AN EMBEDDED ADAPTIVE MULTI-RATE WIDEBAND SPEECH CODER

Alan McCree, Takahiro Unno, Anand Anandakumar, Alexis Bernard[1]*, and Erdal Paksoy*

DSP Solutions R&D Center, Texas Instruments, Dallas, Texas

ABSTRACT

This paper presents a multi-rate wideband speech coder with bit rates from 8 to 32 kb/s. The coder uses a splitband approach, where the input signal, sampled at 16 kHz, is split into two equal frequency bands from 0-4 kHz and 4-8 kHz, each of which is decimated to an 8 kHz sampling rate. The lower band is coded using the Adaptive Multi-rate (AMR) family of high-quality narrowband speech coders, while the higher band is represented by a simple but effective parametric model. A complete solution including this wideband speech coder, channel coding for various GSM channels, and dynamic rate adaptation, easily passed all Selection Rules and ranked second overall in the recent 3GPP AMR Wideband Selection Testing. Besides high performance, additional advantages of the embedded split-band approach include ease of implementation, reduced complexity, and simplified interoperation with narrowband speech coders.

1. INTRODUCTION

Wideband speech coding, using the bandwidth from 0 to 7 kHz, offers the potential for a significant improvement in speech quality over traditional narrowband coding (0-4 kHz) at comparable bit rates (8-32 kb/s). Potential applications for wideband speech coding include Voice over Internet Protocol (VoIP), high-quality audio conferencing, and third-generation wireless communications. In this paper, we present an adaptive multi-rate (AMR) wideband speech coder designed for the Third Generation Partnership Project (3GPP) AMR Wideband standardization, an extension to the recently-completed Adaptive Multi-rate (AMR) narrowband standard.

We have previously demonstrated the feasibility of producing high-quality speech with an embedded, split-band coding scheme based on the ITU standard G.729 Annex E [1] . This paper presents a complete embedded AMR wideband system, based on the AMR narrowband standard [2]. In extensive 3GPP AMR Wideband Selection Testing, this coder easily passed all Selection Rules and ranked second overall. The organization of this paper is as follows: Section 2 provides an overview of the wideband coder. Section 3 presents recent improvements made to the coding scheme: low-delay filterbanks, highband quality improvements, and extensions to higher bit rates. Section 4 describes our AMR Wideband candidate in detail, including source coding, channel coding, and dynamic rate adaptation. Finally, key results of the Selection Testing are provided in Section 5.

[1]Currently with Department of Electrical Engineering, University of California, Los Angeles, CA

2. WIDEBAND CODER OVERVIEW

We use a split-band coding approach, using a high-quality CELP coder based on the AMR narrowband standard for the lower band and a simple parametric coding scheme for the upper band. The AMR narrowband coder was originally standardized for GSM applications, but has also been selected by 3GPP for third generation systems and by TIA for TDMA wireless appplications. This coder provides good quality at a range of bit rates from 4.75 to 12.2 kb/s. By using a parametric coder for the highband information, we are able to maintain high speech quality at a low data rate for the high band (1.35 or 2.3 kb/s), leaving as many bits as possible available for high-quality coding of the lowband signal.

2.1. Embedded Split-band Coder

A block diagram of our coder is shown in Figure 1. The input signal, sampled at 16 kHz, is lowpass filtered from 0-4 kHz, downsampled to 8 kHz sampling rate, and encoded with the AMR narrowband CELP speech coder. The highpass-filtered input signal is also downsampled to 8 kHz and encoded with a parametric coder. The coding of the highband uses information from the lower band original signal, as described in the following section. At the receiver, the CELP decoder generates the coded narrowband speech; this signal is then upsampled and lowpass filtered. The highband decoder uses the coded bitstream as well as information from the lowband coded speech to synthesize a highband signal, which is then upsampled, highpass filtered, and combined with the coded baseband speech to produce wideband speech output.

We use a high-quality filterbank for the filtering decomposition; perfect reconstruction filters are not useful in this context since the highband speech waveform is not preserved by our parametric coder. This implies that the narrowband speech coding output is independent of the highband signal, so that the narrowband bitstream can be embedded in the overall bitstream. Also, this split-band approach ensures that a narrowband analog input signal, such as from a traditional telephone line band-limited to 3.4 kHz, can still be encoded well with the wideband coder.

