OPTIMUM FREQUENCY RESPONSE CHARACTERISTICS FOR WIDEBAND TERMINALS

H.W. Gierlich[1], Silvia Poschen[1], Frank Kettler[1], Alexander Raake[2], Sascha Spors[2]

[1] HEAD acoustics GmbH (Germany), Ebertstraße 30a, 52134 Herzogenrath, Germany
[2] Deutsche Telekom Laboratories (Germany), Ernst-Reuter-Platz 7, 10587 Berlin, Germany

ABSTRACT

In ETSI standard ES 202 739 and ES 202 740 a new testing technique for the measurement of wideband terminals is introduced. Tolerance masks are given for sending and receiving frequency response characteristics. As an important new concept in this standard no longer the ear reference point (ERP) but the free field reference point is used for determining the response characteristics in receiving.

Subjective tests have been carried out in order to derive the impact of non optimum receiving frequency response characteristics on the perceived speech sound quality. Different experiments as conducted are introduced and the results are shown. Based on these frequency response characteristics a tolerance mask based on diffuse field equalization is proposed. The tolerance mask guarantees (based on subjective testing) a minimum speech sound quality in receiving with respect to impairments introduced by different frequency response characteristics only.

In sending direction different subjective experiments were conducted with different types of background noises and different types of wideband terminals. The aim of this investigation was to find desirable sending frequency response characteristics with and without background noise at the near end. Furthermore general recommendations concerning frequency responses in case of speech with near end background noise are desired. The subjective experiments are introduced and the results will be discussed.

1 INTRODUCTION

In ETSI standards ES 202 739 [2] and ES 202 740 [3] a new testing technique for the measurement of receiving frequency response characteristics is introduced. Both standards give requirements for wideband VoIP phones from a users perspective. Especially for the handset testing described in ES 202 739 the testing principle is changed as well as the receiving frequency response tolerance mask for wideband terminals. Instead of referring the measured acoustic sound pressure to the ERP the free field reference point is chosen. In general this technique and the associated tolerance scheme which is mostly flat should lead to a better sound quality as compared to the previously defined ERP testing. Furthermore this testing technique is more compatible to the measurement of headphones and hands-free phones. However, so far the tolerance mask in the standard is not yet based on the speech quality as perceived by the user.

In sending direction typically the question arises whether or not the wideband frequency response characteristics should be narrowed in case of background noise.

In order to find frequency response characteristics tolerance schemes which can be associated to a certain perceived speech quality, subjective tests were carried out at HEAD acoustics in a project for Deutsche Telekom Laboratories including a variety of different frequency response characteristics and different types of wideband terminals including simulations. The results of these tests are introduced.

2 RECEIVING TEST SETUP

2.0 Pretests

In a first test session different types of terminals were provided to different expert users. These different terminals were available as prototypes and covered VoIP as well as cordless terminals. It was the task of the experts to use an software equalizer and adjust the frequency response characteristics in such a way that the optimum frequency response according to their opinion was created. The frequency response was adjustable in 1/3 octave bands. The test signal used for this experiment consisted of two double sentences spoken by a male and a female talker each. In addition to the task to find an optimized speech sound quality the test persons also were asked to adjust the preferred listening level. **Figure 1** shows a typical result of such an experiment. It can be seen that this task leads to highly different results for the individually preferred frequency response characteristics. Furthermore it could be seen that typically the results were not repeatable meaning that there was no unambiguous subjective opinion about the preferred frequency response characteristics based on this experiments.

Therefore a second experiment was conducted which was aimed to find a rank order of the different frequency response characteristics. This test was conducted by seven experts for each of the six terminals individually. The experts were using the actual terminals including the equalization functions as found in the previous experiment. In this experiment it could be found that the preferred frequency response characteristics of the terminals are more similar than in the first experiment however still a clear answer about the optimum frequency response characteristics can not be given from this experiment. As a consequence a listening test was conducted based on ITU-T recommendation P.800 [1].

