A DRUMBEAT-PATTERN BASED ERROR CONCEALMENT METHOD FOR MUSIC STREAMING APPLICATIONS

*Ye Wang, Sebastian Streich**

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

ABSTRACT

This paper presents a novel drumbeat-pattern based error concealment scheme, which detects the drumbeat-pattern of music signals on the encoder side and embeds the beat information as ancillary data in a preceding data unit in the compressed bitstream. The embedded beat information is then used to perform an error concealment task on the decoder side. The proposed method was implemented using an MPEG-4 AAC (Advanced Audio Coding) codec. Informal evaluations have shown that the proposed active error concealment method clearly improved the overall subjective sound quality in comparison with conventional methods if the packet losses include the drumbeats.

1. INTRODUCTION

The transmission of compressed digital audio, such as MP3, over the Internet has already shown a profound effect on the traditional process of music distribution. Recent developments in this field have made possible streaming digital audio using mobile terminals, 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.

The wireless channel is another source of error that can also lead to packet loss. Under such conditions, sound quality may be improved by the application of an error-concealment algorithm. Error concealment is usually a receiver-based error recovery method, which serves as the last resort to mitigate the degradation of audio quality when data packets are lost in audio streaming over error prone channels such as mobile Internet.

We emphasize the importance of error concealment of compressed audio based on the following facts: 1) Streaming uncompressed audio over wireless channel is simply an uneconomic use of the scarce resource. 2) A compressed audio bitstream, after removing most of the signal redundancy and irrelevance, is more sensitive to channel errors in comparison with an uncompressed bitstream.

The most relevant prior arts for error concealment employ small segments (typically around 20 ms) oriented concealment methods including: 1) muting, 2**)** packet repetition, 3) interpolation, 4) time-scale modification, and 5) regeneration-based schemes. The use of packet repetition is recommended as offering a good compromise between achieved quality and excessive complexity [1].

However, a fundamental limitation of packet repetition and other existing error concealment schemes is that they all operate with the assumption that the audio signals are short-term stationary. Thus, if the lost or distorted portion of the audio signal includes a short transient signal, such as a drumbeat, the conventional methods will not be able to produce satisfactory results.

We illustrate possible problems of the simple packet repetition approach in Figure 1. If the drumbeat is replaced with other signals such as singing, the drumbeat is simply missing as in Figure 1 (a). If the drumbeat is copied to its following packet, it may result in a subjectively very annoying distortion, which we define as a *double-drumbeat effect,* as shown in Figure 1 (b). The degree of annoyance of the double-drumbeat effect depends on the distance between the original drumbeat and the generated one due to packet repetition. The double-drumbeat effect becomes imperceptible if the distance is sufficiently small.

We therefore developed a method to exploit the beat pattern of music signals for error concealment purposes

* Sebastian Streich is a MSc student at Ilmenau Technical University, Germany. The work for this paper was performed during an internship at Nokia Research Center.

[2][3]. The rationale for this type of approach was that a segment around a drumbeat is subjectively more similar to a segment around a previous drumbeat than its immediate neighboring segment.

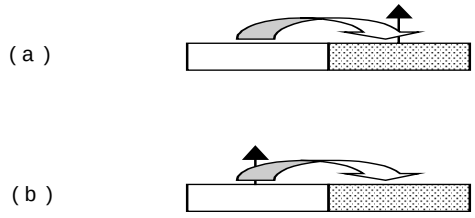

Figure 1. Illustration of possible problems with the simple packet repetition approach. Rectangles filled with dots represent corrupted packets. Blank rectangles represent error-free ones. The thin arrows indicate the drumbeats and the thick arrows indicate packet repetition operations. (a) drumbeat eliminated, (b) double-drumbeat created.

In order to simplify the following discussion we distinguish a *unit* of data from a *packet*. A unit is an interval of audio data such as an MPEG-4 AAC [4] frame, which consists of 1024 frequency components. A packet comprises one or more units, encapsulated for transmission over the network [1]. Because of possible application of interleaving, we use *unit* as the basic data segment in the following sections.

