SUPER-FRAME BASED SOURCE CONTROLLED VARIABLE RATE CODING USING APPROXIMATED TRELLIS DIAGRAM

Kei Kikuiri, Nobuhiko Naka, and Tomoyuki Ohya

Multimedia Laboratories, NTT DoCoMo, Inc.
3-5 Hikari-no-oka, Yokosuka, Kanagawa, 239-8526, Japan

ABSTRACT

This paper proposes a variable rate control method for speech/audio coding; the assumption is that all super-frames passed during the connection have a constant bit rate. The method optimizes the bit rate allocated to each frame in a super-frame by using an approximated trellis diagram that represents a transition of an averaged bit rate. The approximation is to ignore the difference of memories in the encoder among the different modes. Simulations that implemented the method on AMR Wideband (AMR-WB) show that, allowing for a 100 ms additional algorithmic delay, the method achieves a maximum of about 4.3 dB improvement in the perceptual weighted Signal-to-Noise ratio, compared to constant rate coding.

1. INTRODUCTION

Multimedia applications for mobile communications such as video and audio streaming are becoming much more popular but demand is growing for higher sound quality. Since these applications are one-way and not real-time, they have more relaxed delay requirements than conversational applications. This allows the coding parameters to be optimized since longer coding intervals are possible. Because current cellular terminals support various codecs, sound quality can be improved by using different coding schemes as well as different coding bit rates and frame lengths.

Variable rate control schemes for speech coding are categorized as either open-loop schemes with analytical methods or closed-loop schemes like Analysis-by-Synthesis. Open-loop schemes have lower computational complexity than closed-loop schemes and so are more suitable for real-time applications. Most open-loop schemes utilize local speech features, for example voice detection [3][4][5], signal power [6], spectral entropy[7] and multi-level phonetic classification [8][9]. Closed-loop schemes are more interesting since they can develop more optimal bit rates if the appropriate evaluation criterion is used.

A good example of a typical closed-loop scheme is FS-CELP (Finite State CELP). It selects the lowest bit rate which achieves predetermined weighted Signal-to-Noise ratio (WSNR). Eriksson *et al.* proposed a scheme that uses cost function unified segmental SNR and bit rate. Cellario *et al.* developed a hybrid open-loop and closed loop scheme consisting open-loop voicing decision and close-loop bit rate selection with perceptual weighted distortion power. All of these schemes are applied to conversational applications.

This paper proposes a closed-loop variable rate control method using perceptual weighted distortion power without phonetic classification. Since this paper assumes that longer delay is acceptable, the method controls the bit rates of the frames within each super-frame so as to optimize the decoded signal quality within the super-frame. It uses an approximated trellis diagram that represents the transition of averaged bit rate in a super-frame. This paper introduces simulations that apply the method to the AMR Wideband (AMR-WB) encoder standardized by 3GPP as well as by ITU-T G.722.2 [2].

2. VARIABLE RATE CONTROL USING APPROXIMATED TRELLIS DIAGRAM

2.1. Super-frame Structure

This paper assumes that the input signal is encoded so that each super-frame has the same bit rate. Fig. 1 shows that each super-frame consists of N frames. The bit rate of super-frame k is R where

$$R = \frac{1}{N}\sum_{n=0}^{N-1} r_{k,n} \qquad (1)$$

and is constant; $r_{k,n}$ (n = 0, 1, …, N-1) is the bit rate of frame n in super-frame k. We locate the optimum set of $r_{k,n}$ that satisfies formula (1) while optimizing the quality of the decoded signal within a super-frame. This paper uses perceptual weighted distortion power as the criterion for quality optimization. The perceptual weighted distortion power of frame n $WD(n)$ is defined as

$$\begin{aligned} WD(n) &= \sum_{t=0}^{T-1} wd(t)^2 \\ &= \sum_{t=0}^{T-1} \left(ws(t) - w\hat{s}(t) \right)^2 \end{aligned} \qquad (2)$$

where T is the length of a frame, $wd(t)$ is the perceptual weighted distortion, and $ws(t)$ and $w\hat{s}(m)$ are the perceptual weighted input signal and perceptual weighted synthesized signal, respectively.

