pp. 165-168
- 4 EPPS, J., and HOLMES, W.H.: 'A new technique for wideband enhancement of coded narrowband speech'. Proc. IEEE Workshop on Speech Coding, 2000, pp. 174-176
- 5 PARK, K.Y., and KIM, H.S.: 'Narrowband to wideband conversion of speech using GMM based transformation'. Proc. ICASSP 2000, 2000, Vol. 3, pp. 165-168
- 6 ENBOM, N., and KLEIJN, W.B.: 'Bandwidth expansion of speech based on vector quantization of the mel frequency cepstral coefficients'. Proc. IEEE Workshop on Speech Coding, 1999, pp. 171-173
- 7 UNCINA, A., GOBBI, F., and PIAZZA, F.: 'Frequency recovery of narrow-band speech using adaptive spline neural networks'. Proc. ICASSP 2000, 1999, Vol. 2, pp. 997-1000
- 8 CHENG, Y.M., SHAUGHNESSY, D., and MERMELSTEIN, P.: 'Statistical recovery of wideband speech from narrowband speech', *IEEE Trans. Speech Audio Process.*, 1994, 2, pp. 544-548
- 9 YASUKAWA, H.: 'A simple method of broad band speech recovery from narrowband speech for quality enhancement'. Proc. of IEEE Digital Signal Processing Workshop, Sep 1996, pp. 173-175

Measurement-based queue length distribution estimation for power-law traffic

C.M. Leung, J.A. Schormans and A.H.I. Ma

Accurate measurement of queue distributions is vital to guarantee the QoS performance in packet networks, e.g. IP. Measurement is particularly difficult when the packet traffic is governed by extremely high variance power laws. A measurement scheme is proposed to estimate the queue length distribution for aggregate power law ON-OFF sources multiplexed into a FIFO queue. This scheme is carried out on a non-intrusive basis, so in contrast to active measurement methods, it will not create extra traffic. The methodology is validated with simulation, and results show excellent agreement between estimated and actual queue length distribution, even in the presence of traffic with extremely high variance. This can therefore be used as the core mathematical part of a packet network-based assurance technique.

Introduction: Traffic measurement has been widely proposed as a key technique for guaranteeing network performance, e.g. for QoS or SLA validation. At present, the measurement techniques can be classified into two major categories: active and passive [1]. In previous work [2], we proposed a measurement methodology for the end-to-end delay performance validation, which is based on having theoretical knowledge of the 'shape' of the queue length distribution. We estimate the queue length distribution by passively monitoring the queue length during the measuring period (i.e. it can be considered as a passive measurement method as no probing packets are generated). However, in [2], we considered only Markovian traffic (ON/OFF sources with exponentially distributed sojourn and silence time). This traffic model has been extensively employed to model voice traffic, but it is not necessarily right to model data traffic as data traffic may exhibit self-similarity [3]. In this Letter, we propose a further development of our previous approach to work for power-law traffic. This traffic has extremely high variance, and causes problems for measurement-based systems.

Measurement methodology: The measurement methodology is similar to our previous work [2]. The FIFO queue is logically partitioned into $queue_{low}$ and $queue_{high}$ regions via a partition point called q_p . The reason for this logical partition is to separate the packet-scale queuing and the burst-scale queuing region that occurs when multiplexing on-off traffic [4]. Five measurement data are required: $freq_{low}$, $freq_{high}$, q_{low} and q_{high} and B . After a measurement, $freq_{low}$ and $freq_{high}$ represent the frequency of occurrence of $queue_{low}$ and $queue_{high}$ region being hit, respectively. q_{low} and q_{high} are used to estimate the mean queue length in the regions. B is the maximum queue length