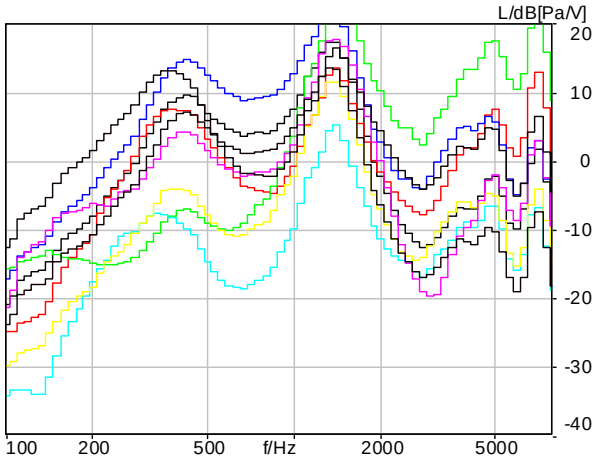

Fig. 1: "Optimum" terminal frequency response characteristics for one terminal individually adjusted by 9 experts including the individual settings of an adjustable equalizer (1/3 octave bands) measured at a free field equalized artificial head, 8 N application force

2.1 Test setup for the ITU-T P.800 listening test

In order to find the relationship between the subjectively perceived sound quality and the corresponding frequency response characteristics an experiment was conducted which included the following frequency response characteristics each for several phones:

- three frequency response characteristics which were found in the previously described ranking experiments to give subjectively a good ranking (favorites 1-3), separately for each terminal
- one frequency response characteristics flat with respect to the ERP
- one frequency response characteristics which equals to a diffuse field equalized terminal
- one frequency response characteristics which equals to a free field equalized terminal

In addition two "references" were integrated. These references were generated by the same speech material as the other listening examples however they were recorded in the so called "orthotelefonic reference position" (see **figure 2**).

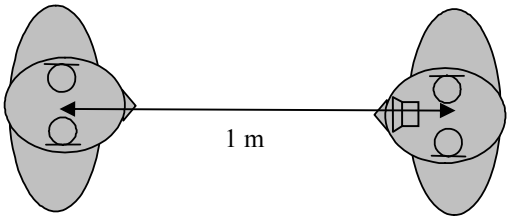

Fig. 2: Orthotelefonic *reference position*

In this position two artificial heads are facing each other in a distance of 1 m. One artificial head (acc. to [6], [7]) is playing back the speech material. The other one is used to record the speech material in anechoic conditions. In one experiment the recording was produced full band, in the second experiment the signal is band limited between 100 Hz and 8 kHz. For the recording the artificial head was free field equalized, for the playback a free field equalized head phone was used. All terminals were tested with G.722 speech codec. All real speech sequences were recorded in anechoic conditions. For the application of the handset to the artificial head 2N, 8N and 13N application force were chosen. The setup is shown in **figure 3.**

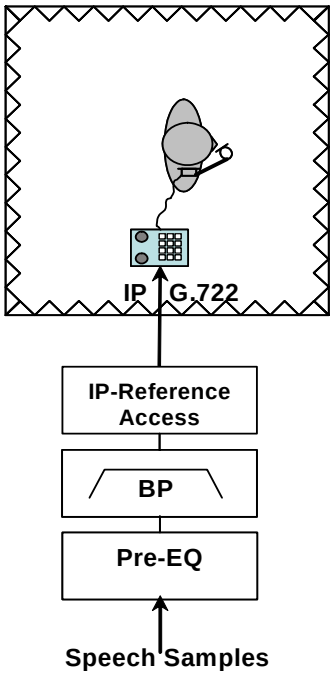

Fig. 3: Test setup with wideband handsets

In total the matrix described in **table 1** gives an overview about the different conditions used the experiments.

The speech material used in this test consists of sequences spoken by two male and two female talkers, each producing a double sentence for the subjective judgment.

In addition to the recorded speech samples as given in **table 1** one additional scenario which was produced offline using an artificial bandwidth extension was included in the test program.

Telephone	Favorit 1	Favorit 2	Favorit 3	Flat with diffusefield equalized HATS	Flat with DRP-ERP correction for HATS	Flat with freefield equalized HATS
Tel 1, 2N	Tel 1 Fav1 2N	Tel 1 Fav2 2N	Tel 1 Fav3 2N	--	Tel 1 Lin opt 2N	Tel 1 FF opt 2N
Tel 1, 8N	Tel 1 Fav1 8N	Tel 1 Fav2 8N	Tel 1 Fav3 8N	Tel 1 DF opt 8N	Tel 1 Lin opt 8N	Tel 1 FF opt 8N
Tel 1, 13N	Tel 1 Fav1 13N	Tel 1 Fav2 13N	Tel 1 Fav3 13N	--	Tel 1 Lin opt 13N	Tel 1 FF opt 13N
Tel 2, 2N	Tel 2 Fav1 2N	Tel 2 Fav2 2N	Tel 2 Fav3 2N	--	--	--
Tel 2, 8N	Tel 2 Fav1 8N	Tel 2 Fav2 8N	Tel 2 Fav3 8N	Tel 2 DF opt 8N	Tel 2 Lin opt 8N	Tel 2 FF opt 8N
Tel 2, 13N	Tel 2 Fav1 13N	Tel 2 Fav2 13N	Tel 2 Fav3 13N	--	--	--
Tel 3, 8N	Tel 3 Fav1 8N	Tel 3 Fav2 8N	Tel 3 Fav3 8N	Tel 3 DF opt 8N	Tel 3 Lin opt 8N	Tel 3 FF opt 8N
Tel 7, 8N	Tel 7 Fav1 8N	Tel 7 Fav2 8N	Tel 7 Fav3 8N	Tel 7 DF opt 8N	Tel 7 Lin opt 8N	Tel 7 FF opt 8N
+ Reference 1 (orthotelephonic reference position), "Ortho_Ref1"						
+ Reference 2 (orthotelephonic reference position + bandpass-filter), "Ortho_Ref2_BP"						
+ artificial bandwidth extension algorithm, "ABE"						

