Speech Quality Evaluation in Telephone Networks

Thorsten Ludwig

Institute for Circuits and Systems Theory,
Faculty of Engineering, Christian-Albrechts University of Kiel,
Kaiserstraße 2, 24143 Kiel, Germany
Phone: +49 (431) 880-6129, Fax: +49 (431) 880-6125
E-Mail: tlu@tf.uni-kiel.de

1. Introduction:
An In-service, Non-intrusive Measurement Device (INMD) [1, 2] extracts quality-defining parameters from an existing telephone link without disturbing the link. Such a device serves as a network monitor. A large number of these devices, placed at switches (on PCM-coded lines) in the whole network, can observe a multitude of telephone calls. A central evaluation gives evidence about the quality-of-service parameters of the network. Classical parameters to be measured are the noise level, the active speech level, and the echo loss and delay. However, in modern networks these typical parameters are not sufficient for reliable statements about the speech quality as there are additional disturbances, e.g. packet loss in IP-telephone services, frame loss in mobile communications, comfort noise inserted by DTX-Systems (Discontinuous Transmission), cascading of different digital transmission systems, and others. This paper proposes two algorithms to measure additional characteristics of telephone links. The first algorithm presented here allows to detect the GSM-FR codec [3] in transmission systems. For this purpose, a spectral attenuation around 2700Hz introduced by the coding principle is evaluated for short signal blocks. The error rate is below 5%. The second algorithm presented here samples frequency points of the background-noise spectrum throughout the duration of speech utterances, by making use of minima statistics in frequency-tracks and speech segments. These frequency points are compared to the noise in speech pauses in a statistical manner to evaluate differences and decide about the occurrence of comfort noise. The error rate for the used data base is below 5%, but further investigations are necessary to verify the algorithm.

2. Detection of the GSM-FR-Coder:
Figure 1 shows the long-term power spectrum densities of two signals, one from a pure fixed-network connection compared to one of a partly mobile (GSM-FR-) connection. The spectral reduction in the area around 2700 Hz is a feature of the GSM-FR-transmission system. The reconstruction technique in the decoder that mirrors the spectral low-intensity part of the received residual signal around 0 Hz to the area around 2700 Hz to form the excitation signal for the following synthesis filters, is responsible for this reduction in intensity. This intensity loss can be used to detect the GSM-FR-transmission system. Therefore each speech-signal block of length 16 ms of the signal under test is transformed by a DFT to the spectral domain to determine the spectral minima around 2700 Hz.

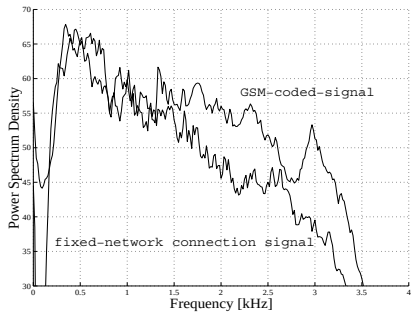

Figure 1: Power spectral density (PSD) of a solely fixed-network connection signal compared to the PSD of a mobile to fixed-network connection signal.

The evaluation of all spectral minima leads to their distribution in the spectral area of interest. Figure 2 shows two examples, one for a connection with an GSM-FR-codec involved and one without.

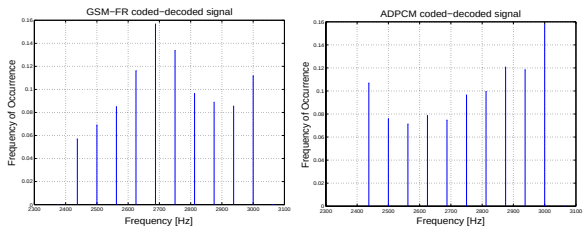

Figure 2: Left: GSM-FR connection. Right: ADPCM connection.

The distribution can be approximated by a polynomial of second order. For the GSM-FR-transmission system the polynomial shows a typical downward-opened parabola with its maximum around 2700 Hz. For other transmission systems the maximum can be found at a higher frequency, as speech has normally low-pass character. This leads to an upward-opened parabola. So the quadratic coefficient of the polynomial decides about the occurrence of the GSM-FR-transmission system. For the test of the GSM-FR detection, a database with approximately 1000 short signals with all kinds of background noise was used. The error rate for the GSM-FR-Coder detection was overall about

4%. Most errors occur in case of additional complex coders following the GSM-FR-transmission system.

