Chapter 6
Wideband Hands-Free in Cars – New Challenges for System Design and Testing

Hans W. Gierlich and Frank Kettler

Abstract Wideband hands-free technology in cars provides the capability to substantially improve the quality of the perceived speech for the driver as well as for the far-end communicational partner. However, in order to achieve a superior wideband speech quality, a variety of requirements – different from narrowband telephony – have to be taken into account. A few important parameters most critical for the success of wideband in cars are discussed. Since wideband transmission is at least partially IP-based, a higher delay can be expected as compared to narrowband calls. The impact of higher delay on the communicational quality is shown, and the different elements contributing to the delay in car hands-free systems are shown. Also, the impact of delay on conversational quality is discussed. The other aspects of wideband communication include speech sound quality in sending and receiving direction. A new objective test procedure 3QUEST for speech quality with background noise and its application to wideband car hands-free is introduced. For echo performance in wideband, new subjective test results are shown, and results of a new objective echo analysis method based on the hearing model "Relative Approach" are shown.

Keywords Human perception • System design • Wideband hands-free technology

6.1 Introduction

The deployment of wideband hands-free technology in cars provides the capability to substantially improve the quality of perceived speech for the driver as well as for the far-end communicational partner. In-vehicle hands-free terminals would benefit from wideband than traditional communication terminals. The difference in sound

H.W. Gierlich (✉) • F. Kettler
HEAD Acoustics GmbH, Herzogenrath, Germany
e-mail: H.W.Giderlich@head-acoustics.de

J.H.L. Hansen et al. (eds.), *Digital Signal Processing for In-Vehicle Systems and Safety*,
DOI 10.1007/978-1-4419-9607-7_6, © Springer Science+Business Media, LLC 2012

quality would immediately be noticeable to the driver since she/he will always have a perceptual comparison of the high-quality audio playback in the car for other media. Speech intelligibility in the car will be significantly increased, which is highly beneficial, especially in background noise situations while driving. As a consequence, the listening effort for the driver is reduced, the distraction from the primary task (driving) will be reduced as well. Thus, the driver's distraction may be reduced substantially if wideband technology is implemented properly. However, in order to achieve a superior wideband speech quality, a variety of requirements different from narrowband telephony have to be taken into account. This includes careful system design of all components involved in the transmission. The impact of delay and the components contributing to delay are described in Sect. 6.2. The listening speech quality analyses for wideband car hands-free systems are described in Sect. 6.3, and the special requirements on echo performance are given in Sect. 6.4.

6.2 Transmission Delay

Since wideband transmission is most likely IP-based when connecting to a fixed line network, a higher delay can be expected as compared to narrowband calls. The higher delay not only contributes to a degraded communicational quality but also requires a more thorough investigation of the echo loss required for wideband systems. This concerns spectral as well as temporal characteristics and is discussed in Sect. 6.4.

An overview of the components of a typical hands-free system and their effect on delay is given in Fig. 6.1.

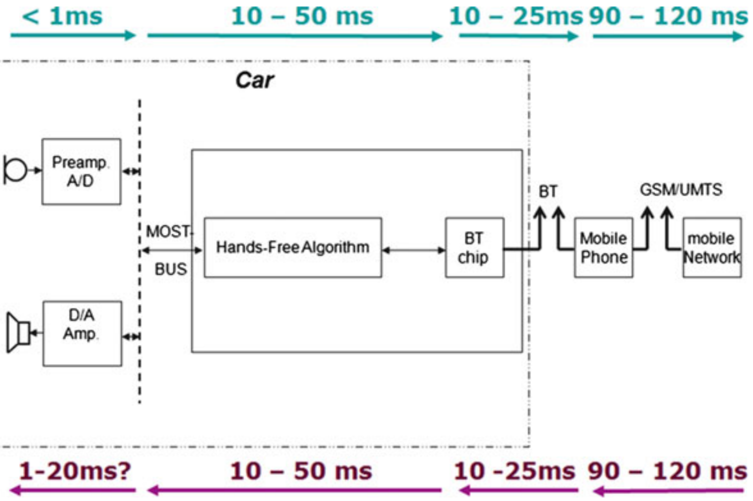

Fig. 6.1 Typical components of a car hands-free system and their contribution to transmission delay

While the microphone and its connection to the in-car audio or in-car bus system typically introduces low delay, the hands-free algorithm in the uplink (sending direction) may introduce a significant transmission delay. In uplink, the most important signal processing is active: echo cancelation and noise cancelation. Both require substantial signal processing capacity, and in wideband systems, it is likely that these algorithms are realized in the frequency domain and/or in sub-bands. These technologies are known not only to provide good performance [1] but also to introduce higher delay – compared to simple LMS-type algorithm.

