HTML Paper

A Beat-Pattern based Error Concealment Scheme for Music Delivery with Burst Packet Loss

Ye Wang
Speech and Audio Systems Laboratory
Nokia Research Center
P.O.Box 100, FIN-33721 Tampere, Finland

ABSTRACT

Error concealment is an important method to mitigate the degradation of the audio quality when compressed audio packets are lost in error prone channels, such as mobile Internet and digital audio broadcasting. This paper presents a novel error concealment scheme, which exploits the beat and rhythmic pattern of music signals. Preliminary simulations show significantly improved subjective sound quality in comparison with conventional methods in the case of burst packet losses. The new scheme is proposed as a complement to prior arts. It can be adopted to essentially all existing perceptual audio decoders such as an MP3 decoder for streaming music.

1. INTRODUCTION

The transmission of audio signals in compressed digital packet formats, such as MP3, has revolutionized the process of music distribution. Recent developments in this field have made possible the reception of streaming digital audio with handheld network communication devices, for example. However, with the increase in network traffic, there is often a loss of audio packets because of either congestion or excessive delay in the packet network, such as may occur in a best-effort based Internet.

Under severe conditions, for example, errors resulting from burst packet loss may occur which are beyond the capability of a conventional channel-coding correction method, particularly in wireless systems such as GSM, WCDMA or Bluetooth. Under such conditions, sound quality may be improved by the application of an error-concealment algorithm. Error concealment is an important process used to improve the quality of service (QoS) when a compressed audio bitstream is transmitted over an error-prone channel, such as found in mobile network communications and in digital audio broadcasts.

The focus of the new scheme is given to bitstream errors in the compressed domain, because a compressed domain bitstream, after removing most of the signal redundancy and irrelevance, is more sensitive to channel errors in comparison with an uncompressed domain bitstream.

With sufficient overhead and cost of the codec, it is theoretically possible to devise a perfect error detection and correction method. However, such a scheme would be impractical and undesirable. A practical error-correction method balances those limitations against the probability of uncorrected errors, and allows severe errors to remain uncorrected [1]. Then, error concealment methods are used as the last resort to mitigate the degradation of the audio quality in case of uncorrected errors.

In principle, all error concealment methods exploit the correlation of the signal and characteristics of human hearing to reduce the effects of uncorrected errors or packet losses. Since all perceptual audio codecs use frame-wise compression of audio signals, the new scheme is designed as segment-oriented error concealment in connection with an audio decoder.

Though error protection (detection/correction) and error concealment methods are closely related, they are, however, different concepts for tackling errors. A good system design should include [2][3]:

- Detailed analysis of the channel status and error pattern. This is the basis for choosing an appropriate error concealment strategy. The error detection is a prerequisite for error concealment.
- Careful consideration of the interdependency among channel coding, source coding and error concealment, in order to find the optimal trade-off between error resilience and bandwidth efficiency.

This paper presents a new error concealment scheme to exploit the beat and rhythmic pattern of music signals. This long-term time domain correlation has not been exploited in any existing perceptual audio-coding algorithm. Our preliminary simulation shows promising results in comparison with conventional methods in the case of long burst packet loss, which does happen now and then in the Internet [4] and Wireless LANs [5]. Even with one or two packet loss, the proposed method may produce better results than prior arts.

This paper is organized as follows. A brief review of the prior arts is given in section 2. Then our new concept and method are described in section 3. Some preliminary evaluations of the new scheme are presented in section 4. Finally, section 5 concludes the paper with some discussions and indicates some future work.

2. PREVIOUS METHODS

A lot of investigations into error concealment have been conducted during the development of a digital audio broadcasting (DAB) system within the EUREKA Project 147 [2][3][6]. A good summary of previous methods can be found in [7]. A more recent method can be found in [8].

The most relevant prior arts for error concealment employ small segments (typically around 20 ms) oriented concealment methods including: 1) muting, 2) repeating prior packet, 3) interpolation, and 4) time-scale modification. However, a fundamental limitation of conventional error concealment systems is that they all operate with the assumption that the audio signals are stationary. Thus, if the lost or distorted portion of the audio signal includes a short transient signal, such as a 'beat,' the conventional system will not be able to recover the signal. This paper presents a first attempt to solve this problem by exploiting the beat and rhythmic pattern of music signals.

3. PROPOSED METHOD

The new error concealment scheme results from the observations that a music signal typically exhibits rhythm and beat characteristics, which do remain fairly constant. This is one of the most important features that makes the music flow unique and differentiates it from other audio signals.

