CODING WIDEBAND SPEECH AT NARROW-BAND BIT RATES

J.R. Epps and W.H. Holmes
School of Electrical Engineering and Telecommunications,
The University of New South Wales

ABSTRACT. The 'muffled' quality of coded speech, which arises from the bandlimiting of speech to 4 kHz, can be reduced either by coding speech with a wider bandwidth or by wideband enhancement of the narrowband coded speech. This paper investigates the limitations of wideband enhancement and possibilities for its improvement. A new wideband coding scheme is proposed that is based on any narrowband coder, but augmented by wideband enhancement plus a few bits per frame of highband information. The scheme thus has a bit rate only slightly greater than that of the narrowband coder. Subjective listening tests show that this scheme can produce wideband speech of significantly higher quality than the narrowband coded speech.

1. INTRODUCTION

The need for wideband speech transmission arises both from an ongoing requirement for improved speech quality in all types of services, and from the specific needs of applications such as hands-free and Internet telephony. One solution is to code the parameters of wider bandwidth speech, which leads to a substantial increase in the bit rate relative to narrowband coders (Schnitzler, 1998).

An alternative is to employ wideband enhancement (Makhoul and Berouti, 1979; Carl and Heute, 1994; Cheng et al., 1994; Yoshida and Abe, 1994; Chan and Hui, 1996; Enbom, 1998; Epps and Holmes, 1998 and 1999; Epps, 2000; Jax and Vary, 2000), a technique which synthesizes wideband speech based on pitch, voicing and spectral envelope information in the narrowband speech. Wideband enhancement requires no increase in bit rate, but the quality of the output wideband speech is poorer than that resulting from wideband coding due to less accurate highband spectral envelope estimates. In this paper, an assessment is made of the limits to the accuracy of highband envelope estimation under realistic test criteria.

A new technique for wideband coding is also proposed which is based upon a combination of the wideband enhancement paradigm with any narrowband coder. Wideband speech of higher quality than that produced by wideband enhancement alone is obtained by allocating just a few bits per frame for highband spectral coding. This means that the new wideband coder has a bit rate only slightly greater than that of the narrowband coder.

Section 2 examines the potential performance of wideband enhancement envelope estimation, section 3 reviews selected literature on very low bit rate wideband spectral coding, section 4 presents a new technique for coding wideband speech at near narrowband bit rates, and section 5 details subjective test results of various schemes.

2. WIDEBAND ENHANCEMENT

Wideband enhancement is a scheme which adds a synthesized highband signal to narrowband speech to produce a wideband speech signal, as shown in Fig. 1. The synthesis of the highband signal is based entirely on the information available in the narrowband speech. Note that the narrowband speech is not re-synthesized, since it is assumed to be of sufficiently high quality.

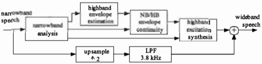

Figure 1. Overall scheme of wideband enhancement

Highband Excitation Synthesis

Previous research (Epps and Holmes, 1998) has shown that a combination of sinusoidal and random excitation can be used to produce high quality highband excitation estimates. In this technique, based on the sinusoidal transform coding (STC) harmonic model (McAulay and Quatieri, 1995), sinusoidal highband excitation is synthesized from the narrowband speech using pitch, voicing and highband spectral envelope estimates. This technique gives perfectly harmonic periodic excitation, with the amplitudes of the sinusoidal components determined directly by the spectral envelope. Random excitation, at an amplitude controlled by the narrowband degree of voicing, is employed to model the highband unvoiced components. This approach was found to accurately represent the voiced/unvoiced mix of a wide variety of different speech frames. The use of STC-derived parameter interpolation methods produced smooth variation of the sinusoid frequencies and phases between frames. These features contributed to a better perceptual perfomance of the novel STC-based excitation than other methods such as spectral folding (Makhoul and Berouti, 1979).

