ETSI TS 103 224 V1.5.1 (2020-03)

TECHNICAL SPECIFICATION

**Speech and multimedia Transmission Quality (STQ);
A sound field reproduction method for terminal testing
including a background noise database**

Reference
RTS/STQ-289

Keywords
noise, quality, speech, terminal

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

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

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 https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables 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 (https://ipr.etsi.org/).

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.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Technical Specification (TS) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

The present document describes a sound field recording and reproduction technique which can be applied for all types of terminals but is especially suitable for modern multi-microphone terminals including array techniques. The present document provides an additional simulation technique which can be used instead of the part 1 of ETSI multi-part deliverable ES/EG 202 396 "Speech quality performance in the presence of background noise", as identified below:

- ETSI ES 202 396-1: "Background noise simulation technique and background noise database" [i.7];
- ETSI EG 202 396-2: "Background noise transmission - Network simulation - Subjective test database and results" [i.8];
- ETSI EG 202 396-3: "Background noise transmission - Objective test methods" [i.9].

The background noise simulation can be used in conjunction with the objective test methods as described in ETSI EG 202 396-3 [i.9] and ETSI TS 103 106 [i.10].

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

Background noise is present in most of the conversations today. Background noise may impact the speech communication performance of 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 conditions. Since modern terminals incorporate more advanced noise cancellation techniques, such as multi-microphone based noise cancellation, the use of microphone-array recording techniques and more realistic noise field simulations (compared to the method described in ETSI ES 202 396-1 [i.7]) are required.

The present document addresses this topic by specifying a methodology for recording and playback of realistic background noise fields under conditions that are well-defined and able to be calibrated 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. Terminals, networks, and system configurations including wideband, superwideband, and fullband speech services can be greatly improved with a proper design of terminals and systems in the presence of background noise. The present document:

- describes a sound field simulation technique allowing to simulate the real environment using realistic background noise scenarios for laboratory use;
- contains a database including relevant background noise samples for subjective and objective evaluation.

The present document describes the recording technique used for the sound field simulation, the loudspeaker setup, and the loudspeaker calibration and equalization procedures. Furthermore the present document specifies the test room requirements for laboratory conditions.

The simulation environment specified can be used for the evaluation and optimization of terminals and of complex configurations including terminals, networks and others. The main application areas are: outdoor, 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 signals 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

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at https://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.

The following referenced documents are necessary for the application of the present document.

Not applicable.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

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

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] Berkhout A. J., de Vries D., & Vogel, P.: "Acoustic control by wave field synthesis", J. Acoust. Soc. Am., p. 2764-2778, Mai 1993.

[i.2] Gerzon, M. A.: "Periphony: With-Height Sound Production", Journal of the Audio Engineering Society 21, 1973.

[i.3] Ward D. B., Abhayapala T. D.: "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers", IEEE transactions on speech and audio processing, Vol. 9, No.6, p. 697-707, September 2001.

[i.4] Kirkeby O., Nelson P. A., Orduna-Bustamante F., Hamada H.: "Local sound field reproduction using digital signal processing", J. Acoust. Soc. Am. 100(3), p. 1584-1593, September 1996.

[i.5] Kirkeby O., Nelson P. A., Hamada H., Orduna-Bustamante F.: "Fast Deconvolution of Multichannel Systems Using Regularization", IEEE transactions on speech and audio processing, VOL. 6, NO. 2, p. 189-195, March 1998.

[i.6] Recommendation ITU-T P.58: "Head and Torso Simulator for Telephonometry".

[i.7] ETSI ES 202 396-1: "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".

[i.8] ETSI EG 202 396-2: "Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 2: Background noise transmission - Network simulation - Subjective test database and results".

[i.9] ETSI EG 202 396-3: "Speech and multimedia Transmission Quality (STQ); Speech Quality performance in the presence of background noise; Part 3: Background noise transmission - Objective test methods".

[i.10] ETSI TS 103 106: "Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods".

