Corrupted encoded data detection for a CELP decoder

Hiroyuki Ehara\textsuperscript{1a)}, Takuya Kawashima\textsuperscript{2}, and Toshiaki Sakurai\textsuperscript{1}
\textsuperscript{1} Panasonic AVC Networks Company, Panasonic Corporation, 600 Saedo-cho, Tsuzuki-ku, Yokohama, Kanagawa 224–8539, Japan
\textsuperscript{2} Panasonic System Networks R&D Lab. Co., Ltd., 7F Confidence Kanazawa, 1–1–3 Sainen, Kanazawa, Ishikawa 920–0024, Japan
\textsuperscript{a)} ehara.hiroyuki@jp.panasonic.com

Abstract: Undetectable data corruption in wireless communication links can occur in some cases, including cases in which a ciphering algorithm can fail to generate the same keystream for encryption at a receiver side as it does for a sender side. In wireless speech communication, such corruption might trigger an explosion of decoded sound, which might in turn hurt a user’s ear. This article presents a simple algorithm for detecting corruption for a typical code-excited linear-prediction (CELP) speech decoder. The algorithm requires no additional bits for corruption detection, but detects it by monitoring the amplitude ratio of decoded adaptive and fixed excitation signals in CELP. Its practicality is demonstrated by simulation.

Keywords: error detection, speech codec, CELP

Classification: Multimedia Systems for Communications

References

[1] 3GPP, “6.6.3 Ciphering method” in 3GPP TS 33.102 V12.2.0, p. 39, 2014.
[2] N. Gortz, “Zero Redundancy Error Protection for CELP Speech Codecs,” Proc. EuroSpeech’97, vol. 3, pp. 1283–1286, 1997.
[3] M. R. Schroeder and B. S. Atal, “Code excited linear prediction: High quality speech at low bit rates,” Proc. IEEE ICASSP-85, pp. 937–940, 1985. DOI:10.1109/ICASSP.1985.1168147
[4] J.-P. Adoul, P. Mabillement, M. Delprat, and S. Morissette, “Fast CELP coding on algebraic codes,” Proc. IEEE ICASSP-87, pp. 1957–1960, 1987. DOI:10.1109/ICASSP.1987.1169413
[5] T. Kawashima and H. Ehara, “Low bit-rate CELP decoder with corrupted encoded-data detection,” Proc. IEICE General Conference, D-14-18, Mar. 2016 (in Japanese).
[6] P. J. Erdelsky, “Rijndael Encryption Algorithm,” http://www.efgh.com/software/rijndael.htm.

1 Introduction

In wireless communication systems, error protection is commonly applied to avoid substantial degradation in communication quality caused by any error in communication links. However, in some cases, such errors are undetectable. For example,
some input parameters to a cipher algorithm can differ between the sender side and the receiver side. Synchronization of an encryption keystream generated by the cipher algorithm might be lost at the decoder side. This problem is known to occur in some mobile communication systems using the ciphering method described in 3GPP specifications [1], which use sequence numbers counted separately at each of the sender and receiver sides as input parameters. In this case, encoded data received by a decoder are corrupted completely, but they are received as correct encoded data. In speech communication systems, such corruption might trigger an explosion of decoded sound, which might in turn hurt a user’s ear when headphone-type listening devices are used. It is therefore desirable to detect such sound-explosion-causing errors using corrupted data with no additional information.

An earlier study used the difference of consecutive decoded parameters for such “zero-redundant” error detection [2]. This approach is based on the assumptions that parameters decoded without error have some correlation in time, and that the correlation diminishes if the data include errors. Errors can be detected by checking whether the difference is outside of a predetermined range. A benefit of this approach is that it enables parameter-by-parameter error detection, although some wrong error detection is tolerated because the assumption is not always true. Our objective is detecting corrupted data that might trigger an explosion of decoded sound. We started with an investigation of the source of the sound explosion and found that it was more appropriate for our objective to check excitation signals reconstructed from the decoded parameters rather than the decoded parameters themselves. By exploiting the excitation signals, we inferred a criterion that was effective for detecting explosions. Using this criterion, we developed an algorithm for corrupted encoded data detection (CED) used in a code-excited linear-prediction (CELP) decoder. Its practicability is demonstrated by experimental simulation.

