Speech Quality Assessment for Wideband Systems

H.W. Gierlich, S. Völl, P. Jax, F. Kettler*

HEAD acoustics GmbH, Ebertstr. 30 a, D-52134 Herzogenrath, H.W.Gierlich@head-acoustics.de
**Institut für Nachrichtengeräte und Datenverarbeitung, RWTH Aachen*

1. Abstract

The integration of wideband communication for future telecommunication systems is of critical importance. Without wideband communication Voice over IP-configurations and future wireless systems might not be able to differentiate from traditional speech services. Besides the use of speech coders already standardized (AMR-WB [5], ITU-T G.722 [1]) bandwidth extension procedures maybe introduced in order to "virtually" increase the perceived bandwidth of the signal. Wideband systems will introduce specific distortions and artefacts which mainly may influence the one way listening speech quality. Furthermore, echo disturbances and background noise are transmitted over the wideband channel. Speech quality determination is more complex than in narrowband systems and needs to consider the different perception of the communication transmitted over the wideband channel. Objective measurements are currently investigated in order to provide appropriate analysis tools for wideband scenarios accordingly. The contribution discusses parameters, testing and analysis methods with focus on the transmission quality of background noise.

2. Introduction

Conversational tests were conducted with four wideband scenarios (see [1], [4], [5] and [8]) in order to determine the relevant parameters for speech quality assessment. Similar to narrowband communication systems these parameters were identified as

- one way listening speech quality,
- echo and double talk performance and
- the quality of background noise transmission.

For determination of one way listening speech quality objective methods are available including wideband options [7]. The double talk performance mainly depends on the implemented nonlinear processor working in conjunction with the echo canceller. The focus of these investigations here was on the determination of disturbances due to audible echoes and the assessment of background noise transmission quality. Furthermore an informative intelligibility test on a logatom basis was conducted to compare narrowband and wideband systems. Two G.722 hands-free systems were used as acoustical interfaces. The different coders were therefore always cascaded with the G.722 coder.

3. Intelligibility Test

The logatom test was chosen due to its high sensibility. The informative character of this test limited the vocabulary to 30 logatoms from one male voice. 12 subjects assessed the intelligibility. Two wideband scenarios (PCM, 16 bit, 50-7000 Hz; AMR-wideband 12.65 kbit/s) and an ISDN connection were compared.

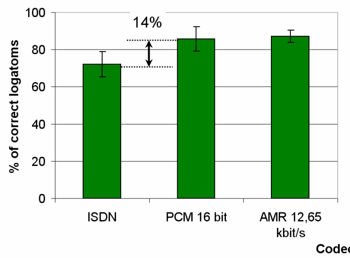

Figure 1: Intelligibility differences

The tests indicate a higher intelligibility for both wideband scenarios compared to the traditional telephone bandwidth.

The 14% difference in figure 1 is significant.

4. Disturbances Caused by Echoes

A listening-only test was conducted to determine echo related parameters. In order to avoid influences due to delay differences between the different setups a constant delay of 170 ms was introduced in all listening recordings. The listening examples were generated using artificial head technology, thus considering aspects like self-masking due to the acoustically coupled signal from the artificial mouth to the artificial ear.

The test results, derived from ratings of 18 subjects, pointed out that the echo disturbances are mainly influenced by the echo level, the spectral echo attenuation and the temporal structure. Due to the constant delay in all scenarios, this influencing factor was eliminated for the comparison of the different setups. Echo signals significantly exceeding the masking threshold lead to disturbances. Further information can be found in [9]. Basically these parameters are known from narrowband scenarios and appropriate measurement techniques are available. This transmission aspect was therefore not further evaluated in detail.

5. Quality of Background Noise Transmission

The listening tests concerning the quality of transmitted background noises used a voice babble and music. An organ music was transmitted via 16 wideband and narrowband connections. The transmitted signal was binaurally recorded in the hands-free configuration and judged by the test persons on a 5 point MOS-Scale. Figure 2 shows the average results from 18 subjects. All codecs (indicated on the x-axis) were cascaded with a G.722 coder used in the hands-free interface.