2.2. Highband Coding Method

The highband signal is generated using a modulated noise excitation signal with linear predictive (LP) synthesis [1]. The modulation signal, based on the time envelope of the 3-4 kHz region of the decoded baseband signal, introduces a time-domain pitch structure to improve the perceived quality of voiced speech. In addition, a high-frequency reversal technique makes the expected distribution of the high-frequency LP coefficients similar to that of narrowband LPC, allowing the highband LPC quantizer to re-use the baseband coder line spectral frequency (LSF) quantization tables, simplifying the overall coder implementation.

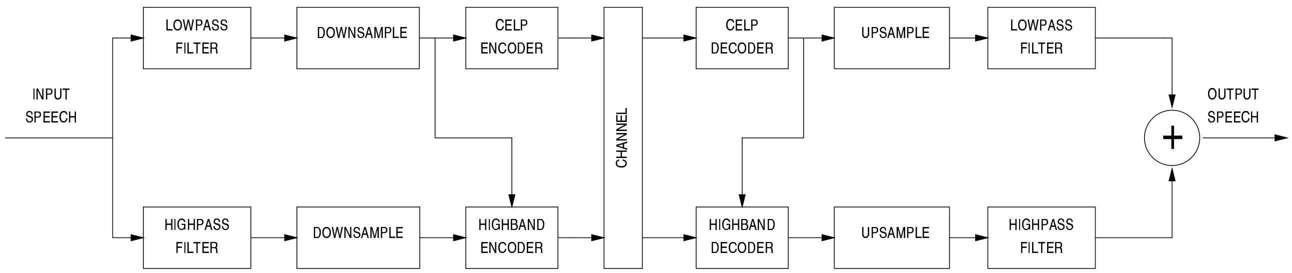

Fig. 1. Block diagram of the split-band coder.

2.3. Advantages of the Embedded Approach

The embedded nature of this coder presents several advantages: ease of implementation due to the reuse of the AMR narrowband code, high-quality encoding of narrowband sources, potential tandem-free operation between AMR wideband and narrowband, and low complexity, since the computationally-intensive CELP search routines are run at only an 8 kHz sampling rate.

3. IMPROVEMENTS TO THE WIDEBAND CODER

Our wideband coder includes three significant improvements over the baseline method. First, a low-delay filterbank reduces filtering delays without degrading subjective quality. Second, noise smoothing applied to the highband signal significantly improves the perceptual quality in acoustic background noise. Finally, new baseband coders provide better speech quality at higher bit rates. These improvements are described in this section.

3.1. Low-delay Filterbank

In our earlier work, high-order linear phase FIR filters were used. These provided high speech quality, but with a total filtering delay approaching 10 ms. For some applications, such as wireless communications, this delay is excessive. Therefore, we have designed a low-delay IIR filterbank using 12th-order Butterworth lowpass and highpass filters. As can be seen from the lowpass filter magnitude response shown in Figure 2, these filters have good stopband attenuation with reasonably sharp transition bands. Other IIR filter designs, such as elliptical or Chebyshev, provide sharper transitions, but have poles that are very close to the unit circle. The milder poles of the Butterworth design provide a smooth phase response in the passband, as shown by the group delay plot in Figure 3. This figure also shows that the group delay in the passband region is approximately 4 samples at the 16 kHz sampling rate, so that total delay through the filterbank is about 0.5 ms. An additional benefit of these weaker poles is that implementation in fixed-point arithmetic is straightforward.