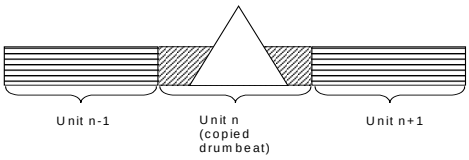

Figure 2. Illustration of the *spectral fine structure disruption effect.* The rectangle filled with upward diagonals represents the harmonic structure around the copied drumbeat. The triangle represents the drumbeat. The rectangles filled with horizontal lines represent the harmonic structures in the neighboring units around the missing unit.

Although the scheme in [2][3] has shown better results than simple packet repetition, it still has some serious limitations, especially when we focus on single packet loss. These limitations are illustrated with the help of Figure 2. First, as the lost unit n is replaced by a previous drumbeat, it is likely to create a distortion, which we define as a *spectral fine structure disruption effect*. The spectral fine structure is particularly important for the harmonic and melodic parts of the music, such as singing. Although a typical drumbeat lasts about 100 ~ 200 ms, it is not reasonable to assume that a drumbeat in the entire duration of unit n is always loud enough to mask other signals such as singing. The masking effect depends on the relative strength of the drumbeat and the singing. Thus, if there is singing around the drumbeat, we can expect such discontinuity on the unit boundaries with our previous method [2]. Second, the double-drumbeat effect is reduced but not eliminated with our receiver-based method. These problems become even more evident as we move from MP3 to AAC with the unit duration almost doubled from 576 to 1024 MDCT (Modified Discrete Cosine Transform) coefficients, which affect about 26 ms and 46 ms time domain samples respectively if the sampling frequency is 44.1 kHz.

In the following sections we describe some new approaches, which are introduced to overcome above mentioned limitations of our previous method.

2. PROPOSED METHOD

While our previous work was concentrated on receiver-based error concealment with burst packet losses, this paper focuses on a sender-based active error concealment designed mainly to deal with single packet losses. The rationale for this shift is as follows. First, the majority of packet losses in streaming applications are single packet losses [5][6]. But even these single packet losses can result in significant degradation in the subjective audio quality. Second, the receiver-based approach has limited the potential performance of the proposed scheme and its computational complexity was significantly higher than a simple packet repetition. Therefore, we aimed to achieve better error concealment performance, while reducing the decoder complexity to the level similar to a simple packet repetition. These objectives were achieved by adopting a sender-based approach.

2.1. System overview

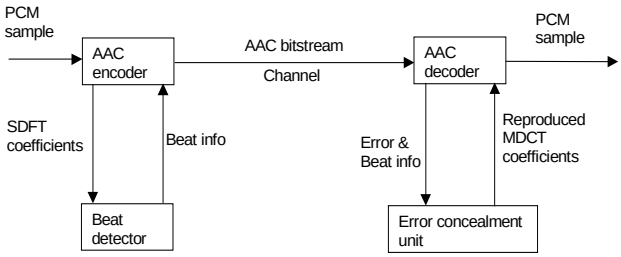

Figure 3. System overview.

The overall system comprises the blocks illustrated in Figure 3. The implementation is based on an MPEG-4 AAC codec. An incoming musical signal in PCM format is fed to the AAC encoder. The AAC performs a frequency analysis in a form of SDFT (Shifted Discrete Fourier Transform) [7]. The beat detector uses SDFT based FV (Feature Vector) to detect beats and then embeds the beat

information within the compressed bitstream as ancillary data at a preceding data unit. If the data unit of the beat is lost in the transmission channel, its position can still be determined by the beat information embedded in a previous data unit, since the probability of the simultaneous loss of two separate data units on the beats is extremely low. The error concealment unit on the decoder side uses the embedded beat information and error information to reconstruct the lost MDCT coefficients. The reconstructed MDCT coefficients are then sent to the AAC decoder to produce the final PCM musical samples.

2.2. Beat detection

The beat detector used here is similar to that in [3]. However, we have improved it in several ways. First, we employed the SDFT coefficients alone as the input instead of both the window types and the MDCT coefficients. Second, we have used the subband energy slope (derivative) as FV.