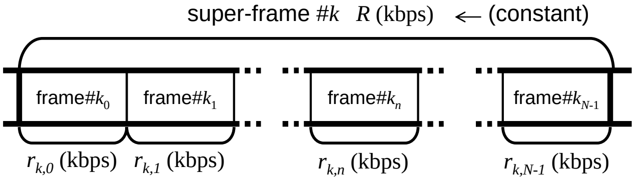


Fig.1. Super-frame structure.

2.2. Approximated Trellis Diagram Representing Transition of Averaged Bit Rate

This paper proposes a variable rate control method that uses a trellis diagram to represent the transition of averaged bit rate in each super-frame. The trellis diagram for the parameter set shown in Table 1 is shown in Fig. 2. In this figure, the nodes denote the averaged bit rates after coding the corresponding frames. The branches and the paths denote the coded bit rates and the set of bit rates, respectively. This paper uses the Viterbi algorithm with the metric of the perceptual weighted distortion power to locate the best path (the set of bit rates from the first to the last frame). Since the proposed method determines the best path super-frame by super-frame, its algorithmic delay is one super-frame.

An important point is that this trellis diagram is approximated so that the paths, whose averaged bit rates are equal, degenerate to one node. The memories in the encoder are truly different among the bit rate sets. For example, when we accurately illustrate the trellis diagram in Fig. 2, the node of bit rate 20 kbps at frame 3 is as described by Fig.3 (a). In comparison, the approximated trellis diagram (Fig.3 (b)) has fewer branches. Therefore, using the approximated trellis diagram reduces the number of encoding procedures.

Without the approximation technique, the trellis diagram is equivalent to exhaustive search. In the case of Table 1, all possible sets of $r_{k,n}$ total 75. In order to calculate the sum of the perceptual weighted distortion power for each $r_{k,n}$ set, the exhaustive search method needs to process the encoding procedure of 450 frames because one super-frame consists of 6 frames in this case. This imposes excessive computational costs. On the other hand, the approximated trellis diagram has 53 branches. Consequently, the proposed method processes the encoding procedures of 53 frames to identify the best path. This represents an 88% reduction in computational complexity.

Due to this approximation, however, the best path is not strictly optimal. The next section uses computer simulations to evaluate the impact of the approximation.

Table 1. Example of parameters.

Bit rate R (kbps)	19.85
Super-frame length N (frame)	6
Selectable bit rates at each frame $r_{k,n}$ (kbps)	23.05 19.85 18.25

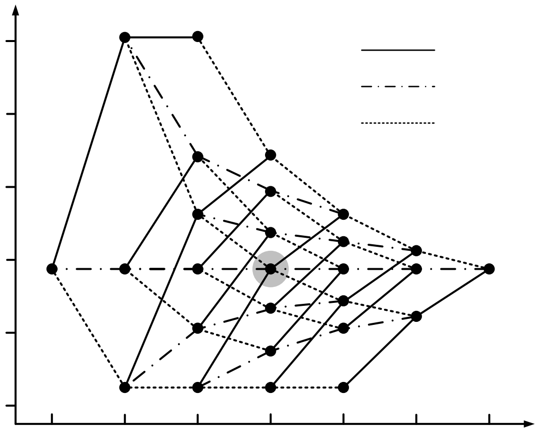
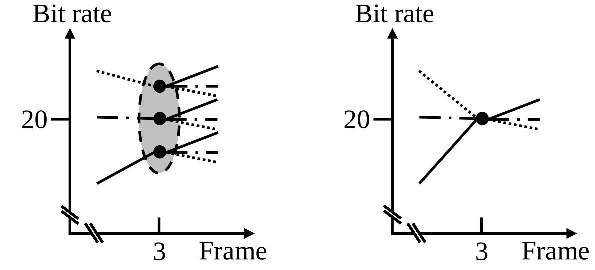


Fig.2. Trellis diagram.