Table 1: *Matrix of the 43 listening examples used in the subjective test*

As a result for each condition for the male as well as for the female speakers 24 judgments (24 different subjects, normal hearing) were collected in the listening tests. As usual in advance to the actual test twelve conditions were rated for conditioning the test persons. In these twelve conditions the complete range of quality was represented.

The test was conducted separately for male and female talkers. However, each condition was judged by the same test person for male and female talkers.

The test was organized as a listening only test described in ITU-T recommendation P.800 series, the listening level was adjusted to 73 dB_{SPL}. The listening examples were presented diotically (the same signal on both ears). The reference recordings in the orthotelephonic reference position were presented binaurally.

The judgment was made according to ITU-T recommendation P.800 [1] using a 5-point ACR scale.

3 RESULTS OF THE LISTENING TESTS FOR RECEIVING

The results of the listening tests (average over the four speakers) are given as an overview in **figure 4**.

Although wideband speech is included in the test the maximum MOS achieved in this test is MOS 4.3. This effect seems to be similar to the tests which have been conducted with narrow band scenarios in the past and seems to be due to the fact that intuitively test persons might expect even higher qualities than the ones presented in the listening test. As expected, the subjectively rated ortotelephonic reference positions lead to a fairly high MOS rating although it is not the highest rating of all conditions in the test.

For terminal 1 and 2 the free field and the diffuse field equalized response characteristics were judged best. For terminals 3 and 7 the free field and diffuse field equalized frequency response characteristics are judged worse however the main reason for this worse judgment is the fairly high noise level which was produced by these phones. This noise mainly influences the judgment of the sound quality. Thus, these results were not taken into account for further analyses.

Further analyses aiming to find differences between average male and average female talkers do not show significant differences between male and female talkers in most of the cases. In some conditions the difference between the two male or female talker may be up to 1 MOS. This is most probably due to the different speech characteristics (spectral characteristics) of the different voices which – due to differently shaped frequency response characteristics – may be more or less pleasant. However, these differences do not occur for the conditions which were rated with high MOS scores. In case of high ratings mostly no talker dependency on the MOS-scores was observed.

Summarizing, it can be concluded, that a clear ranking can be found for the different frequency response characteristics influencing the subjectively perceived speech sound quality over a wide range.

4 DEVELOPMENT OF A NEW RECEIVING TOLERANCE SCHEME

For the development of an "optimum frequency response characteristics" for a wideband terminal the results of each talker were averaged per talker and summarized in the table below. Only those scenarios were considered which lead to an MOS judgment higher than 3.3 for each speaker separately. These results are shown in **table 2**.

ranking	male 1		male 2	
1	Ortho_Ref2_BP	4.25	Tel 2 FF opt 8N	4.42
2	Tel 2 Diff opt 8N	4.25	Ortho_Ref1	4.33
3	Tel 2 FF opt 8N	4.17	Ortho_Ref2_BP	4.33
4	Tel 1 Diff opt 8N	3.92	Tel 1 Diff opt 8N	4.25
5	Tel 1 Fav2 13N	3.83	Tel 1 FF opt 2N	4.08
6	Ortho_Ref1	3.67	Tel 1 FF opt 8N	4.08
7	Tel 1 Lin opt 2N	3.58	Tel 1 Fav2 13N	3.75
8	Tel 1 Fav2 8N	3.58	Tel 1 Fav2 8N	3.75
9	Tel 1 FF opt 2N	3.58	Tel 1 Lin opt 2N	3.75
10	Tel 1 FF opt 8N	3.42	Tel 2 Diff opt 8N	3.75
11	Tel 1 Fav2 2N	3.42	Tel 7 Fav1 8N	3.58
12	Tel 7 Fav1 8N	3.33	Tel 1 Fav2 2N	3.50
13	---		Tel 1 FF opt 13N	3.33
14	---		---	---

