WIDEBAND ECHO PERCEPTION

Silvia Poschen[1], *Frank Kettler*[1], *Alexander Raake*[2], *Sascha Spors*[2]

[1] HEAD acoustics GmbH (Germany), Ebertstrasse 30a, 52134 Herzogenrath, Germany
{Silvia.Poschen, Frank.Kettler}@head-acoustics.de
[2] Deutsche Telekom Laboratories (Germany), Ernst-Reuter-Platz 7, 10587 Berlin, Germany
{Alexander.Raake, Sascha.Spors}@telekom.de

ABSTRACT

The migration of telecommunication towards wideband transmission may also introduce new types of degradations. This paper focuses on the influence of echoes. The different frequency content of un-cancelled residual echoes in wideband transmission, especially the signal energy in higher frequency ranges above 3 kHz, significantly changes the echo perception and echo disturbance compared to the narrowband case. The results of subjective tests presented in this paper point out that the spectral content of residual echoes may influence the quality rating by more than 2 MOS on a Degradation Category Rating (DCR) scale. Especially the frequency range between approximately 3 and 6 kHz needs to be addressed in wideband communication in order to avoid echo disturbances. The results presented here motivate new tolerance schemes for the echo analysis which differ from commonly used methods, as e.g. given in standards like ETSI 202 739 or TIA 920 for wideband telephony.

Index Terms — Wideband transmission, echo perception, echo attenuation, subjective testing

1. INTRODUCTION

Telecommunication is going to change significantly with the introduction of wideband capable networks and terminals. Speech coders for mobile communication are available, first terminals are announced. Beside the expected benefit of an improved listening speech quality and intelligibility "side effects" like the occurrence of wideband disturbances need to be considered. For instance, the detection and annoyance caused by audible residual echoes is influenced by the spectral content of the echo signal and needs to be carefully addressed. Customers' acceptance highly depends on quality and a perceived benefit compared to the reference narrowband telephony.

One dimensional values like the weighted terminal coupling loss (TCL_w) for terminals can not adequately describe echo disturbances. Current standardization documents like ETSI ES 202 739 [1] for handset respectively ES 202 740 [2] for hands-free applications are established and cover wideband scenarios. However, due to the lack of subjective test results the requirements are often copied or extrapolated from the narrowband scenario without taking into account human perception. The tolerance scheme for the spectral echo attenuation in the current version of ETSI ES 202 739 and ES 202 740 is adapted for wideband scenarios by linear extrapolation for the high frequency range above 3.4 kHz from the narrowband test case. Furthermore these standards -like also TIA 920 [3]- calculate the echo attenuation using the spectral weighting function given in ITU-T G.122 [4]. Again the weighting factors originally derived for narrowband scenarios are linear extrapolated from the narrowband case (see figure 1.1). The extrapolation is only applied above 3.4 kHz up to 6.7 kHz, weighting factors for the low frequency range below 300 Hz are missing.

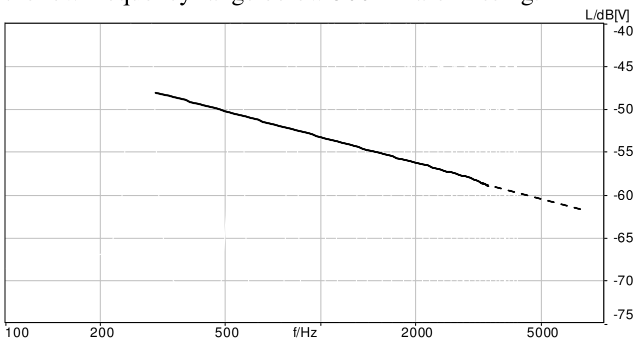

Figure 1.1: TCL_w weighting function (from [4]); dashed line: extension for wideband TCL_w calculation

Speech quality and the compliance with requirements derived from subjective testing are crucial for terminals and networks. This compliance is typically verified by objective laboratory tests. Hence, a subjective test of wideband echo scenarios was carried out in order to derive reliable requirements. The test design and scenarios are described in Section 2, the test results are outlined in Section 3, and a new echo tolerance scheme is derived in Section 4.