Signal processing in downlink may also introduce more delay than known in narrowband systems. This is caused, e.g. by advanced adaptive signal enhancement techniques, such as adaptive equalization or compression, and especially by wideband extension techniques. Such techniques can be used to generate a pseudowideband signal from narrowband speech and would help to minimize the perceived speech sound quality between wideband and narrowband calls (see [2, 3]). An additional source of delay might be the audio processor which is used to enhance the audio presentation of other audio sources in the car.

The Bluetooth® connection is the most typical link between the hands-free system and the mobile phone today. Currently, the Bluetooth® wideband specification is not yet available. In order to achieve a superior speech sound quality in conjunction with a low delay, tandem-free coding would be desirable. This would require the support of the AMR wideband transmission over the Bluetooth® link and the realization of speech coding and decoding in the hands-free system. However, an additional coder for the Bluetooth® link is in discussion. This would introduce additional distortion to the speech signal and increase significantly the overall delay in a connection. For a superior wideband service, such implementation is not desirable.

Summing up the delays assumed from Fig. 6.1, the transmission delay would be around 200 ms from car to car in the best case. Assuming an average Bluetooth® delay of about 30 ms and a fixed network delay of 50 ms, it is quite likely that the transmission delay in such a connection exceeds 400 ms.

The effect of delay in transmission systems is well known and described in ITU-T Recommendation G.131 [4] and G.107 [5]. While in [4], the impact of delay on the required echo loss is described, ITU-T Recommendation G.107 [5] gives an insight of delay on users' satisfaction. Although these investigations are still based on narrowband transmission, a similar impact can be expected in wideband systems. Figure 6.2 shows the impact of delay on user satisfaction [5], assuming ideal performance of all components in a connection except echo loss.

It can be seen that even with perfect echo loss, many users will be dissatisfied when exposed to a transmission delay of 400 ms or more. This is clearly not advisable for a superior service. But even with lower transmission delays, an excellent echo loss is required in order to achieve good users' satisfaction.

As a consequence, any component in a car hands-free system should be designed in such a way that only a minimum of delay is inserted.

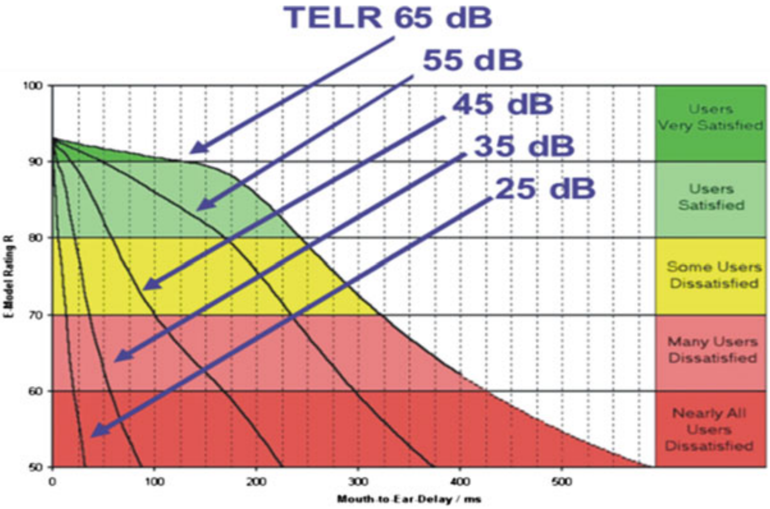

Fig. 6.2 Users' satisfaction depending on delay and TELR (TELR = SLR + RLR + Echo Loss) from [5]

6.3 Listening Speech Quality

The performance requirements for the speech quality in receiving are probably easiest to fulfill due to the high quality of built-in car audio systems. For aftermarket hands-free systems, this is much more challenging. The extension of the frequency range in sending direction not only provides better representation of the low-frequency components of the transmitted speech but also increases the amount of noise transmitted by the microphone. This is of particular importance because the in-car noise is dominant in the low-frequency range. It imposes additional quality requirements on all speech enhancement techniques such as beamforming for microphones, noise cancelation, and others.