ranking	female 1		female 2	
1	Tel 2 Diff opt 8N	4.33	Tel 2 FF opt 8N	4.42
2	Ortho_Ref12	4.25	Tel 1 FF opt 8N	4.08
3	Tel 1 Lin opt 2N	4.17	Ortho_Ref2_BP	4.08
4	Tel 2 FF opt 8N	4.17	Tel 1 FF opt 2N	3.92
5	Tel 7 Fav1 8N	3.83	Tel 1 Diff opt 8N	3.83
6	Ortho_Ref2_BP	3.75	Tel 1 Lin opt 2N	3.83
7	Tel 1 Fav2 8N	3.75	Ortho_Ref13	3.83
8	Tel 1 FF opt 2N	3.75	Tel 1 Fav2 13N	3.58
9	Tel 1 Diff opt 8N	3.67	Tel 7 Fav1 8N	3.58
10	Tel 1 FF opt 8N	3.58	Tel 1 Fav2 2N	3.50
11	Tel 1 Fav2 2N	3.58	Tel 1 Fav3 8N	3.50
12	Tel 1 Fav3 8N	3.50	Tel 1 FF opt 13N	3.50
13	Tel 1 Fav2 13N	3.33	Tel 2 Diff opt 8N	3.42
14	---	---	Tel 1 Fav3 13N	3.33

Table 2: *Conditions with MOS (averaged per talker) ≥ 3.3*

From this overview the conditions were extracted where for all talkers in one condition an MOS value > 3.3 was achieved. **Table 3** is listing these conditions together with the averaged MOS for all talkers.

Ortho_Ref1	4.0	Tel 2 Diff opt 8N	3.9
Ortho_Ref2_BP	4.1	Tel 2 FF opt 8N	4.3
Tel 1 Diff opt 8N	3.9	Tel 1 Fav2 8N	3.6
Tel 1 FF opt 8N	3.8	Tel 1 Fav2 2N	3.5
Tel 1 FF opt 2N	3.8	Tel 1 Fav2 13N	3.6
Tel 1 Lin opt 2N	3.8	Tel 7 Fav1 8N	3.6

Table 3: *Conditions with MOS > 3.3 for all talkers*

Based on the corresponding frequency responses now the mask for the frequency response characteristics can be developed, see **figure 5**. This mask would ensure, that the perceived sound quality of a wideband phone is typically higher than MOS 3.6. This clearly would help to demonstrate the user the benefit of wideband transmission as compared to narrowband transmission and thus certainly would help to promote wideband technology. It may be discussed whether or not the tolerance for lower frequencies may be widened a bit like indicated by the dotted line in **figure 5**. Also it may be discussed whether the lower cut off frequency may be shifted towards higher frequencies in order to take into account the physical capabilities of loudspeakers integrated in telephones.

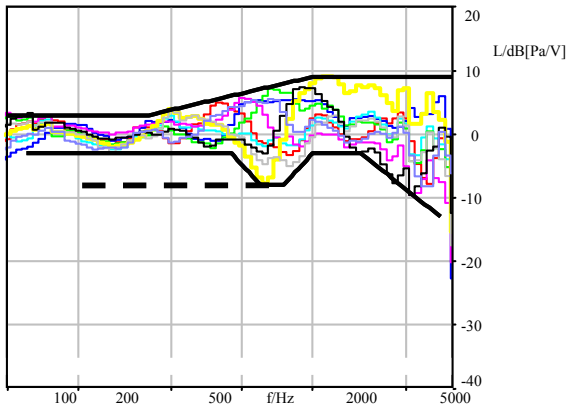

Fig. 5: Proposal for a new tolerance mask for wideband handset terminals in receiving, based on diffuse field equalized HATS used for the tests, ensuring MOS-LQSw of typically ≥ 3.6

5 INVESTIGATION OF THE FREQUENCY RESPONSE CHARACTERISTICS IN SENDING DIRECTION

In sending direction all tested phones meet the tolerance scheme defined in [2, 3] and also provide a good listening speech quality. A potential risk may be the use of wideband phones in noisy environments. Here the background noise is transmitted from 50 to 7000 Hz simultaneously with the speech signal and thus, may lead to a poor speech sound and a reduced speech intelligibility.