The results are of relevance for the development and verification of echo cancellation algorithms because they can serve as a basis to define new requirements.

2. TEST DESIGN

Talking and listening tests as described in ITU-T P.831 [5] are an appropriate method for the subjective assessment of echo disturbances. Test persons are asked to talk into a phone or terminal and judge the echo. Self-masking due to own voice is considered in a realistic way in this scenario. If the utterances for the test persons are limited to a short but sufficient duration (typically a few seconds), the tests can be carried out in a very efficient way and provide a high amount of results data within a limited time.

2.1 Test setup

The tests were conducted using a wideband echo simulator. The principle is shown in figure 2.1. A headset was connected to the internal sound card as acoustical interface. The frequency response in sending direction and receiving direction as well as the appropriate gains (sending loudness rating, receiving loudness rating) can be adjusted using online filters and gain settings. The echo can be simulated by a variable delay and spectral shaping filters.

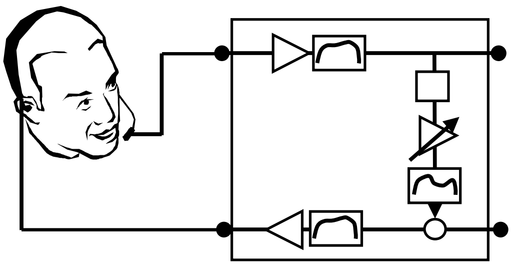

Figure 2.1: Block diagram of echo simulator

For the selection of an appropriate acoustical interface three different commercially available supra-aural headsets were pre-selected and characterized by objective measurements. The sidetone characteristics (frequency response, Sidetone Masking Rating STMR acc. to ITU-T P.79 [6] Annex A) were measured using an artificial head measurement system according to ITU-T P.58 [7] equipped with two type 3.4 artificial ears according to ITU-T P.57 [8]. The intention of these tests was the selection of a device that most realistically reproduces the "natural" sidetone between mouth and ear of humans (not covered by any headphone or handset). Figure 2.2 compares the measured curves of the three headphones to the natural sidetone (open ear of the artificial head, black line in figure 2.2). Results showing the similarity between the sidetone measured with this artificial head without headsets and test persons can be found in [9].

The supra-aural headphones provide very similar sidetone characteristics below approximately 2 kHz but significantly attenuate the speech coupling between 2 and 6 kHz. The device represented by the grey dashed line (see arrow) represents the closest approximation of the natural sidetone and was therefore selected for the tests (Sennheiser PC130).

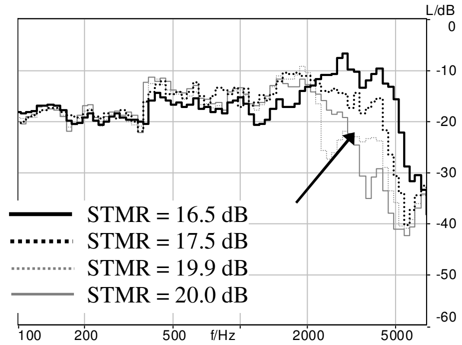

Figure 2.2: Sidetone and STMR (grey solid, dashed, dotted: headsets; black: natural sidetone)

The sending and receiving sensitivity was adjusted in order to reproduce recommended values for a wideband telephone according to [1] (weighted attenuations: Sending Loudness Rating SLR=8.1, Receiving Loudness Rating RLR=2.1). The frequency responses were adjusted by correction filters in order to fulfill the tolerance schemes currently suggested in [1] and [10] (see figure 2.3).