3.2. High-Band Quality Improvements

In the baseline embedded wideband coder, the highband excitation signal is generated by sample-by-sample multiplication of a random noise source by the envelope of the 3-4 kHz region of the baseband decoder output. This method produces high-quality output for clean input speech signals. However, for lower bit rates in the presence of acoustic background noise, the lowband coder does not always code the 3-4 kHz band accurately. As a result,

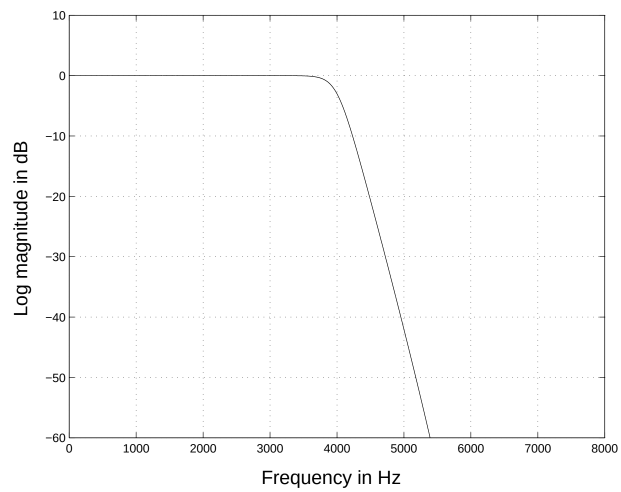

Fig. 2. Log magnitude spectrum in dB for Butterworth lowpass filter.

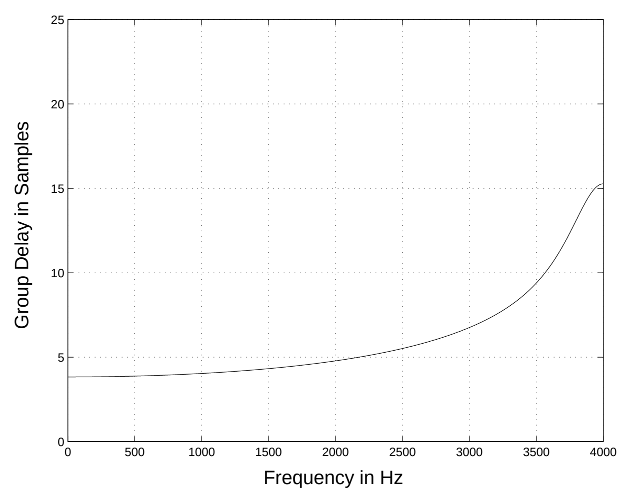

Fig. 3. Group delay in samples (at 16 kHz) for Butterworth lowpass filter.

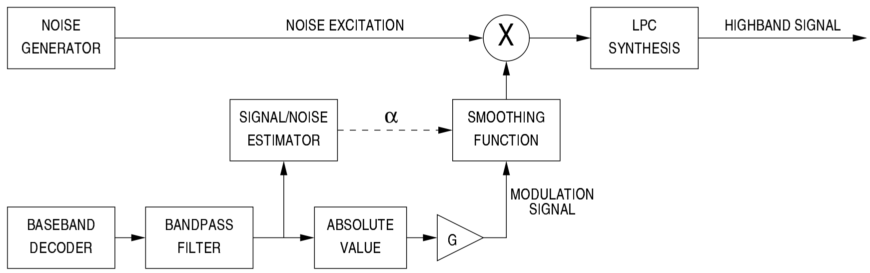

Fig. 4. Block diagram of the highband decoder.

the output time-domain signal in this band shows more rapid time variation than the input. This, in turn, causes the high-band signal to have rapid amplitude variations, which are perceived as busy high-frequency noise upon listening.

Our solution for this problem is to smooth the modulating signal at the decoder, as shown in Figure 4. This creates a more slowly varying envelope signal in the presence of background noise, and reduces the annoying "busy" noise. The equation for this smoothing is as follows:

$$e_{sm}[n] = \alpha * e_{sm}[n-1] + (1-\alpha) * e[n] \qquad (1)$$

where $e[n]$ is the envelope signal prior to smoothing, $e_{sm}[n]$ is the envelope signal after smoothing, and α is a smoothing coefficient between 0 and 1.

To ensure that this smoothing is only applied in the presence of noise, signal and noise level estimators are used to adaptively adjust the coefficient of the smoothing function. If the signal is much louder than the noise level, no smoothing is applied. If it is slightly louder than the noise level, a moderate level of smoothing is applied. If the signal is completely buried in noise, heavy smoothing is applied. The signal and noise level estimation is performed on the 3-4 kHz decoded signal.

