10

Evaluation of Hands-free Terminals

Frank Kettler and Hans-Wilhelm Gierlich

HEAD acoustics GmbH
Aachen, Germany

The "hands-free problem" describes the high acoustical coupling between the hands-free loudspeaker and microphone and the resulting acoustical echo for the subscriber on the far end of this connection. It is basically caused by the high distance between the technical interface, i.e. hands-free loudspeaker and microphone, and the human interface, i.e. mouth and ear. The high playback volume – necessary to provide a sufficient playback level at the users ear – and the high sensitivity of the microphone – necessary to amplify the users voice from far distance – leads to a strong coupling – the acoustic echo. Echo cancellers instead of level switching devices are standard today. Moreover, noise reduction and other algorithms further improve the speech quality in noisy environment. On the other hand the use of mobile hands-free telephones in a wide and important application field, i.e. in vehicles, was further enforced by legislation. Hands-free telephones are standard in the automotive industry – at least in middle and upper class vehicles. As a consequence the test procedures and results described in this section mainly focus on the test of hands-free terminals installed in vehicles.

10.1 Introduction

This chapter gives an overview about current evaluation procedures for hands-free terminals, both subjective and objective methods. Sec. 10.2 outlines the principles that need to be considered when testing quality aspects of hands-free terminals. The relevant speech quality parameters are briefly introduced. Sec. 10.3 describes subjective test methods as they have been developed during the last years – from well-known listening-only tests to specific double talk performance tests. The test environment, test signals and analysis methods are introduced in Secs. 10.4 and 10.5. Practical examples of measurement results on different hands-free implementations are used to show the significance of objective laboratory tests. This directly leads to another important aspect, the appropriate summary and representation of the multitude of results, necessary for an in-depth analysis of hands-free implementations. A graphical representation – best described as a "quality pie" – bridges the gap between the complexity and multitude of tests on one side and the need for a quick and comprehensive summary on the other side. It is introduced and discussed in Sec. 10.6. The last section ends up with a discussion of ideas and related aspects in speech communication over hands-free phones and quality testing.

This chapter and especially the practical examples focus on mobile hands-free implementations – simply due to three facts:

- They are standard in the automotive industry today and probably the most rapidly growing hands-free market over the last and coming years.
- Furthermore, the ambient conditions in a driving car, like vehicle noise as one of the crucial parameters for echo cancellation algorithms in identifying the impulse responses, are very critical.
- Last but not least the costs of these systems increase the user's expectation on quality – but do not always satisfy it.

10.2 Quality Assessment of Hands-free Terminals

Hands-free implementations with their typical components like microphone arrays, echo cancellation, noise reduction and speech coders are highly non linear, time variant, speech controlled devices. The development of both subjective and objective quality assessment methods requires a deep understanding of the complexity of each signal processing component and especially the interaction between them. A principal block diagram can be found in Fig. 10.1.

The sending direction (uplink transmission path) typically comprises the microphone or microphone array with its associated algorithms for beamforming. In addition, acoustic echo cancellers (AEC) combined with additional post processing – often also designated as echo suppression or non linear processor – or automatic gain control (AGC) provide the main functionality for reducing the acoustically coupled echo. The block diagram in Fig. 10.1 represents a general example.

In principal different combinations of beamforming and AEC can be realized ("AEC first", "beamforming first", see [39]). Hands-free algorithms and microphone solutions are typically not provided by the same manufacturer, microphone solutions might even change between a single microphone solution and an array during the life cycle of a vehicle type. Consequently there is a high demand for flexible implementations.

Noise reduction algorithms shall further improve the near end signal by algorithmically reducing the added noise from the near end speech signal. Speech coders then provide the signal conditioning for RF transmission.[1]

[1] *RF* abbreviates *radio frequency.*

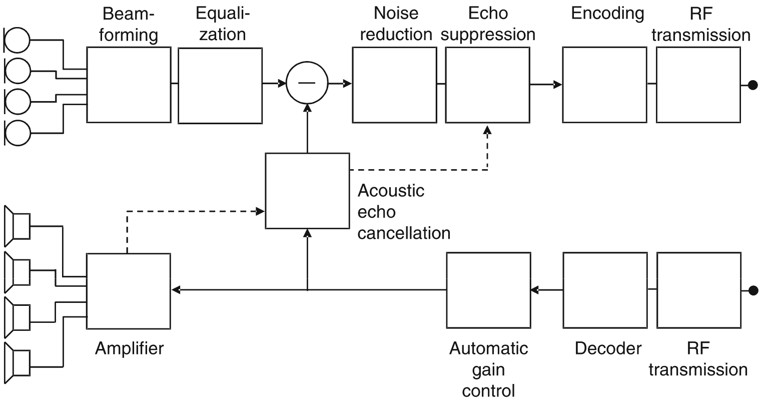


Fig. 10.1. Block diagram of a hands-free implementation with typical components like microphone arrays, echo cancellation, noise reduction and speech coders.

The receive direction (downlink transmission path) typically consists of the speech decoder, potential AGC, e. g. to adapt the volume automatically to the speed and background noise level in the driving car, the playback system and the loudspeakers itself.

The listening speech quality in receiving direction can typically be assessed by subjective listening tests (see Sec. 10.3.5). Objective methods typically reproduce the same situation, the analysis of recorded test signals or real speech at the driver's position using artificial head recording systems on the drivers' seat [25, 48]. The quality is influenced by the loudness of the transmitted speech, the frequency content, the signal to noise ratio in the driving car, the intelligibility and the absence of additional non linear disturbances. The built-in loudspeakers are typically used for playback. In contrary to the sound system used for CD or radio playback, only the front speakers – installed in the door in the driver's and co-driver's footwell – sometimes combined with center speakers are used.

In a similar way the speech quality in sending direction can also be assessed subjectively by listening tests. This transmission path is especially important because all relevant signal processing like microphone characteristics, potential beamforming algorithms and noise reduction are implemented in sending direction. This transmission path is extremely critical in terms of signal to noise ratio caused by the high distance between microphone and driver's mouth. In a technical sense the relevant parameters are the microphone position, the microphone frequency response, the signal to noise ratio, the frequency content of speech and noise, the intelligibility, artifacts like musical tones or other non linear distortions. Objective measures are therefore again based on analyses of transmitted speech or test signals in sending direction with and without background noise.

The two transmission paths can not be regarded as independent from each other. It is obvious that hands-free implementations do not only influence the one-way transmission quality. A comprehensive quality assessment either subjectively or objectively therefore needs to consider all conversational aspects including echo performance, double talk capability – both subscribers act at the same time – and the quality of background noise transmission.

The echo performance is typically assessed by so-called "talking and listening tests" (see Sec. 10.3.4). Test persons assess the echo while they are using a common telephone handset providing standard characteristics at the far end. It is also possible to judge the echo disturbance in listening tests if the self masking effect is considered. Up to a certain extent, this can be reproduced using artificial head technology at the far end [31]. The echo performance is technically influenced by the echo attenuation expressed in dB values, the delay, echo fluctuations vs. time (temporal echo aspects), the spectral content of echo and the "intelligibility" or "clearness" of echoes.

The double talk performance requires conversational tests with two participating subjects at the same time (see detailed description in Sec. 10.3.3). However even this conversational situation can – up to a certain extent – be reproduced by a listening test on simulated conversations using two artificial head testing systems. The double talk capability is mainly determined by three parameters:

- Audible level variations and modulations in sending direction typically introduced by AEC post processing or AGC,
- audible modulation in receiving direction and
- the echo disturbance during double talk.

Last but not least the transmission quality of background noise plays a very important role for mobile hands-free implementations. These devices are typically used in driving cars, thus the background noise situation is critical, the level is rather high. Consequently, this transmission aspect cannot be disregarded any longer. It is important that the background noise itself is not only regarded as a disturbing factor. It carries important information for the conversational partner at the far end side. Consequently, a pleasant and smooth transmission quality needs to be ensured.

10.3 Subjective Methods for Determining the Communicational Quality

The main purpose of telecommunication systems is the bi-directional exchange of information. This has to be considered in the design of terminals and networks. In the first step, quality assessment is always subjective – taking into account the quality as perceived by the user of a communication system including all aspects of realistic conversational situations. When assessing hands-free systems the special conditions where the system is used have to be identified.

For mobile hands-free terminals typical acouctical environmental condtions such as telephoning on a street at the airport or other typical situations have to be considered. For car hands-free systems the acoustical conditions in a car have to be taken into account. Since all quality parameters are based on sensations perceived by the subjects using the communication system, the basis of all objective assessment methods are subjective tests. They have to be defined carefully in order to reflect the real use situation as closely as possible and, on the other hand, to provide reproducible and reliable results under laboratory conditions.

As for other telecommunication systems and services the basic description of the subjective assessment of speech quality in telecommunication is ITU-T Recommendation P.800 [29]. The methods described here are intended to be generally applicable whatever type of degradation factors are present. The ITU-T Recommendation P.832 [32] is of special importance since it addresses the special requirements and test scenarios for hands-free systems. ITU-T Recommendation P.835 [33] is the most relevant recommendation when assessing the speech quality in the presence of background noise which is of major importance in car hands-free systems. It is especially useful when optimizing the design and parameterization of noise canceling techniques based on subjective judgments.

10.3.1 General Setup and Opinion Scales Used for Subjective Performance Evaluation

Subjective testing requires an exactly defined test setup, a well defined selection procedure of the test subjects participating in the tests as well as exact and unambiguous scaling of the scores derived from the subjects participating in subjective experiments.

The setup of the test depends on the type of test to be conducted. Independent of the type of test, as a general rule the test setup should be as realistic as possible. For car hands-free testing the environment chosen for the tests should be “car-like” when evaluating parameters relevant for the user in the car. Other applications require the simulation of their typical use conditions.

Furthermore, the results of a subjective test highly depend on the type of test subjects participating in the test. According to ITU-T Recommendation P.832 the following types of test subjects can be identified:

- **Untrained subjects**

  Untrained subjects are accustomed to daily use of a telephone. However, they are neither experienced in subjective testing nor are they experts in technical implementations of hands-free terminals. Ideally, they have no specific knowledge about the device that they will be evaluating.

- **Experienced subjects**

Experienced subjects (for the purpose of hands-free terminal evaluation) are experienced in subjective testing, but do not include individuals who routinely conduct subjective evaluations. Experienced subjects are able to describe an auditory event in detail and are able to separate different events based on specific impairments. They are able to describe their subjective impressions in detail. However, experienced subjects neither have a background in technical implementations of hands-free terminals nor do they have detailed knowledge of the influence of particular hands-free terminal implementations on subjective quality.

- **Experts**

  Experts (for the purpose of hands-free terminal evaluation) are experienced in subjective testing. Experts are able to describe an auditory event in detail and are able to separate different events based on specific impairments. They are able to describe their subjective impressions in detail. They have a background in technical implementations of hands-free implementations and do have detailed knowledge of the influence of particular hands-free implementations on subjective quality.

For the identification and general evaluation of parameters influencing the communicational quality in hands-free terminals typically untrained subjects are used. Experts and experienced subjects typically are used in order to optimize the performance of a hands-free terminal in a very efficient way. Experts may be used for all types of tests. Care should be taken in case only experts are used in a test since they may focus on parameters not of significance for the average user while missing other parameters average users may find significant. Typically the expert's judgement is validated by untrained subjects representing the average user group the set is intended to be used for.