(a) Without approximation. (b) With approximation.

Fig. 3. Example of trellis diagram with/without approximation.

3. COMPUTER SIMULATIONS AND RESULTS

3.1. Computer Simulations

The computer simulations implemented the proposed method on AMR-WB with the parameters shown in Table 2. Since the method needs no modification of AMR-WB bit streams, output bit streams can be decoded by existing AMR-WB decoders. The quality of the method was compared with that of the constant rate coding (the bit rate of which equaled that of the super-frames) and the exhaustive search method.

Both modes used the trellis diagrams in Fig.2. In mode 2, however, the dashed lines and the dotted lines correspond to 18.25 kbps and 15.85 kbps respectively. The criterion to optimize the bit rate sets is the perceptual weighted distortion power calculated at 12.8 kHz sampling frequency; this is already used in the AMR-WB encoder.

The performance of proposed method is evaluated by the segmental WSNR which is calculated at 12.8 kHz sampling frequency. The segmental WSNR $WSNR_{seg}$ is defined as

$$WSNR_{seg} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{l=0}^{L-1} ws(l)^2}{\sum_{l=0}^{L-1} wd(l)^2} \quad (3)$$

where M is the number of segments and L (6 frames in these simulations) is segment length.

Table. 2. Parameters for the computer simulations.

	Mode 1	Mode 2
Bit rate R (kbps)	18.25	19.85
Super-frame length N (frame)	6	
Selectable bit rates at each frame $r_{k,n}$ (kbps)	23.05 18.25 15.85	23.05 19.85 18.25
Materials	Speech (female and male), speech with BGM, flute, piano, instrumental pop	

3.3. Results and Discussions

Figs. 4 - 9 show the segmental WSNR for each material. They show that the proposed method performs significantly better than constant rate coding, especially for speech signals. The maximum improvement was 4.3 dB. It offers a smaller improvement for the music signals, but this appears to be due to the poor capability of AMR-WB in handling music signals.

In comparison to the exhaustive search method, the segmental WSNR of the proposed method is degraded by only 0.5 dB in the worst case. However, the computational complexity is reduced by 88%as mentioned in section 2.2. We can see that the approximation of the trellis diagram significantly reduces the computational complexity at the cost of an insignificant degradation in decoded signal quality.

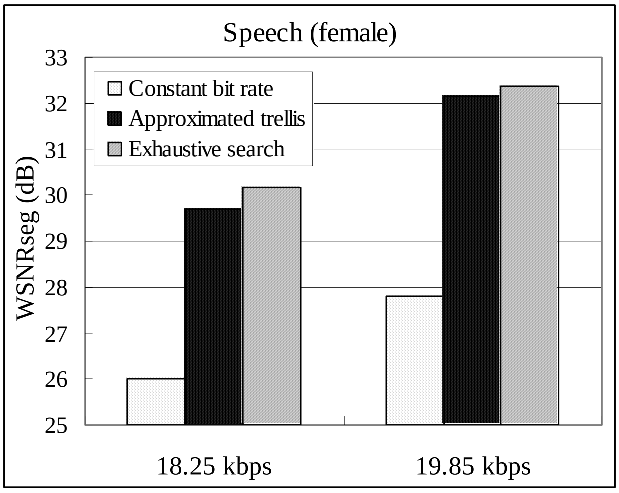


Fig.4. Speech (female).