Highband Envelope Estimation

Different techniques for estimating the shape of the highband spectral envelope have also been considered, with codebook mapping (Gersho, 1990; Carl and Heute, 1994) performing well under a spectral distortion comparison (Epps and Holmes, 1999). As seen in Fig. 2, codebook mapping estimates the highband spectral envelope by selecting the highband code vector whose corresponding narrowband code vector has the most similar envelope shape (in a spectral distortion sense) to the input narrowband envelope. Details of the codebook design from training data can be found in the work of Carl and Heute (1994) and Epps (2000). Other methods of highband envelope estimation include statistical recovery (Cheng et al., 1994), codebook mapping with codebooks split by voicing (Epps and Holmes, 1999), and codebook mapping based upon hidden Markov models (Jax and Vary, 2000).

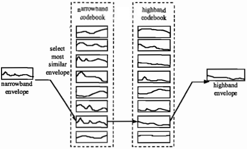

Figure 2. Codebook mapping for highband envelope estimation

Narrowband-Highband Envelope Continuity

Preserving the continuity of the spectral envelope between the narrowband envelope and the estimated highband envelope is an important perceptual requirement, however in instances where the accuracy of the highband envelope estimation is poor, the resulting highband spectral distortion can be quite large. Typically the estimated highband envelope is matched to the narrowband envelope either at a single frequency or over a range of frequencies, depending on the size of the overlap between the two envelopes, but there are alternative techniques (Epps, 2000).

Limits to Wideband Enhancement

Highband envelope estimation is based upon the assumption that two wideband spectral envelopes with similar narrowband envelope shapes will also have similar highband envelope shapes. One method for testing the validity of this assumption is to select two pairs of narrowband-highband spectral envelopes from independent speech databases. Their narrowband spectral distortion is then calculated and their highband spectral distortion is also computed, after ensuring that the second highband envelope is properly matched to (i.e. continuous with) the first narrowband envelope. This procedure is then repeated for all combinations of envelope pairs from the two speech databases.

The resulting data can be used to gain an idea of the distribution of continuity-adjusted highband spectral distortion results which could be expected from any codebook mapping scheme with a given maximum narrowband spectral distortion. Figure 3 illustrates these distributions, showing that highband spectral distortion is weakly correlated with narrowband spectral distortion. Present wideband enhancement techniques produce average continuity-adjusted highband spectral distortions of around 6.4 dB (Epps, 2000) using narrowband codebooks with a maximum narrowband distortion of around 5.8 dB, and are thus slightly better than the median expected performance.

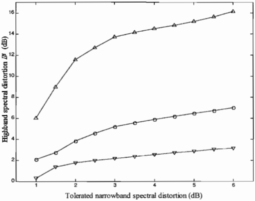

Figure 3. Distribution of continuity-adjusted highband spectral distortion as a function of tolerated (maximum) narrowband spectral distortion. Results are based on around 5(106 data points. The percentile contours shown are the 10% (-), 50% or median (o), and 90% ((). Note that there are relatively few data points for small values of tolerated narrowband spectral distortions.

Good performance (median highband spectral distortion 2 dB) with codebook mapping schemes is therefore possible only if the narrowband spectral distortion can be contained to around 1 dB. This would require codebooks consisting of around 2^{60} vectors, a size which is not feasible to implement under present storage constraints. It is concluded that the practical performance of highband envelope estimation methods is only likely to be improved, compared to existing methods, by allowing some knowledge of the original highband speech, rather than relying entirely on narrowband information. This is the subject of the following sections.

3. EXISTING TECHNIQUES FOR VERY LOW BIT RATE WIDEBAND CODING

Low Order Highband LP Coding

In some previous coder implementations (McElroy *et al.*; 1993, Seymour and Robinson, 1997), the highband (4-8 kHz) spectral envelope was coded using a 2nd order LP envelope. The LP parameters and highband gain are quantized using 440-500 bit/s (plus quantized highband excitation), which is all additional to the narrowband coding scheme.