Subjective testing requires scales easily understandable by the subjects, unambiguous and widely accepted. The design and wording of opinion scales, as seen by subjects in experiments, is very important. Different ways of scaling and different types of scales may be used. Here those most often used in telecommunications are described. More information can be found in the ITU-T P. 800 series Recommendations and [44] and [37].

One of the most frequently used rating is a category rating obtained from each subject at the end of each subjective experiment (see [29, 30]) which is typically based on the following question: Your opinion of (the overall quality, the listening speech quality, etc.) of the connection you have just been using:

1 – excellent
2 – good
3 – fair
4 – poor
5 – bad

The averaged result of a this so-called ACR (absolute category rating) test is a mean opinion score MOS. If applied in a conversational test, the result is MOSc (Mean opinion score, conversational). If the scale is used for speech quality rating in listening tests, the result is called MOS (mean listening-quality opinion score).

DCR (degradation category rating) tests are used if degradation occurs in a transmission system. The scale used is given as follows (see [29]):

5 – degradation is inaudible
4 – degradation is audible but not annoying
3 – degradation is slightly annoying
2 – degradation is annoying
1 – degradation is very annoying

The quantity derived from the scores is termed DMOS (degradation mean opinion score).

Sometimes only small differences in quality need to be evaluated. This is relevant e.g. for system optimization or when evaluating higher quality systems. In such experiments a comparison between systems is made and comparison rating is used. The scale used in CCR (comparison category rating) tests is given as follows (see [29]):

The quality of the second system compared to the quality of the first one is

3 – much better
2 – better
1 – slightly better
0 – about the same
−1 – slightly worse
−2 – worse
−3 – much worse

Especially in conversation tests sometimes a binary response is obtained from each subject at the end of a conversation, asking about talking or listening difficulties over the connection used. In such conditions the score is simply "Yes" or "No". Further information can be acquired if the experimenter carefully tries to identify the type of difficulties experienced by the subject in cases where the subject indicates difficulties.

More scales are known, additional information about scales and testing can be found in [28–33].

10.3.2 Conversation Tests

The most realistic test known for communication systems is the conversation test. Both conversational partners exchange information, both act as talker and as listener. Ideally, the test is set up in a way that subjects behave very similar to a real conversation. Therefore the tasks chosen for a conversation test should be mostly natural with respect to the system evaluated. In car hands-free evaluations a situation should be chosen where at least one of the conversational partners is immersed in a (simulated) driving situation. The task chosen for the experiment should be easy to perform for each subject and avoid emotional involvement of the test subjects. Furthermore, the test should be mostly symmetrical (the contribution of each test subject to the conversation should be similar) and it should be independent of the individual personal temperament (the task must stimulate people with low interest in talking and must reduce the engagement of people who always like to talk). Examples for tests used in telecommunication are the so-called “Kandinski tests” (see [31, 32]) or the so-called “short conversational tests” (see [32, 44]).

The “Kandinsky test” is based on pictures with geometrical figures including numbers at different positions in the picture. Each subject has the same picture in front of him but with the numbers at different positions in the picture. The subjects are asked to describe to their partner the position of a set of numbers on a picture. For the subjective evaluation of car hands-free systems this test could be used only in situations where the simulation of the driving task is of minor importance and the focus of the test is mainly on the conversational quality without taking into account additional tasks.

In the so-called “short conversational tests”, the test subjects are given a task to be conducted at the telephone similar to a daily-life situation. Finding a specific flight or train connection, ordering a pizza at a pizza service are examples of typical tasks. These tasks can also be performed in the driving situation.

It is advisable to include a sufficient number of talkers in the conversational tests to minimize talker/speaker-dependent effects. The test persons used should be representative with respect to gender, age etc. for the user group of the system evaluated.

Due to the complexity of the task itself, subjects mostly rate their opinion about the overall quality of a connection based on the ACR scale. Often they are asked about difficulties in talking or listening during the conversation. Careful investigation of the nature of these difficulties may require more specialized tests than described below. A more detailed parameter investigation can only be made if experienced subjects or experts are used in the conversation test.

10.3.3 Double Talk Tests

The ability to interact in all conversational situations and especially to interact during double talk with no audible impairments is of critical importance in car hands-free communication. Due to the difficult acoustical situation especially in a car a variety of measures are implemented in a hands-free terminal which may impair the speech quality during double talk. Double talk tests may help to evaluate the system performance under such conditions. The double talk testing method (see [14, 15, 31]) is designed especially for the quality assessment during double talk periods. The test duration is very short, they are very efficient in subjectively evaluating this very important quality aspect in detail.

In general the test setup is the same as for conversational tests. Double talk tests involve two parties. In double talk tests untrained subjects are used when it is important to get an indication of how the general telephone using population would rate the double talk performance of a car hands-free telephone. The test procedure is sensitive enough to let untrained subjects assess the relevant parameters even during sophisticated double talk situations. Experienced subjects are used in situations where it is necessary to obtain information about the subjective effects of individual degradations.

During double talk tests, two subjects read a text. The texts differ slightly. Subject 1 (talking continuously) starts reading the text. It consists of simple, short and meaningful sentences. Subject 2 (double talk) has the text of subject 1 in front of him, follows the text and starts reading his text simultaneously at a clearly defined point. Clearly this situation is less realistic than in a conversation test. Even if the text is very simple, the subjects have to concentrate in a different way compared to a free conversation.

Parameters which are assessed typically using double talk tests are: the dialog capability, the completeness of the speech transmission during double talk, echo and clipping during double talk. In most of the tests ACR or DCR scales are used.

10.3.4 Talking and Listening Tests

Nowadays many hands-free terminals are often used in mobile or IP based transmission systems. As a consequence the delay introduced in a transmission link increases. Complex signal processing in the terminals may add additional significant delay. Therefore the investigation of talking-related disturbances like echo or background noise modulation is of critical importance. In order to investigate such types of impairments in more detail, talking and listening tests can be used. Such tests are mainly used to investigate the performance of speech echo cancellers (EC) and the noise canceller (NC) integrated in the car hands-free terminal. All aspects of EC and NC functions that influence the transmission quality for subscribers while they are either talking-and-listening are covered by this procedure.

The EC and NC implementations of a hands-free terminal are the focus of the setup – it is found on one side of a (simulated) connection. The performance of this implementation is judged by a test subject placed at the opposite end of the (simulated) connection. From the subjects point of view, this is the far-end echo and noise canceller.

A potential far-end subscriber can be simulated by an artificial head if double-talk sequences are required in the test. In this case the artificial mouth is used to produce exactly defined double talk sequences. The environmental conditions used at the far end side (e.g. car-hands-free) should correspond to the typical environmental conditions found in a car especially with respect to background noise.

The test procedure may focus on the examination of the initial performance of a hands-free terminal, e.g., the convergence of an echo canceller or in a second part on the evaluation of the performance under steady-state conditions.

When testing the initial performance, subjects answer an incoming telephone call with the same greeting: e.g. '[name], [greeting]'. After the greeting, the call is terminated and subjects give their rating.

When testing "steady-state conditions", the algorithms of the car hands-free system should be fully converged. Subjects are asked to perform a task, such as to describe the position of given numbers in a picture similar to the "Kandinsky" test procedure described for the conversational tests. An artificial head can be used to generate double talk at defined points in time in order to introduce interfering signal components for the speech echo canceller and test the canceller's ability to handle double talk. After the termination of the call, the subjects are asked to give a rating. The scales used are typically ACR or DCR scales. More information can be found e.g. in [31].

10.3.5 Listening-only Tests (LOT) and Third Party Listening Tests

The main purpose of listening-only tests and third party listening tests is the evaluation of impairments under well-defined and reproducible conditions in the listening situation. Their application for the evaluation of hands-free systems is most useful when evaluating the sending direction of the hands-free system. It should be noted that listening tests are very artificial. Listening-only tests are strongly influenced by the selection of the speech material used in the tests; the influence of the test stimuli is much stronger than e.g. in conversation tests. The tests must be designed carefully including the appropriate selection of test sequences (phoneme distribution), talkers (male, female, age, target groups) and others. A sufficient number of presentations must be integrated into a test, ranging from the best to the worst-case condition of the impairment investigated in the test. Reference conditions (simulated, defined impairments with known subjective rating results) may be included in order to check the validity of the test. More detailed descriptions of the requirements and rules how to conduct listening-only tests for various purposes are found in [29–33].

Pre-recorded, processed speech is presented to the subjects. For car hands-free applications these speech sequences are either recorded in the car (when assessing the listening speech quality in the car) or they are recorded at the output of the hands-free terminal in sending direction. In general two possibilities exist for the presentation of prerecorded speech material. Either reference handset terminals are used in case the sending direction of the car hands-free terminal is judged simulating a handset connection at the far end side. Alternatively third party listening tests can be used. In third party listening tests [31, 32] the speech material is recorded by an artificial head which is used to record the complete acoustical situation including the background.

This procedure can be applied to assess the listening speech quality in the car as well as assessing the speech quality in sending direction.

With this procedure, all types of handset, headset, and hands-free configurations can be evaluated in a listening-only test including the environmental conditions at the terminal location. For playback, equalized headphones are used. The equalization must guarantee that during playback the same ear signals are reproduced which were measured during recording. Thus a binaural reproduction (for details see [9]) is possible which leads to a close-to-original presentation of the acoustical situation during recording.

Furthermore the third party listening setup allows to use this type of test for investigating conversational situations by third parties. Therefore, a complete conversation is recorded and presented to the listeners. Although the listeners are not talking themselves but listening to other persons' voices, these tests have proven their usefulness in investigating conversational impairments in a listening test.

In listening tests all scales are used, mostly ACR or DCR scales. Loudness preference scales can be used as well. More information can be found e.g. in [31] and [32]. Instead of ACR or DCR tests, also CCR (comparison category rating) is used which offers a higher sensitivity and may be used for the quality evaluation of high-quality systems or for the optimization of systems. CCR tests are based on paired comparisons of samples.

10.3.6 Experts Tests for Assessing Real Life Situations

Sometimes besides objective testing of hands-free telephones complementary subjective performance evaluation may be useful. Especially for car hands-free systems a lot of experience has been gained with these types of complementary tests. The general considerations when conducting additional subjective tests are given here with the example of car hands-free systems. Supplementary subjective tests are targeted mainly to “in situ” hands-free tests for optimizing hands-free systems in a target car and under conditions which are not covered by objective test specifications. The main purpose is to investigate the hands-free performance in real live conditions including networks and car to car communication. They are of diagnostic nature and not suitable for parameter identification and value selection. Generally, such tests are based on tests as described above and are found in the ITU-T P.800 series Recommendations but not intended to replace tests as described in the ITU-T P.800 series Recommendations.

For conducting the tests the hands-free system under test is installed in the target car, which is referenced as near-end. The far-end is either a landline phone or an observing car also equipped with the hands-free system under test (car-to-car test). It is recommended to not only test the hands-free system in a landline connection but also in a car-to-car connection because the latter case can be regarded as a worst case scenario resulting in worse hands-free quality compared to landline connections.

The evaluation of the hands-free performance should be done in different driving conditions including different background noise scenarios, different driving speeds, different fan/defrost settings, etc.

Since conversational tests are rather time consuming most of the hands-free tests are conducted as single-talk and double-talk tests as described above. Evaluations are done at the far-end and/or the near-end, depending on the type of impairment to be evaluated.