An objective measure 3QUEST according to ETSI EG 201 396-3 [6] is capable of determining the speech, noise, and overall quality, and such can be used in the optimization of wideband hands-free systems. The algorithm calculates correlation between the processed signal – typically recorded in sending direction of a hands-free system (uplink) – and two references, the original clean speech signal and the signal recorded close to the hands-free microphone. This signal consists of the near-end speech and the overlapped in-car noise. The algorithm is described in [6] and [7] in detail. Statistical analyses lead to a one-dimensional speech quality score (S-MOS), a noise quality score (N-MOS), and an overall quality score representing the general impression (G-MOS). The algorithm is narrowband and wideband capable and provides correlations in the range of >0.91 to the results of subjective tests.

The model was developed and trained with a certain amount of given randomized data (179 conditions). The rest of the databases were used for own validation only. During the development of the algorithm in the STF 294 project [8], the subjective S-, N-, and G-MOS results of 81 conditions remained unknown until the end of the algorithm development.

The 179 different test conditions included existing hands-free terminals and hands-free simulations in combination with different background noise scenarios such as in-car noise and outdoor road noise. The following plots show a very small amount of these data comparing subjective and objective results for the narrowband and wideband test case in hands-free conditions.

The subjective and objective results (S-MOS, N-MOS, and G-MOS) do not differ by more than 0.5 MOS in the narrowband case (see Figs. 6.3–6.5). This can be regarded as very reliable, especially when considering the complexity of this listening situation and amount of signal processing typically involved. The same can be analyzed for wideband scenarios, as shown in Figs. 6.6–6.8.

The correlation coefficient and root mean square error (RMSE) between the subjective and objective MOS data are shown in Table 6.1 for the entity of all 179 wideband test conditions.

This analysis method provides comprehensive quality scores for uplink transmission quality. It needs to be combined with further detailed parameter analyses like measurements of loudness ratings, frequency responses, signal-to-noise ratio, and others in order to provide the "whole picture" for a given implementation. Furthermore, the combination of comprehensive quality scores, on one hand, and detailed parameter analyses, on the other, may provide important hints for quality improvement and tuning, if necessary.

6.4 Echo Performance

Conversational aspects of wideband communication are important as well for the success of wideband services. Therefore, the requirements for conversational parameters such as double-talk capability and echo performance are to be revisited with respect to different perceptions between narrowband and wideband telephony.

As seen before, the delay plays a crucial role for echo perception. Furthermore, the extended transmission range in wideband scenarios and the spectral content of echoes strongly influence echo perception. This also demands new analysis techniques and requirements for wideband echo perception.

Current echo analyses combine various single measurements like echo attenuation or spectral echo loss and verify the compliance to requirements and tolerances. These parameters are incomplete, neither perception-oriented nor aurally adequate. They do not appropriately consider wideband-specific aspects. New investigations on wideband echo perception further point out that the spectral echo content in the frequency range between 3.1 and 5.6 kHz is especially crucial for echo disturbance [9]. New tolerances for the spectral echo attenuation have therefore been introduced in [9].

