Some considerations on the replacement of artificial test signals by speech

H.W. Gierlich

HEAD acoustics GmbH, 52134 Herzogenrath, Email: H.W.Gierlich@head-acoustics.de

Introduction

During the recent 20 years a lot of effort has been put into development of artificial, speech like test signals for the test of non-linear and time invariant systems used in speech communication. Associated with this kind of test signals analysis methods had been defined which as close as possible allowed prediction of speech quality parameters. However, due to the non-linear and time variant nature of the modern signal processing this approach is no longer satisfactory.

A new discussion had been started on the type of test signals used for modern communication systems and associated analysis methods. This discussion also has to anticipate the deployment of wideband and super-wideband transmission technologies which are introduced into the networks.

General considerations on speech as a test signal

Based on various discussions in standard organizations (ITU-T, ETSI, 3GPP…), based on the discussions with algorithm developers and manufacturers it can be included that the only future-proof way to realize new types of test signals is the use of speech as a test signal. Any other artificial speech like test signal may potentially fail with upcoming new algorithms not known today. When creating new speech like test signals it is clear, that the test signals to be created may be representative only for e.g. one language and even within this language were may be only partially representative. However, a technical implementation should work on all languages and therefore it may be sufficient, to firstly restrict tests to English samples which is anticipated to be one of the most often used language. So far the work in ITU-T SG12, Q.6 has agreed upon this approach. Clearly this does not exclude the creation of test samples in other languages. Other approaches with e.g. had been taken by the hearing aid industry in developing an international speech test signal (ISTS) [1] do not sufficiently represent all the required features. Furthermore it is desirable to use speech because it can be used for subjective judgments as well. When selecting speech signals to be used for objective analyses the following points should be taken into account:

- **Phonetical balance**

Preferably there selected speech sequences should be phonetically balanced, different speakers should be considered, sufficient amount of test sentences should be available for long term single talk measurements

Spectral distribution: Speech signals to be created need to be full band. Furthermore they should be created to provide sufficient energy for the measurements in the low frequency domain (< 200 Hz) and in the high frequency domain (> 4 kHz). The more energy in those frequency bands is available the higher measurement dynamic range is.

Signal to noise ratio: Signal to noise ratio of the new test signals must be as high as possible.

Crest factor: The crest factor (peak to RMS ratio) of the test signals reduces the dynamic range of the measurement. Therfore speech samples with low crest factor are desirable. It is still under discussion whether also signal compression should be used to further reduce the crest factor.

Test sequences for conversation simulation: For application in speech communication the test signals should give the possibility to simulate a conversational conversation. This includes speech sequences suitable for single talk tests and conversational tests including different types of double talk simulations.

Proposals for speech sequences

Based on the considerations mentioned above ITU-T Q.6/12 is working on speech sequences for the different conversation applications. The current proposals for speech signals are shown below.

Figure 1: Single talk speech sequence consisting of 6 male and 6 female talkers derived from Nokia and Spytechnics database

Single talk sequence shown in **figure 1** mostly fulfills all the properties required: crest factor in average is a 17.4 dB, each sentence is phonetically balanced, the sequence provides sufficient energy in low and high frequency and is applicable for full band systems.

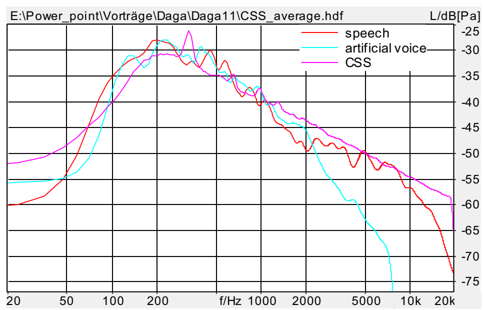

Figure 2: Power density spectra of the new speech single talk sequence compared to artificial voice (P.50) and CSS (P.501)

Figure 2 shows the comparison of the power density spectrum of the new developed single talk sequence [2] in comparison with artificial voice (ITU-T P.50[4]) and the speech like CSS signal (ITU-T P.501[5]). It can be seen clearly that especially in the high frequency area the energy provided is closed to what can be achieved with the CSS signal.

Besides the single talk sequence a double talk sequence which allows to give simulated conversational situations/double talk situations has been derived. **Figure 3** gives an overview about this sequence.