The performance evaluation of the hands-free system typically covers categories like

- echo cancellation (echo intensity, speed of convergence, etc.),
- double talk performance (echo during double talk, speech level variation, etc.),
- speech and background noise quality in sending direction (level, level variation, speech distortion, etc.),
- speech quality in receiving direction (level, level variation, speech distortion),
- stability of the echo canceller for "closed loop" connection during car-to-car hands-free communication.

The evaluation has to be done by experts who are experienced with subjective testing of hands-free systems. During the tests the signals on near-end and far-end may be recorded to be used for third-party listening evaluation later on. More detailed information can be found in [36].

10.4 Test Environment

In general the evaluation of hands-free terminals is made in a lab-type environment which is simulating the acoustical conditions close to the real use conditions. The test environment described focusses on car hands-free system since car hands-free systems are dedicated to be used in cars only and the car is a quite a special environment from the acoustic point of view the test environment has to be selected carefully. Different approaches can be taken starting from a digital simulation of the transmission paths in the car (mouth to microphone, loudspeaker to the drivers ear and loudspeaker to the microphone) up to the use of a car cabin for installing and testing the hands-free device in a car which is the approach taken in [36] and [47]. The relevant transmission paths in a car are shown in Fig. 10.2.

It is common to most test setups to use a car type environment under lab conditions in order to control the influence of the network as well as the background noise conditions as exactly as possible. Certainly the hands-free evaluation can also be done in real driving situations. However, due to the highly uncontrolled environment (time-variant driving noise, pass-by traffic, unpredictable environmental conditions, unknown network conditions), this type of evaluation is not recommended except for validation tests and design

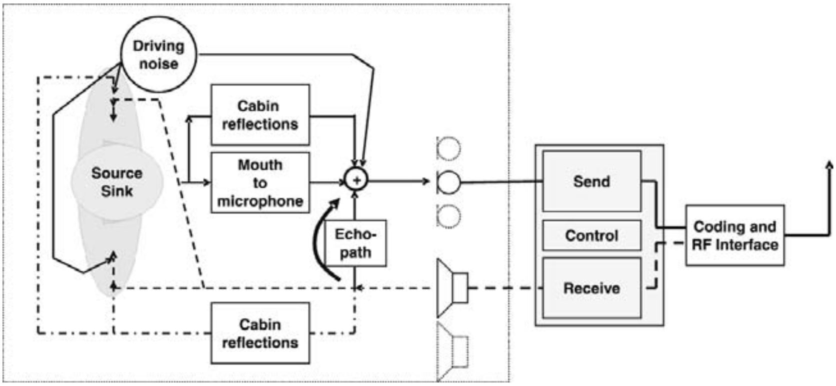


Fig. 10.2. Transmission paths in a car cabin.

optimization taking into account additional influencing factors which could not be simulated in the lab environment.

10.4.1 The Acoustical Environment

The acoustical environment of a car cabin is rather complex: different materials ranging from hard reflecting surfaces such as windows or glass roofs to highly absorbing surfaces such as seats lead to the fact that the simulation of a car type environment is rather difficult. Furthermore different car types have to be considered. Compact cars show completely different properties than luxury cars, trucks, vans or sports cars. Therefore, the coupling between the car hands-free microphone(s) and the hands-free loudspeaker(s) also highly depends on the individual design of the car and the positioning of the microphones and loudspeakers inside the car. The positioning of the hands-free microphone is of special importance. The microphone should be positioned as close as possible to the talker but also in such a way that the coupling between the car hands-free loudspeakers and the car hands-free microphone(s) is minimized. Consequently, it is the easiest solution for most test setups if the actual target car is used when testing complete hands-free systems. This is the best representation of all transmission paths relevant to the hands-free implementation which will finally give the performance of the hands-free system in the target car.

10.4.2 Background Noise Simulation Techniques

Background noise is one of the most influencing factors in hands-free systems but especially in car hands-free systems. Consequently in order to simulate a realistic driving situation, background noise has to be simulated as realistically as possible even in a lab type environment. Background noise simulation techniques are described in [11,36,48,49]. Typically a 4-loudspeaker arrangement with subwoofer is used. This background noise setup is available for simulations under laboratory conditions as well as for car cabins. In Fig. 10.3 the simulation arrangement for car hands-free systems is shown. In order to use this arrangement prior to the tests the background noise produced by a car has to be recorded. This is done under real driving conditions typically using a high quality measurement microphone positioned close to the hands-free microphone. In general all different driving conditions can be recorded. For built-in systems the background noise of the target car is recorded, for after-market systems the background noise of one or more cars considered to be typical target cars is used. If possible the output signal of the hands-free microphone can be used directly. In such a case structure borne noise which might be picked up by the microphone can also be considered in the simulation.

The loudspeaker arrangement used for playback of the recorded background noise signals is equalized and calibrated so that the power density spectrum measured at the microphone position is equal to the recorded one. For equalization either the measurement microphone or the hands-free microphone used for recording is used. The maximum deviation of the A-weighted sound pressure level is required to be less than 1 dB. The third octave power density spectrum between 100 Hz and 10 kHz should not deviate by more than 3 dB from the original spectrum. A detailed description of the equalization procedure as well as a database with background noises can be found e.g. in [11] and [48].

10.4.3 Positioning of the Hands-Free Terminal

The hands-free terminal is installed either as described in the relevant standards (see e.g. [25,27]) or according to the requirements of the manufacturers. In cars the positioning of the microphone/microphone array and loudspeaker are given by the manufacturer. If no position requirements are given, the test lab has to choose the arrangement. Typically, the microphone is placed close to the in-door mirror, the loudspeaker is typically positioned in the footwell of the driver or the co-driver. In any case the exact location has to be noted. Hands-free terminals installed by the car manufacturer are measured in the original arrangement.

Headset hands-free terminals are positioned according to the requirements of the manufacturer. If no position requirements are given, the test lab has to choose the arrangement. Further information is found in [25,36,48].

10.4.4 Positioning of the Artificial Head

The artificial head (HATS Head and Torso Simulator according to ITU-T Recommendation P.58 [21]) is placed as described in the relevant standards

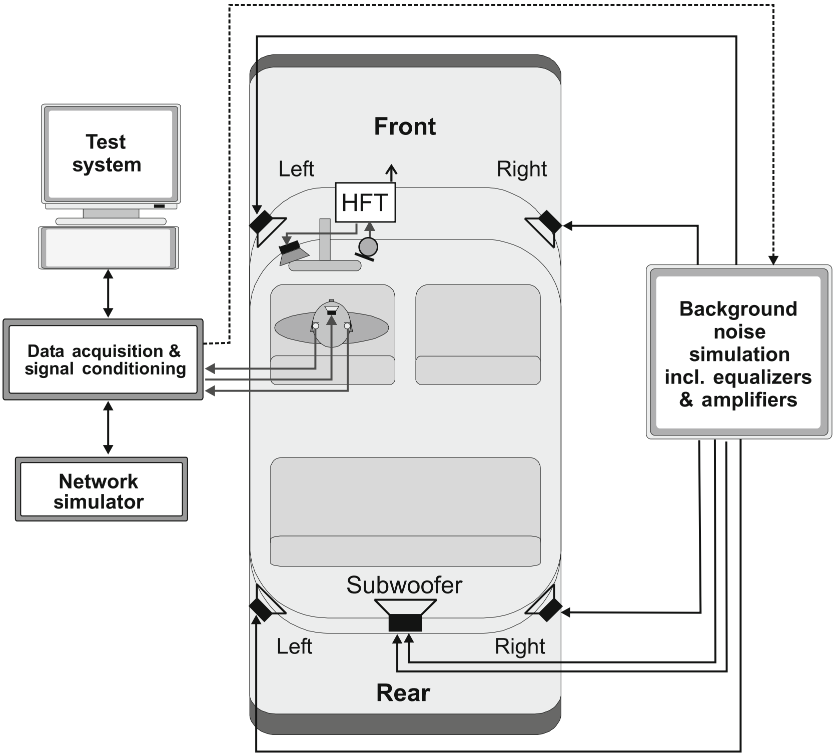

Test system
Data acquisition & signal conditioning
Network simulator
Front
Left
Right
HFT
Background noise simulation incl. equalizers & amplifiers
Subwoofer
Left
Right
Rear

Fig. 10.3. Test arrangement with background noise simulation.

(see e.g. [25, 27]). In cars it is installed at the driver's seat for the measurement. The position has to be in line with the average user's position. Clearly there may be different locations for different users which may be taken into account in addition to the average users position. The position of the HATS (mouth/ears) within the placing arrangement is chosen individually for each type of car. The position used has to be described in detail by using suitable measures (marks in the car, relative position to A-, B-pillar, height from the floor etc.). The exact reproduction of the artificial head position must be possible at any later time. If no requirements for positioning are given, the distance from the microphone to the MRP [20, 21] is defined by the test lab.

The artificial head used should conform to ITU-T Recommendation P.58. Before conducting tests the artificial mouth is equalized at the MRP according to ITU-T Recommendation P.340 [27], the sound pressure level is calibrated at the HATS-HFRP (HATS-hands-free reference point) so that the average level at HATS-HFRP is -28.7 dB_{Pa}. The detailed description for equalization at the MRP and level correction at the HATS-HFRP can be found in ITU-T Recommendation P.581 [25]. For assessing the hands-free terminals in receiving direction the ear signal of the right ear of the artificial head is used (for cars where the steering wheel is on the right hand side, the left ear is used). The artificial head is free-field equalized as described in ITU-T Recommendation P.581.

10.4.5 Influence of the Transmission System

Measurements may be influenced by signal processing [1–4] (different speech codecs, DTX[2], comfort noise insertion. etc.) depending on the transmission system and the system simulator used in the test setup. In general, a network simulator (system simulator) is therefore used in the test in order to provide a mostly controlled network environment. All settings of the system simulator have to ensure that the audio signal is not disturbed by any processing and the transmission of the signal (in cases of mobile networks especially the radio signal) is error-free. DTX, VAD and other network signal processing is switched-off. In case of different speech coders available in a network the one providing the best audio performance is typically used. E.g. for measurements with AMR-codec [3] the highest bitrate of 12.2 kb/s is used. Nevertheless, there may be tests where lower bitrates providing less speech quality are used e.g. in order to evaluate the listening speech quality of the complete hands-free system in more detail. Except conditions which are targeted to investigate the influence of transmission errors such as packet loss or jitter no network impairments should influence the tests.