5.1 Test Setup

Lombard speech [4] was created in order to provide a realistic test environment including background noises. Therefore several realistic background noises used in the experiment were presented binaurally to 2 talkers (1 male, 1 female) via closed headphones. Consequently the Lombard effect was initiated. Both talkers uttered two sentences,

which were recorded synchronously to the background noise by an omni-directional microphone.

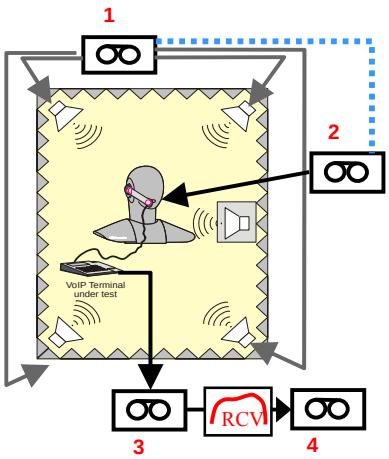

Fig. 6: Test arrangement for transmission of realistic background noises and Lombard speech via wideband terminals

The background noises (1) and the Lombard speech recordings (2) were played back via a background noise simulation system and an artificial mouth respectively, both installed in an anechoic chamber, see **figure 6**. Speech and noise are then transmitted via several wideband phones and recorded at the electrical point of interconnection (POI). In order to simulate a receiving handset, all recordings were filtered with the "winner" frequency response of the previous test in receiving direction.

5.2 Experts Tests

These recording were then presented to expert listeners. Again they were asked to modify the frequency characteristic by a software equalizer in 1/3 octaves between 100 and 8000 Hz in order to achieve the perceptually best speech sound. Two results for one phone and two transmitted background noises are shown in **figure 7**. The following conclusion from the expert's test can be drawn:

- The expert settings are similar for all combinations of phones and transmitted background noises.
- For high background noise level the experts adjusted significant high- and low pass characteristics. Proceeding interviews showed that the experts tried to reduce the annoyance of the background noise and to increase the speech intelligibility.
- The settings are similar for all phones transmitting the same background noise.

With these indications a formal listening test was setup in order to confirm the expert's findings.

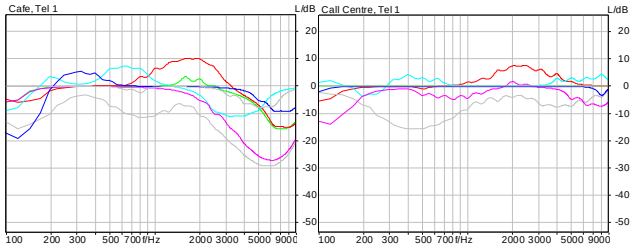

Fig. 7: Expert's equalizer settings for one phone and transmitted café noise (left, non-stationary, 68dB(A)) and call centre noise (right, non-stationary, 58dB(A))

6 LISTENING TESTS IN SENDING

6.1 First Listening Test

Based on the experts pretest different sending frequency response characteristics shown in **figure 8** were used in the formal listening tests:

- Full band (100 Hz – 8 kHz) (A)
- High pass (100 Hz – 200 Hz, 10 dB), with very moderate low pass (5 kHz – 8 kHz, 5 dB) (B)
- Strong high pass (100 Hz – 300 Hz, 20 dB), with very moderate low pass (5 kHz – 8 kHz, 5 dB) (C)
- Moderate high pass (50 Hz – 200 Hz, 5 dB), with very moderate low pass (5 kHz – 8 kHz, 5 dB) (D)
- Medium high pass (50 Hz – 200 Hz, 10 dB) with moderate low pass (4 kHz – 8 kHz, 10 dB) (E)
- Strong high pass (50 Hz – 300 Hz, 20 dB), with medium low pass (4 kHz – 8 kHz, 13 dB) (F)
- High pass rising 2 dB per octave from 100 Hz to 3 kHz (G)
- Medium high pass (50 Hz – 200 Hz, 10 dB), with emphasize around 2 kHz (8 dB) (H)

In the formal listening test the speech was transmitted via one wideband phone (the one providing the flattest frequency response with a "traditional" big handset, no noise cancellation). The same five background noises were used as for the expert test:

- non stationary call center noise (58 dB(A))
- non stationary living room noise (58 dB(A))
- stationary car noise (69 dB(A))
- non stationary cafeteria noise (78 dB(A))
- non stationary pub noise (74 dB(A))

All conditions were filtered with eight filters extracted from the expert's pretests. For listening the "winner" frequency response derived from the listening test in receiving direction was used. Due to the Lombard effect up to 15 dB difference speech levels exists between the listening samples. Therefore individual level adjustment was made in order to create the same loudness for all samples.