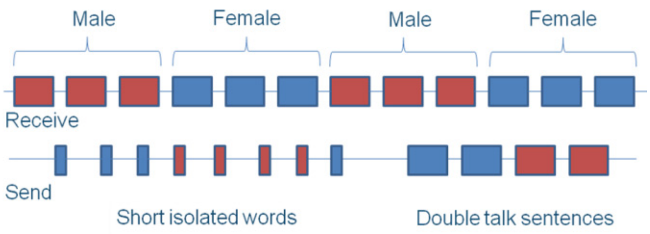

Figure 3: Conversational/double talk signal proposed in ITU-T Q.6/12

The conversational/double talk signal allows the investigation of different types of double talk: short double talk which may occur in special situations when just one party is answering, double talk which is fully masked by the single talk, double talk which is mostly in the pauses between single talk, and finally double talk sequences which partially overlap and which is similar to the timing of the P.501 double talk sequences [5].

Besides those conversational sequences different conditional sequences are developed which can be used to precondition a system under test before doing the actual measurement. This preconditioning is highly required in order to settle all signal processing elements in modern telecommunication systems.

Results

Figure 4 gives the frequency response measured of a modern wideband mobile terminal in receiving. Differences between the measurement with the artificial voice and the new speech signals are seen in the whole frequency range. In the frequency range above about 3 kHz the measurement results deviate substantially which may be due to the either insufficient speech energy of the artificial voice signal or due to coding effects. Measurement for frequencies higher than 8 kHz are clearly more reliable with the speech signal due to much higher SNR ratio.

Figure 5 gives the measurement applying the double talk sequence. The measurement shows the signal level variation of the double talk speech sequence in blue, the reference signal at the input of the phone in sending direction is displayed, in red the transmitted signal in sending direction is shown. All double talk signals are transmitted which indicates an excellent double talk performance which also can be verified by listening to the sequences. However, this analysis also indicates the need of advanced procedures for double talk testing - just the level difference is no good indicator for the double talk performance of the phone. A proposal for advanced double talk evaluation has been introduced in [6] and has been annexed to ITU-T Recommendation P.502[6]. Further investigation on this topic is needed.

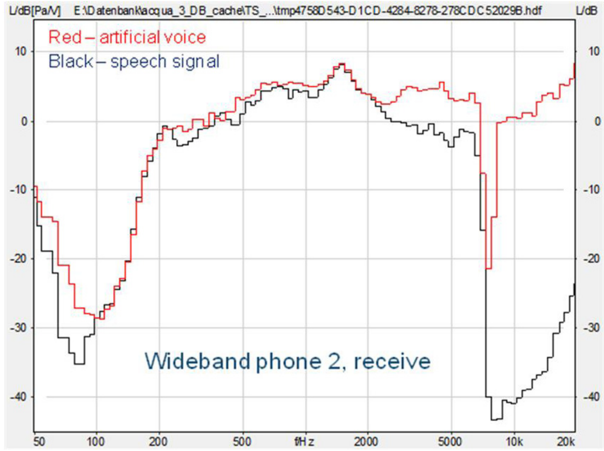

Figure 4: Comparison of frequency response char. measurements in receive for a wideb. mobile phone

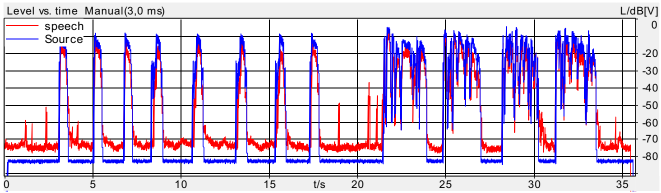

Figure 5: Double talk test – blue original DT-signal, red – transmitted DT-signal

Conclusions

The new proposed speech test signals provide adequate properties for the evaluation of long term characteristics of telecommunication equipment. Further investigation on the use and applicability of the signals for the evaluation of short term characteristics is needed. Testing methodologies and requirements have to be adapted to the new test signals. Currently also no Lombard effect is considered which especially in noisy environments would be extremely beneficial. In general the new speech signals will help to measure modern speech communication devices more realistically.

References

[1] Holube et. al.: Development and analysis of an International Speech Test Signal (ISTS), International Journal of Audiology, December 2010, Vol. 49, No. 12
[2] ITU-T COM12- C186: Proposals for new full-band speech samples and sequences within ITU-T Rec. P.501, Nokia Corporation, Geneva Jan. 2011
[3] ITU-T COM12- C204: Real speech signals for telephonometry - some considerations, Telefon AB – LM Ericsson, Geneva Jan. 2011
[4]ITU-T Rec. P.50: Artificial Voices
[5]ITU-T Rec. P.501: Test Signals for Use in Telephonometry
[6] ITU-T COM 12-C107: Automated double talk analysis procedure, Geneva, May 2010
[7]ITU-T Rec. P.502: Objective test methods for speech communication systems using complex test signals