10.5 Test Signals and Analysis Methods

The choice of the test signal as well as the analysis method depends on the application. On the one hand, speech sequences are best suited as test signal for hands-free devices incorporating algorithms, which are optimized based on specific speech characteristics. But the dynamics of speech, the multitude of different languages with their specific characteristics, the directly related question of robustness and reproducibility of analyses make it difficult to come to a common agreement in standardization. On the other hand, artificial test signals providing speech-like properties have the advantage of not being limited to a specific language. These signals can be optimized to measure specific parameters and provide a high reproducibility of results, e.g. in different labs. However, it is also obvious, that typically a large number of test signals is needed – each designed for specific purposes. Furthermore, it is generally recommended to verify test results by speech recordings and listening examples. Test methods applicable for car hands-free evaluation can be separated in two main categories – the "traditional" analysis methods and the advanced test methods. The "traditional" analysis methods focus on the basic telephonometry parameters and include:

[2] The term *DTX* stands for *discontinuous transmission.*

- **Loudness Rating** calculations [26] which are the basis for setting the correct sensitivities in the hands-free terminals in order to ensure seamless interaction with the networks and the far end terminals.
  The Loudness Rating then is defined as:

  $$LoudnessRating = -\frac{10}{m}\log_{10}\left\{\sum_{i=1}^{N} 10^{-\frac{m}{10}(L_{\mathrm{UME},i}-\overline{L_{\mathrm{RME}}}+W_i)}\right\}. \quad (10.1)$$

  W_i are weighting factors as defined in [26], different for SLR, RLR, STMR, LSTR. L_{RME} represents the mouth-to-ear transmission loss of the reference speech path (IRS speech path [18]). m is a constant in the order of 0.2, different for the different loudness ratings. For a given telephone or transmission system, the values of L_{UME} can be derived from the measurement different sensitivities S_{MJ} (mouth-to-junction) for calculation of the SLR (sending loudness rating), from S_{JE} (junction-to-ear) for the calculation of the RLR (receiving loudness rating) or from S_{ME} (mouth-to-ear) for the overall loudness rating OLR.
  In a similar manner, the sidetone paths can be described: STMR is the Sidetone Masking Rating describing the perceived loudness of the user's own voice and LSTR (Listener Sidetone Rating) describes the perceived loudness of room noise coupled to the user's ear.
- Requirements for **frequency response characteristics** [27, 48] in sending and receiving in order to ensure a sufficient sound quality and intelligibility. In Sending a rising frequency response characteristics with a high pass characteristics at around 300 Hz is recommended to ensure sufficient intelligibility and the reduction of low frequency background noise. In receiving a most flat frequency response characteristics is advisable.
- **Echo loss requirements** [16] ensuring an echo free connection under different network conditions. Envisaging that delay is inserted
  - by the mobile network itself,
  - by the hands-free terminal where advanced signal processing leads to higher delay compared to standard handset terminals,
  - by connecting networks which increasingly insert VoIP transmission which adds additional delay in the transmission link,

  the echo loss requirements for car hands-free terminals are high. The terminal coupling loss (TCL) required is at least 40 dB, typically however higher (46 dB to 50 dB, see [36, 48]) in order to prevent the far end partner form the car hands-free echo. The basic information about transmission delay and the echo loss required can be found in [16].
- **Delay requirements** [48] referring to the processing delay introduced by the car hands-free terminal to ensure a minimum delay introduced by these terminals for the benefit of the overall conversational quality.

More details on these tests can be found in e.g. [36, 48].

10.5.1 Speech and Perceptual Speech Quality Measures

Hearing model based analysis methods like PESQ$^{\text{TM}}$ [34,35] and TOSQA2001 [7, 8] calculate estimated mean opinion scores. These objective scores represent the listening speech quality in a one-way transmission scenario with a high correlation to the results of a subjective listening test. The test signal used by such methods is speech. Due to the different characteristics of the different languages it is difficult to define an "average" speech signal to be used in conjunction with these methods. Therefore ITU-T has defined in Recommendation P.501 [22] a set of reference speech samples for different languages which can be used. The ITU-T recommended PESQ$^{\text{TM}}$ has not been validated for acoustic terminal and handset testing, e.g. using HATS [34] and does therefore not play a practical role in testing hands-free implementations. TOSQA2001 is validated for terminal testing at acoustical interfaces [8] and is therefore also used for testing hand-free devices. The method estimates the listening speech quality (TMOS, TOSQA2001 mean opinion score) by using reference and degraded speech samples. Frequency content of the transmitted speech, loudness and noise, additive disturbances or non-linear coder distortions contribute to speech quality degradations and influence the TMOS score accordingly. These results are very useful in terminal testing, but provide only very limited information about the reason for unexpected quality degradations.

10.5.2 Speech-like Test Signals

Different test signals with different levels of complexity are available and have been evaluated for different types of applications. A comprehensive description of the most important signal can be found in ITU-T Recommendation P.501 [22]. ITU-T Recommendation P.502 [23] describes appropriate analysis methods for each signal. The most complex speech-like signal in telephonometry is the artificial voice as described in ITU-T Recommendation P.50 [19]. This signal provides a statistical representation of real speech. The signal duration amounts to 10 s, it is suited and often used to measure long-term or average parameters like frequency responses or loudness ratings [26].

An important signal for laboratory quality testing of hands-free telephones is the composite source signal (CSS) ([14, 22], see Fig. 10.4). It is composed in the time domain and consists of different parts like voiced and unvoiced segments and a pause. Due to its short duration of the active signal part (approximately 250 ms) it is well suited to measure short term parameters, e.g. switching behaviour of AEC between single and double talk sequences. Parameters in the frequency domain such as frequency response, loudness ratings etc., as well as parameters in the time domain such as switch-on times can be determined.

Fig. 10.5 shows the combination of two uncorrelated CSS (composite source signals) to simulated a double talk sequence. The power density spectra are given in Fig. 10.6.

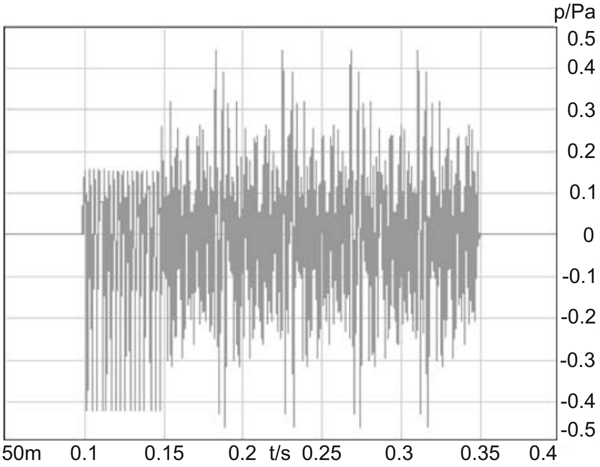


Fig. 10.4. Composite source signal. The signal consists of different parts like voiced and unvoiced segments and a pause.

The simulated double talk starts with a CSS burst applied in receiving direction (dark gray signal) followed by a near end double talk burst (light gray bursts). This sequence is then periodically repeated. The typical test signal levels are -4.7 $\mathrm{dB_{Pa}}$ for the near end signal at the mouth reference point (MRP) and -16 $\mathrm{dB_{m0}}$ in downlink direction. Measurements in sending direction of an HFT (hands-free terminal) typically analyze the transmitted

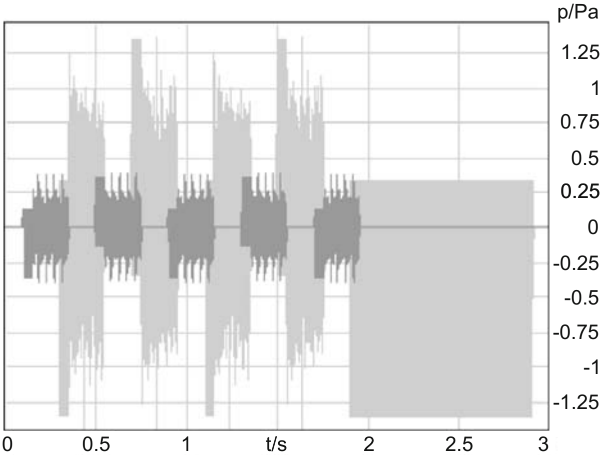


Fig. 10.5. Simulated double talk sequence. The simulated double talk starts with a CSS burst applied in receiving direction (dark gray signal) followed by a near end double talk burst (light gray bursts).

uplink signal referred to the near end test signal. The resulting sensitivity curve can be used to detect level modulations typically introduced by AEC post processing. Figs. 10.7 and 10.8 show two examples. The HFT implementation represented by the analysis curve in Fig. 10.7 introduces an attenuation of approximately 15 dB in the microphone path during the double talk sequence. The driver's voice is partly attenuated during a double talk sequence using real speech over this implementation. Vice versa the sending direction can be regarded as nearly transparent in the analysis curve in Fig. 10.8.

In the same way, the analysis can be carried out in receiving direction in order to verify if the implementations do not insert attenuation in this transmission path during double talk.

A third parameter determining the double talk capability of a hands-free implementation is the echo attenuation during double talk. Subjective test results are available comparing the echo attenuation during single and double periods [27, 40]. The challenge for measurement technique is to separate the near signal from the echo components in the send signal. A suitable test signal that provides this characteristic consists of two uncorrelated AM/FM modulated signals [22]. The time signal is shown in Fig. 10.9. The two signals show comb-filter spectra as given in Fig. 10.10, which are necessary to distinguish between the double talk signal (coming from the near end) and the echo signal (coming from the echo path as a reaction on the receive signal). The power density spectra of both signals calculated by Fourier transformation are given in Fig. 10.10. Echo components during double talk can be detected in the send signal by comparison of the uplink signal and the original downlink signal. The near signal components can easily be removed by appropriate filtering.

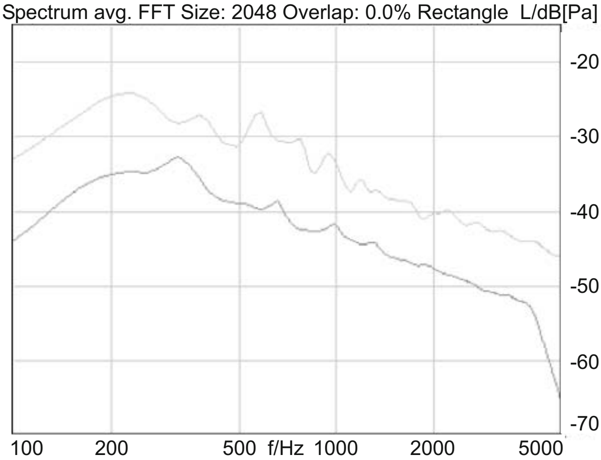


Fig. 10.6. Power density spectra of the signals shown in Fig. 10.5 (dark gray: far end, light gray: near end).

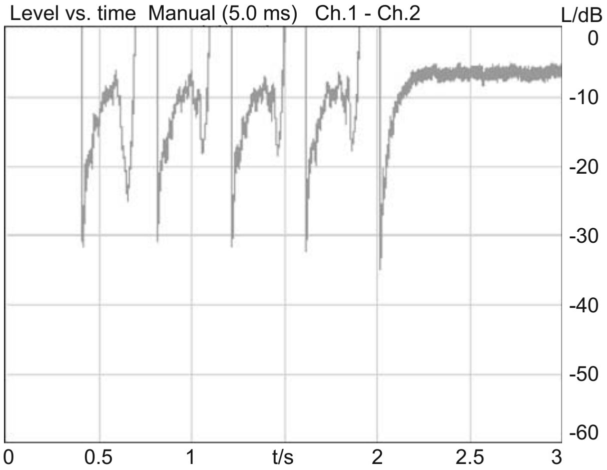


Fig. 10.7. Uplink sensitivity during double talk – Hands-free terminal 1. This terminal introduces an attenuation of approximately 15 dB in the microphone path during the double talk sequence.

