Final draft ETSI EG 202 396-1 V1.2.3 (2009-01)

ETSI Guide

Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database

Reference

REG/STQ-00139

Keywords

performance, quality, speech

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This ETSI Guide (EG) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ), and is now submitted for the ETSI standards Membership Approval Procedure.

The present document is part 1 of a multi-part deliverable covering Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise, as identified below:

Part 1: "Background noise simulation technique and background noise database";

Part 2: "Background noise transmission - Network simulation - Subjective test database and results";

Part 3: "Background noise transmission - Objective test methods".

Introduction

Background noise is present in most of the conversations today. Background noise may impact the speech communication performance to terminal and network equipment significantly. Therefore testing and optimization of such equipment is necessary using realistic background noises. Furthermore reproducible conditions for the tests are required which can be guaranteed only under lab type condition.

The present document addresses this issue by describing a methodology for recording and playback of background noises under well defined and calibratable conditions in a lab-type environment. Furthermore a database with real background noises is included.

1 Scope

The quality of background noise transmission is an important factor, which significantly contributes to the perceived overall quality of speech. Existing and even more the new generation of terminals, networks and system configurations including broadband services can be greatly improved with a proper design of terminals and systems in the presence of background noise. The present document:

- describes a noise simulation environment using realistic background noise scenarios for laboratory use;
- contains a database including the relevant background noise samples for subjective and objective evaluation.

The present document provides information about the recording techniques needed for background noise recordings and discusses the advantages and drawbacks of existing methods. The present document describes the requirements for laboratory conditions. The loudspeaker setup and the loudspeaker calibration and equalization procedure are described. The simulation environment specified can be used for the evaluation and optimization of terminals and of complex configurations including terminals, networks and other configurations. The main application areas should be: office, home and car environment.

The setup and database as described in the present document are applicable for:

- Objective performance evaluation of terminals in different (simulated) background noise environments.
- Speech processing evaluation by using the pre-processed speech signal in the presence of background noise, recorded by a terminal.
- Subjective evaluation of terminals by performing conversational tests, specific double talk tests or talking and listening tests in the presence of background noise.
- Subjective evaluation in third party listening tests by recording the speech samples of terminals in the presence of background noise.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

Not applicable.

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area. For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] Surround Sound Past, Present, and Future: "A history of multichannel audio from mag stripe to Dolby Digital", Joseph Hull - Dolby Laboratories Inc.

[i.2] AES preprint 3332 (1992): "Improved Possibilities of Binaural Recording and Playback Techniques", K. Genuit, H.W. Gierlich; U. Künzli.

NOTE: See at http://www.aes.org/e-lib/browse.cfm?elib=6801.

[i.3] AES preprint 3732 (1993): "A System for the Reproduction Technique for Playback of Binaural Recordings", N. Xiang, K. Genuit, H.W. Gierlich.

NOTE: See at http://www.aes.org/e-lib/browse.cfm?elib=6501.

[i.4] NTTAT Database: "Ambient Noise Database CD-ROM".

NOTE: See at http://www.ntt-at.com/products_e/noise-DB/index.html.

[i.5] ISO 11904-1: "Acoustics - Determination of sound immission from sound sources placed close to the ear - Part 1: Technique using a microphone in a real ear (MIRE technique)".

[i.6] Spatial Hearing: "The psychophysics of human sound localization", J. Blauert.

[i.7] ITU-T Recommendation P.57: "Artificial ears".

[i.8] ITU-T Recommendation P.58: "Head and torso simulator for telephonometry".

[i.9] ITU-T Recommendation P.340: "Transmission characteristics and speech quality parameters of hands-free terminals".

[i.10] ITU-T recommendation P.64: "Determination of sensitivity/frequency characteristics of local telephone systems".

[i.11] ITU-T Recommendation G.722: "7 kHz audio-coding within 64 kbit/s".

[i.12] Genuit, K.: "A Description of the Human Outer Ear Transfer Function by Elements of Communication Theory (No. B6-8)".

