ETSI TS 103 224 V1.6.1 (2022-03)

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

Reference

RTS/STQ-300

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

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

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

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], ETSI TS 103 106 [i.10] and ETSI TS 103 281 [i.12].

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, super-wideband, 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.

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

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

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] Void.

[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".

[i.12] ETSI TS 103 281: "Speech and multimedia Transmission Quality (STQ); Speech quality in the presence of background noise: Objective test methods for super-wideband and fullband terminals".

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:

DRP	Drum Reference Point
DUT	Device Under Test
EEP	Ear canal Entrance Point
FFT	Fast Fourier Transform
HATS	Head And Torso Simulator
IR	Impulse Response
LFE	Low-Frequency Extension
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.

A generic variant for flexible microphone and loudspeaker arrangements is described in clause 7.

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 to be tested. 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 of several microphone arrangements 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 [1] and for hands-free testing. The procedure described here can be followed in the same way for other microphone setups.

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:

$$f_{lim} = \frac{c}{2d_{max}} \tag{1}$$

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

EXAMPLE: For the eight microphones in Figure 1, *flim* depends 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 *flim*.

5.1.2 Microphone calibration

In order to yield an accurate 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 shall provide identical frequency response, phase response and level calibration.

The supplier of such devices shall provide information regarding the sensitivity vs frequency of the individual microphones of the array for verification purposes as specified in Table 0.

Table 0: Accuracy requirements vs frequency of array microphones

Frequency range	Octave band resolution	Accuracy [dB]	Comment
f_{min} to 50 Hz	1/12th	±3,0	Only applicable in case the setup is used with extended frequency range (see clause 6.2.4)
50 Hz to 3 kHz	1/12th	±0,5	
3 kHz to 10 kHz	1/3rd	±0,5	
10 kHz to 20 kHz	1/3rd	±3,0	
250 Hz or 1 kHz	-	±0,1	Overall calibration at single frequency

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 [1]. This position is called "fine tuning set" and is used for optimization and verification of the equalization.