[i.11] ISO 3382-1: "Measurement of room acoustic parameters -- Part 1: Performance spaces".

3 Definition of terms, symbols and abbreviations

3.1 Terms

Void.

3.2 Symbols

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

c	Sound velocity
C	Matrix of FFT coefficients of Compensation Filters
H	Matrix of FFT coefficients of Impulse Responses

3.2 Abbreviations

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

DUT	Device Under Test
EEP	Ear canal Entrance Point
FFT	Fast Fourier Transform
HATS	Head And Torso Simulator
IR	Impulse Response
MLS	Maximum Length Sequence
MOS	Mean Opinion Score
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level

4 Methods for realistic sound reproduction

For reproduction of real world sound fields there exists a variety of different methods, two of them are wave field synthesis [i.1] and Ambisonics [i.2]. Both methods, however, require a large number of microphones and loudspeakers to achieve a sound field reproduction which is sufficiently good for testing purposes. The Wave-Field synthesis setup is that complex and expensive that it can be neglected for laboratory purposes. Ambisonics, for example, has to be performed using 43 microphones and 43 loudspeakers to reach a good sound field reproduction up to 2 kHz in a sweet spot with radius 15 cm (using the rule of thumb in [i.3]). It furthermore cannot consider individual room characteristics or insufficiencies, but is only designed for rooms offering pure free field conditions. If, e.g. for testing purposes a HATS is positioned in the artificial noise field, the reproduction quality is reduced by an unknown amount. In summary, the Ambisonics approach is due to its design not feasible for the intended testing scenario.

The present document introduces an alternative least mean squares method [i.4], which requires eight recording channels and eight loudspeakers in order to achieve reasonably good reproduction results. The method is based on eight sweet spots at important testing positions e.g. near the HATS, mainly at the microphone positions of modern phones.

A reasonable reproduction of the recorded sound field at the corresponding eight points in the reproduction situation also yields good reproduction accuracy in between these points. This well-known property of sound fields is limited to an upper cut-off frequency which depends on the distances between the recording microphones (see clause 5.1.1).

In clause 5, the recording technique required for this new method is described, while the setup allowing the reproduction in laboratories and the different steps of the equalization procedure are introduced in clause 6.

5 Recording arrangement

5.0 General

The sound field recording technique (Multi-point sound field recording technique) is based on optimization of the sound field reproduction at different points in space. The optimization criterion is based on minimization of the reproduction error at each microphone position. Based on this principle the microphone locations and as a consequence the points in space for which the sound field reproduction is mostly accurate can be chosen in a wide range. The advantage of the method is that these locations can be adapted to the type of device which is to be tested. E.g. if the Device Under Test (DUT) incorporates a microphone array of the Multi-point sound field recording microphones can be positioned in the area of the microphones of the DUT. If a hands-free device is to be tested the Multi-point sound field recording microphones are positioned in the area of the hands-free device.

The setup described in detail in clause 5 is optimized for the testing of handset or headset terminals using HATS according to Recommendation ITU-T P.58 [i.6] and for hands-free testing. The procedure described here can be followed in the same way for other microphone setups.

In this clause the setups for the microphone arrangements as used in the present document are described. The background noise recordings based on these different recording setups are described in clause 8.

5.1 Microphone array setup

5.1.1 Principle limitations

With a perfect sound field reproduction at two closely spaced points, the cut-off frequency up to which the sound field in between those two points is also correctly reproduced depends on their distance. This upper cut-off frequency can be estimated as:

$$f_{lim} = \frac{c}{2d_{max}} \qquad (1)$$

where d_{max} is the maximum distance between two microphones and c is the sound velocity.

EXAMPLE: For the eight microphones in Figure 1, f_{lim} is dependent on the distance of the microphone pair considered and is about *1,7 kHz* in the region of sparsely spaced microphones and approximately *3 kHz* in the region of densely spaced microphones. Note, that at the microphone positions itself the reproduction quality is optimal across the whole frequency range. In between of these positions the accurate spatial reproduction can only be guaranteed up to f_{lim}.