Flat Highband Envelope Coding

If the narrowband is defined to be the 0-6 kHz frequency range, a 6-7 kHz highband envelope can be sufficiently well represented by a flat spectrum at the correct gain. In (Schnitzler, 1998) this gain is encoded on a sub-frame basis using 3 bit MA prediction (an additional 1.2 kb/s). Random excitation was used in the highband.

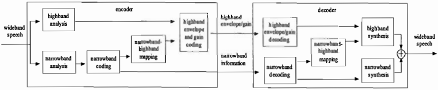

Figure 4. Proposed wideband encoder and decoder based on wideband enhancement

Wideband Enhancement and Gain Coding

The use of highband gain coding in conjunction with wideband enhancement was suggested in (Enbom, 1998). Here the (4-8 kHz) highband envelope was estimated from the narrowband envelope using codebook mapping. The gain was coded as the difference G_{HB} - G_{NB} (in dB) between the highband and narrowband gains G_{HB} and G_{NB}, which were quantized on a linear scale and coded using 4 bits per frame (an additional 700 bit/s). In this case, highband excitation was provided by spectral folding (Makhoul and Berouti, 1979).

4. A NEW WIDEBAND CODER

Overview

The proposed structure of the new wideband speech coder is based around any narrowband coder, as seen in Fig. 4. The decoded narrowband information, in addition to some coded highband spectral envelope and gain information, is used to synthesize the wideband signal at the decoder. In the highband envelope and gain coding technique presented in section 4.3, narrowband-highband mapping is combined with coding. Highband excitation synthesis is discussed in section 4.2. This configuration is well suited to coders which must primarily satisfy a set of narrowband performance criteria, but which can accommodate a few bits per frame of highband information. It also allows bit allocation trade-offs to be made between the narrowband and highband in a similar fashion to sub-band or split band coding.

Highband Excitation Synthesis using a Sinusoidal Model

Earlier research has shown that high quality mixed voiced/unvoiced highband excitation may be synthesized using the STC-based approach discussed in section 2.1. In informal listening tests this approach was considered superior in quality to spectral folding (Makhoul and Berouti, 1979) or purely random excitation in the highband. If a sinusoidal narrowband codec is employed, the highband excitation could be efficiently synthesized at the decoder concurrently with the narrowband excitation.

Wideband Enhancement with Classified Highband Codebooks

A new method for highband spectral coding employs vector quantization of the highband envelope and gain using a small partition of a full highband codebook, where the selection of vectors comprising the partition is based upon the shape of the narrowband envelope. The index i of the narrowband code vector most similar to the input narrowband spectral envelope is first determined. Each narrowband code vector contains 2^n indices to the highband codebook, where n is the number of highband bits employed. The input highband envelope and highband gain are compared with all 2^n vectors in the highband codebook whose indices are contained in the narrowband code vector with index i, and the most similar highband code vector is chosen. The index of the chosen highband code vector is then coded using n bits. This method is illustrated in Fig. 5. The use of as few as one, two or three highband bits to select between many highband candidate envelopes substantially reduces the highband spectral distortion resulting from the codebook mapping scheme discussed in sections 2.2 to 2.4.

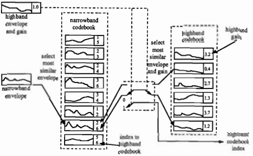

Figure 5. Block diagram example of a classified codebook mapping-based highband envelope and gain encoder using one highband bit per frame.

5. SUBJECTIVE ASSESSMENT

18 listeners (16 male and 2 female, between the ages of 20 and 35) were each presented with 70 randomized samples of speech prepared using various methods for determining the highband excitation and envelope. Codebooks of size 1024 were employed in wideband enhancement, while speech coded according to section 4 used a narrowband codebook of size 1024 and a highband codebook of size 8192 (i.e. three highband bits per frame). The resulting preliminary mean opinion scores (MOS) are shown in Table 1, where the 95% confidence interval is (0.15.