A second source of quality improvement in the high-band signal is a smoothing applied to the high-band LSF's. In the presence of stationary noise, low-resolution quantization of the high-band LSF's can sometimes cause the quantized spectrum to oscillate in the neighborhood of the desired value, contributing to the audible "busy" character of the signal. This problem is solved by applying reduced-fluctuation quantization [3] to the encoding of the high-band LSF parameters during background noise segments. This speech/noise detection is performed on the highband input signal.

3.3. Higher Bit Rates

While the quality of the AMR narrowband coder at its highest bit rate of 12.2 kb/s is high, it is still not as good as the original signal. We find this difference to be significant in the context of embedded wideband speech coding, especially in the presence of acoustic background noise. In a formal listening test, we have found that replacing the 12.2 kb/s AMR narrowband coder with direct (uncoded) speech in the lowband while continuing to use the parametric coding scheme in the highband provides significantly improved performance. For both MOS testing of clean speech and DMOS testing of noisy speech, the performance of this coding scheme was virtually identical to that of direct wideband speech. We conclude that in our embedded wideband coder, improved quality at higher rates can be achieved by using additional bits to increase the quality of the baseband coder.

Mode	Baseband	Highband	Total
Mode 1	6.70 kb/s	1.35 kb/s	8.05 kb/s
Mode 2	7.95 kb/s	1.35 kb/s	9.30 kb/s
Mode 3	10.20 kb/s	1.35 kb/s	11.55 kb/s
Mode 4	13.05 kb/s	1.35 kb/s	14.40 kb/s
Mode 5	14.60 kb/s	1.35 kb/s	15.95 kb/s
Mode 6	21.70 kb/s	2.30 kb/s	24.00 kb/s
Mode 7	29.50 kb/s	2.30 kb/s	31.80 kb/s

Table 1. Baseband, highband, and total bit rates for wideband coder modes.

Therefore, in our design of higher rate modes, the high band coding bit rate is only slightly increased. For the baseband, we have created new coding modes, based on the AMR narrowband standard, by increasing the number of pulses in the algebraic codebooks and reducing the subframe sizes. The pulse signs are encoded using an efficient low-complexity technique which requires only 1 bit per track regardless of the number of pulses, by inferring the signs from the relative pulse position code ordering within a codeword. In this method, an extension of that used in the existing AMR narrowband coder, only the sign of the pulse with the smallest index is transmitted. All pulses with negative sign are sent prior to the smallest indexed pulse, while all pulses with positive sign are sent after it.

4. THE AMR WIDEBAND CANDIDATE

This section presents the details of our candidate for Selection Testing of the 3GPP AMR Wideband Standard.

4.1. Source Coding

Our AMR wideband candidate uses seven modes designed for different channel applications. Table 1 summarizes the bit allocation of the modes used in the coder.

The baseband coders in modes 1-3 are modes MR67, MR795, and MR102 of the AMR narrowband standard. For these modes, the bitstream for the AMR narrowband coder modes is embedded in the bitstream of the TI AMR wideband candidate. Modes 4-7 are newly designed modes where the baseband coder reuses much of the AMR narrowband software, with changes to the subframe size and the fixed excitation codebooks. In the first five modes, the highband signal is represented every 20 ms by one 4-bit gain factor and 23 bits for LSF's. For the two highest bit rate modes, the highband signal parameters are updated every 10 ms, with two

4-bit gain factors and a 38-bit matrix quantization of the two sets of LSF's in each 20 ms frame. In addition to these tested modes, the coder also supports all five other AMR narrowband modes.

Voice activity detection (VAD), discontinuous transmission (DTX) and comfort noise generation (CNG) were required for the AMR wideband standardization process. The existing AMR narrowband VAD was reused on the baseband signal. The AMR narrowband DTX/CNG scheme was slightly modified by adding 4 bits for the highband gain in the silence description update frames.

4.2. Channel Coding and Adaptation

The AMR wideband coder will be able to operate on a number of different mobile channels, including the current GSM full rate channel (22.8 kb/s), higher-bit-rate EDGE full rate (68.4 kb/s) and half rate (34.2 kb/s) channels, and the upcoming third-generation (3G) channels (at source rates up to 32 kb/s). For GSM and EDGE channels, we designed a combined speech and channel coding system for each desired bit rate, as in the current narrowband standards. This channel coding consists of rate-compatible punctured convolutional codes for error correction and cyclic redundancy coding for error detection.