NOTE: Proceedings of the 12th International Congress on Acoustics. Toronto published on behalf of the Technical Program Committee by the Executive Committee of the 12th International Congress on Acoustics.

[i.13] IEC 60050-722: "International Electrotechnical Vocabulary - Chapter 722: Telephony".

[i.14] "Wellenfeldsynthese - Eine neue Dimension der 3D-Audiowiedergabe"; Fernseh- und Kino-Technik, Nr. 11/2002, pp. 735-738.

[i.15] "The Iosono Sound Difference".

NOTE: http://www.iosono-sound.de

[i.16] "Ein neues Verfahren der raumbezogenen Stereophonie mit verbesserter Übertragung der Rauminformation"; P. Scherer, Rundfunktechnische Mitteilungen, 1977, pp. 196-204.

[i.17] ETSI EG 202 396-1 (V.1.1.2): "Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database".

[i.18] ETSI TS 151 010-1: "Digital cellular telecommunications system (Phase 2+); Mobile Station (MS) conformance specification; Part 1: Conformance specification (3GPP TS 51.010-1)".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

crosstalk: appearance of undesired energy in a channel, owing to the presence of a signal in another channel, caused by, for example induction, conduction or non linearity

NOTE: See IEC 60050-722 [i.13].

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

CD	Compact Disc
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
HATS	Head And Torso Simulator
IIR	Infinite Impulse Response
MIRE	Microphone In Real Ear
NTT	Nippon Telegraph and Telephone Corporation
SLR	Send Loudness Rating
VHF	Very High Frequency

4 Overview of existing methods for realistic sound reproduction

4.1 Introduction

In general the existing methods for close to original sound recording and reproduction aimed for different applications:

- Techniques intending to reproduce the actual sound field.
- Techniques providing hearing adequate (ear related) signals in the human ear canal.
- Techniques generating artificial acoustical environments.

Within this clause the different methods are briefly described and their applicability for close to original sound-filed reproduction is discussed. A variety of methods have been studied, in the following a summary of the most important ones relevant to the present document is given. The different methods were analyzed on the basis of the following requirements:

- The background noise recording technique should be:
  - easy to use;
  - easy to calibrate;
  - capable of wideband recording;
  - available at reasonable costs;
  - mostly compatible to existing standards and procedures used in telecommunications testing;
  - applicable to different environments (at least office, home and car).

- The background noise simulation arrangement should:
  - be easy to setup;
  - not require any specific acoustical treatment for the simulation requirement;
  - provide a mostly realistic background noise simulation for all typical background noises faced with in telecommunication applications;
  - be easy to calibrate;
  - be mostly insensitive against the positioning of (test)-objects in the simulated sound field;
  - be applicable to all typical terminals used in telecommunication;
  - be available at reasonable costs.

4.2 Surround Sound Techniques

The basics of surround techniques are found in cinema applications. The virtual image provided by stereophonic presentation of sounds seemed not to be sufficient for the large screen display in cinema. In the 1950s 4-channel and 6-channel soundtracks recorded on magnetic stripes associated to the films were developed, 4-channel and 6-channel loudspeaker systems were installed in cinemas to reproduce the multichannel sounds. The newer techniques were mostly developed and marketed by Dolby® [i.1]: Dolby Surround, Dolby Surround Pro Logic, Dolby Digital and Dolby Digital Surround are examples for the techniques introduced more recently. The most common configuration is the "5.1-configuration" used in cinema but in home applications as well. The reproduction system consists of left and right channel, a centre speaker, two surround channels (left and right, arranged in the back of the listener) and a low frequency channel for low frequency effects.

The aim of all surround system is to create an artificial acoustical image in the recording studio rather than recording a real acoustical scenario and providing true to original playback possibilities.

On the recording side special surround encoders are used allowing the 5-channel signal to be encoded from a special mixing console to the 5.1 digital data stream. The playback system consists of a special decoder allowing to separate the 5 channels again and distribute them on the 5.1 loudspeaker playback system. The systems are mono and stereo compatible and can handle the older 4 channel surround techniques by a specific decoder.