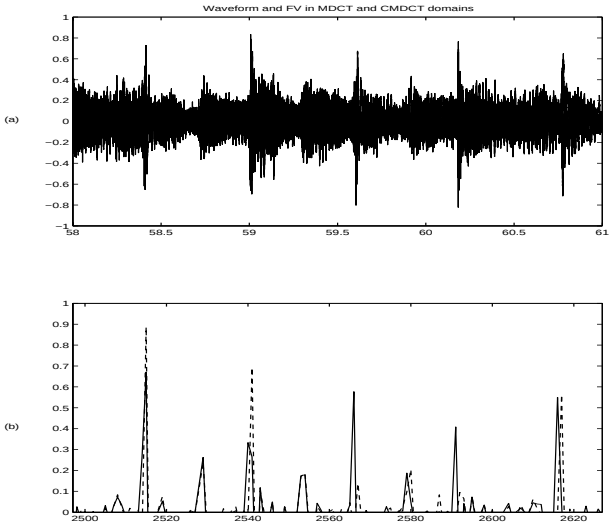

Figure 4. Music waveform and its corresponding AAC FV. (a) Music waveform versus time in seconds, (b) FVs versus AAC frame index. FVs in MDCT domain (dashed) and SDFT domain (solid).

It should be noted that the receiver-based approach has some inherent weakness. That is, the two possible inputs to the beat detector, which can be decoded from the bitstream, are the window types and the MDCT coefficients. However, both inputs can be lost in the transmission channel. In addition, MDCT does not obey Parseval's theorem, e.g. it does not preserve time domain energy [7]. This compromises the FV quality in two ways. First, the MDCT based FV fluctuates excessively over time (see the dashed line in Figure 4(b)). This makes it difficult to set a proper threshold for selecting beat candidates. Second, the maximum positions of the MDCT based FV over time jitter around the real beats by one AAC frame, while the SDFT based FV is far more stable and consistent with the position of the real beat (see Figure 4(b)). This was the rationale for us to use the SDFT based FV alone. As a result of the improved time resolution, the double-drumbeat effect is notably reduced.

The detected beat position is embedded in a previous data unit (AAC frame) for the application in the decoder. If we focus on single packet loss, only 1 bit is needed for each data unit of the audio stream to indicate whether the following data unit is on a drumbeat. If we want more protection of the beat position to tackle burst packet loss, the beat position can be embedded in two separate previous data units, which can be the preceding unit and a preceding drumbeat. In this case, some additional bits are needed to embed the position of drumbeat 3 as ancillary data in the frame at the position of drumbeat 1. Likewise, the position of drumbeat 4 is embedded in the frame at the position of drumbeat 2 as shown in Figure 5.

2.3. Error concealment

We assume the time signature of 4/4, which is valid for most intended music signals such as pop, dance and march music. According to their window types, the decoder saves the MDCT coefficients on the drumbeats in two drum-buffers for the bass drum and snare drum respectively (see Figure 5). The drum-buffers are updated if no error is detected on the drumbeat. When a packet loss is detected, the error concealment unit first checks the embedded beat information and the window types of the neighboring units. If the lost data unit (an AAC frame) is on the beat, it fetches the saved frame with a correct window type from the corresponding buffer as shown in Figure 5. This effectively eliminates the *window type mismatch phenomenon* [3].

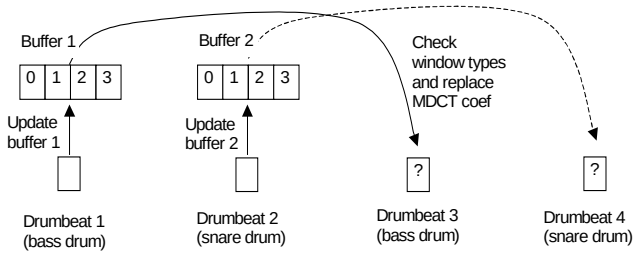

Figure 5. Illustration of the proposed error concealment operation based on drumbeat pattern. The numbers in the two drum buffers indicate the window types. The four window types (long, long-to-short, short and short-to-long) are indexed with 0, 1, 2, 3 respectively.

In order to effectively reduce the *spectral fine structure disruption effect,* we have adopted a subband approach instead of the fullband approach in [2][3]. The new subband approach is illustrated with the help of Figure 6.