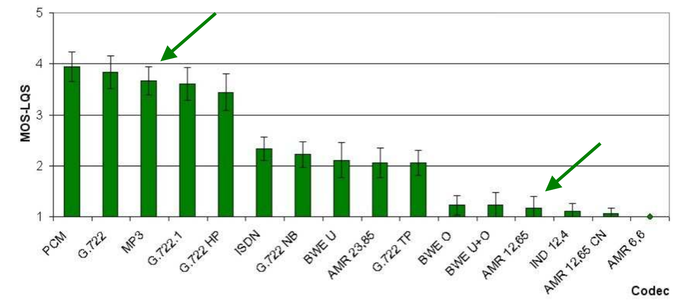

Figure 2: Quality of background noise transmission *(BWE U: bandwidth extension <300 Hz, BWE O: >3.4 kHz, HP: high pass 300 Hz, TP: low pass 3.4kHz, IND 12.4: AMR codec acc. to [4], 12.4 kbit/s)*

The subjectively relevant parameter could be derived as the clearness and possible distinction between single tones, "artefacts" and the different bandwidths.

For an objective quality analysis a hearing model [6] based 3D-power density spectrum was chosen. This hearing model considers the time and spectral resolution of the human hearing and has a clear advantage compared to a standard spectral transformation especially for the hearing adequate distinction between single tones.

The transmitted music signals were analysed accordingly. Figure 3 shows two examples. The analysis shown on the left hand side was assessed with MOS 3.6, the right one with MOS 1.2 (see green arrows in figure 2). As indicated by the white circle, the distinction between single tones (around 4 kHz in this example, see frequency scale on the y-axis) is better for the left hand example compared to the one on the right hand side. In addition, this analysis detects temporal disturbances which are produced by the codec and which led to low MOS score in the listening test (1.2 MOS for the right hand example).

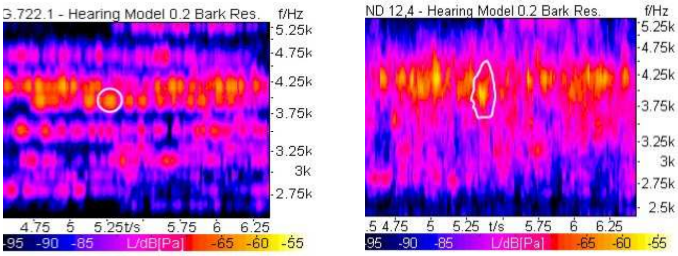

Figure 3: Transmitted music analysed with 3D power density spectrum based on the hearing model [6]; left: G.722.1 codec, right: IND AMR-wideband codec [4], 12.4 kbit/s.

For a better detection and analysis of signal components introduced by the codec or components not being transmitted, the same 3D power density spectra were referenced to the spectra of the G.722 codec. Note that this codec was chosen as a reference providing the most transparent signal transmission amongst the different setups. Figure 4 shows these spectra for the same examples as already analyzed in figure 3. The magenta-coloured areas indicate an undisturbed transmission compared to the G.722 link. Orange or yellow colour codes mark signal components added by the codec. Correspondingly signal components not being transmitted are given in dark colours (blue or black). Due to the non-linear characteristics of these speech coders it is necessary to analyze the spectral differences on a hearing model basis. A linear transformation would not guarantee reliable results.

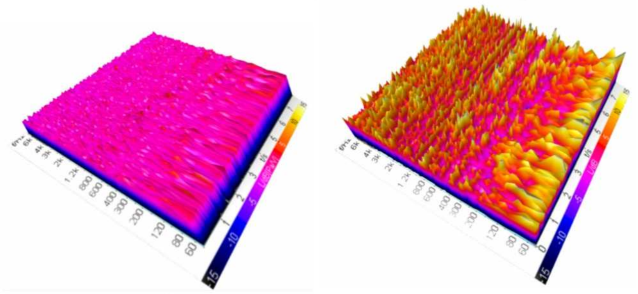

Figure 4: 3D power density spectrum referenced to the spectrum of the G.722 codec result; left: G.722.1 codec, right: IND AMR-wideband codec [4], 12.4 kbit/s

The example shown on the right hand side of figure 4 indicates significant – added – components compared to the G.722 reference. Due to the hearing adequate character of this analysis, these components can be reliably regarded as disturbing (the MOS-LQS was 1.2) whereas a linear transformation could only analyze differences without taking into account the human perception.