3. Detection of Comfort Noise Inserted by DTX-Systems:
DTX-systems insert comfort noise in speech pauses to save bandwidth and/or energy (in mobile communications). The comfort noise should fit to the real background noise in a way that the subscribers are not disturbed by the insertion. Otherwise it is disturbing/distinguishable and should be detectable. One approach to detect comfort noise is to compare the background noise of speech pauses with the background noise during speech. To isolate background noise during speech, the following algorithm can be used: The first step is to separate speech and pause segments of the examined signal by means of a voice-activity detector (e.g. GSM VAD). After excluding uncertain segments and introducing safety margins before and after speech segments, both parts are transformed with a DFT in terms of short blocks to the spectral domain. To isolate background noise during speech in each sub-band, it is assumed that a defined part (the spectral minima) of the spectral points in each subband and/or each speech segment belongs to the background noise. Figure 3 shows as an example a signal clip with the speech/pause segmentation and the isolated spectral frequency points during speech segments and pause segments. In pause segments 1% of the spectral peaks are dropped. In a next step a 2D-histogram of the residual spectral frequency points, separated for speech and pause segments, is generated. Figures 4 and 5 show two examples. The two 2D-historgrams of the signal without comfort noise show a higher correlation than the other two, where comfort noise is inserted in pause segments. To evaluate the two 2D-histograms of one signal, the correlation ρ_k between chosen magnitude bins is calculated and averaged:

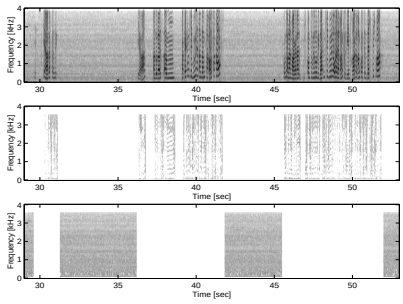

Figure 3: Upper: Signal clip. Middle: Isolated background noise frequency points during speech. Lower: Pause segments.

$$\rho = \frac{1}{N} \sum_k \rho_k \,.$$

A low correlation $\rho < 0.6$ indicates the detection of Comfort Noise. For a database with 106 signals (90 signals with comfort noise, 16 without) and all kinds of background noise (clear, office, car, street, Hoth), there were only 3 wrong decisions (3%). For a correlation threshold of $\rho < 0.55$, there is only one additional error.

5. Conclusion:
In this paper, two algorithms to measure additional characteristics of telephone links are proposed. The first one uses spectral attenuation introduced by the GSM-FR codec to detect this codec in a transmission system and works with high accuracy. The second one proposes a method to detect comfort noise in telephone links. Therefore spectral points during speech which belong to the background noise are localized. These frequency points are compared to the noise in speech pauses in a statistical manner to decide about the occurrence of comfort noise. For the tested database, the results with error rates below 4% are very promising. Nevertheless a larger database is necessary to verify the algorithm.
The author would like to thank T-Nova Innovationsgesellschaft mbH, Berlin, for the support of this research work.

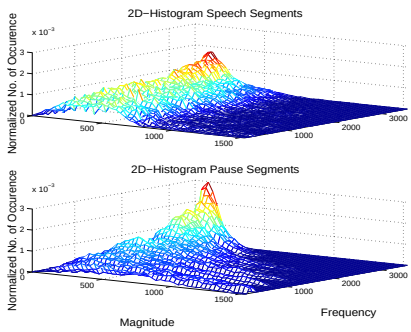

Figure 4: Upper: 2D-histogram of spectral frequency points during speech segments. Lower: 2D-histogram of spectral frequency points during pause segments (no comfort noise).

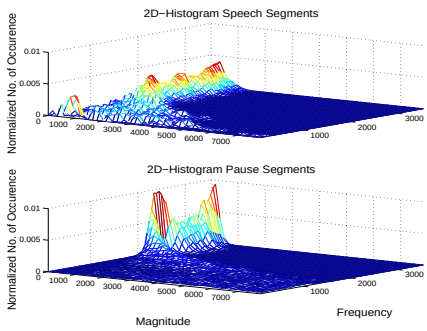

Figure 5: Upper: 2D-histogram of spectral frequency points during speech segments. Lower: 2D-histogram of spectral frequency points during pause segments (comfort noise).

References

[1] ITU-T Recommendation P.561, In-Servive, Non-Intrusive Measurement Device – Voice Service Measurements.

[2] T. Gänsler, G. Salomonsson, Nonintrusive Meausrements of the Telephone Channel, IEEE Transactions on Communications, Vol.47, No.1, January 1999.

[3] ETSI, European digital celluar telecommunications system (Phase 2); Full rate speech processing functions (GSM 06.01).