April 2001

# Temporal decomposition for low rate wideband speech compression

C. H. Ritz
*University of Wollongong*, critz@uow.edu.au

I. Burnett
*University of Wollongong*, ianb@uow.edu.au

# Temporal decomposition for low rate wideband speech compression


## Abstract

An investigation into low bit rate wideband speech coding for applications such as unicast streaming is presented. Wideband spectral parameters are quantised below 1 kbit/s using temporal decomposition (TD) applied to the line spectral frequencies. Quantisation using TD performs significantly better than split vector quantisation at an equivalent bit rate.



## Disciplines

Physical Sciences and Mathematics

## Publication Details

This paper originally appeared as: Ritz, CH and Burnett, IS, Temporal decomposition for low rate wideband speech compression, Electronics Letters, 12 April 2001, 37(8), 542-543. Copyright IEEE 2001.

lower performances than both the MD- and the HD-based method for low $E_d$s; (ii) the HD-based method exhibits significant degradation of the performances for an increase of $E_d$; and (iii) the MD-based method produces, regardless of the VAD performance, robust and superior performances in comparison with both the HD- and SD-based methods. Note that for very low $E_d$s, i.e. $0.0 \leq E_d < 0.1$, the performances of the MD and HD are slightly degraded compared with the case of $E_d = 0.2$. This is caused by less frequent adaptation of the noise frames due to the increased false alarm rate of VAD. In other words, VAD produces the low $E_d$ at the expense of the increased false alarm rate at speech pauses.

Experimental results using various noise sources, such as helicopter and HMMWV with levels of 0, 5, and 10 SNR dB, exhibit performance patterns similar to those shown in Figs. 1 and 2, despite differences in the absolute values being measured.

*Conclusion:* The MD-based noise adaptation method has been proposed for robust estimation of noise variance. From the experiments, it has been shown that the MD-based method gives performances superior to both the SD- and HD-based methods.


*6 February 2001*
*Electronics Letters Online No: 20010368*
*DOI: 10.1049/el:20010368*

Y.D. Cho, K. Al-Naimi and A. Kondoz (*Centre for Communication Systems Research (CCSR), University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom*)

E-mail: Y.Cho@eim.surrey.ac.uk

**References**

1 SOHN, J., and SUNG, W.: 'A voice activity detection employing soft decision based noise spectrum adaptation'. Proc. Int. Conf. Acoust., Speech, Signal Processing, May 1998, pp. 365–368
2 KIM, N.S., and CHANG, J.H.: 'Spectral enhancement based on global soft decision', *IEEE Signal Process. Lett.*, 2000, **7**, pp. 108–110
3 EPHRAIM, Y., and MALAH, D.: 'Speech enhancement using a minimum mean square error short-time spectral amplitude estimator', *IEEE Trans. Acoust., Speech, Signal Process.*, 1984, **ASSP-32**, pp. 1109–1120
4 LIM, J.S., and OPPENHEIM, A.L.: 'All-pole modelling of degraded speech', *IEEE Trans. Acoust., Speech, Signal Process.*, 1978, **ASSP-26**, pp. 197–210
5 CHO, Y.D., AL-NAIMI, K., and KONDOZ, A.: 'Improved voice activity detection based on a smoothed statistical likelihood ratio' to appear in Proc. Int. Conf. Acoust., Speech, Signal Processing, May 2001

## Temporal decomposition for low rate wideband speech compression

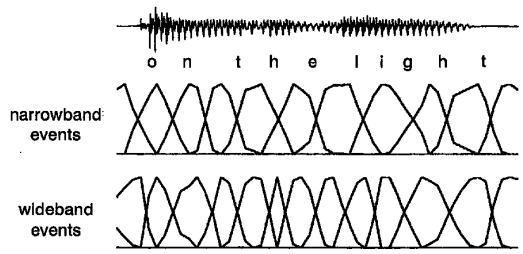
C.H. Ritz and I.S. Burnett

An investigation into low bit rate wideband speech coding for applications such as unicast streaming is presented. Wideband spectral parameters are quantised below 1kbit/s using temporal decomposition (TD) applied to the line spectral frequencies. Quantisation using TD performs significantly better than split vector quantisation at an equivalent bit rate.

*Introduction:* Wideband speech (up to 8kHz bandwidth) offers a significant improvement in the subjective quality over narrowband speech (up to 4kHz bandwidth) [1]. Most of the current proposals for wideband speech coding achieve bit rates of 8kbit/s and above, and are targeted at real-time applications such as teleconferencing [1]. Low rate (4–8kbit/s) applications for wideband speech coding include high-quality voicemail and paging services and low rate mobile Internet applications including streaming media and speech storage for online teaching material and news bulletins.

A high proportion of the bit rate used by a speech coder is allocated to the spectral parameters. One technique proposed in narrowband speech coding for significantly reducing this bit rate is temporal decomposition (TD) [2 – 4]. Note that this technique introduces encoding delay of up to a few hundred ms. For the above applications, however, this is acceptable.