These results show that with only a few bits per frame, wideband spectral coding can be achieved at close to the quality of the original wideband speech. While the sinusoidal-based synthetic excitation performed reasonably well when combined with the original highband spectral envelopes, this excitation generally attracted lower scores, indicating that more attention still needs to be paid to perceptual artifacts arising from its implementation.

Table 1. MOS results

Condition	MOS
Original wideband speech	4.25
Wideband coded speech, 3 highband bits per frame, original highband excitation	3.99
Wideband enhanced speech, original highband excitation, synthetic highband envelope	3.65
Wideband enhanced speech, original highband envelope, synthetic highband excitation	3.31
Wideband coded speech, 3 highband bits per frame, synthetic highband excitation	2.83
Wideband enhanced speech, synthetic highband excitation and envelope	2.78
Original narrowband speech	2.74

6. CONCLUSION

A new scheme for wideband speech coding at a bit rate only slightly greater than that of narrowband coding has been presented. This scheme is based on wideband enhancement techniques, but improves upon these by transmitting a small amount of highband spectral envelope and gain information. Listening test results show that a significant quality improvement over narrowband speech can be achieved using this scheme.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Motorola Australia Pty Ltd, and wish to express their thanks to Dr M.M. Thomson for his advice and encouragement.

REFERENCES

Carl, H., and Heute, U. (1994). "Bandwidth enhancement of narrowband speech signals", *Signal Processing VII, Th. and Appl.*, EUSIPCO, **2**, 1178-1181.

Chan, C-F., and Hui, W-K. (1996). "Wideband enhancement of narrowband coded speech using MBE re-synthesis", *Proc. Int. Conf. on Signal Processing, ICSP* **1**, 667-670.

Cheng, Y.M., O'Shaughnessy, D., and Mermelstein, P. (1994). "Statistical recovery of wideband speech from narrowband speech", *IEEE Transactions on Speech and Audio Processing*, **2**(4) 544-548.

Enbom, N. (1998). *Bandwidth Expansion of Speech.* Thesis, Royal Institute of Technology (KTH).

Epps, J.R., and Holmes, W.H. (1998). "Speech enhancement using STC-based bandwidth extension", *Proc. ICSLP* (Sydney, Australia), 519-522.

Epps, J.R., and Holmes, W.H. (1999). "A new technique for wideband enhancement of narrowband coded speech", *Proc. IEEE Workshop on Speech Coding* (Porvoo, Finland), 174-176.

Epps, J.R. (2000). *Wideband Enhancement of Narrowband Speech*, Ph.D. Thesis (submitted), The University of New South Wales.

Gersho, A. (1990). "Optimal nonlinear interpolative vector quantization", *IEEE Trans. Comm.*, **38**(9), 1285-1287.

Jax, P., and Vary, P. (2000). "Wideband extension of telephone speech using a Hidden Markov Model", *Proc. IEEE Workshop on Speech Coding* (Delavan, Wisconsin).

Makhoul, J., and Berouti, M. (1979). "High frequency regeneration in speech coding systems", *Proc. ICASSP* (Washington D.C.), 428-431.

McAulay, R.J., and Quatieri, T.F. (1995). "Sinusoidal coding", in *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliwal (Eds), Elsevier, Amsterdam, Chapter 4, pp. 121-173.

McElroy, C., Murray, B., and Fagan, A.D. (1993). "Wideband speech coding in 7.2 kb/s", *Proc. ICASSP*, **2**, 620-623.

Schnitzler, J. (1998). "A 13.0 kbit/s wideband speech codec based on SB-ACELP", *Proc. ICASSP*, **1**, 157-160.

Seymour, C.W., and Robinson, A.J. (1997). "A low-bit-rate speech coder using adaptive line spectral frequency prediction", *Proc. EUROSPEECH* (Rhodes, Greece) **3**, 1319-1322.

Yoshida, Y., and Abe, M. (1994). "An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping", *Proc. ICSLP* (Yokohama, Japan), pp. 1591-1594.