In addition, for these channels the new standard will utilize the AMR paradigm where the network monitors the state of the communication channel and directs the coders to adjust the allocation of bits between source and channel coding accordingly [2, 4]. For clean channels, most of the available bits are used for source coding with only minimal error protection, while for poor channels a lower rate speech coder is protected by stronger channel coding. This is implemented by an adaptation algorithm whereby the network selects one of a number of available coder modes, each with a predetermined source/channel bit allocation. In our coder, the dynamic rate adaptation algorithm is based on an estimate of the C/I value, which is obtained based on the accumulated metric given by the soft-input/hard-output Viterbi channel decoder. AMR signaling is accomplished by reserving 8 bits for the GSM full rate channel, 12 bits for the GSM EDGE half rate channel, and 24 bits for the GSM EDGE full rate channel.

4.3. Complexity and Delay

The coder was implemented in fixed point C code using the ETSI fixed point libraries. The maximum source coding complexity for the candidate was measured at 39 weighted MOPS, less than the 40 WMOPS requirement. Much of this complexity was due to the double precision implementation of certain coder blocks, including the IIR filters for subband decomposition. Later optimizations have made it possible to implement the coder using around 25 WMOPS.

One of the main advantages of the embedded coder is the ability to reuse all the program code and quantization tables of AMR narrowband. The embedded AMR wideband coder requires only a 20% increase in program ROM size and a 10% increase in RAM size with respect to the AMR narrowband coder.

The coder has a frame size of 20 ms, an LPC analysis lookahead of 5 ms, and a filtering delay of 0.5 ms, resulting in a total algorithmic delay of 25.5 ms.

5. TEST RESULTS

For AMR Wideband Selection, five candidates were tested over a wide range of wireless channels in multiple languages with and without acoustic background noise. Our coder easily passed all the Selection Rules for this competition. For overall selection criteria, such as number of requirements met and total ΔMOS relative to the requirements, our coder was ranked second. This shows that high performance be attained while retaining the advantages of an embedded coding method.

Here are some key results for our coder as compared to the reference ITU wideband standard G.722. Our 16 kb/s coder was typically equivalent to G.722 at 64 kb/s in clean speech and in street noise. In the difficult case of car noise, it was only equivalent to G.722 at 48 kb/s in one of two tests. The 24 and 32 kb/s coders had progressively higher performance in car noise, and were roughly equivalent to G.722 at 56 kb/s. At 8 kb/s, despite the influence of residual GSM channel errors, our coder was equivalent to G.722 at 48 kb/s in one of two clean speech tests. For both street and car noise, our 8 kb/s coder was equivalent to G.722 at 48 kb/s in one of two tests when using the 10% Poor or Worse criterion.

6. CONCLUSION

We have presented an embedded multi-rate wideband speech coder based on the AMR narrowband standard. A complete AMR solution based on this coder easily passed all Selection Rules and ranked second overall in the recent 3GPP AMR Wideband Selection Testing. Besides high performance, additional advantages of the embedded split-band approach include ease of implementation, reduced complexity, and simplified interoperation with narrowband speech coders.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of Vishu Viswanathan throughout this work. In addition, Wai-Ming Lai assisted in the Qualification and Selection Testing procedures.

8. REFERENCES

[1] A. McCree, "A 14 kb/s Wideband Speech Coder with a Parametric Highband Model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Istanbul, 2000, pp. 1153–1156.

[2] E. Ekudden, R. Hagen, I. Johansson, and J. Svedberg, "The Adaptive Multi-rate Speech Coder," in *Proc. IEEE Workshop on Speech Coding*, 1999, pp. 117–119.

[3] E. Paksoy, A. McCree, and V. Viswanathan, "A Variable-Rate Multimodal Speech Coder with Gain-Matched Analysis-by-Synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Munich, 1997, pp. 751–754.

[4] E. Paksoy, J. C. DeMartin, A. McCree, C. G. Gerlach, A. Anandakumar, W.-M. Lai, and V. Viswanathan, "An Adaptive Multi-Rate Speech Coder For Digital Cellular Telephony," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, 1999, pp. 193–196.