Fig. 6.3 S-MOS, narrowband HFT

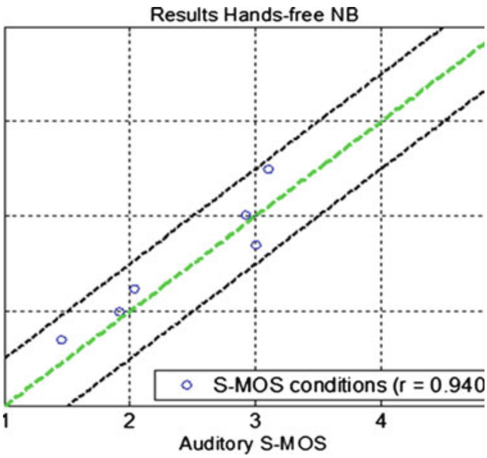

Fig. 6.4 N-MOS, narrowband HFT

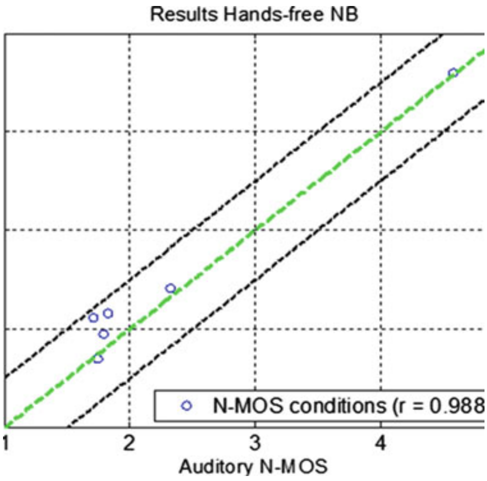

Fig. 6.5 G-MOS, narrowband HFT

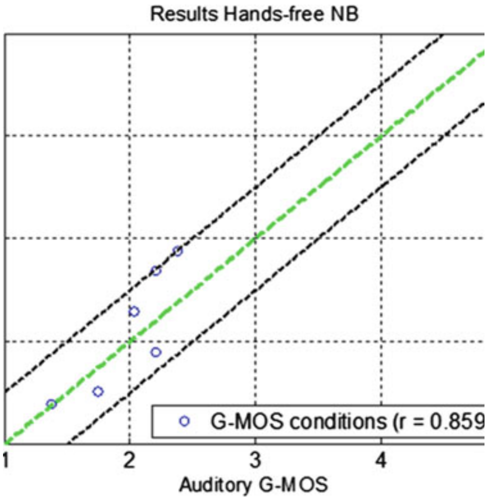

Fig. 6.6 S-MOS, wideband HFT

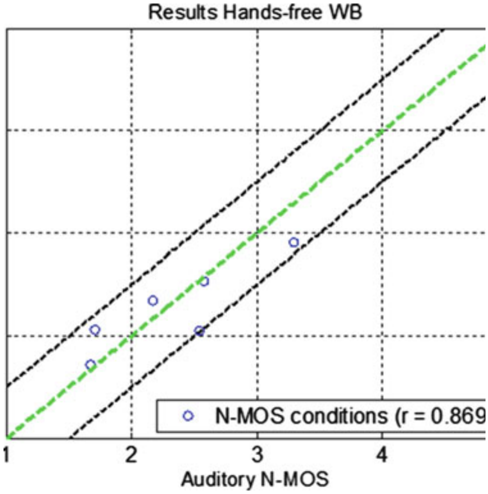

Fig. 6.7 N-MOS, wideband HFT

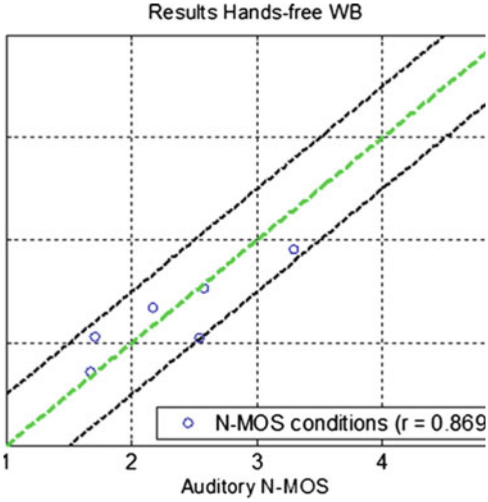

Fig. 6.8 G-MOS, wideband HFT

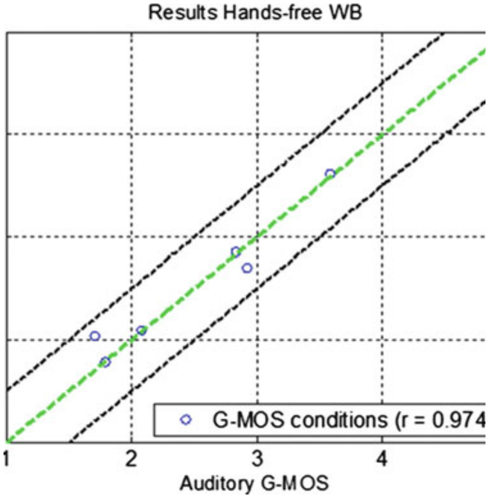

Table 6.1 Correlation and RMSE of prediction for wideband database

	Training		Validation	
	corr.	RMSE	corr.	RMSE
S-MOS	91.2%	0.37	93.0%	0.33
N-MOS	94.3%	0.27	92.4%	0.32
G-MOS	94.6%	0.25	93.5%	0.28