For further abstraction some important characteristics of the organ music, which was used during the subjective tests, were reproduced by a mathematically defined signal. This signal is composed by consecutive major chords as shown in the spectrogram in figure 5. Measurements on all connections were repeated using this signal. The same results were audible and visible as with the "original" organ music. The signal is therefore suited to reproduce these effects (distinction between tones,…) and provides advantages in terms of measurement techniques. The benefit is the priori information about the expected tones. This can be used to distinguish between different kinds of errors determining the transmission quality.

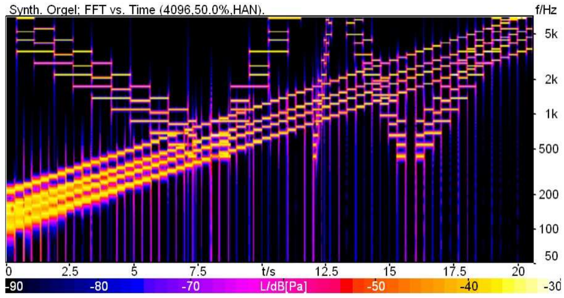

Figure 5: Mathematically defined signal based on major chords

The AMR wideband codec analyzed in the example on the right hand side in figure 6 produces noisy components "around" the tones (see black arrows). The tones themselves are accurately reproduced at the receiving side leading to the magenta coloured lines in the analysis (high similarity between reference signal and transmitted signal over AMR connection, see blue arrows).

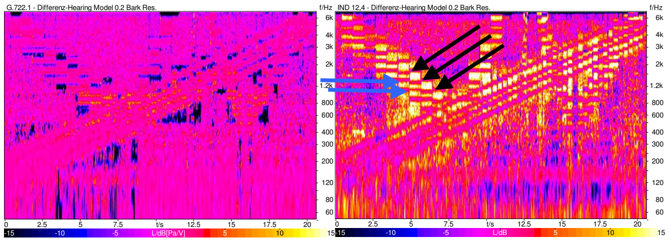

Figure 6: 3D power density spectrum (referenced to G.722) use of test signal; left: G.722.1 codec, right: IND AMR-wideband codec [4], 12.4 kbit/s

The combination of the hearing model based analysis and the choice of this test signal provides not only a high correlation to the subjective impression. It is also suited to analyze the kind of disturbance in detail.

6. Conclusion

The new proposed test signal in combination with the hearing model based reference analysis approach is very promising for the evaluation of tonal background noise. First work conducted to form a single value out of the spectral analysis –based on the results of the subjective listening tests- was started.

7. References

[1] ITU-T Recommendation G.722.2, Wideband coding of speech at around 16 kbit/s using Adaptive Multi-rate Wideband (AMR-WB)

[2] ITU-T Recommendation G.722.1, Coding at 24 and 32 kbit/s for Hands-free Operation in Systems with Low Frame Loss

[3] "MP3", T. Painter, A. Spanias, "Perceptual Coding of Digital Audio", Proceedings of the IEEE, 2000, vol. 88, pp. 449-513

[4] C. Erdmann, P. Vary, K. Fischer, W. Xu, M. Marke, T. Fingscheidt, I. Varga, M. Kaindl, C. Quinquis, B. Kövesi, D. Massaloux, "A candidate proposal for a 3GPP adaptive multi-rate wideband speech codec", IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Salt Lake City, UT, USA, 2001

[5] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, K. Järvinen, "The Adaptive Multirate Wideband Speech Codec (AMR-WB) (ITU-T Recommendation G.722.2)", IEEE Transacti ons On Speech And Audio Processing, 2002, vol. 10, Nr. 8, pp. 620-636

[6] R. Sottek, „Modelle zur Signalverarbeitung im menschlichen Gehör", Chapter 4, 1993, Verlag M.Wehle, Bonn, Germany

[7] J. Berger, „Instrumentelle Verfahren zur Sprachqualitätsschätzung - Modelle auditiver Tests", Chapter 5, Shaker Verlag, 1998

[8] P. Jax, P. Vary, "On artificial bandwidth extension of telephone speech", Signal Processing, 2003, vol. 83, pp. 1707-1719

[9] Völl, S.: "Sprachqualitätsbeurteilung für Breitbandsysteme" Diploma theses, RWTH Aachen, 2004