Applications:

Typical applications for surround systems are cinemas and home theatres. The source material is produced by professional recording studios using multi-channel mixing consoles and specific 5.1 decoding techniques. In mostly all cases virtual environments are created which support the visual image by an appropriate acoustical image.

Conclusion:

Surround techniques are designed for creating acoustical images rather than for close to original recording and reproduction. Although the spatial impression provided by surround techniques is sometimes remarkable the acoustical image created is always artificial. Due to the lack of easy to use recording techniques allowing a spatial recording of a sound field surround sound techniques are not suitable for creation of a background noise database with realistic background noises and calibrated background noise simulation in a lab.

4.3 IOSONO

The IOSONO® sound system (see [i.14] and [i.16]) is based on the Wave-Field Synthesis. It employs Huygens principle of wave theory. Applied to acoustics this principle means that it is possible to reproduce any form of wave front with an array of loudspeakers, so that virtual sound sources can be placed anywhere within a listening area. For practical use it is necessary to position loudspeakers all-round the playback room. In order to generate realistic sound fields the input signal for each loudspeaker has to be calculated separately. For this purpose each single sound source (e.g. voices) has to be recorded individually. If the recordings are done in a room, the characteristics (like reverberation) of the recording room also have to be recorded separately. All resulting sound tracks are then mixed and manipulated during the post-editing process and the reproduction.

The natural and realistic spatial sound reproduction is then achieved in a wide area of the play back room. Common 5.1 stereo systems achieve a "realistic" sound reproduction only in a small area of the reproduction room.

Applications:

Typical applications are sound systems for home use, cinemas and other entertainment events. The IOSONO sound system is also able to play back recordings made in common stereo or 5.1 stereo techniques.

Conclusion:

The drawbacks of this method are the components needed: a sophisticated recording system, a powerful computing unit for real-time mixing the large number of recorded sound tracks and the number of loudspeakers that have to be installed in the listening room. In a common size cinema for example about 200 loudspeakers are needed.

The advantage is that with the IOSONO sound system a very realistic sound reproduction is possible, but it requires an enormous effort, which is too high for daily use in laboratories.

4.4 Eidophonie

This method (see [i.17]) was developed for realistic sound reproduction using the VHF transmission technique. The main principle is to separate the base signal from the part of the signal, which contains the information about the direction of sound incidence.

For recording a 1st order gradient microphone with a cardioid directivity is used. During the recordings its directivity rotates with 38 kHz in the recording plane. This "turning microphone" provides an amplitude-modulated signal at its electrical output. The resulting side bands are out of the transmitted frequency range. But these side bands contain the information of the direction of sound incidence. Using the VHF- transmission techniques this phase information can be transmitted within the 2nd audio-frequency channel.

The sound reproduction is made by a spatial demodulation: a switch is positioned before each loudspeaker and each switches synchronously with the turning directivity. So a low pass filtered short-term section of the signal containing the information of the direction of sound incidence is played back on each loudspeaker. The loudspeakers are positioned all around the playback room.

Applications:

Eidophonie was developed to provide a realistic sound environment using a signal received from a VHF broadcast station. With this technique the common stereo sound reproduction should be improved. Nevertheless Eidophonie is also compatible to common mono and stereo recordings.

Conclusion:

Benefits of this system are that three loudspeakers are sufficient to produce a realistic sound field. Using more loudspeakers (e.g. 16) the spatial sound reproduction gets more and more independent from the listening position. Moreover the independency of the transmitted sound from the acoustics of the reproduction room increases with the number of loudspeakers used. But there are significant limitations of the method: The microphone directivity is frequency dependent and not ideal. Therefore the interference between the different channels is created. A second problem is the loudspeaker directivity, which does not fit the microphone directivity. This problem could be reduced if the number of channels would be increased. This however is not possible due to the limited directivity of the microphone arrangement used.

Localization of sound sources is hardly possible due to the interference effects of the microphone signals and the loudspeakers. At close to original reproduction depends on the number and distribution of sound sources present. For most of the sound source combinations this goal cannot be achieved.