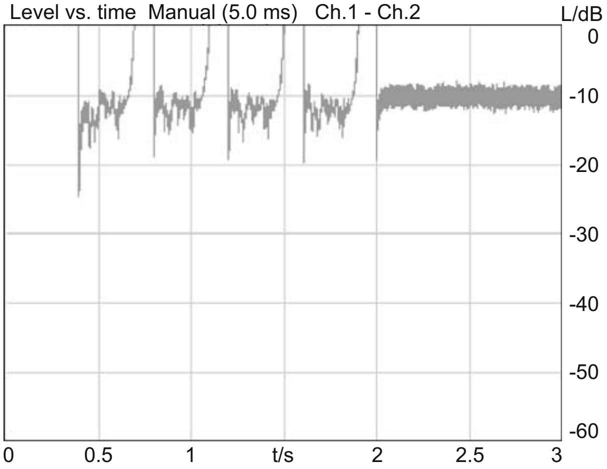


Fig. 10.8. Uplink sensitivity during double talk – Hands-free terminal 2. The sending direction can be regarded as nearly transparent in the analysis curve.

ITU-T Recommendation P.340 [27,40] defines different types of double talk performance for hands-free implementations. The characterization is based on the measured attenuation between the single and double talk situation inserted in sending and receiving direction (a_{HSDT}, a_{HRDT}). The third parameter is the echo attenuation during double talk (EL_{DT}).

The three parameters are measured independently. However, the worst result determines the characterization. A "type 1" implementation provides

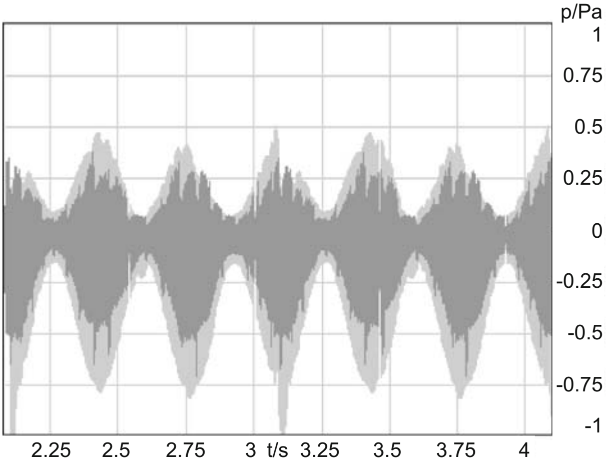


Fig. 10.9. AM/FM modulated test signals (dark gray: far end, light gray: near end) for determining the double talk capability of a hands-free implementation.

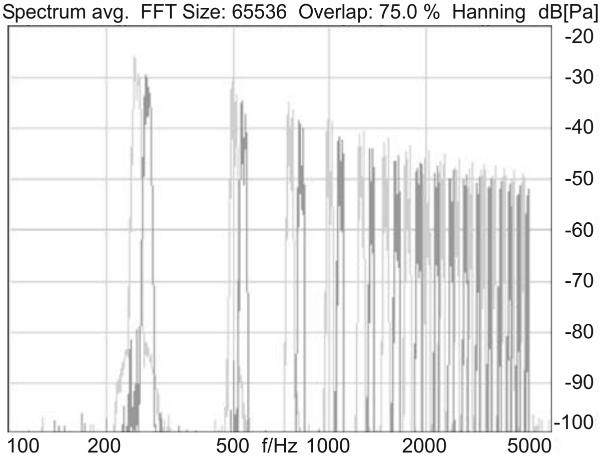


Fig. 10.10. Power density spectra of the signals presented in Fig. 10.9 (dark gray: far end, light gray: near end). Echo components during double talk can be detected in the send signal by comparison of the uplink signal (not depicted) and the original downlink signal (dark gray).

full duplex capability; “type 2a”, “2b” and “2c” devices are partial duplex capable, a “type 3” characterization indicates no duplex capability.

10.5.3 Background Noise

Besides speech and artificial test signals the transmission quality of the background noise present e.g. in the driving vehicle needs to be evaluated in detail.

Table 10.1. Duplex capability.

Characterization	Type 1	Type 2a	Type 2b	Type 2c	Type 3
a_{HSDT}	≤ 3 dB	≤ 6 dB	≤ 9 dB	≤ 12 dB	> 12 dB
a_{HRDT}	≤ 3 dB	≤ 5 dB	≤ 8 dB	≤ 10 dB	> 10 dB
EL_{DT}	≥ 27 dB	≥ 23 dB	≥ 17 dB	≥ 11 dB	< 11 dB

The driving noise can not only be regarded as a disturbing signal for an HFT implementation, but carries important information for the far end subscriber. It is typically processed through noise reduction algorithms, might be modulated by echo suppression or partly substituted by comfort noise, if a downlink signal is applied. Related quality parameters range from D-value calculation, comparing the sensitivity of the microphone path on speech and on background noise, the signal to noise ratio, if near end signals are transmitted together with background noise or the modulation of transmitted background noise by echo cancellation or echo suppression.

A very promising method to analyze the performance of noise reduction algorithms is the *Relative Approach* [13, 47]. This method takes into account the sensitivity of the human ear on unexpected events both in the time and in the spectral domain. In contrary to all other methods the Relative Approach does not use any reference signal. The signal is band filtered (1/12 octave) and a forward estimation based on the signal history is calculated in order to predict the new back-ground noise signal value. Values between the frequency bands are interpolated.

The predicted signal pattern is compared to the actual signal characteristic and the deviation in time and frequency is displayed as an "estimation error". Thus instantaneous variations in time and dominant spectral structures are found based on the human ear sensitivity on these parameters. Typical disturbances produced by noise reduction algorithms like musical tones can be detected and verified, if these components lead to speech quality degradations. A typical example is shown in Fig. 10.11. It analyzes the adaptation phase of a noise reduction algorithm. Disturbing artefacts as detected by the Relative Approach are indicated by the arrows.

A more advanced test procedure is described in ETSI EG 202 396-3 [12]. The model described here is a perceptual model again based on the Relative Approach. The model is applicable for speech in background noise at the near end of a terminal and provides an estimation of the results that normally would be derived from a subjective test made using ITU-T recommendation P.835 [33]. Three MOS scores are predicted:

- S-MOS (speech MOS), describing the quality of the speech signal as perceived by the listener,

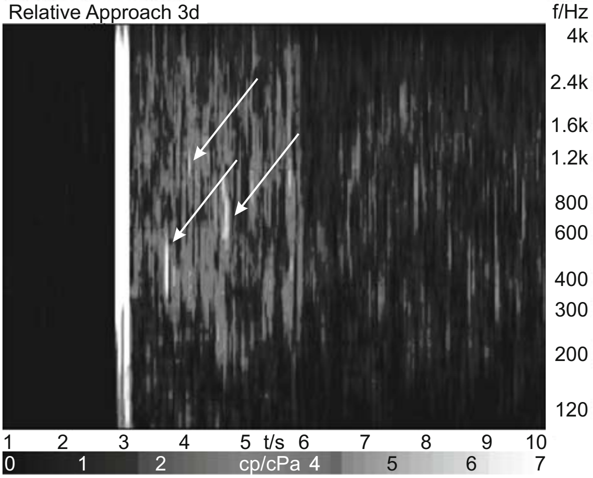


Fig. 10.11. Relative Approach analysis of an adaptation phase of noise reduction.

- N-MOS (noise MOS), describing the quality of the transmitted background noise, and
- G-MOS (global MOS), describing the perceived overall quality of the transmitted speech plus background noise signal.

Currently the test method is applicable for:

- Wideband handset and wideband hands-free devices (in sending direction),
- noisy environments (stationary or non-stationary noise),
- different noise reduction algorithms,
- AMR [3] and G.722 [17] wideband coders,
- VoIP networks introducing packet loss.

However the extension of this method to narrowband terminals and systems is already on ETSI's roadmap. Different input signals are required for the model and subsequently are used for the calculation of N-MOS, S-MOS and G-MOS. Beside the signals processed by the terminal or the near end device two additional signals are used as a priori knowledge for the calculation:

1. The "clean speech" signal, which is played back via the artificial mouth.
2. The "unprocessed signal", which is recorded close to the microphone position of the handset or the hands-free terminal.

Both signals are used in order to determine the degradation of speech and background noise due to the signal processing as the listeners did during the listening tests. The principle of the method is shown in Fig. 10.12. Further information and details about the algorithm can be found in [12].

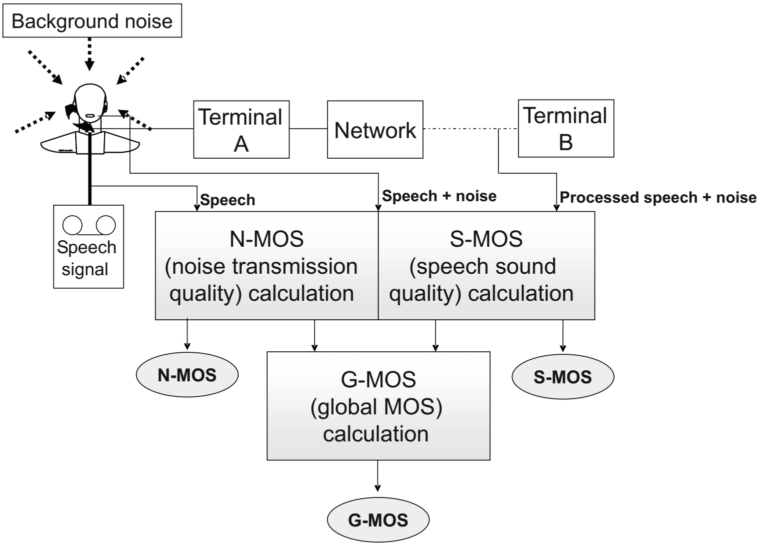

Background noise
Terminal A
Network
Terminal B
Speech
Speech + noise
Processed speech + noise
Speech signal
N-MOS (noise transmission quality) calculation
S-MOS (speech sound quality) calculation
N-MOS
G-MOS (global MOS) calculation
S-MOS
G-MOS

Fig. 10.12. Principle of the S-MOS, N-MOS and G-MOS prediction as described in [12].

10.5.4 Applications

A typical application example for these test signals and the interaction between the results is shown by a comparison analysis of two different HFT aftermarket implementations. Both devices are measured via Bluetooth connection to a commercially available 2G mobile phone. The GSM full rate speech coder is used.

The frequency response in Fig. 10.13 is relatively balanced without showing a strong high pass characteristic. The sending loudness rating [26] of 13.3 dB for this implementation absolutely meets the recommended range of 13 ± 4 dB according to [48]. The TMOS of 3.3 confirms the high uplink quality (recommended $\geq$ 3.0 TMOS [48]).

The signal-to-noise ratio estimated from the measurement result based on composite source signal bursts transmitted together with background noise (simulated 130 km/h background noise) is very high (approximately 27 dB, Fig. 10.14). The near end test signal is also affected by the uplink signal processing and attenuated. The D-value comparing the sensitivities of the uplink transmission path on speech and on background noise of +3.5 dB is also extremely high (recommended ≥ -10 dB). Both results are consistent. However, these parameters are extremely high especially when considering the balanced frequency response without a strong high pass characteristic. This indicates a very aggressive noise reduction algorithm. The undesired side effect is an unpleasant metallic speech sound and disturbing, artificial musical tones in the transmitted background noise.