5.1.2 Microphone calibration

In order to yield a good sound field reproduction at the defined positions, the microphone array for recording of the real sound field and the microphone array for equalization and calibration of the reproduction setup have to match. In detail, the frequency/phase response and the directional sensitivity of the corresponding microphones of the two arrays has to be identical. As a consequence, each microphone has to be calibrated individually with regard to frequency response, phase response and level.

The supplier of such devices should provide information regarding the sensitivity of the individual microphones constituting the microphone array for verification purposes. The calibration data provided need to be suitable to ensure a proper phase calibration up to at least 3 kHz, a proper frequency response calibration in the frequency range between 50 Hz and at least 3 kHz with an accuracy of < 0,5 dB in 1/12th octave, between 3 kHz and 10 kHz with an accuracy of < 0,5 dB in 1/3rd octave, between 10 kHz and 20 kHz with an accuracy of < 3 dB in 1/3rd octave and a proper level calibration (at 250 Hz or 1 kHz) with an accuracy of < 0,1 dB.

5.2 Microphone array setup for handset-type and headset terminals

Figure 1 shows the configuration of microphones located around an artificial head. The locations of the microphones define the sweet spots where the reproduction of the recorded signals is optimal for all frequencies. In consequence the majority of these points are at relevant positions where the microphones of the test devices are usually located (see Figure 1, top left). The exact positions for the eight recording microphones are given in Figure 1 (bottom). Eight additional positions are defined by clockwise rotation of the microphone array by 10 degrees. (Figure 1, top right, in dark) around the axis of rotation of the HATS as defined in Recommendation ITU-T P.58 [i.6]. This position is called "fine tuning set" and is used for optimization and verification of the equalization.