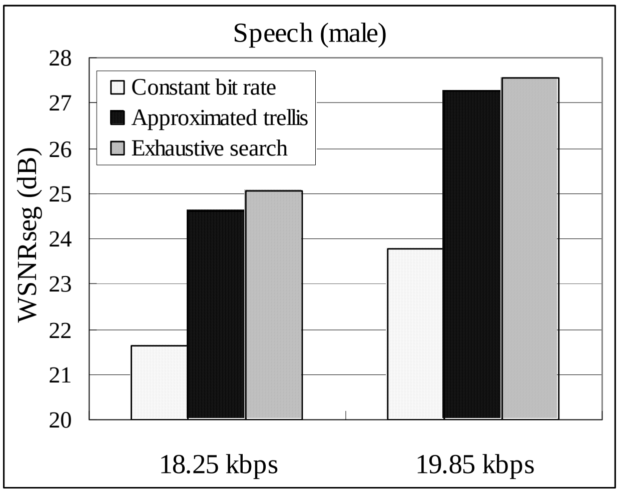


Fig.5. Speech (male).

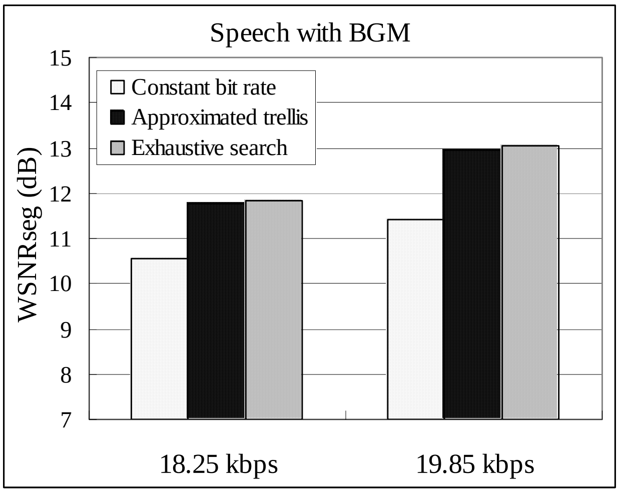


Fig. 6. Speech with BGM.

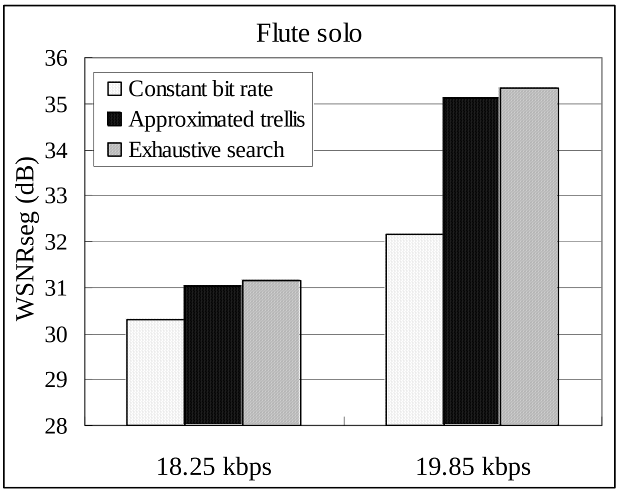


Fig. 7. Flute.

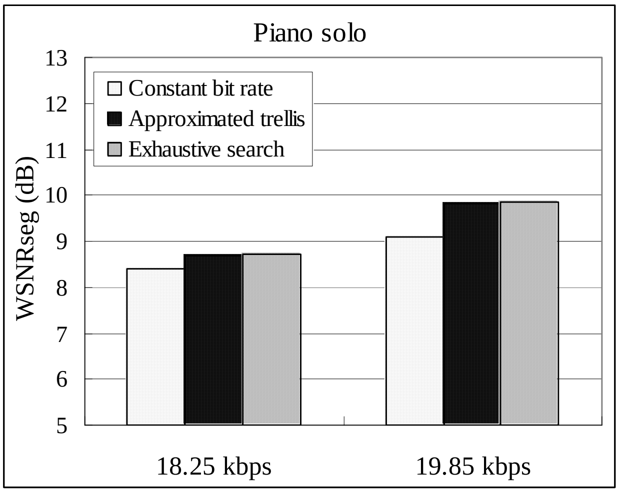


Fig. 8. Piano.

