Application of the Relative Approach to Optimize Packet Loss Concealment Implementations

Frank Kettler, H.W. Gierlich, Frank Rosenberger; HEAD acoustics GmbH, Herzogenrath

1. Introduction

Data transmission over packet based networks is typically affected by packet loss or delay jitter. The resulting impairments can easily be accounted for by appropriate, long term buffering and repeating the lost packets. If the network is used for real time speech communication like in voice over IP (VoIP) applications the tolerable delay underlies restrictions [1]. Conversational quality is degraded by conversational dynamics impairment or echo. The maximum buffer size is therefore limited, consequently packet loss occurs in the transmitted signal leading to audible disturbances like gaps or crackling sounds. The listening speech quality is significantly impaired.

Appropriate measures (packet loss concealment, PLC) have to be taken against these disturbances. Current implementations range from simple fading to avoid crackling sounds, the repetition of previously received packets or more sophisticated speech interpolation algorithms. If not properly implemented and optimized the improvement in terms of minimizing the audible disturbances of lost packets are limited. Objective optimization criteria are therefore necessary.

The Relative Approach [2], a hearing model based analysis method, is a suitable test method to optimize PLC implementations. This analysis method has already been successfully used during the 2nd ETSI Speech Quality Test Event comparing various implementations of PLC of different VoIP gateway and terminal manufacturers [3].

This contribution discusses the principles of this method and gives guidelines to analyze typical results. Comparative results of different VoIP gateway implementations can be found in [3], [4].

2. Principle of the Relative Approach

Human hearing is particularly sensitive to time changes in acoustic signals, or spectral structures which contain characteristic peaks and notches. The idea behind the Relative Approach analysis is to determine an estimate value for the current signal from the signal history and to subsequently compare this estimate and the current signal value. The basis of this approach is an aurally-adequate spectroscopy. A time/frequency analysis is required reproducing the properties of human hearing. The non-linear relation between sound pressure and perceived loudness is considered by the application of aurally-equivalent level transformation [5]. In contrast to other analysis methods the algorithm does not use any reference signal. Only the transmitted signal is analyzed. The algorithm generates an "internal reference" which can best be described as a forward estimation.

The estimated value can initially be regarded as a mean value for former signal values (analysis window length 200 ms). The difference between the current signal value (window length 5 ms) and the estimated value is a measure of signal change.

Due to the time variant characteristics, speech can not directly be assessed by the Relative Approach. A suitable test signal is therefore required providing speech-like characteristics on the one hand and stationary characteristics on the other hand. An appropriate test signal consists of the periodical repetition of a voiced sound with a decreasing level versus time. This voiced sound is described in ITU-T Recommendation P.501 [6]. The complete duration of the test signal is 5 s. Two kinds of analysis are applied to the transmitted signal:

- the Relative Approach to analyze audible disturbances in the time and frequency domain and
- a cross correlation analysis between the transmitted signal and the original test signal to show the technical implementation of PLC and jitter buffer management.

The following Figures 1 and 2 show the analysis for an (undisturbed) reference signal. The upper window in both figures show the measured time sequence in green. The sequence length is 5 s in Figure 1 whereas Figure 2 shows an enlarged sequence of 400 ms.

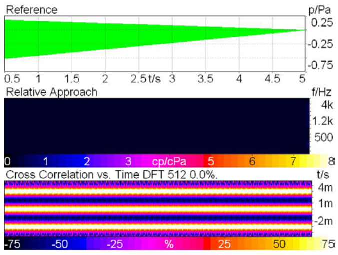

Fig. 1: Reference condition (5 s)

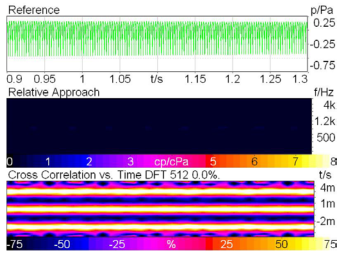

Fig. 2: Enlarged sequence (0.4 s)

The middle window shows the Relative Approach analysis result vs. time (x-axis) and frequency (y-axis between 100 Hz and 5 kHz). As described above this analysis calculates a forward estimation and compares it with the current signal. The estimation error is then color coded and displayed. The warmth of the color correlates to the estimation error, small errors are displayed in dark colors. As demonstrated in both figures no audible disturbances in the time or frequency domain are detected by this algorithm for this reference sequence.