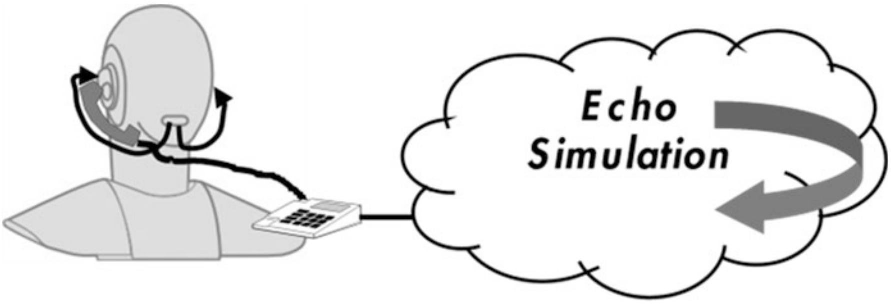

Fig. 6.9 Principle of binaural recordings for third-party listening tests (Type A [15, 16])

A consequent next step in the field of analysis techniques is the development of an objective model providing one-dimensional values with high correlation to the MOS results from subjective tests. Models providing good correlations for echo assessment have already been evaluated for narrowband telephony, distorted sidetone, and room reverberations [10]. A new model based on the Relative Approach [11] may be applicable for narrowband and wideband telephony and may deliver hints for improvement of devices under test such as acoustic or network echo cancellers. The Relative Approach method is especially sensitive to detect unexpected temporal and spectral components and can be used as an aurally adequate analysis to assess temporal echo disturbances [12–14].

The Third-Party Listening Tests were carried out with 20 subjects in total, 14 naïve and 6 expert listeners. The speech material consists of male and female voices.

The basis for a new echo model – like for all other objective analyses – must be the subjective impression of test subjects. Therefore, subjective echo assessment tests were carried out first under wideband conditions. In principle, these tests can be conducted as so-called Talking-and-Listening Tests according to ITU-T P.831 [15] or as Third-Party Listening Tests based on artificial head recordings (ITU-T P.831, Test A [15, 16]). The principle of the recording procedure is shown in Fig. 6.9. A wideband-capable handset was simulated at the right ear of the HATS [17]. Besides the more efficient test conduction – a group of test subjects can perform the tests at the same time – the listening tests provide the advantage that the same audio files, as assessed in the subjective test, can be used for the objective analyses.

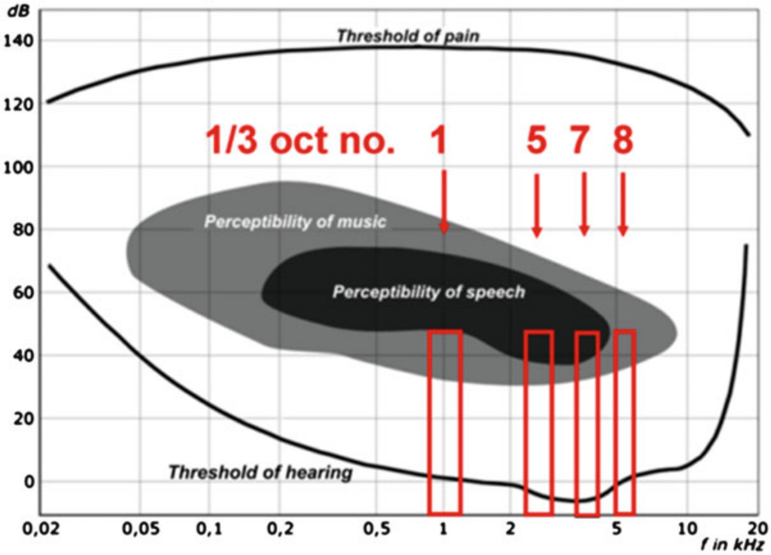

Fig. 6.10 Filter characteristics (subset of test conditions)

A total number of 33 test conditions, including the reference scenarios (infinite echo attenuation) and different combinations of delay, echo attenuation, and spectral shaping, were included:

- Round-trip delays between 100 and 500 ms
- Echo attenuation between 35 and 55 dB
- Simulation of nonlinear residual echoes

The spectral echo content was realized by the following filter characteristics (subset of test conditions):

- NB: narrowband filter, 300–3.4 kHz
- HF1: 3.1–5.6 kHz
- HF2: 5.2–8 Hz
- 1/3 oct.no 1: 900–1,120 Hz
- 1/3 oct.no 5: 2.24–2.8 kHz
- 1/3 oct.no 7: 3.55–4.5 kHz
- 1/3 oct.no 8: 4.5–5.6 kHz

The 1/3 octave filter characteristics are shown in Fig. 6.10 together with the hearing and speech perception threshold. These filters seek a more detailed analysis of the critical frequency range between 1 and 5 kHz which provides the highest sensitivity for sound and speech perception.

A 5-point annoyance scale was used (5 points: Echo is inaudible, . . ., 1 point: Echo is very annoying [18]). The stimuli were presented without pair comparison. The results were analyzed on a MOS basis together with confidence intervals based

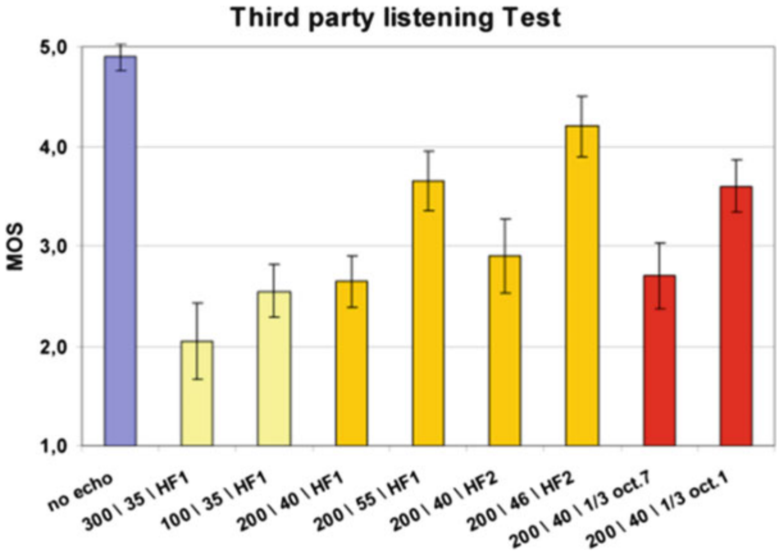

Fig. 6.11 Subset of test results [14]

on a 95% level. Its first analysis pointed out that the quality rating for both groups (naïve, expert listeners) was very similar. The results were therefore combined.

A small subset of results from the listening-only test is shown in Fig. 6.11. The blue bar indicates the echo-free test condition. The rating of 4.8 MOS must be expected under this condition.

One example proving the importance of spectral content on echo perception is given by the red bars in Fig. 6.11. Both conditions represent a 200-ms round-trip delay in combination with a 40-dB echo attenuation. The two different filter characteristics "1/3 oct.1" and "1/3 oct.7" are introduced in Fig. 6.10. The results differ by approximately 1 MOS and point out the strong influence of spectral echo shaping on subjective assessment.

Figure 6.12 shows an example of a Δ 3D Relative Approach between the echo signal *e* and the reference ear signal *r*. The echo signal is recorded at the artificial ear of the HATS. The reference signal *r* represents the sidetone signal in the artificial ear as a combination of acoustical sidetone from mouth to ear and electrical sidetone via microphone and loudspeaker of a wideband-capable handset.

In the first approach, the two-dimensional mean value mΔRAe-r is calculated according to formula:

$$m\Delta RA_{e-r} = \frac{1}{KM} \sum_{k=1}^{K} \sum_{m=1}^{M} \Delta RA_{e-r}(k,\ m) \tag{6.1}$$

where K = no. of freq. bands and M = no. of samples per band.