In general the coding technique needed to record the sound field by a "turning microphone", is complicated and not available commercially. A further drawback of this method is the complicated decoding technique needed on the reproduction side, which is also not commercially available.

4.5 Four-loudspeaker arrangement for playback of binaurally recorded signals

This reproduction procedure was originally investigated to reproduce binaurally signals recorded using artificial head technology. It improves the impressions of direction and distance. Four loudspeakers are typically positioned in a square formation around a central point (listening point) equidistantly e.g. 2 m. The binaural recordings are played back as follows: the two left-hand loudspeakers receive the same free-field equalized artificial head signal of the left-hand channel only. The right-hand side is arranged similarly. For equalization the transfer function from the two left-hand loudspeakers is measured at the artificial heads left ear channel. With this result IIR and FIR filters are designed, with which the input signal of the left-hand loudspeakers during the play back is filtered in such a way that the transfer function then measured at the artificial heads left-hand channel is spectrally flat. The equalization for the right-hand loudspeakers is done similarly.

The equalization procedure does not take into account the correction of crosstalk. This means, the left-hand channel of the artificial head is only equalized for the left-hand loudspeakers, but during the reproduction this left-hand channel will also receive a signal from the right-hand loudspeakers. But despite this simplification the equalization procedure provides a realistic binaural listening impression.

Investigations [i.2], [i.3] carried out in different rooms have shown that directional hearing and distance localization by sound reproduction with this four-loudspeaker arrangement are comparable to those with sound reproduction by headphones.

For equalization there are several strategies. The equalization can either be done for each loudspeaker individually or by pairs left - right or by pairs front - rear.

Applications:

A practical application is the sound reproduction for binaural recordings in a typical office-type room, e.g. for listening tests but objective tests as well. The investigations shown in [i.2], [i.3] indicate that the subjective impression provided by the arrangement corresponds to the headphone reproduction of binaural recordings with respect to the perception of the sound colour, the distance perception and (with some limitations) with respect to the sound source localization. The data provided indicate that the setup could be used for the objective measurements of e.g. telecommunication terminals.

This four-loudspeaker arrangement is also used in advanced driving simulators - which typically provide a visual simulation of the driving situation in addition to the acoustical playback system.

Conclusion:

The advantage of this arrangement on the recording side is the compatibility to standard mono and stereo recordings. Due to this it is easily possible to playback either binaural recordings for subjective and objective experiments or mono/stereo recordings in cases where a less realistic reproduction is sufficient. With slight modifications the geometrical setup can be transferred to other environments like cars for example. Another benefit of this arrangement is the moderate hardware effort.

Concerning the sound reproduction a drawback of this arrangement is that due to the superimposed loudspeaker signals mostly in anechoic rooms interferences appear. This effect obliged test subjects to keep their heads in a mostly fixed position during hearing tests, but also means that an exact sound reproduction is only possible in a small area in anechoic rooms. Another drawback for binaural reproduction is the fact that no exact crosstalk cancellation is possible with this arrangement. However in general this technique seems to be the most promising under the restrictions given (moderate hardware effort on the recording and especially on the reproduction side, close to original reproduction of the scenarios recorded without additional adjustment of the reproduction arrangement).

4.6 NTT Background-Noise Database

The NTT Background Database [i.4], which is commercially available from NTT, is typically used for codec tests. The database contains noise files, which were recorded with a 4-channel recording using 4 directional microphones. The microphones were arranged in an angle of 90 degrees with 70 cm diagonal. Although the original signals were recorded with 20 kHz bandwidth, the signals commercially available are specially coded on a CD providing a bandwidth of 11 kHz/15 bit for each channel. A special decoder is needed if the signals are to be presented acoustically over loudspeakers. The loudspeaker arrangement is suggested in a condition list. For calibration a calibration tone is provided, only level calibration is performed, no equalization procedure is described.

For the electrical evaluation of systems a downmix of the 4-channel recording to a mono channel with 8 kHz sampling rate is available. This signal is mostly used for the evaluation of new speech codecs in order to evaluate the influence of background noise on the coder performance.