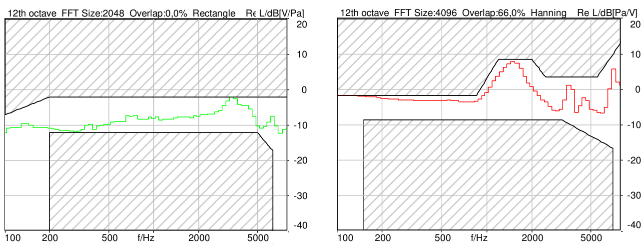

Figure 2.3: Frequency responses of selected headset and tolerance schemes; sending (left, [1]) and receiving (right, [10])

2.2 Test conditions

The test conditions combined round trip delays between 100 ms and 500 ms (four settings), four echo attenuation settings between 25 and 55 dB and 4 different echo filters of different shapes. The corresponding gain functios are shown in figure 2.4. Beside a pure attenuation over the whole frequency range (denoted as "WB") different implementations were tested. The filter denoted as "WB_F1" limits the frequency content to a frequency range between approximately 100 Hz up to 1.3 kHz. A narrowband scenario (300 Hz to 3.4 kHz, designation "NB") and two different wideband scenarios ("WB_F2", 3.1 kHz to 5.6 kHz; "WB_F3", 5.2 kHz to 8 kHz) were included.

The echo attenuation adjustment in the simulator was determined by the unweighted level difference between the send and echo signals. A total number of 129 different echo conditions were tested.

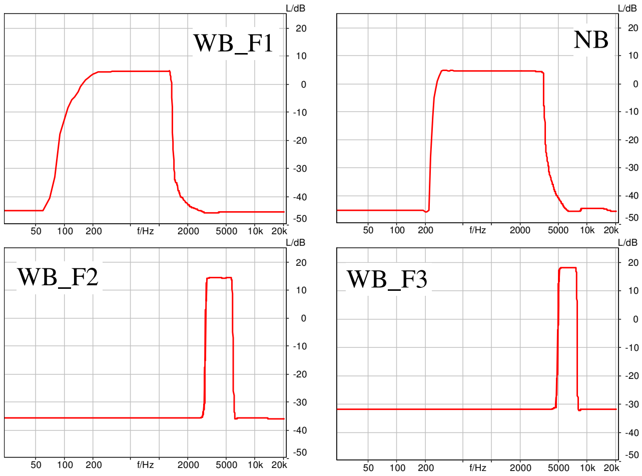

Figure 2.4: Gain functions of echo shaping filter

2.3 Test procedure

The talking and listening tests were conducted with six expert listeners and 13 naïve test subjects. All participants were native German speakers and audiometrically checked. The headset microphone was positioned relative to the test persons' mouth in a similar way for all participants. The echo signal was monaurally displayed on the right headset loudspeaker in order to simulate a monaural telephone situation.

The task for the test persons was to simulate a salutation during a telephone call. The sentences were provided in printed form: "Schönen guten Tag, Frau Schmitz. – Mein Name ist <individual name> von der Firma HEAD acoustics." The intention behind this phrase was to achieve a limited duration, to use utterances with fricatives at word ends ("Schmitz", "acoustics"), and to allow using the individual names. Furthermore this task is easy to conduct and represents a realistic telephone situation.

The echo disturbance was assessed using a five point DCR scale according to ITU-T P.800 [11]: 5 – echo is inaudible; 4 – echo is audible but not annoying; 3 – echo is slightly annoying; 2 – echo is annoying; 1 – echo is very annoying. The individual ratings are averaged to form mean opinion scores (MOS). The confidence interval (CI) is calculated based on a 95 % confidence level.

3. RESULTS

A first analysis of the results pointed out that the expert listeners rated very similarly to the naïve test persons group. The results can therefore be combined for further study. Furthermore the MOS analysis of the entire test conditions demonstrated that the test persons used the whole quality scale from MOS 1 up to MOS 5. Reference conditions without any echo were assessed with 5.0 points on the MOS scale indicating that the task was easy to conduct for test persons. This could also be confirmed by the individual interviews after the tests.

Figure 3.1 depicts the MOS results together with the CIs for echo attenuations of 35 dB, 40 dB and 55 dB, shown in groups for the different echo shaping filters. The filters are denoted on the x-axis. The bars represent round trip delays of 100 ms, 200 ms, 300 ms and 500 ms respectively.