![Fig. 1. Block diagram of a CELP decoder with CED.](image)

### 2 CELP decoder with corrupted encoded data detection

CELP [3], an algorithm for encoding speech signals at low bit-rates, is deployed widely in wireless speech communication systems. It is based on a speech synthesis model. Fig. 1 presents a block diagram of a CELP decoder with CED. After an excitation signal is passed through a linear predictive synthesis filter, a synthesized speech signal is outputted from the synthesis filter. Summing an adaptive codebook (ACB) vector and a fixed-codebook (FCB) vector generates the excitation signal.
The ACB, a buffered sequence of the excitation signal generated in the past, is used to represent periodic components in the excitation signal. The FCB contains several pre-fixed vectors. Such vectors can be represented in various ways. For example, in widely deployed algebraic CELP (ACELP) [4], the pre-fixed vectors are generated by combining some unit pulses located at pre-determined positions. The CED is performed using both ACB and FCB vectors. Once the corrupted encoded data are detected, attenuation is applied to the excitation signal input to the synthesis filter to avoid explosion of synthesized sound [5]. The CED algorithm and the attenuation of the excitation signal are explained herein in the following sections.

3 Criterion for detecting corrupted encoded data

The objective of this study is detecting corrupted encoded data for avoiding sound explosion generation. Therefore, we have started with investigation of the sources of sound explosions generated from corrupted encoded data. For generating corrupted encoded data, an Advanced Encryption Standard (AES) cipher is used: the Rijndael Encryption Algorithm [6]. A CELP-encoded bitstream is ciphered using 256-bit AES. Then the ciphered bitstream is decrypted using a different keystream from that used in encryption. The decrypted bitstream is then decoded by the CELP decoder. Typical sound-explosion cases are selected and investigated closely. Results show that the amplitude of the ACB vector (ACV) tends to be remarkably large in those cases. However, a remarkable ACV amplitude is found even in error-free conditions. Therefore, one cannot simply rely on the ACV amplitude to detect sound-explosion cases. In our experience, the ratio between the ACV amplitude and corresponding FCB vector (FCV) amplitude is useful as a criterion for detecting abnormalities causing possible sound explosion. In CELP, in general, the periodic component is represented with ACV. The non-periodic component is represented with FCV. However, even if the input signal is a perfectly periodic signal, FCV cannot be zero because a quantization error of the periodic component always exists. Therefore, a practical upper limit of the ACV/FCV amplitude ratio must exist. We have examined the ACV/FCV amplitude ratio using an approximately 3081-s database containing 8-language 800 short sentences with no background noise, 8-language 80 short sentences with various background noises (e.g. street noise, babble noise, car noise, train noise), 8-language 160 short sentences at different input levels, and 13 music samples. Results show that the ACB/FCB amplitude ratio never exceeds 50 for our developed CELP case. However, when the corrupted encoded data are decoded, the ACV/FCV amplitude ratio frequently exceeds 100. Fig. 2 presents histograms of the ACB/FCB amplitude ratio for both cases. From Fig. 2, it is apparent that corrupted encoded data segments are detectable by checking whether the ACV/FCV amplitude ratio exceeds, for example, 64. Although the sound explosion is not always observed in such cases, the amplitude ratio is useful as a reliable criterion for detecting the corrupted encoded data. The amplitude ratio is defined by Eq. (1). \(ACV[i]\) and \(FCV[i]\) respectively signify the decoded ACV and the decoded FCV, which are scaled by corresponding quantized gains. \(L\) represents the vector length.
4 Corrupted encoded data detection