A segment of audio data lost from one defined interval can be replaced by a segment of audio data from a corresponding preceding interval. By exploiting the beat pattern of music signals, error concealment performance can be significantly improved, especially in the case of long burst packet loss.

In western music, especially pop music, it is well known that beat patterns are composed of regularly spaced strong and weak beats. For the sake of brevity, we have considered only pop music with clear time signature of 4/4 in this paper. The block diagram of the proposed system is shown in Figure 1. An MP3 decoder is used to perform all simulations.

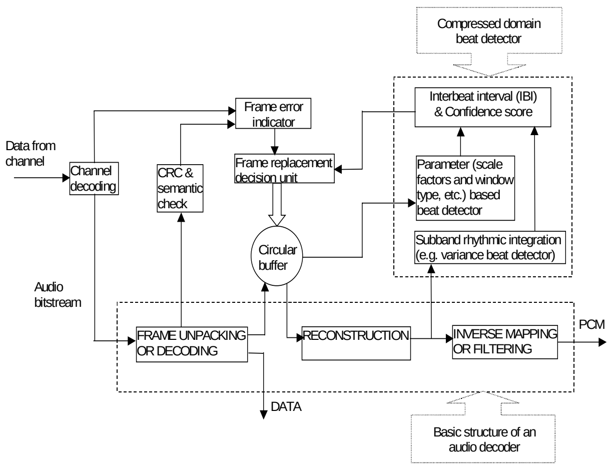

Figure 1. Block diagram of an extended audio decoder system including an error detection section, a compressed domain beat detector and a circular FIFO buffer in accordance with the proposed error concealment algorithm.

3.1 Frame Error Detection

The channel decoder is able to derive some information concerning the reliability of the received frames (status-bit). The Frame-CRC and semantic check in MPEG audio can also provide frame error information. In the case of packet-based network, the time stamp of the packet is a reliable cue for missing packets. The frame error indicator in Figure 1 analyzes the type, structure and duration of the errors. The determination of the suitable error concealment technique is based on these results. Optionally, the encoder should provide the determination and transmission of some concealment control information. It could directly point to the best error concealment strategy for a defined error situation [2].

3.2 Compressed Domain Beat Detector

Beat refers to a perceived pulse marking off equal duration units [9]. Beats are usually created by certain instruments such as drums and bass guitars.

The beat detector tries to determine the beat location, beat width and inter-beat interval. A detailed description of the compressed domain beat detector will be published elsewhere. The key ideas are summarized in this section.

In this paper, beat detection is accomplished by two methods. The more reliable method uses the energy of the music signal, which is derived from decoded Modified Discrete Cosine Transform (MDCT) coefficients available in an MP3 decoder. This method detects primarily strong beats. An adaptive statistical model is employed for improved detection accuracy. The second method uses a window-switching pattern to identify the beats present. The window-switching method detects both strong and weak beats. However, the window-switching method alone is not reliable, thus must be applied together with other more reliable methods.

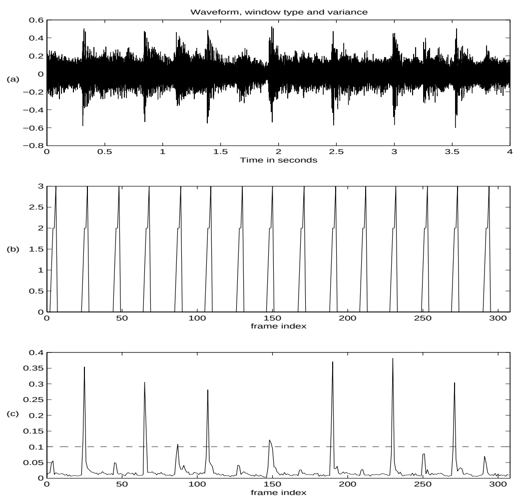

Figure 2. A sample of a pop music recording from ABBA. (a) time domain waveform, (b) window switching pattern, the vertical axis values indicate the window types: 0 – long window, 1 – long to short window, 2 – short window, 3 – short to long window, (c) energy of the signal and a threshold for beat detection.

In accordance with the energy method, the energy $EN(\tau)$ of the music signal at time τ is calculated directly by summing the squares of the decoded MDCT coefficients to give:

$$EN(\tau) = \sum_{j=0}^{575} [X_j(\tau)]^2 \qquad (1)$$

where $X_j(\tau)$ is the jth MDCT coefficient decoded at time τ. The location of the beats are determined to be those places where $EN(\tau)$ exceeds a pre-determined threshold value (see dashed line in Figure 2 (c)).