3. Analysis examples

The following examples show analysis results for typical impairments like packet loss, fading, substitution of previous packets and jitter. In order to demonstrate the analysis purposes and the sensitivity of the Relative Approach the disturbances are simulated in these examples. Packet loss is simulated in Figure 3 and 4 by inserting a 20 ms signal gap. These 20 ms can be seen as a typical packet length used in VoIP applications [3].

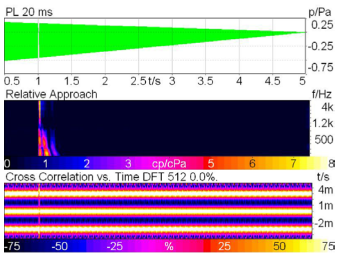

Fig. 3: Packet loss (20 ms)

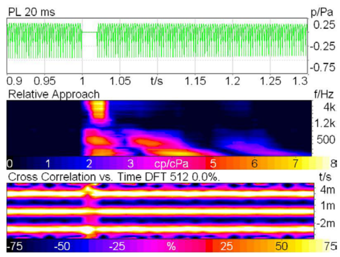

Fig. 4: Enlarged sequence

The resulting gap leads to a discontinuity in the cross correlation analysis in the lower window of Figure 3. The Relative Approach analysis indicates a wide band disturbance (see the middle window in Figure 3). Figure 4 shows the enlarged sequence demonstrating the audible disturbances in the higher and lower frequency range. The disturbances decrease more slowly in the lower frequency range compared to the higher frequency components. This is considered in the algorithm by the longer lasting excitation of the aurally adequate filter bank in the lower frequency bands used in the hearing model.

Improvements can already be achieved by applying simple fading as demonstrated in the Figures 5 and 6. In this example a 10 ms linear fading is used. Consequently the effective duration of the disturbance is slightly higher (20 ms packet length plus 10 ms fading in). The information about the signal phase is partially lost as indicated by the cross correlation analysis during the enlarged sequence in Figure 6. But the audible disturbance is significantly lower especially in the lower frequency range as indicated by the Relative Approach analysis in the middle window in Figure 6.

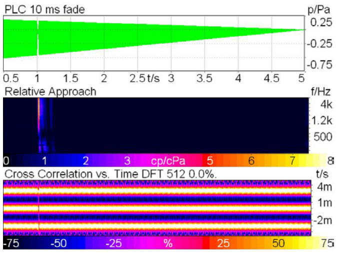

Fig. 5: Packet loss and fading

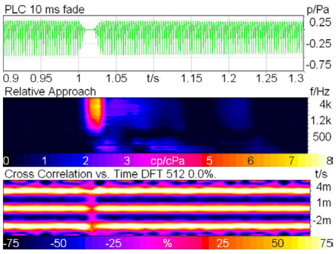

Fig. 6: Enlarged sequence

Figures 7 and 8 show the analysis results for a relatively simple but frequently implemented packet loss concealment based on a substitution of the lost information by a previously received packet. Due to this substitution process the signal phase between the substituted packet and the adjacent signal is not properly interpolated leading to strong audible disturbances mainly in the frequency range below 800 Hz. The phase difference between the substituted packet and the transmitted signal is demonstrated in the cross correlation analysis (lower window in Figure 8).

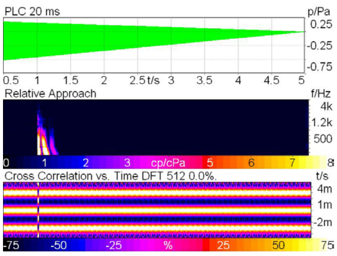

Fig. 7: Packet loss, substitution

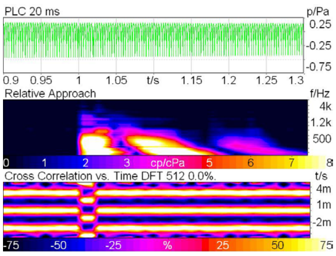

Fig. 8: Enlarged sequence

An example of the principal disturbance caused by jitter is shown in Figures 9 and 10. This is simulated by removing a 20 ms signal packet and shifting the following signal. The phase shift can easily be analyzed as demonstrated by the cross correlation analysis in the lower window of Figure 9. The audible disturbance is comparable to a packet substitution (see Figure 7) but, taking into account that the phase shift occurs only once in this simulated jitter example, the resulting disturbance is lower compared to the substituted packet with two signal phase adjustments between the adjacent signal periods.