As expected, the MOS values increase for all filters with higher echo attenuation. But the results also demonstrate that the MOS values are significantly lower for the filter characteristic denoted as WB_F2. Especially the results for the relatively high echo attenuation of 46 dB demonstrate that echoes of frequency content between 3.1 kHz and 5.6 kHz seem to be easier to detect, and therefore are more annoying for test subjects compared to the other test conditions (see middle figure).

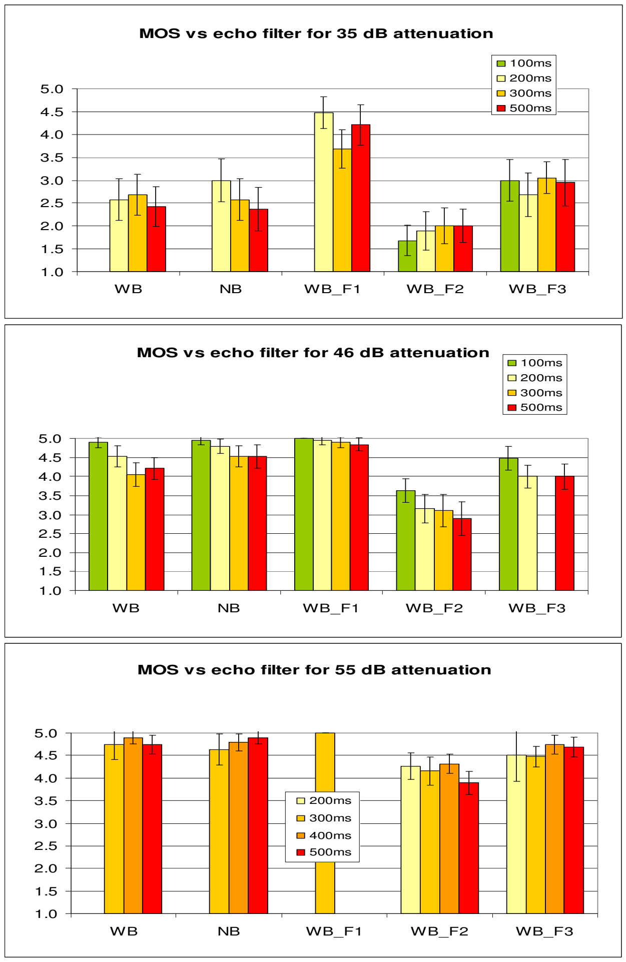

Figure 3.1: MOS results vs. echo shaping filter

MOS differences of up to 2 points occur between low frequency echo disturbances (WB_F1), narrowband scenarios (NB) and high frequency echoes (WB_F2). This is especially relevant for a realistic test case of 46 dB echo attenuation as this is a value often required for terminals.

A similar effect can already be observed for the very high echo attenuation of 55 dB (lower figure).. The MOS scores vary between approximately 4.5 and 5 for the pure attenuation over the entire frequency range ("WB"); the results for the "WB_F2" filter lie around 4.0 MOS indicating that the echo is audible, but not annoying. It should be considered that terminals for VoIP applications need to fulfill the requirement of 55 dB echo attenuation [3].

A similar analysis is shown in figure 3.2 for 200 ms round trip delay and different echo attenuation (between 35 and 46 dB). The low pass filter between approximately 200 Hz up to 1.3 kHz ("WB_F1") is less critical compared to the other test cases. This may be explained by the higher self-masking due to own's voice in the low frequency range.