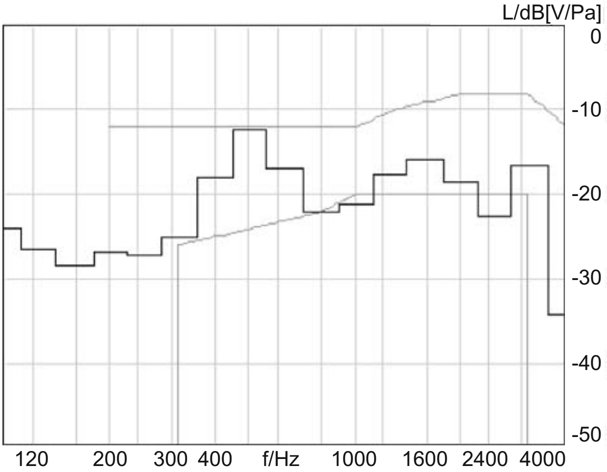


Fig. 10.13. Sending frequency response, hands-free terminal 1. The frequency response is relatively balanced without showing a strong high pass characteristics (compare with Fig. 10.15).

In comparison the analysis in Fig. 10.15 shows a frequency response providing a clear, distinct high order high pass around 300 Hz. The curve only slightly violates the tolerance scheme, which can practically be neglected. All other parameters like the SLR of 12.7 dB meet the requirements. The TMOS of 3.0 still indicates a sufficient listening speech quality – although the frequency response provides the strong high pass characteristic.

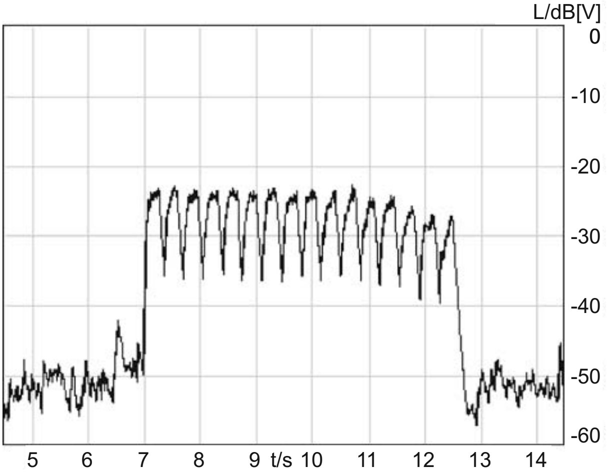


Fig. 10.14. Transmission of background noise and near end signal (level vs. time), hands-free terminal 1.

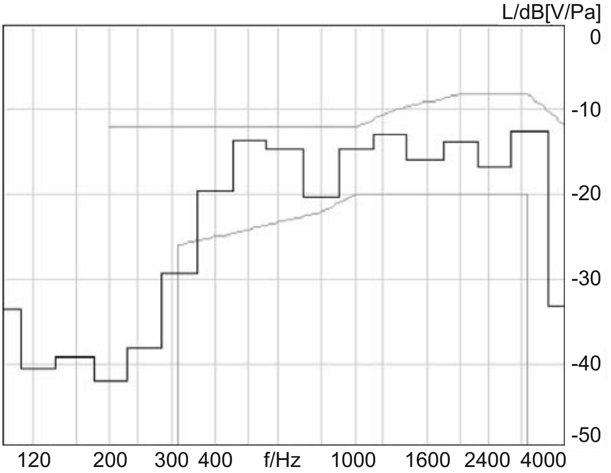


Fig. 10.15. Sending frequency response, hands-free terminal 2. This terminal has a frequency response providing a clear, distinct high order high pass around 300 Hz (compare with Fig. 10.13).

The signal-to-noise ratio as estimated from the analysis curve in Fig. 10.16 is 16 dB. The near end composite source signal bursts are only slightly distorted in this level analysis. This indicates that the noise cancellation algorithm does not significantly affect the near end test signal when transmitted together with background noise. The signal-to-noise ratio at the algorithm input seems to be high enough to clearly distinguish between both signals. The 16 dB signal-to-noise ratio estimated from this analysis and a reasonable D-value of −6.8 dB are consistent. The good quality for the uplink transmission can be confirmed by the listening example of the transmitted speech together with background noise.

The strong high pass as indicated above significantly contributes to a high signal-to-noise ratio in sending direction. Audible disturbances like musical tones are minimized, thus indicating that a high order microphone high pass significantly improves the performance in the presence of background noise. An acoustical tuning already at the microphone is a good compromise though even it might slightly degrade the listening speech quality under silent conditions.

10.6 Result Representation

The complexity of in-depth quality testing of hands-free implementations and the multitude of results acquired during laboratory tests require an appropriate result representation. An overall quality score that covers all conversational aspects is not yet available. Moreover, such a one-dimensional score

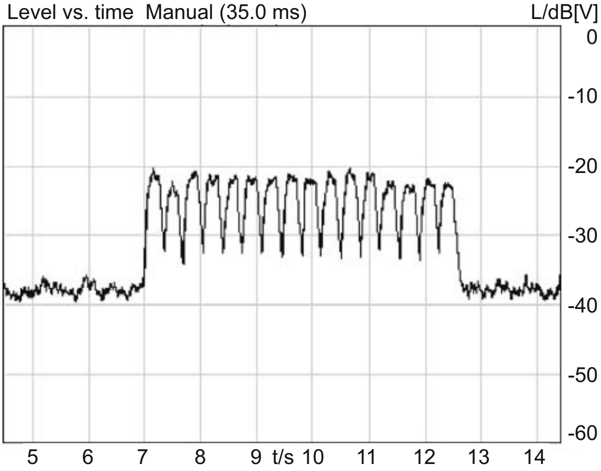


Fig. 10.16. Transmission of background noise and near end signal (level vs. time), hands-free terminal 2.

might even be misleading and therefore fail in practice, because completely different implementations might be represented by the same score. Such a score does not represent the acoustical "fingerprint" of an individual implementation.

The ITU-T Recommendation P.505 [24] provides a new representation methodology – best described as a "quality pie" – that bridges this gap. The circle segments and displayed parameters can be selected and adapted to the application, i.e. the device under test. An example of a hands-free implementation with parameter selection according to the VDA specification [48] is shown in Fig. 10.17.

The focus of this representation is to provide

- a "quick and easy to read" overview about the implementation for experts and non-experts including strengths and weaknesses,
- a comparison to limits, recommended values or average results from benchmarking tests, and
- detailed information for development to improve the performance.

10.6.1 Interpretation of HFT "Quality Pies"

The hands-free "quality pie" shown in Fig.10.17 does not represent an existing implementation. It is only used here for explanation purposes. In general the 12 segments – which can be regarded as a maximum suitable number being visualized in one diagram – can be subdivided into three groups covering different conversational aspects. The first 5 segments – clockwise arranged – represent one-way transmission parameters. The sending direction is

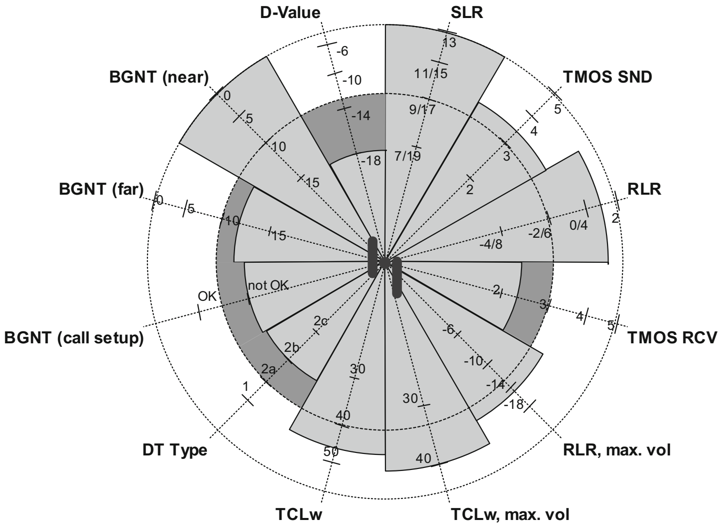


Fig. 10.17. Hands-free "quality pie" according to [24]. The following abbreviations were used: *SLR* means sending loudness rating, *TMOS SND* stands for TMOS value in sending direction, *RLR* abbreviates receiving loudness rating, *TMOS RCV* is short for the TMOS value in receiving direction, *RLR, max. vol* is the receiving loudness rating at maximum volume, *TCLw, max. vol* is the terminal coupling loss measured at maximum volume, *TCLw* abbreviates the terminal coupling loss measured during standard terminal operation, *DT type* means double-talk type, and the different *BGNT* slices show the background noise transmission in different situations.

covered by the sending loudness rating and the TMOS. The following two slices represent the receiving loudness rating (RLR) and the TMOS in receiving direction. The fifth segment represents the RLR value at maximum volume.

The following 3 segments indicate the echo attenuation expressed through the parameter weighted terminal coupling loss according to ITU-T Recommendation G.122 [16] measured at maximum volume ("$\mathrm{TCL_W}$(max.vol.)"), at nominal volume ("$\mathrm{TCL_W}$") and the double talk performance ("DT type"). The last 4 segments represent parameters concerning the quality of background noise transmission.

The following general assumptions are made for the quality pie representation: Each parameter is represented by a pie slice. The size of each slice directly correlates to quality. The gray color indicates a quality higher than the requirement for this specific parameter. Interaction aspects between single parameters are not considered. An inner circle (dark gray) indicates the minimum requirement for each parameter. For those parameters that should be within a range, like the sending loudness rating (SLR) of 13 ± 4 dB [48] the axis is double scaled. It raises from the origin of the diagram radial to the outside up to the recommended value (13 dB for the SLR in this example) and in addition radial to the inside. Other axes like the background noise transmission quality after call setup ("BGNT call setup") are scaled only between two states (ok, not ok).

10.6.2 Examples

The significance of this representation, e.g. in tracing different development phases can best be shown on a practical example. Fig. 10.18(a) represents an early quality status of a hands-free implementation during development.

The left pie chart points out the following:

- The SLR of 13 dB indicates a sufficient loudness, the TMOS score (parameter "TMOS SND") significantly exceeds the limit in sending direction.
- The D-value of −18 dB is too low, the inner dark gray circle represents the limit of −10 dB and gets visible. The sensitivity on background noise needs to be reduced or the sensitivity on speech increased.
- The echo attenuation is too low at maximum volume, the TCL_W requirement is violated under this condition (parameter "TCL_W(max. vol.)").
- Significant impairments could also be observed in background noise transmission during the application of far end signals (parameter "BGNT(far end)"). The background noise is completely attenuated by echo suppression, comfort noise is not inserted. The resulting modulation in the transmitted background is very high, gaps occur. The maximum acceptable level modulation of 10 dB for this parameter is exceeded.