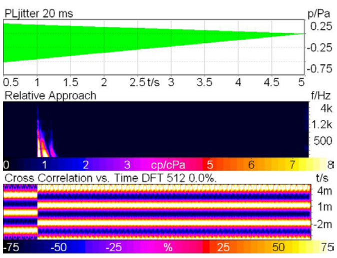

Fig. 9: 20 ms jitter

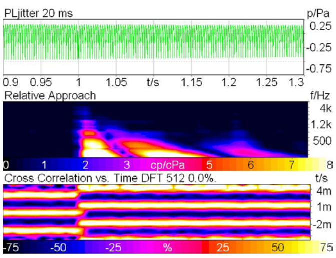

Fig. 10: Enlarged sequence

In addition to these simulated examples Figure 11 shows the result measured for a gateway to gateway connection using a G.711 speech coder [3]. The simulated IP network introduced 5% packet loss in this test. The disturbances as principally shown above for the simulated examples occur in this analysis in Figure 11:

- Packet loss concealment realized by substitution using previously received packets can be detected in the cross correlation analysis (see the phase shifts in the lower window). The resulting disturbances occur mainly in the lower frequency range (see the Relative Approach analysis).
- Speech gaps introducing audible disturbances in the whole frequency range.
- Phase shifts, probably introduced by jitter buffer alignment causing low frequency disturbances (see the Relative Approach analysis).

Speech transmission over this implementation leads to audible quality degradation requiring improvement for the current implementation.

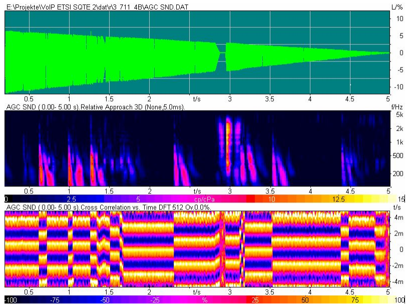

Fig. 11 Gateway-to-gateway connection, G.711 speech coder, 5% packet loss [3]

3. Correlation to listening speech quality

The Relative Approach has been successfully used during the 2nd ETSI Speech Quality Test Event for VoIP equipment [3]. These results measured for VoIP gateways and terminal equipment of different manufacturers demonstrate a high correlation to objective listening speech quality tests using PESQ [7] and TOSQA2001 [8]. Comparison tests demonstrate differences up to 1 estimated MOS caused by different PLC implementations [3].

4. Summary

Listening speech quality in VoIP scenarios is impaired by speech coders, packet loss, jitter and other parameters. Objective analysis methods like PESQ or TOSQA2001 consider this but provide only limited possibilities for optimization criteria for each individual parameter. The Relative Approach, a hearing model based analysis method, as a suitable test method to optimize PLC implementations and jitter buffer management can be combined with these measures.

5. References

[1] ITU-T Recommendation G.114 One-way Transmission Time

[2] Genuit, K.: Objective Evaluation of Acoustic Quality Based on a Relative Approach, InterNoise '96, Liverpool, UK

[3] Anonymized Test Report, 2nd ETSI Speech Quality Test Event, ETSI Plugtests, HEAD acoustics, T-Systems Nova, April 2002, Sophia Antipolis, France

[4] Kettler, F.; Gierlich, H.W.; Rosenberger, F.; Klaus, H.; Berger J.; Eisfeldt, O.; Cousin, P.; Freisse, J.-L., 2nd Speech Quality Test Event for Voice over IP, DAGA 2003, 18.-20.03.2003, Aachen

[5] Sottek, R.: Modelle zur Signalverarbeitung im menschlichen Gehör, PHD thesis RWTH Aachen, 1993

[6] ITU-T Recommendation P.501, Test Signals for Use in Telephonometry

[7] ITU-T Recommendation P.862, Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs

[8] ITU-T Contribution, Results of Objective Speech Quality Assessment including Receiving Terminals using the Advanced TOSQA2001, COM 12-20-E, December 2000, T-Nova Deutsche Telekom Innovationsgesellschaft mbH Berkom