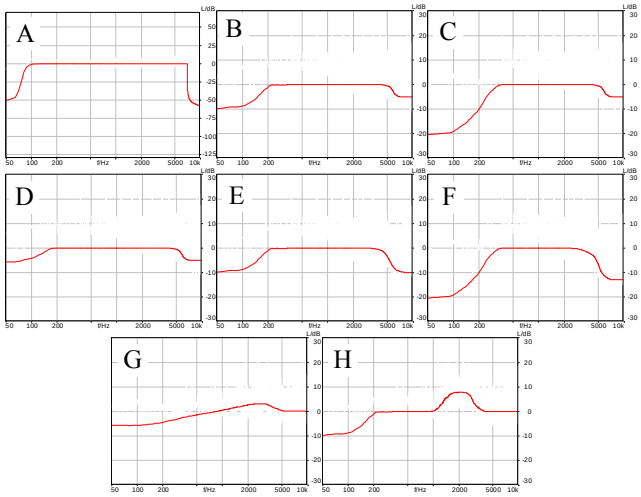

Fig. 8: Sending frequency responses used in the formal listening test

Two speakers, one male and one female were used in the experiment. As a result 36 conditions with two speakers were assessed each by 16 naïve test persons. The presentation was diotic. The assessment was based on ITU-T Recommendation P.800 [1] "overall quality" using a 5 point MOS scale.

Based on these subjective tests the following conclusions can be drawn:

- The higher the background noise level is the stronger the Lombard effect is. As a consequence the MOS scores are lower.
- For some background noises the results for males are better than for females (female speakers tend to sound shrill and sharp for higher background noises in conjunction with the Lombard effect)
- No clear preference for one response characteristics was found for any type of noise. **Only tendencies could be found.**
- The full bandwidth is preferred for most noises.
- Strong high pass characteristics are not preferred by the listeners.

6.2 Second Listening Test

Since no clear recommendation could be given based on this test a second formal listening test was conducted. In this test three different noises were assessed separately. Thus the test subjects could better concentrate on the differences just due to filtering. The same phone as previously mentioned was used. However the numbers of filters per noise were reduced. SNR conditions with 5 dB and 10 dB worse SNR ("SNR-5", "SNR-10" in **figure 9** to **11**) for each noise and filter were added. Furthermore the naïve test persons listened three times to the sample and assessed three parameters during the listening test:

- listening effort
- speech sound quality
- overall quality

This type of test is similar to the one as described in ITU-T Recommendation P.835 [5].

As an example the results of the second formal listening tests for the "café noise" are shown in **figures 8 to 10**. The listening effort, the speech sound and the overall quality are displayed for the male and female voice separately.

Despite the fact that there are small differences between male and female voices only slight differences in MOS scores can be found with respect to the full band transmission for mostly all conditions. The strongly band-pass filtered version leads to a similar listening effort compared to the full-band version, but to a reduced speech sound (and also overall) quality – especially for the male voice. Since this background noise has no dominant low frequency components, the limited bandwidth impairs the speech quality instead of reducing the annoyance due to the background noise.

Based on the other background noise conditions in general the following conclusions can be drawn from this experiment:

- All conditions (except those with 10 dB lower SNR were perceived in the range of "good" → high MOS scores
- Strongly band-pass filtered version with 10 dB worse SNR than original got worst results for all noises and test parameters → no further analysis of this condition
- For most of the samples the MOS score for listening effort, speech sound and overall quality were similar
- Two separate groups of test persons were found:
  1. Noticed nearly no difference between the samples → MOS scores always in the range of 4 to 5
  2. Realized that there were differences between samples → MOS scores between 1 and 5, relatively high confidence intervals

Finally no real "favourite" response characteristics (except the full-band version) was identified by the naïve test persons. This was possible only for experts. Since the effect of noise reduction algorithms was not investigated in this test, it may be helpful to implement an "acoustical noise reduction" by applying a moderate or – depending on the noise – even medium band-pass filter. This would "acoustically" reduce the noise by a linear filter and only slightly affect the speech sound. The implemented noise reduction algorithms may then be adjusted less aggressive. Thus impairments on the speech quality due too strong noise reduction can be avoided. A proposal for such a response characteristics is shown in **figure 12.**