TD models the trajectory of the speech spectral parameters as a weighted sum of interpolating functions. The interpolating functions are commonly known as event functions and the weights as target vectors. Quantising and transmitting these parameters instead of the line spectral frequency (LSF) vectors leads to significant bit rate reductions [2]. In this Letter, TD is applied to the LSF vectors derived for wideband speech by adapting the TD approach for narrowband speech outlined in [3]. We refer to TD for narrowband speech as narrowband temporal decomposition (NBTD) and TD applied for wideband speech as wideband temporal decomposition (WBTD).




009/1

**Fig. 1** *Event functions derived for narrowband and wideband LSFs for section of male speech*

*TD parameter quantisation:* Event functions have both duration and shape; examples derived for wideband and narrowband speech are shown in Fig. 1. The maximum event duration is ten frames; hence, they are scalar quantised with four bits. Event shapes are quantised using vector quantisation (VQ) following linear interpolation of the event shape vector to a fixed length. Target vectors are modified versions of the LSF vectors and can be quantised using VQ. However, to reduce complexity they are quantised using split VQ.

*Evaluating NBTD and WBTD:* All speech used in this work was taken from the ANDOSL 20kHz sampled speech database [5]. Narrowband and wideband speech is formed by re-sampling to 8 and 16kHz, respectively. LSFs were generated from tenth and twentieth order linear prediction (LP) analysis of narrowband and wideband speech, respectively, using 30ms windows and 25ms frames. VQ codebooks for event functions and target vectors were trained using approximately 31000 event functions and target vectors derived from 30min of male and female speech. NBTD and WBTD were applied to a separate set of 12 male and female sentences. To evaluate modelling distortion we used the average log spectral distortion (avLSD) (eqn. 1) [6], which measures the average difference over $N$ frames between the original and modified LPC power spectra, $P(\omega)$ and $P'(\omega)$, respectively, and where $\omega$ ranges over 4kHz for narrowband and 8kHz for wideband speech, respectively. Event function quantisation was evaluated by using the mean squared error (MSE) while target vector quantisation was evaluated by measuring the average log spectral distortion (eqn. 1).

$$avLSD = \frac{1}{N}\sum_{N}\sqrt{\left(\sum_{\omega}\left[10\log_{10}\frac{P(\omega)}{P'(\omega)}\right]^2\right)} \quad (1)$$

The overall quantisation scheme was evaluated by measuring the average log spectral distortion between LSFs modelled from unquantised and quantised WBTD parameters. The total distortion introduced from both modelling and quantisation was evaluated by measuring the average log spectral distortion between original LSFs and LSFs modelled by quantised WBTD parameters.

*Results:* Table 1 shows the results for the modelling distortion (eqn. 1), between original LSFs and the LSFs reconstructed using NBTD and WBTD without quantisation. To indicate perceived distortion [6], outlier percentages are also shown. Since both techniques achieved the same average spectral distortion values and

similar outlier percentages we propose that the TD model is equally suited to both wideband and narrowband speech.

**Table 1:** Average log spectral distortion (avLSD) for model.

| Narrowband | | | Wideband | | |
|---|---|---|---|---|---|
| avLSD | 2 < SD ≤ 4 | SD > 4 | avLSD | 2 < SD ≤ 4 | SD > 4 |
| 1.6 dB | 25.2% | 3.3% | 1.6 dB | 24.1% | 3.6% |

Percentage outliers from 2–4 dB and greater than 4 dB also shown

**Table 2:** MSE between original event functions and quantised event functions

| Codebook size | Narrowband | Wideband |
|---|---|---|
| 256 | 0.5 | 0.4 |
| 128 | 0.6 | 0.5 |
| 64 | 0.9 | 0.8 |
| 32 | 1.5 | 1.1 |
| 16 | 2.2 | 2.0 |

The mean squared errors between the original and quantised event shapes are shown in Table 2 for various size codebooks. Results show that event functions derived using WBTD and NBTD can be quantised using approximately the same number of bits. These results are explained by considering Fig. 1 and the proposal in [4]. Fig. 1 shows that although WB LSF vectors have twice the order and cover twice the frequency range, only a few extra event functions are generated compared to NBTD. In [4] it was proposed that event functions correspond closely to the phonemes in speech, which are independent of the frequency content. Hence the near equality of bit rates for event functions appears reasonable.

Target vectors derived using WBTD were found to require twice the bit rate for quantisation as those derived using NBTD to achieve similar spectral distortion and outlier numbers. Since this is to be expected, detailed results are not shown.