The entire frequency band is divided into 3 parts. The frequency band between F1 and F2 represents the most relevant harmonic and melodic parts. The low and high frequency bands are more relevant for drumbeats. By copying the stochastic parts (drums) from a previous beat and copying the spectral fine structure from the neighboring data unit, we have achieved a very satisfactory overall subjective quality in the case of packet loss on the drumbeat.

F1 and F2 were about 344 Hz and 4500 Hz respectively. They were chosen empirically based on the spectrogram observation of the test signals and the constraints of the AAC standard. In the case of a long window, F1 corresponds to the 16th MDCT coefficient, and F2 corresponds to the 208th MDCT coefficient. In the case of the short window, F1 corresponds to the 2nd MDCT coefficient, and F2 corresponds to the 26th MDCT coefficient.

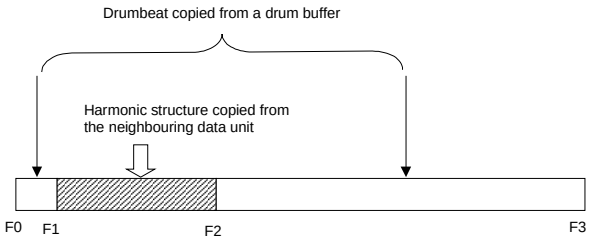

Figure 6. Illustration of the subband based error concealment. The rectangle filled with upward diagonals represents the fine spectral structure copied from a neighboring frame. The two blank rectangles represent the drum data copied from a drum buffer.

3. EXPERIMENTAL RESULTS

Informal evaluations were performed by the authors and several expert listeners, who have extensive experience in performing formal subjective listening tests. These results showed that in comparison with existing methods such as muting, simple repetition and frequency domain interpolation our new method was clearly preferred if the packet loss includes drumbeats. In comparison with our pervious receiver-based method, the proposed method significantly reduced the *spectral fine structure disruption effect* and the *double-drumbeat effect*.

4. CONCLUSION AND FUTURE WORK

In this paper, we have described a novel error concealment technique for streaming music via error prone channels. Some further optimization of the frequency band that is the most relevant harmonic and melodic parts may be required based on larger amount of test samples. The case of two-frame loss should be further investigated. We plan to perform some network simulations employing the proposed technique.

Although the new technique has demonstrated its great potential, a formal listening test is required to verify its superiority in comparison with existing techniques.

5. ACKNOWLEDGEMENT

The Academy of Finland and Nokia Foundation are acknowledged for providing the first author scholarships to initiate and to conduct this research. We thank Dr. Simon Dixon (Austrian Research Institute for Artificial Intelligence), Dr. Masataka Goto (National Institute of Advanced Industrial Science and Technology, Japan), Mr. Anssi Klapuri (Tampere University of Technology, Finland) for stimulating discussions.

6. REFERENCES

[1] Perkins, C., Hodson, O., Hardman, V., "A Survey of Packet Loss Recovery Techniques for Streaming Audio," IEEE Network, Sept/Oct 1998

[2] Wang, Y., "A Beat-Pattern based Error Concealment Scheme for Music Delivery with Burst Packet Loss", IEEE International Conference on Multimedia and Expo (ICME2001), August, 2001, Tokyo, Japan.

[3] Wang, Y., Vilermo, M., "A Compressed Domain Beat Detector Using MP3 Audio Bitstreams," Proc. of the 9th ACM International Conference on Multimedia, pp. 194-202, September 30-October 5, 2001, Ottawa, Canada.

[4] ISO/IEC JTC1/SC29/WG11 (MPEG-4), Coding of Audio-Visual Objects: Audio, International Standard 14496-3, 1999.

[5] Bolot, J.C., Crepin, H., Garcia, A.V., "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, April 1995, Durham.

[6] http://www1.acm.org/sigs/sigmm/MM2000/ep/mckinley/

[7] Wang, Y., Vilermo, M., Isherwood, D. "The Impact of the Relationship Between MDCT and DFT on Audio Compression: A Step Towards Solving the Mismatch", The First IEEE Pacific-Rim Conference on Multimedia (IEEE-PCM2000), pp. 130-138, December 13-15, 2000, Sydney, Australia.