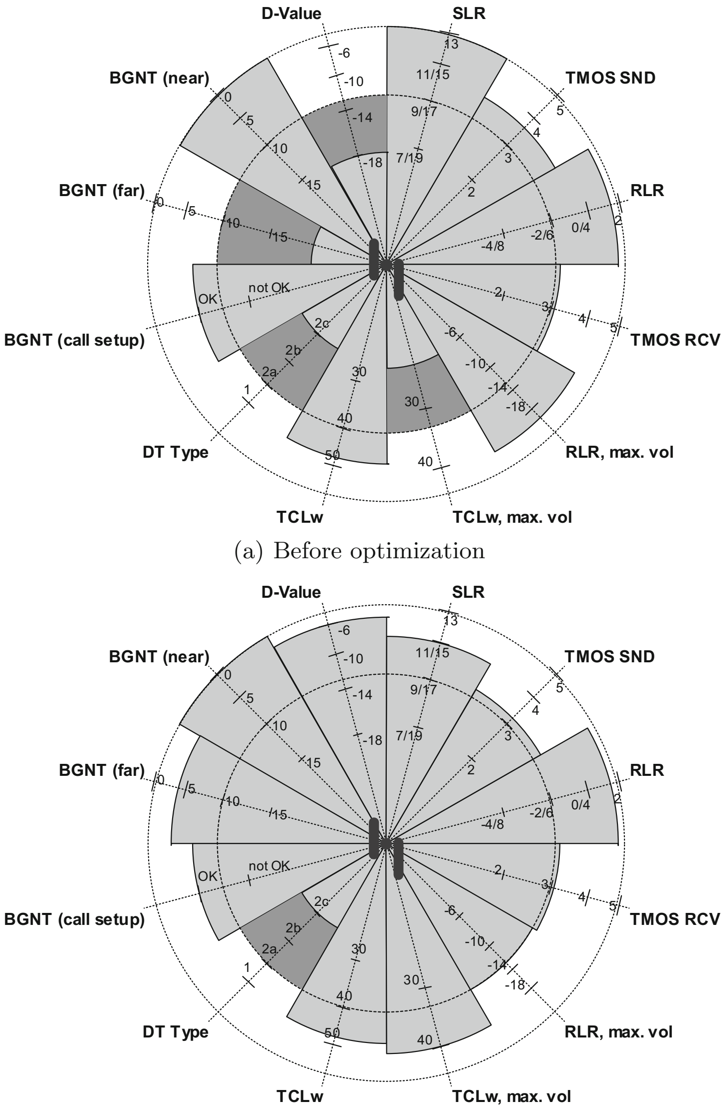
The quality pie in Fig. 10.18(b) indicates a significantly improved performance compared to the previous status represented in part (a). However, the next step that should be addressed by tuning echo cancellation, echo suppression, double talk detection and the associated control parameters is the double talk performance. "Type 1" implementations, i.e. full duplex capable hands-free implementations are available today. It should be noted that it is not always recommended to tune the algorithms to full duplex capability, especially not for the price of lower robustness. Partial duplex capable HFTs ("type 2a" or even "2b") may sometimes be a preferable solution.

In summary, it can be stated that the quality pie representation simplifies the performance discussion. This representation can serve as a basis for commercial decisions, but still provides enough detailed information to discuss possible next optimization steps for speech quality. Important features like interaction aspects between single parameters are explicitly not considered yet and require further investigations.

10.7 Related Aspects

10.7.1 The Lombard Effect

The Lombard effect – also designated as Lombard reflex emphasizing more its intuitive character – describes the result of speech transformation under the



(a) Before optimization

(b) After optimization

Fig. 10.18. Quality pie before (a) and after (b) first optimization step. For the meaning of the different abbreviations see the caption of Fig. 10.17.

influence of a reduced acoustical feedback, e.g. for hearing impaired people or under the influence of noise and stress. However, the Lombard effect is not only a physiological effect. In practice the main intention for the modification of speech production in a conversation is to be more intelligible to others. It can therefore be assumed that the “naturalness” of Lombard speech cannot be completely reproduced in laboratory testing by recording speech samples that are read from a list. Furthermore, different studies showed that multi-talker babble noise led to different Lombard speech characteristics, e.g. larger vowel duration as compared to stationary noise. In the same way, there is a dependence of the Lombard effect on the noise frequency distribution [38]. Lombard speech recorded under the influence of non-stationary noise provides a higher dynamic compared to Lombard speech produced under stationary noise conditions.

Databases available today do not always consider the mentioned aspects. They are typically recorded with test persons reading predefined sentences. These data are of course valid to be used for certain applications, however, the restrictions need to be known and considered.

It is obvious and reasonable to consider Lombard speech characteristics not only for testing speech recognition systems (e.g. [43]) but also for hands-free terminal testing instead of using neutral voice. An appropriate method is to play back these recordings via artificial head systems in a driving car or in a driving simulator [43]. Furthermore, it is important to analyze Lombard speech in order to verify, if important characteristics need to be considered in objective speech quality tests and analyses.

There are different simulation techniques in use providing a recording scenario for Lombard speech. Test persons typically wear equalized closed headphones during noise playback while their Lombard speech is recorded [10, 41]. These headphones lower the perception of the own voice, thus introducing already the Lombard effect. This can be minimized by introducing a feedback path between the microphone and the headphones itself, thus playing back simultaneously the recorded speech via the headphones [45]. Comparison tests with and without this feedback path indicated that the Lombard effect introduced by the headset itself can be neglected compared to the Lombard effect introduced by the background noise scenario e.g. simulating a driving car [45].

Recording scenarios for Lombard speech under the influence of driving noise are described e.g. in [10, 41]. The setup used during own tests is shown in Fig. 10.19. The recordings were carried out in a driving simulator consisting of a real car cabin equipped with an acoustical background noise simulation system.

It is important to reproduce not only the driving situation acoustically during this kind of speech recordings but also the concentration for a typical driving situation and the impression of having a real conversation over a hands-free system. The driving simulator is therefore operated interactively. The speed is indicated on a speedometer and the test persons are instructed to keep a constant speed. Furthermore, a typical hands-free microphone is installed visible near the interior mirror. The test persons were instructed that they should imagine the conversational situation of having a telephone conversation over a hands-free system.

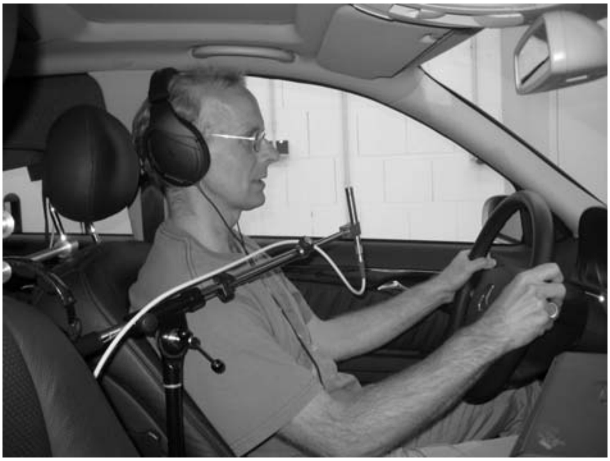


Fig. 10.19. Setup for Lombard speech recordings.

This suitability of this scenario was verified by recording Lombard speech of eight test persons, four male and four female speakers. Different speech material like free utterances, given test sentences to be read from a list and command words were recorded and analyzed. The speech level analyses first demonstrated that the influence of the headphones can practically be neglected. The average speech levels for the different speech materials increased by less than 1 dB if the test persons wore headphones.

Fig. 10.20 shows the average speech levels for free utterances (dark gray bars) and command words (light gray bars). The speech level under quiet conditions was determined to approximately -1 $\mathrm{dB_{Pa}}$ at the mouth reference point (MRP) of the test persons for the free utterances. A standardized test signal level for objective terminal testing is -4.7 $\mathrm{dB_{Pa}}$ at the MRP. However, it is reported that people tend to increase their speech level by approximately 3 dB when using hands-free devices [27]. The resulting level of approximately -1.7 $\mathrm{dB_{Pa}}$ at the mouth reference point is rather accurately confirmed by the measured level of -1 $\mathrm{dB_{Pa}}$ for the free utterances. These speech recordings confirm the tendency given in [27], the analyses of speech material recorded from eight test persons are not representative in a statistical sense.

The Lombard recordings were carried out for three different speeds and levels of 50 km/h (49 $\mathrm{dB_{SPL(A)}}$), 130 km/h (69 $\mathrm{dB_{SPL(A)}}$) and 200 km/h (79 $\mathrm{dB_{SPL(A)}}$). Fig. 10.20 shows the average speech levels for the command words and the free utterances. An offset of approximately 2.5 dB can be measured for the two speech materials. The command words are more pronounced and therefore provide a higher level compared to the free speech.

The regression further points out that the speech level increases by approximately 0.4 $\mathrm{dB/dB_{(A)}}$ for driving situations with a background noise level between approximately 55 $\mathrm{dB_{(A)}}$ and 70 $\mathrm{dB_{(A)}}$. Similar results are

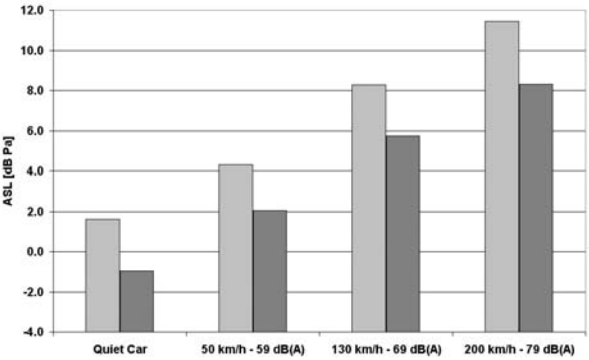


Fig. 10.20. Active speech levels (ASL) at different simulated conditions (light gray: command words, dark gray: free utterances).

reported in [10]. For higher speed the speech level increases by approximately 0.3 $dB/dB_{(A)}$ for the command words and 0.25 $dB/dB_{(A)}$ for the free utterances.

Important conclusions can be drawn from such investigations for objective laboratory tests because they again raise the question of adapting test signal levels during hands-free telephone tests. These results support the idea of increasing the test signal levels for all objective tests by approximately 3 dB at the artificial mouth of an artificial head measurement system simulating the driver's voice. Furthermore the Lombard effect depending on the different background noise scenarios simulated during laboratory tests should be considered and can be estimated from data as analyzed above.

10.7.2 Intelligibility Outside Vehicles

The intelligibility of telephone conversations outside the vehicle is a very important aspect but users are not always aware of this situation. The reason for this undesired effect is elementary: the downlink signal of a hands-free telephone conversation in a vehicle, typically played back via the built-in loudspeakers in the front door, exciting the door structure. The whole surface emits the audible sound outside the vehicle.

This implies, besides the privacy aspect, also a political aspect: a huge effort is taken by legislation in order to lower the external vehicle sound produced e.g. by motors, exhaust systems and tires [5]. The aspect of sound played back via the internal audio systems has – so far – not been addressed. The acoustical coupling between the loudspeakers and the chassis needs to be evaluated in detail in order to identify the transmission paths and individual contributions.

Combined electro-acoustic measures, intelligibility and perceptual analyses on the one hand and vibration analyses on the other hand are necessary in

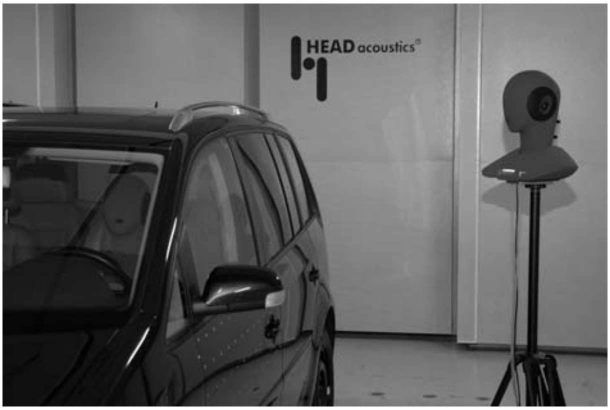


Fig. 10.21. Intelligibility outside the vehicle.

order to document the status, evaluate the transmission paths and verify the effectiveness of modifications [42]. The acoustically relevant parameters can realistically be measured by using two artificial head measurement systems on the driver's seat and outside in a predefined distance and position, e.g. 1 and 2 m from the B-pillar.