**Table 3:** Average log spectral distortion (avLSD) and percentage of outliers for quantised wideband LSFs

| | Average bit rate | avLSD | 3 < SD ≤ 5 | SD > 5 |
|---|---|---|---|---|
| | bit/s | dB | % | % |
| WBTD (modelled) | 906 | 1.4 | 1.3 | 0.2 |
| WBTD (total) | 906 | 2.2 | 15.8 | 2.0 |
| Split VQ | 1200 | 2.2 | 9.3 | 1.3 |
| Split VQ | 1000 | 2.5 | 20.4 | 1.7 |

Table 3 shows average spectral distortion results for quantisation of wideband LSFs using both WBTD and split VQ. Also shown are outlier percentages with ranges of 3 to 5dB and greater than 5dB. These ranges provide a better indication of perceptual quality for wideband speech quantisation [7] than those suggested for narrowband speech quantisation [6]. WBTD (modelled) refers to distortions between LSFs reconstructed from unquantised and quantised WBTD parameters, while WBTD (total) refers to distortions between original LSFs and LSFs reconstructed from quantised WBTD parameters. Event shapes were quantised with eight bits and target vectors were quantised using split VQ with five split vectors of sizes (3, 3, 4, 4, 6) and eight bits per split. Split VQ of the LSFs is designed using five split vectors of sizes (3, 3, 4, 4, 6) and both five bits and six bits per split.

In [7] it is suggested that an average distortion of 1.6 dB with 4% of outliers in the range 3 to 5dB and no outliers greater than 5dB produces transparent quantisation. Table 3 shows that LSFs modelled from WBTD easily meet these requirements, except for a small number of outliers above 5dB. Table 3 also shows that the wideband LSFs can be quantised at 906bit/s using WBTD with approximately 25% less bits than split VQ (without increasing the average distortion). Results also show that LSF quantisation using split VQ at 1000bit/s introduces a higher average distortion and more outliers than those for WBTD. Informal listening tests found speech synthesised from LSFs quantised at 906bit/s using WBTD sounded significantly better than speech synthesised from LSFs quantised at 1000bit/s using split VQ and more natural than speech synthesised from LSFs quantised using split VQ at 1200bit/s. We propose that this is due to the nature of WBTD, which produces smoother LSF tracks than LSFs quantised using split VQ. This is consistent with the findings in [8] that more natural sounding speech results from smooth LSF tracks.

*Conclusions:* Wideband LSFs can be modelled by TD as accurately as narrowband LSFs. It was also found that WBTD requires no increase in transmission bit rate for event function quantisation compared to NBTD. Overall, wideband LSFs resulting from WBTD can be quantised below 1kbit/s with results (from spectral distortion measurements) meeting the proposed transparency requirements suggested in [7]. While the total distortion resulting from LSF quantisation using WBTD did not meet these transparency requirements, we propose that this is due to the model and not the quantisation scheme. Hence, by improving the model (such as suggested in [9]) it should be possible to obtain transparent quantisation of wideband LSFs at bit rates below 1kbit/s. This makes WBTD a promising approach to low bit rate wideband speech compression where encoding delay can be tolerated.

*Acknowledgments:* C.H. Ritz is in receipt of an Australian Postgraduate Award and a Motorola (Australia) Partnerships in Research Grant.


*20 February 2001*
*Electronics Letters Online No: 20010349*
*DOI: 10.1049/el:20010349*

C.H. Ritz and I.S. Burnett (*Whisper Labs, University of Wollongong, Northfields Avenue, Wollongong, NSW, 2522, Australia*)

E-mail: chritz@st.elec.uow.edu.au

**References**

1 SALAMI, R., LAFLAMME, C., and ADOUL, J.P.: 'Real-time implementation of a 9.6kbit/s ACELP wideband speech coder'. Proc. GLOBECOM'92, 1992, Vol. 1, pp. 447–451
2 ATAL, B.S.: 'Efficient coding of LPC parameters by temporal decomposition'. Proc. ICASSP'83, Boston, USA, 1983, pp. 81-84
3 KIM, S.J., LEE, S., HAN, W.J., and OH, Y.H.: 'Efficient quantization of LSF parameters based on temporal decomposition'. Proc. ICSLP'98, 1998
4 VAN DIJK-KAPPERS, A.M.L., and MARCUS, S.M.: 'Temporal decomposition of speech', *Speech Commun.*, 1989, **8**, pp. 125–135
5 Australian National Database of Spoken Language (ANDOSL), CD ROM
6 PALIWAL, K.K. and KLEIJN, W.B.: 'Quantization of LPC parameters' *in* KLEIJN, W.B. and PALIWAL, K.K. (Eds.): 'Speech coding and synthesis' (Elsevier, 1995), p. 443
7 FERHAOUI, M., and VAN GERVEN, S.: 'LSP quantization in wideband speech coders'. Proc. IEEE Workshop on Speech Coding, June 1999, pp. 25–27
8 KLEIJN, W.B., and HAGEN, R.: 'On memoryless quantization in speech coding', *IEEE Signal Process. Lett.*, 1996, **3**, (8) pp. 228–230
9 ATHAUDAGE, C.N., BRADLEY, A.B., and LECH, M.: 'Optimization of a temporal decomposition model of speech'. Proc. ISSPA'99, 1999, pp. 471–474

## Active noise controller with fuzzy filtered-U algorithm

Cheng-Yuan Chang, Kuo-Kai Shyu and Tzu-Neng Chuang

A method that protects active noise control systems against unstable poles such as occur in conventional filtered-U design is proposed. The performance is improved and, because the proposed algorithm uses input-output pairs to construct the control rules, the design complexity is reduced.

*Introduction:* Although the filtered-U algorithm [1, 2] is widely used in active noise control (ANC) systems, it still poses several