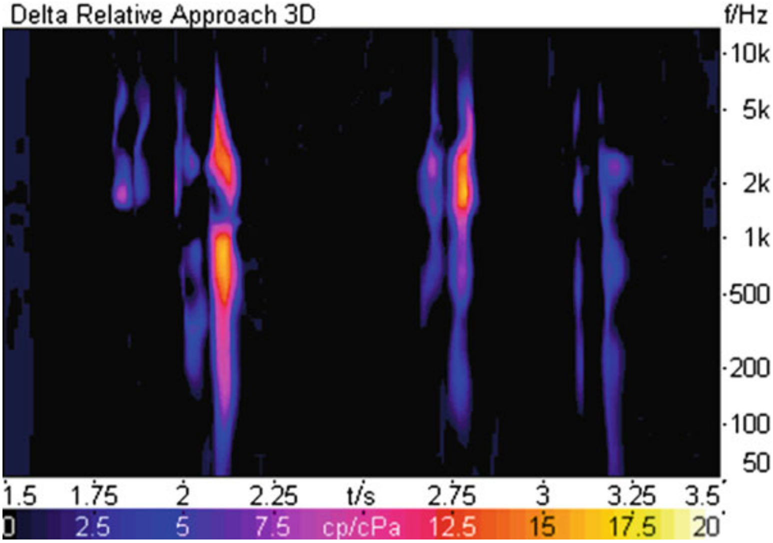

Fig. 6.12 Δ 3D Relative Approach ΔRAe-r(t,f) between the echo signal *e* and the reference ear signal *r*

The parameters echo loss, echo delay, and mΔRAe-r are used as input signal for a linear regression in order to correlate the objective results to the subjective MOS for the echo model.

In the first step, only the two parameters echo loss and echo delay were used in the regression. The result is shown in the left-hand scatterplot in Fig. 6.13. A correlation of $r = 0.80$ is achieved, but the comparison of auditory MOS and objective MOS shows systematical errors: clusters of identical objective MOS occur in Fig. 6.13 (see arrows), which spread over a wide range of auditory MOS (between approximately 1.7 and 3.7 MOS). This can be explained by the different spectral content of these echo signals leading to significant different echo ratings in subjective tests – although the objective parameters (echo delay, echo attenuation) are identical.

The plot on the right-hand side in Fig. 6.13 shows the correlation between the auditory MOS and the objective results based only on the two-dimensional mean value *mΔRAe-r*. The correlation factor increases to $r = 0.84$. The systematical error is implicitly solved using the Relative Approach–based analysis. In principal, this could be expected because the Relative Approach considers the sensitivity of human hearing, especially for different frequency characteristics of transmitted sounds.

The combination of the three parameters *mΔRAe-r*, echo loss, and echo delay to the objective MOS further increases the correlation ($r = 0.90$). The scatterplot is shown in Fig. 6.14 (left-hand side) together with the error distribution in the right-hand picture. The residual error between objective and auditory MOS is below 0.5 MOS in 84% of test conditions.

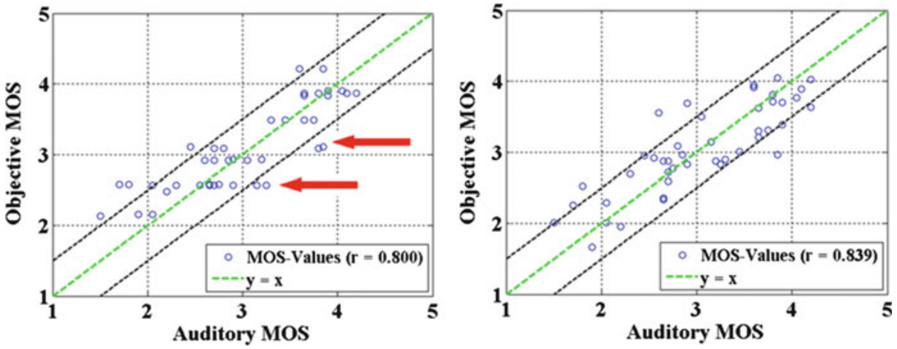

Fig. 6.13 Objective vs. auditory MOS; *left*: input echo loss and echo delay right: *input* mΔRAe-r