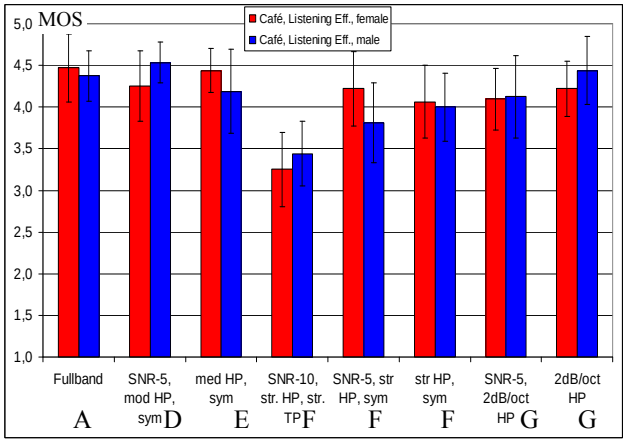

Fig. 9: Listening effort café noise, filter indication acc. to **figure 8**

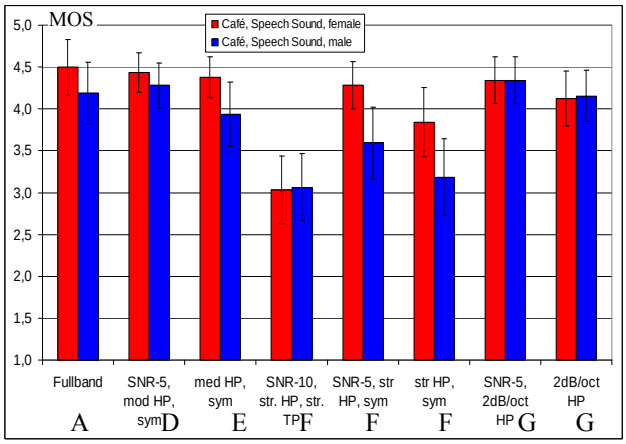

Fig. 10: Speech Sound quality café noise, filter indication acc. to **figure 8**

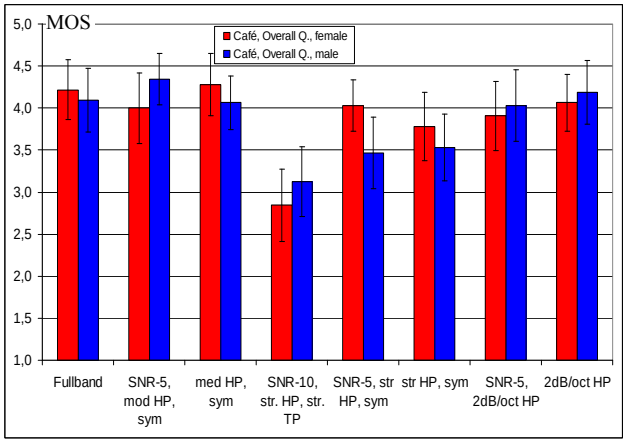

Fig. 11: Overall quality café noise, filter indication acc. to **figure 8**

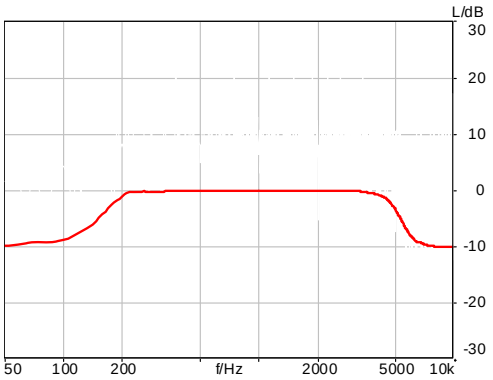

Fig. 12: Shaping curve for the frequency response characteristics in sending to be adaptively inserted depending on the background noise (see also figure 8, E)

7 SUMMARY & CONCLUSION

The evaluation of today's wideband measurement standards as currently defined in ETSI ES 202 739 and ES 202 740 can be improved based on the results described in this paper.

In receiving direction, a tolerance mask was derived based on subjective tests which will lead to a good speech sound quality for wideband phones. It is recommended to use the diffuse field equalization of the HATS in conjunction with the artificial ear used.

The subjective experiments conducted in sending direction do not give a clear and unambiguous answer to the question about the optimum frequency response characteristics in the presence of background noise. Most naïve listeners seem to prefer the full wideband transmission even in the case of background noise. However, it might be useful to adaptively and carefully narrow the bandwidth in sending direction depending on the level and type of background noise. This should be studied further especially in conjunction with noise canceling techniques.