A confidence score on beat detection is included to the audio decoder system in Figure 1 to prevent erroneous beat replacement. The confidence score measures how reliably beats can be detected within an observation window. Accordingly, a threshold value is specified. If the confidence score is above the threshold value, the beat replacement is enabled. Otherwise, the beat replacement is disabled.

3.3 Error Concealment

After the error type and duration has been determined, and the beat pattern has been detected, the error concealment is fairly straightforward.

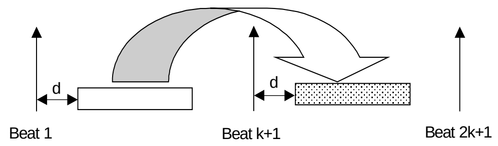

Figure 3. The replacement of an erroneous audio segment in an inter-beat interval using the system of Figure 1. k is a positive integer.

Figure 3 illustrates the replacement procedure. In this case, the audio frames making up the first inter-beat interval have been found error-free. If errors are detected between beat (k+1) and (2k+1) by the frame error indicator, the erroneous segment will be replaced by a corresponding segment from the first inter-beat interval as indicated by the arrow in Figure 3.

For music signals with time signature of 4/4, the error concealment can be performed in consecutive bars as indicated in Figure 4.

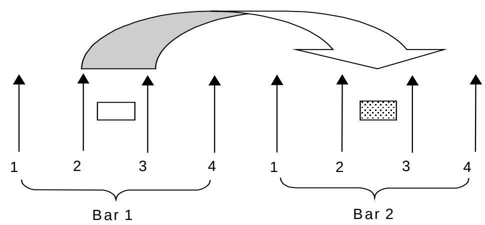

Figure 4. The replacement of an erroneous audio segment in a bar of music using the system of Figure 1.

The above error concealment configuration would require considerable memory consumption and delay in the decoder. In order to save memory and to restrict delay, an alternative configuration stores only selected audio frames around beats rather than every audio frame in the coming bitstream as illustrated in Figure 5. When the reduced memory capacity is used, only the beat structure is preserved. In this case, it is desirable to combine the new method and the conventional method to achieve better error concealment.

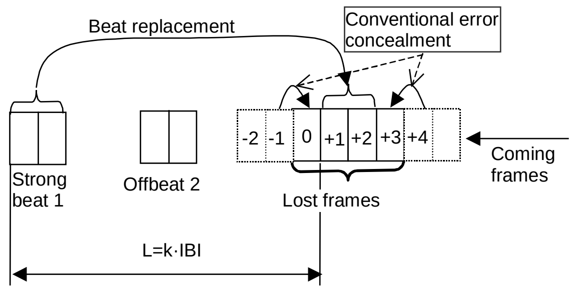

Figure 5. A scenario of burst error concealment with both the new and conventional methods. IBI indicates the inter-beat interval and k is a positive integer.

4. PRELIMINARY EVALUATIONS

As the aim of the concealment process itself is to avoid the degradations in signal quality perceived by the listener, the performance criterion can be formulated as the best restoration of the distorted signals in terms of subjective signal quality [7].

The proposed method belongs to the non-estimating algorithm, which does not attempt to give an optimum estimate but uses a replacement signal which is close enough to the original data in its structure [7].

The non-estimating technique substitutes a whole period of audio data by some other, more or less similar segment, which is available to the algorithm. Thus the processed signal is not intended to approximate the original one and a measurement of the output signal in respect to the reference signal makes no sense.

Nevertheless, we have performed some informal listening tests to evaluate the new algorithm in comparison to conventional ones.

Figure. 6 presents a comparison of the new method with some conventional methods. An error-free audio segment is represented in the top graph by two consecutive inter-beat intervals.

Consider an audio data loss between the two dotted lines, it corresponds to an interval approximately 520 ms in duration (i.e., approximately 20 MP3 audio data frames). Because most conventional error-concealment methods are not intended to deal with errors longer than one audio frame length used in the applied transfer protocol in duration, the conventional error concealment method will not produce satisfactory results. One

0-7695-1198-8/01/$10.00 (C) 2001 IEEE

conventional approach, for example, is to mute the entire segment, as shown in the next graph. Unfortunately, this waveform will be objectionable to a listener as there is an abrupt transition, and the second strong beat is missing.