A threshold is set on the amplitude ratio defined by Eq. (1). 64 is chosen as the threshold based on the observation in Fig. 2. However, as presented in Fig. 2, corrupted data do not always give a high amplitude ratio. The decoded sound explosion was sometimes observed after the segment in which the amplitude ratio exceeds the threshold. Therefore we have additionally introduced a heuristic approach for improving the detection performance. The decision related to the currently decoding segment can be based on the past segment as follows. The decoding segment is defined as a corrupted encoded data segment if the ratio exceeds the threshold somewhere in a past predetermined period including the currently decoded segment. To ascertain an appropriate time length for the predetermined period, we have examined the number of undetected sound explosion segments using various time lengths. Corrupted 3081-s data were used for the experiment. Results show that, of all explosion segments, fewer than 1% were undetected when the length was set to 80 ms. No undetected explosion segment was observed when the length was set to 240 ms. Therefore, the practical duration can be set between these time lengths. It is possible to avoid the occurrence of undetected explosion segments if a long time length is adopted as the predetermined period. However, more non-explosion segments are detected as explosion segments in such cases. When the data are always corrupted, such misdetection is not harmful because the decoded signal is a meaningless signal, even though it is a non-explosion signal. However, if the data can be changed from a corrupted period to a non-corrupted period, a longer predetermined period causes extension of the detected corrupted period after switching to the non-corrupted period from the

\[
\sqrt[\frac{1}{2}]{\frac{1}{L} \sum_{i=0}^{L-1} (ACV[i])^2} \div \sqrt[\frac{1}{2}]{\frac{1}{L} \sum_{i=0}^{L-1} (FCV[i])^2}.
\]
corrupted period. For consideration of such negative effects of detection performance, the F-measure was calculated for various time lengths for the predetermined period using mixed data containing both non-corrupted and corrupted periods. Corrupted data appear with 4 s intervals and have 2 s duration. Therefore, the percentage of corrupted data is 50%. Again, the 3081-s data were used for preparing mixed data. The F-measure was calculated using Eq. (2):

\[
\text{Recall} = \frac{\text{correctly detected “corrupted segments”}}{\text{actual “corrupted segments”}}
\]

\[
\text{Precision} = \frac{\text{correctly detected “corrupted segments”}}{\text{detected “corrupted segments”}}.
\]

\[
\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}.
\]

The results are depicted in Fig. 3. As depicted in Fig. 3, the highest performance exceeding 0.95 is obtainable when the time length is set around 120–160 ms. Therefore we can set the time length around this range for the best detection performance. As an attenuation factor, 0.1 is multiplied to the excitation signal at the attenuator in Fig. 1 during the corrupted encoded data segment detection. It is noteworthy that the optimal time length might depend on the CELP codec frame/sub-frame length. In this experiment, the CELP codec operates on a 40 ms frame that has five 8 ms sub-frames. Because the abnormally large amplitude ratio does not occur in the non-corrupted encoded data case, as shown in Fig. 2, CED does not work on non-corrupted encoded data and therefore does not affect decoding performance under the error-free condition. In the communication system explained in Section 1, non-synchronization of the counters counting the sequence numbers causes sustained data corruption unless the counters are reset. Aside from CED, some signaling mechanism for resetting the synchronization between the sender side and the receiver side is necessary for such systems.

Fig. 3. F-measure results calculated for various time lengths.
5 Conclusion

This article proposes a detection algorithm using the amplitude ratio of adaptive-codebook to fixed-codebook vectors in a code-excited linear prediction (CELP) decoder for detecting corrupted encoded data segments without sending additional information for error detection. The algorithm can apply to typical CELP-based decoders without affecting the performance in an error-free condition. Moreover, it can be integrated into an ordinary CELP decoding process without requiring a specific signaling mechanism. It can therefore be introduced independently to any decoder equipment without changing the entire communication system chain.