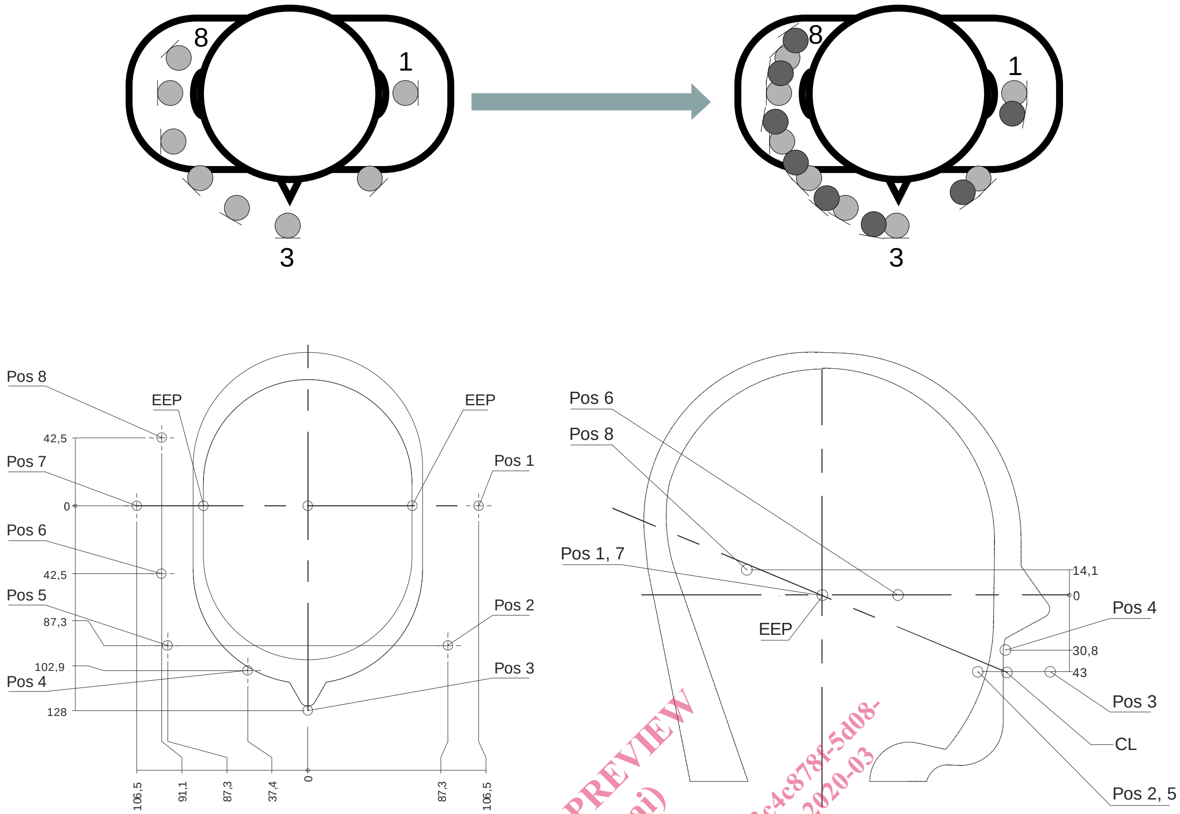

Figure 1: Positions of the recording microphones
Vertical positions are related to the vertical position of the EEP

5.3 Microphone array setup for hands-free terminals

In general, different microphone arrays could be used for hands-free terminals as well as for handsets and headsets. However, to increase reusability and reduce efforts, the same microphone array can be used in both cases. The setup of the array for measuring hands-free terminals is shown in Figure 2.

For the hands-free equalization, the DUT is first positioned at its testing position, which is defined in the relevant standards. Then, the main microphone position of the terminal is determined. In the case of terminals using multi microphone techniques terminals the main microphone is chosen, and in case of array techniques the acoustical centre of the array (typically identical to the centre of the array) is used.

In the setup for hand-held and tablet terminals, the microphone array is positioned such that microphone 5 is in top view right-angled in front of the main microphone position in *25 mm* distance (Figure 2, right) and microphone 6 is at the height of the main microphone position (Figure 2, left).

For desktop operated hands-free terminals, the microphone 5 of the array is positioned right-angled in front of the main microphone position in *25 mm* distance (Figure 3, right) and *25 mm* above the table (Figure 3, left).

Note that the DUT is absent during the equalization procedure itself.

The "fine-tuning set" is realized the same way as described in clause 5.2, rotating the microphone array clockwise by 10 degrees.