8 REFERENCES

[1] ITU-T Recommendation P.800: Methods for subjective determination of speech quality, International Telecommunication Union, Geneva, 2003

[2] ETSI ES 202 739: Transmission requirements for wideband VoIP terminals (handset and headset) from a QoS perspective as perceived by the user, Oct. 2007

[3] ETSI ES 202 740: Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user, Oct. 2007

[4] The influence of acoustics on speech production: A noise-reduced stress phenomenon known as Lombard reflex, Jean-Claude Junqua, Speech communication 20 (1996), p. 13-22

[5] ITU-T P.835: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm, 2006

[6] ITU-T P.58, Head and Torso Simulators for Telephonometry, 1996
[7] ITU-T P.57, Artificial Ears, 1996

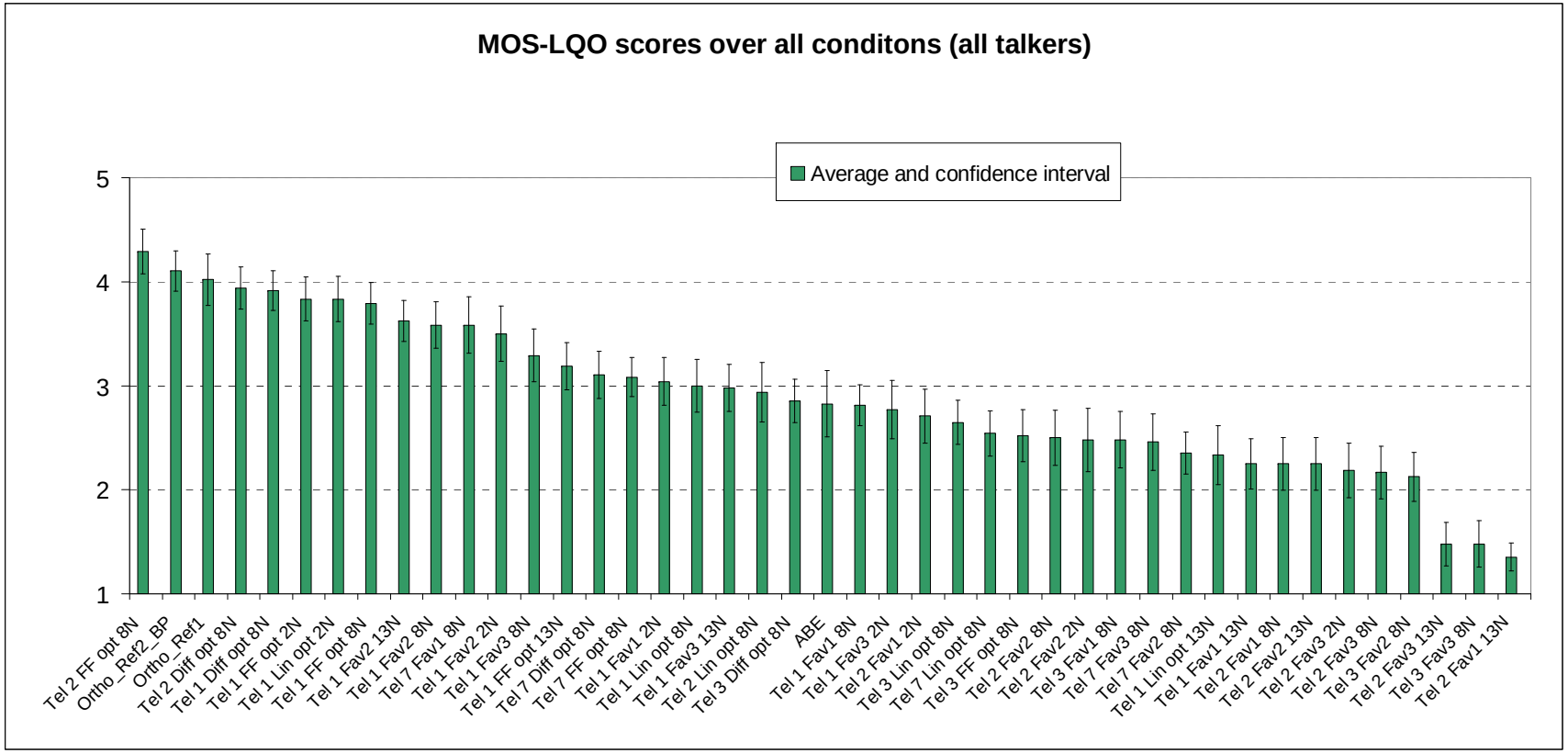

Figure 4: MOS-LQSw results for all 43 scenarios averaged over four talkers