The speech intelligibility index SII [6], can – in principle – be used and calculated for different noise scenarios. But the intelligibility of speech highly depends on the test corpus. The SII calculation is based on a weighted spectral distance between average speech and noise spectra. However, the sentence intelligibility is significantly higher than the SII due to its context information [46].

A more analytical analysis is given by the calculation of the attenuation provided by the car chassis. Fig. 10.22 shows the spectral attenuation between the inside HATS at the driver's position and outside in a distance of 1 m from the B-pillar. The curve indicates a strong low frequency coupling between the loudspeaker and the chassis. The attenuation of the high frequencies above approximately 1 kHz is around 20 dB to 25 dB higher.

Besides the speech-based analyses, a vibration analysis (laser scan of the driver's door, see Fig. 10.23) links the intelligibility to the technical source of the emitted signal. The oscillation amplitude is color coded. The complete door is excited by the acoustic signal. Further tests with different loudspeaker modifications (mechanically decoupling loudspeakers from door structure, use of damping material) in one test car showed that the main factor for the outside intelligibility is caused by airborne coupling between loudspeaker and door. Structure borne coupling played a minor role.

The efficiency of modifications is vehicle dependent. Acoustical coupling typically can be significantly reduced only by new loudspeaker positions and mountings or a complete encapsulation. Both would require an enormous effort in modifying vehicle design. The need and motivation for modifications is

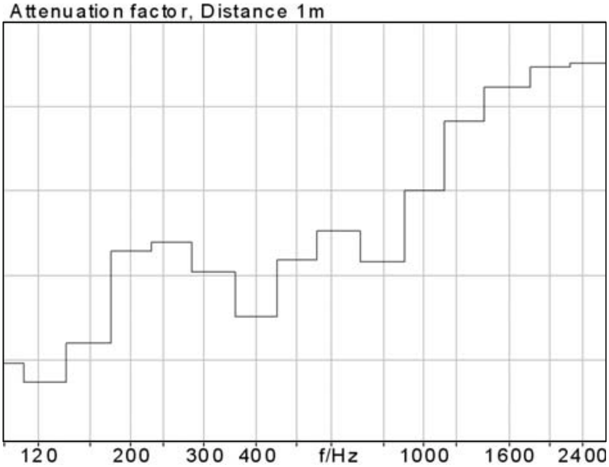


Fig. 10.22. Attenuation "Inside to outside" in 1 m distance, B-pillar.

probably driven by customer's expectations and complaints. A suggestion for reasonable limits for the outside intelligibility can be derived from practical approaches: the intelligibility of the driver's voice outside the vehicle or the intelligibility when using external loudspeakers for playback, e.g. positioned in the drivers and co-drivers footwell.

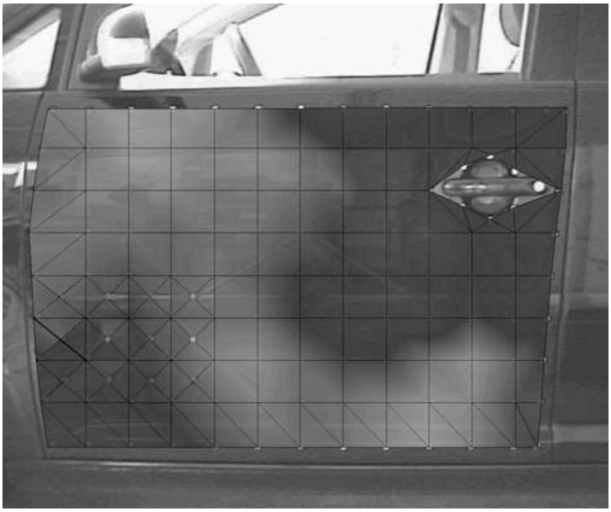


Fig. 10.23. Laser scan, vibration of door structure (excitation frequency 336 Hz (example)).

References

1. 3GPP:TS 46.010: Full rate speech encoding, *Third Generation Partnership Project (3GPP)*, 2002.
2. 3GPP:TS 46.051: GSM Enhanced full rate speech processing functions: General description, *Third Generation Partnership Project (3GPP)*, 2002.
3. 3GPP : TS 46.090: AMR speech codec: Transcoding functions, *Third Generation Partnership Project (3GPP)*, 2002.
4. 3GPP TS 26.077: Technical specification group services and system aspects; Minimum performance requirements for noise suppresser; Application to the adaptive multi-rate (AMR) speech encoder, *Third Generation Partnership Project (3GPP)*, 2003.
5. 70/157/EWG: Richtlinie des Rates zur Angleichung der Rechtsvorschriften der Mitgliedsstaaten über den zulässigen Geräuschpegel und die Auspuffvorrichtung von Kfz, 6. Feb. 1970 (in German).
6. ANSI S3.5-1997: Methods for calculation of the speech intelligibility index.
7. J. Berger: Instrumentelle Verfahren zur Sprachqualitätsschätzung – Modelle auditiver Tests, Ph.D. thesis,, Kiel, 1998 (in German).
8. J. Berger: Results of objective speech quality assessment including receiving terminals using the advanced TOSQA2001, ITU-T Contribution COM 12-20-E, Dec. 2000.
9. J. Blauert: *Spatial Hearing: The Psychophysics of Human Sound Localization,* Cambridge, MA, USA: MIT Press, 1997.
10. M. Buck, H.-J. Köpf, T. Haulick: Lombard-Sprache für Kfz-Anwendungen: eine Analyse verschiedener Aufnahmekonzepte, *Proc. DAGA '06*, Braunschweig, Germany, 2006 (in German).
11. ETSI EG 202 396-1: Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database, 2005.
12. ETSI EG 202 396-3: Speech quality performance in the presence of background noise; Part 3: Background noise transmission – objective test methods, 2007.
13. K. Genuit: Objective evaluation of acoustic quality based on a relative approach, *Proc. Internoise '96*, Liverpool, UK, 1996.
14. H. W. Gierlich: The auditory perceived quality of hands-free telephones: Auditory judgements, instrumental measurements and their relationship, *Speech Communication*, **20**, 241–254, October 1996.
15. H. W. Gierlich, F. Kettler, E. Diedrich: Proposal for the definition of different types of hands-free telephones based on double talk performance, *ITU-T SG 12 Meeting*, COM 12-103, Geneva, Switzerland, 1999.
16. ITU-T Recommendation G.122: Influence of national systems on stability and talker echo in international connections, *International Telecommunication Union*, Geneva, Switzerland, 1993.
17. ITU-T Recommendation G.722: 7 kHz audio-coding within 64 kbit/s, *International Telecommunication Union*, Geneva, Switzerland, 1988.
18. ITU-T Recommendation P.48: Specification for an intermediate refernce system, *International Telecommunication Union*, Geneva, Switzerland, 1993.
19. ITU T Recommendation P.50: Artificial voices, *International Telecommunication Union*, Geneva, Switzerland, 1999.
20. ITU T Recommendation P.51: Artificial mouth, *International Telecommunication Union*, Geneva, Switzerland, 1996.

21. ITU-T Recommendation P.58: Head and torso simulator for telephonometry, *International Telecommunication Union*, Geneva, Switzerland, 1996.
22. ITU T Recommendation P.501: Test signals for use in telephonometry, *International Telecommunication Union*, Geneva, Switzerland, 2000.
23. ITU T Recommendation P.502: Objective test methods for speech communication systems using complex test signals, *International Telecommunication Union*, Geneva, Switzerland, 2000.
24. ITU-T Recommendation P.505: One-view visualization of speech quality measurement results, *International Telecommunication Union*, Geneva, Switzerland, 2005.
25. ITU T Recommendation P.581: Use of head and torso simulator (HATS) for hands free terminal testing, *International Telecommunication Union*, Geneva, Switzerland, 2000.
26. ITU-T Recommendation P.79: Calculation of loudness ratings for telephone sets, *International Telecommunication Union*, Geneva, Switzerland, 2000.
27. ITU-T Recommendation P.340: Transmission characteristics and speech quality parameters of hands-free telephones, *International Telecommunication Union*, Geneva, Switzerland, 2000.
28. ITU-T Recommendation P.800.1: Mean opinion score (MOS terminology), *International Telecommunication Union*, Geneva, Switzerland, 2003.
29. ITU-T Recommendation P.800: Methods for subjective determination of speech quality, *International Telecommunication Union*, Geneva, Switzerland, 2003.
30. ITU-T Recommendation P.830: Subjective performance assessment of telephone-band and wideband digital codes, *International Telecommunication Union*, Geneva, Switzerland, 1996.
31. ITU-T Recommendation P.831: Subjective performance evaluation of network echo cancellers, *International Telecommunication Union*, Geneva, Switzerland, 1998.
32. ITU-T Recommendation P.832: Subjective performance evaluation of hands-free terminals, *International Telecommunication Union*, Geneva, Switzerland, 2000.
33. ITU-T Recommendation P.835: Subjective performance of noise suppression algorithms, *International Telecommunication Union*, Geneva, Switzerland, 2003.
34. ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, *International Telecommunication Union*, Geneva, Switzerland, 2001.
35. ITU-T Recommendation P.862.1: Mapping function for transforming P.862 raw result scores to MOS-LQO, *International Telecommunication Union*, Geneva, Switzerland, 2003.
36. ITU-T Focus Group FITcar: Draft specification for hands-free testing.
37. U. Jekosch: *Voice and Speech Quality Perception,* Berlin, Germany: Springer, 2005.
38. J.-C. Junqua: The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex, *Speech Communication*, **20**, 13–22, 1996.
39. W. Kellermann: Acoustic echo cancellation for beamforming microphone arrays, in M. Brandstein, D. Ward (eds.), *Microphone Arrays*, Berlin, Germany: Springer: 2001.

40. F. Kettler, H. W. Gierlich, E. Diedrich: Echo and speech level variations during double talk influencing hands-free telephones transmission quality, *Proc. IWAENC '99*, Pocono Manor, PA, USA, 1999.
41. F. Kettler, M. Röber: Generierung von Sprachmaterial zum realitätsnahen Test von Freisprecheinrichtungen, *Proc. DAGA '03*, Aachen, Germany, 2003 (in German).
42. F. Kettler, F. Rohrer, C. Nettelbeck, H. W. Gierlich: Intelligibility of hands-free phone calls outside the vehicle, *Proc. DAGA '07*, Stuttgart, Germany, 2007.
43. M. Lieb: Evaluating speech recognition performance in the car, *Proc. CFA/DAGA '04*, Strasbourg, France, 2004.
44. S. Möller: *Assessment and Prediction of Speech Quality in Telecommunications,* Boston, MA, USA: Kluwer Academic Press, 2000.
45. C. Pörschmann: Eigenwahrnehmung der Stimme in virtuellen akustischen Umgebungen, *Proc. DAGA '98*, Zürich, Switzerland, 1998 (in German).
46. J. Sotschek: Methoden zur Messung der Sprachgüte I: Verfahren zur Bestimmung der Satz- und Wortverständlichkeit, *Der Fernmelde-ingenieur*, 1976 (in German).
47. R. Sottek, K. Genuit: Models of signal processing in human hearing, *International Journal of Electronics and Communications*, 157–165, 2005.
48. VDA-Specification for Car Hands-Free Terminals, Version 1.5, VDA, 2005.
49. N. Xiang, K. Genuit, H. W. Gierlich: Investigations on a new reproduction procedure for binaural recordings, *Proc. AES 95th Convention*, Preprint 3732 (B2-AM-9), New York, NY, USA, 1993.