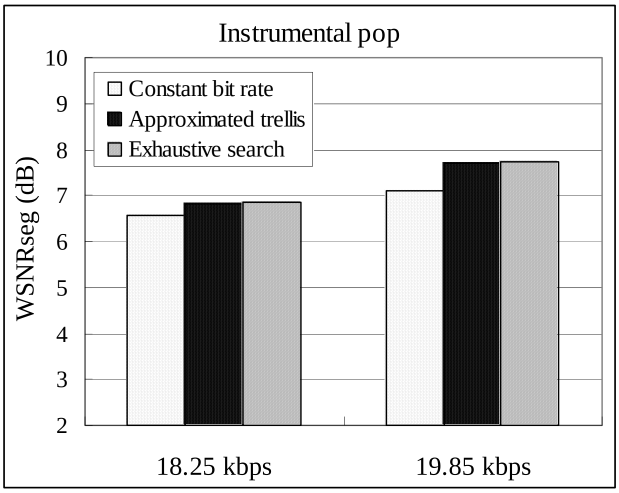


Fig. 9. Instrumental pop.

4. CONCLUSION

This paper proposed a variable rate control method that uses an approximated trellis diagram to optimize the bit rate allocation for frames in a super-frame; the assumption is that each super frame has the same bit rate. The trellis diagram approximation is to ignore the difference of the memories in the encoder among the modes. This approximation greatly reduces the computational complexity and the quality degradation is insignificant. Computer simulations of an AMR-WB implemented confirmed that the method offers significantly better performance than constant rate coding and achieved a maximum of 4.8 dB improvement in segmental WSNR. It can be applied to other codecs and also supports variable frame length control and variable coding scheme control.

REFERENCES

[1] 3GPP TS 26.190, “AMR Wideband Speech Codec; Transcoding Functions,” Feb. 2002.

[2] ITU-T G.722.2, “Wideband Coding of Speech at around 16 kbit/s using Adaptive Multi-rate Wideband (AMR-WB),” Jan. 2002 (pre-published).

[3] W. B. Kleijn, “Continuous Representations in Linear Predictive Coding,” *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 1, pp.201-204, 1991.

[4] Y. Shoham, “High-quality Speech Coding at 2.4 to 4.0 kbps based on Time – frequency Interpolation,” *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 2, pp. 167-170, 1993.

[5] J. –M. Muller and B. Wachter, “A Codec Candidate for the GSM Half Rate Speech channel,” *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 1, pp. 257-260, 1994.

[6] A. DeJaco, W. Gardner, P. Jacobs and C. Lee, “QCELP: the North American CDMA Digital Cellular Variable Rate Speech Coding Standard,” *Proc. IEEE Workshop on Speech Coding for Telecom.*, pp. 5-6, 1993.

[7] S. McClellan and J. D. Gibson, “Variable-Rate CELP based on Subband Flatness,” *IEEE Trans. Speech and Audio Process.*, vol. 5, No. 2, pp. 120-130, Mar. 1997.

[8] E. Paksoy, K. Srinivassan and A. Gersho, “Variable Rate Speech Coding with Phonetic Segmentation,” *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 2, pp. 155-158, 1993.

[9] A Das and A. Gersho, “A Variable-Rate Natural-Quality Parametric Speech Coder,” *Proc. IEEE Int. Conf. on Commun.*, vol. 1, pp. 216-220, 1994.

[10] S. V. Vaseghi, “Finite State CELP for Variable Rate Speech Coding,” *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, pp.37-40, 1990.

[11] T. Eriksson and J. Sjoberg, “Evolution of Variable Rate Speech Coders,” *Proc. IEEE Workshop on Speech Coding for Telecom.*, pp. 3-4, 1993.

[12] L. Cellario, D. Sereno, M. Giani, P. Blocher and K. Hellwig, “A VR-CELP Codec Implementation for CDMA Mobile Communications,” *Proc. IEEE Int. Conf. Accoust., Speech, Sig. Process.*, vol. 1, pp. 281-284, 1994.

[13] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*, Elsevier Science B.V., Amsterdam, 1995.