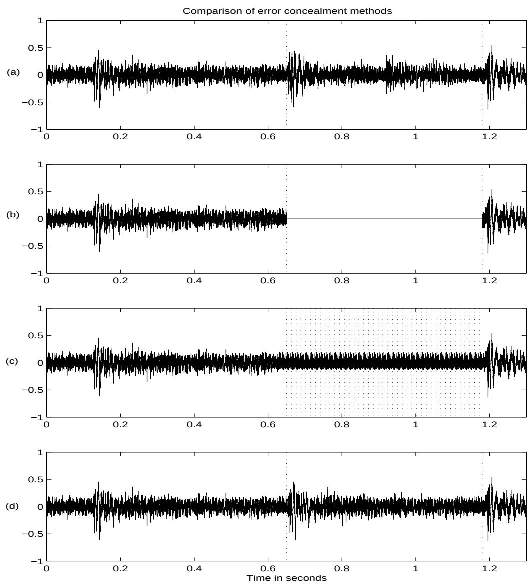

Figure 6. Comparison of the new error concealment method with some conventional methods. (a) waveform of original music signal; (b) muting the long burst errors between the two dotted lines; (c) repeating the previous MP3 frame in case 20 consecutive MP3 frames are lost; (d) beat-pattern based error concealment method.

In another conventional approach, shown in the underlying graph, an audio data frame occurring just before the lost segment is repeatedly copied and added to fill the entire interval, resulting in a monotonic waveform in Figure 6(c). This configuration will also be objectionable to a listener, as there is little if any musical content in the monotonic waveform, and the second strong beat is also missing.

The proposed method exploits the beat pattern knowledge and substitutes the missing audio segment from the previous inter-beat interval as shown in the bottom graph. By casual evaluation by some researchers at Nokia Research Center, the new method provides very promising results.

5. DISCUSSION

In this paper we have described a new error concealment technique for streaming music via error prone channels. The new method has demonstrated its capability to recover burst packet loss, which may include transient parts such as beats in music signals.

The experiments with different audio material have revealed that the current algorithm is quite effective for pure music signals with an obvious and constant beat pattern. If the signal does not have a clear beat structure or if speech and singing are considered, the current system cannot guarantee satisfactory results, because the beat-pattern does not give sufficient information about the similarity between different speech and singing segments. However, it can serve as a basis for future work in this direction.

Future research may include:

- The compressed domain beat detector employed in the proposed system may be generalized into a multi-band approach for improved beat detection.
- A segment similarity measure may be introduced in the encoder side in order to reduce the "blind" segment replacement. A good audio similarity measure should take not only music but also speech and singing sounds into account.
- An intelligent selection agent may be developed to choose the right error concealment method in a given error condition.
- Simulations with a realistic streaming channel such as mobile Internet to give some quantitative information about the subjective performance of the method.

6. ACKNOWLEDGEMENT

The author wishes to thank P. Haavisto, N. Courtis, M. Väänänen, M. Vaalgamaa, M. Vilermo, J. Tian, N. Zacharov, D. Isherwood, J. Ojanperä (Nokia Research Labs) for helpful discussions. Academy of Finland and Nokia Foundation are acknowledged for funding support.

7. REFERENCES

[1] K.C. Pohlmann, "Principles of Digital Audio," 3rd Ed., NY, McGraw-Hill, Inc. 1995
[2] D. Wiese, "Error Concealment Strategies for Digital Audio Broadcasting," 92nd AES Convention, Vienna, 1992, preprint 3264
[3] D. Wiese, "Error Concealment for DSB: Impact on Channel and Source Coding," 94th AES Convention, Berlin, 1993, preprint 3467
[4] J.C. Bolot, H. Crepin, A.V. Garcia, "Analysis of Audio Packet Loss in the Internet," Proc. of 5th Int. Workshop on Network and Operating System Support for Digital Audio and Video, pp. 163-174, Durham, April 1995
[5] http://www1.acm.org/sigs/sigmm/MM2000/ep/mckinley/
[6] M. Barberis, E. F. Schroeder, "Burst Error Concealment for Digital Audio Tape and Broadcast Application," 90th AES Convention, Paris, 1991, preprint 3012
[7] J. Herre, E. Eberlein, "Evaluation of Concealment Techniques for Compressed Digital Audio," 94th AES Convention, Berlin, 1993, preprint 3460
[8] A. Stenger, K.B. Younes, R. Reng, B. Girod, "A New Error Concealment Technique for Audio Transmission with Packet Loss," Proc. EUSIPCO-96, pp.1965-1968, Trieste, Italy, 1996
[9] W.J. Dowling, D.L. Harwood, "Music Cognition," Academic Press, 1986