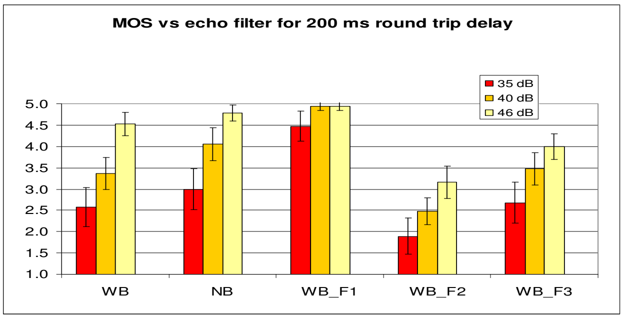

Figure 3.2: MOS results for 200 ms round trip delay

4. NEW TOLERANCE SCHEME

The results of the subjective tests are used to derive a new spectral echo attenuation tolerance scheme. The tolerance scheme needs to reflect the observed sensitivity differences of the human ear with regard to the spectral echo content into account. Furthermore it shall ensure high quality, i.e. a sufficiently high echo attenuation over the entire wideband frequency range. The following assumptions are made:

Instead of using weighting factors extrapolated from ITU-T G.122 for the TCLw calculation, the echo attenuation is determined by unweighted level differences between the test signal and the echo signal. The level analysis shall be applied on real speech signals using an appropriate average time during level calculations (e.g. minimum sequence length 5s). The scheme is derived based only on the test results showing at least 4.5 MOS.

The analysis is carried out individually for different round trip delays between 100 ms and 500 ms. If the round-trip delay is not known for a given connection (or terminal) under test, the "worst case" tolerance level shall be applied.

Figure 4.1 shows the tolerance schemes as they were derived from the subjective data. The frequency range between approximately 3.1 and 5.6 kHz seems to be most critical and needs to be attenuated most. Network components like network echo cancellers and terminals using acoustic echo cancellers may insert comfort noise during the presence of a receive test signal. Hence, it needs to be verified by appropriate settings on the test device or appropriate correlation analyses whether a tolerances scheme violation is caused by the echo or an inserted comfort noise signal.

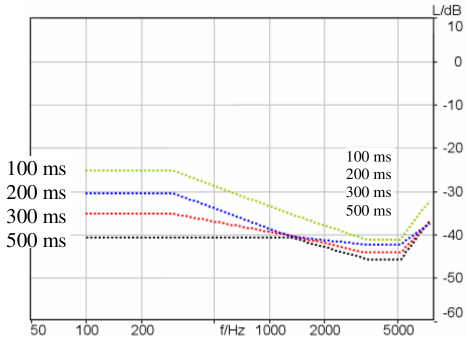

Figure 4.1: Suggestion for spectral tolerance schemes for different round trip delays

5. CONCLUSION

Subjective tests carried out under different talker echo conditions provide insights into the perception of residual echos. The spectral content of echoes, especially in the frequency range between 3.1 and 5.6 kHz is found to be crucial for echo disturbance, requiring an adaptation of existing tolerance schemes as proposed in this paper.

The tests were conducted within a project funded by Deutsche Telekom Laboratories.

6. REFERENCES

[1] ETSI ES 202 739, STQ; Transmission requirements for wideband VoIP terminals (handset and headset) from QoS perspective as perceived by the user, 2007

[2] ETSI ES 202 740, STQ; Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user, 2007

[3] TIA 920, Transmission Requirements for Wideband Digital Wireline Telephones, 2002

[4] ITU-T G.122, Influence of national systems on stability and talker echo in international connections, 1993

[5] ITU-T P.831, Subjective performance evaluation of network echo cancellers, 1998

[6] ITU-T P.79, Calculation of Loudness Ratings for Telephone Sets, Annex A, 1999

[7] ITU-T P.58, Head and Torso Simulators for Telephonometry, 1996

[8] ITU-T P.57, Artificial Ears, 1996

[9] Eigenwahrnehmung der Stimme in auditiven virtuellen Umgebungen, Ph.D. VDI Fortschritt-Berichte, Nr. 666, 2001, Christoph Pörschmann

[10] Testing Wideband Terminals, S. Poschen[1], F. Kettler[1], A. Raake[2], S. Spors[2], [1]HEAD acoustics, Herzogenrath, Germany, [2]Deutsche Telekom Laboratories, Berlin, Germany, DAGA, Jahrestagung der Deutschen Gesellschaft für Akustik, 2008,

[11] ITU-T P.800, Methods for Subjective Determination of Transmission Quality 1996