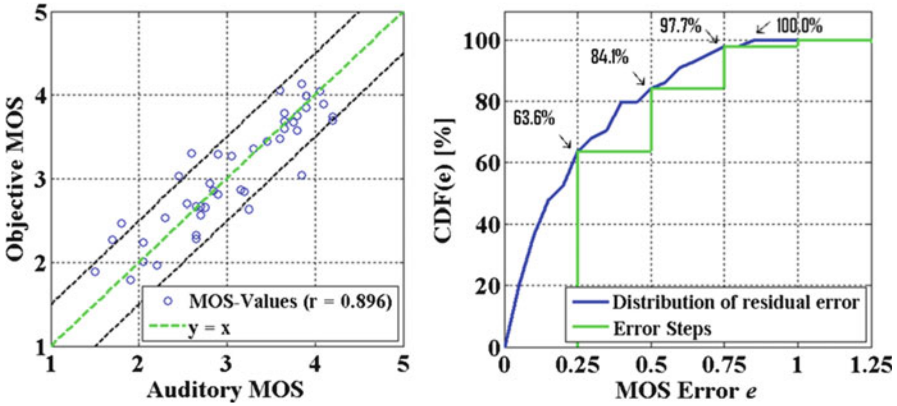

Fig. 6.14 Objective vs. auditory MOS and residual error distribution; input parameter mΔRAe-r, echo loss, and echo delay

Next steps during the development of the echo model are the further adaptation of the Relative Approach on speech characteristics and the application of postprocessing on the resulting Δ 3D Relative Approach $\Delta RAe\text{-}r(t,f)$.

6.5 Conclusions

This chapter introduces several parameters critical to the success of wideband hands-free communication in cars. The impact of delay is shown and discussed. New test results and analysis techniques based on hearing model approaches are given for the speech quality analysis in background noise as well as for echo performance.

Further work is required to derive new analysis techniques and performance criteria for double-talk in wideband systems. It is also clear that the narrowband performance requirements and testing techniques would benefit from such work.

References

1. Hänsler E, Schmidt G (ed) (2008) Speech and audio processing in adverse environments. Springer, Berlin. ISBN:978-3-540-70601-4
2. Vary P, Jax P (2003) On artificial bandwidth extension of telephone speech. Signal Process 83:1707–1719, ISSN 0165-1684
3. Havelock D, Kuwano S, Vorländer M (eds) (2008) Handbook on signal processing in acoustics. Springer, New York. ISBN:978-0-387-77698-9
4. ITU-T Recommendation G.131 (2003). Talker echo and its control
5. ITU-T Recommendation G.107 (2008) The E-model: a computational model for use in transmission planning
6. ETSI EG 202 396-3 V.1.2.1 (2008–11) Speech processing, transmission and quality aspects (STQ); Speech quality performance in the presence of background noise; Part 3: Background noise transmission – Objective test method
7. Gierlich HW, Kettler F, Poschen S, Reimes J (2008) A new objective model for wide – and narrowband speech quality prediction in communications including background noise. In: Proceedings of the EUSIPCO 2008, Lausanne
8. STF 294 project. http://portal.etsi.org/STFs/STF_HomePages/STF294/STF294.asp
9. Poschen S, Kettler F, Raake A, Spors S (2008) Wideband echo perception. In: Proceedings of the IWAENC, Seattle
10. Appel R, Beerendts J (2002) On the quality of hearing one's own voice. JAES 50:237
11. Genuit K. (1996) Objective evaluation of acoustic quality based on a relative approach. In: Proceedings of the Inter-Noise 1996, Liverpool
12. Kettler F, Poschen S, Dyrbusch S, Rohrer N (2006) New developments in mobile phone testing. In: Proceedings of the DAGA 2008. Dresden
13. Lepage M. Evaluation of aurally-adequate analyses for assessment of interactive disturbances. Diploma thesis, IND Aachen
14. Kettler F, Lepage M, Pawig M (2009) Evaluation of aurally-adequate analyses for echo assessment. In: Proceedings of the NAG/DAGA 2009 Rotterdam
15. ITU-T Recommendation P.831 (1998) Subjective performance evaluation of network echo cancellers. International Telecommunication Union, Geneva
16. Kettler F, Gierlich HW, Diedrich E, Berger J (2001) Echobeurteilung beim Abhören von Kunstkopfaufnahmen im Vergleich zum aktiven Sprechen. In: Proceedings of the DAGA 2001, Hamburg
17. ITU-T Recommendation P.58 (1996) Head and torso simulator for telephonometry. International Telecommunication Union, Geneva
18. ITU-T Recommendation P.800 (1996) Methods for subjective determination of transmission quality. International Telecommunication Union, Geneva