Applications:

The NTT-database is mostly used for the evaluation of speech coders using the 8 kHz down sampled signals. When using the acoustical 4-channel playback the limitation is mostly due to the bandwidth limitation of 11 kHz, which may be not sufficient for future wideband applications.

Conclusion:

The disadvantage of this arrangement is found on the recording and on the playback side. For the recording a special microphone arrangement and a special coding technique is needed. The signal on the reproduction side is band limited and may not be used in future wideband applications. Currently it is not clear how the playback system can be calibrated and equalized in order to achieve a close to original sound field and which procedure should be followed during recording in order to get the right calibration and setup for the recording. Furthermore microphone arrangement chosen seems to be impractical for the recording of scenarios in small rooms e.g. in a car.

Another drawback is that the background noise database needs to be purchased for each application including a special decoder. It is not clear to what extent the recording and coding technique is commercially available.

4.7 General conclusions

Although a variety of reproduction techniques exists the usability of such methods within the constraints for laboratory use namely:

- easy and well described recording technique;
- easy to install and easy to use playback technique at reasonable costs;
- installation of the playback technique in a variety of different rooms with different acoustical conditions;

is limited. A four-channel loudspeaker setup with associated subwoofer based on binaurally recorded material is selected as the basis for the ETSI background noise simulation arrangement.

5 Recording arrangement

5.1 Binaural equalization

The sound field simulation technique described in the present document is generally based on the binaural recording and reproduction technique as has been known for many years (see [i.6]). The general principle of the recording and playback technique is to provide a recording instrument (artificial head or test subject wearing binaural probe microphones) which allows the recording of the ear signals typically received by the human user in a sound field as close to the original as possible. When an artificial head (HATS- head and torso simulator) is used it should represent the "average human". A suitable description of the properties of the artificial head to be used for the recording techniques can be found in ITU-T Recommendation P.58 [i.8], which describes the head and torso simulator for use in telephonometry. Since the directivity characteristics of an artificial head is strongly directionally dependent, the output signals of the microphones of an artificial head cannot be directly compared to a standard measurement microphone and an equalization procedure has to be applied in order to get comparable output signals. Various types of equalization procedures for artificial head recordings can be used. The following are known:

Free Field Equalization

The artificial head is equalized in such a way that for frontal sound incidence in anechoic conditions the frequency response of the artificial head is flat.

Diffuse Field Equalization

The artificial head is equalized in a diffuse sound field. For random sound incidence the frequency response of the artificial head is flat (see also microphone diffuse field equalization).

Independent of Direction Equalization

This technique is based on the fact that the transmission characteristics of an artificial head is influenced by components independent of direction (cavum concha resonance, ear canal transfer function) and components which contribute depending on the direction of sound incidence to the measured transfer function. The equalization considers only those components of the head related transfer function which are independent of direction.

5.2 The equalization procedure

Generally the artificial head (HATS) used should comply with the requirements as defined in ITU-T Recommendation P.58 [i.8]. Although all equalization methods described below can be used, the preferred method is the diffuse field equalization or (if available) the independent of direction equalization. This is because there are the less sharp peaks and dips in the transfer function of the artificial head compared the freefield transfer function (see also ITU-T Recommendation P.58 [i.8]).

Free-Field Equalization

The free-field equalization is made for this reference position in an anechoic room. The reference source is placed on the reference axis of the HATS, at a minimum distance of 1,5 m from the HATS lip-ring. When using a binaural microphone this point corresponds to the lip plane of a human user wearing the binaural microphone. The measured free-field response of the HATS is:

$$H_{ff}(0°, 0°, f)$$

From this the free-field equalization is calculated:

$$H_{ff\text{-}EQ}(0°, 0°, f) = 1/H_{ff}(0°, 0°, f)$$

$H_{ff\text{-}EQl}(0°, 0°, f)$ is called free-field equalization of left ear, $H_{ff\text{-}EQr}(0°, 0°, f)$ is called free-field equalization of right ear.