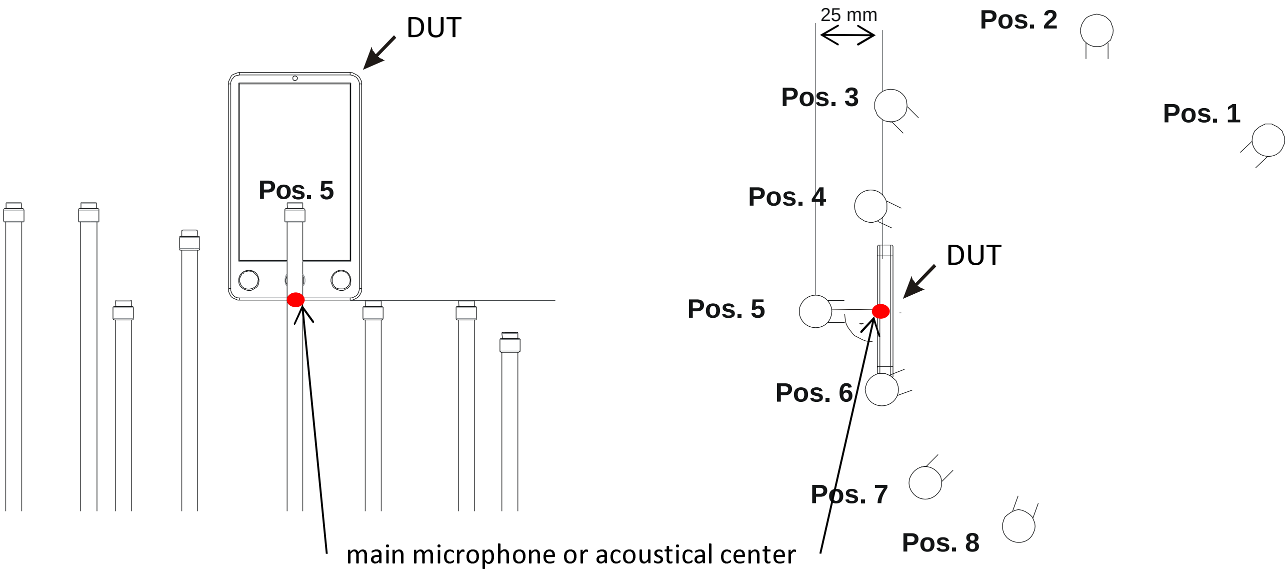

Figure 2: Positions of the recording microphones in a hands-free setup for hand-held and tablet terminals

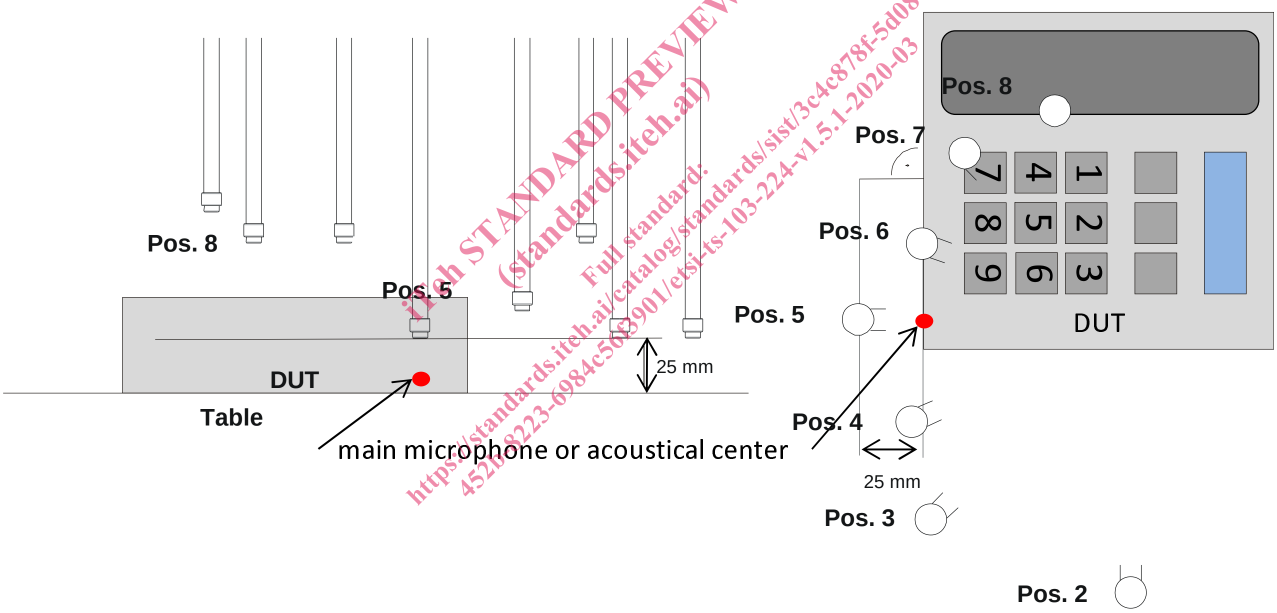

Figure 3: Positions of the recording microphones in a hands-free setup for desktop operated hands-free terminals

5.4 Microphone array setup for binaural applications

Figure 4 shows the configuration of microphones located around an artificial head. The locations of the microphones define the sweet spots where the reproduction of the recorded signals is optimal for all frequencies. In consequence these points are in the direct vicinity of the ears where the microphones of binaural test devices are usually located. The exact positions for the eight recording microphones are given in Figure 4.