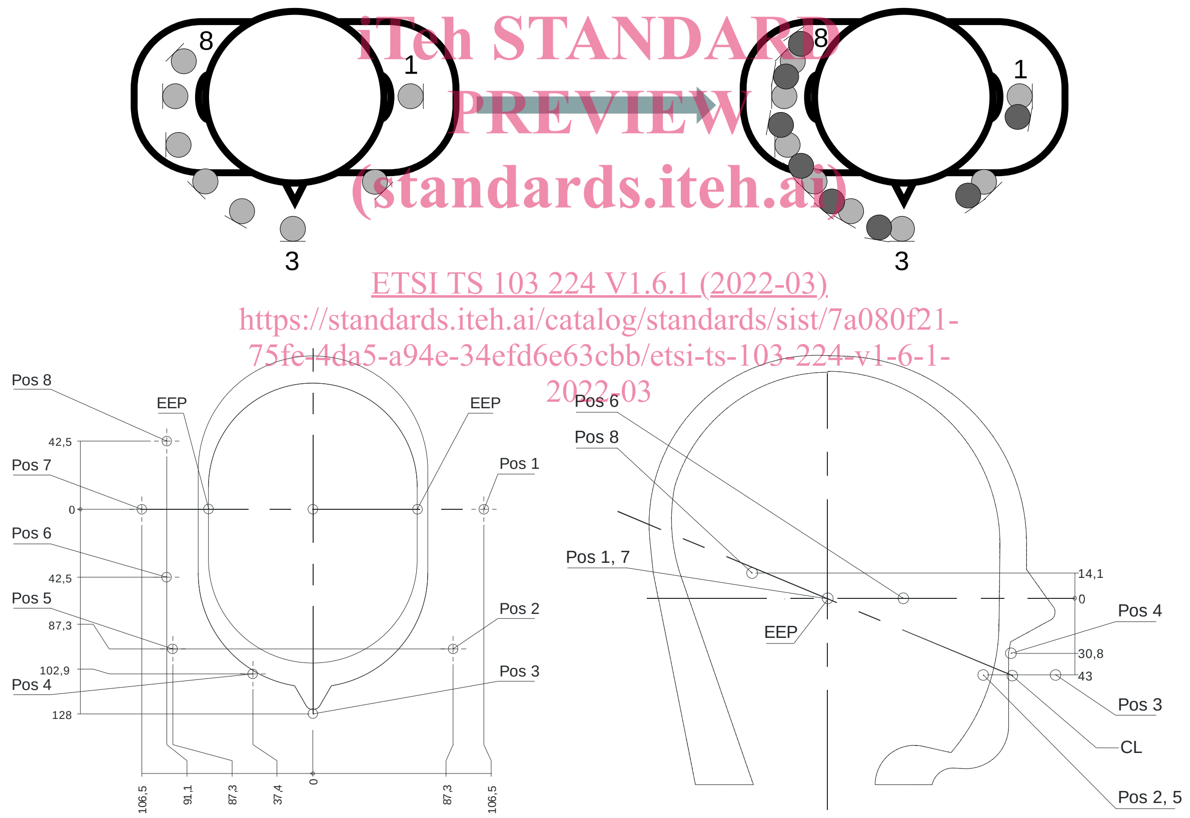

Figure 1: Positions of the recording microphones
Vertical positions are related to the vertical position of the EEP

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.

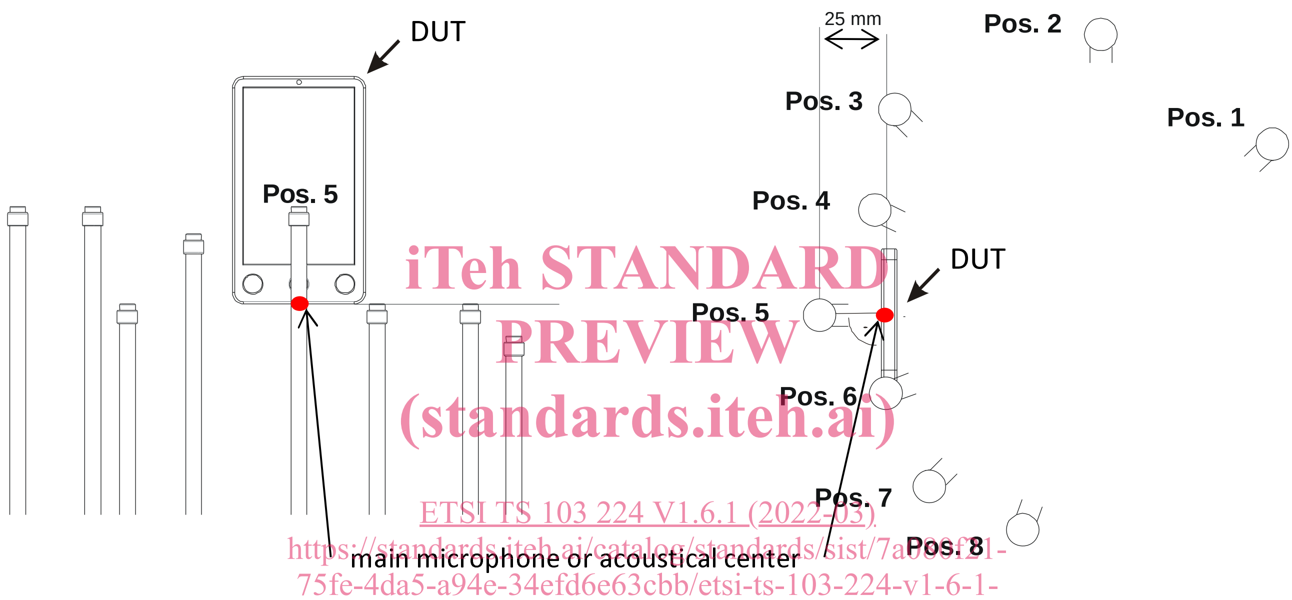

Figure 2: Positions of the recording microphones in a hands-free setup for hand-held and tablet terminals

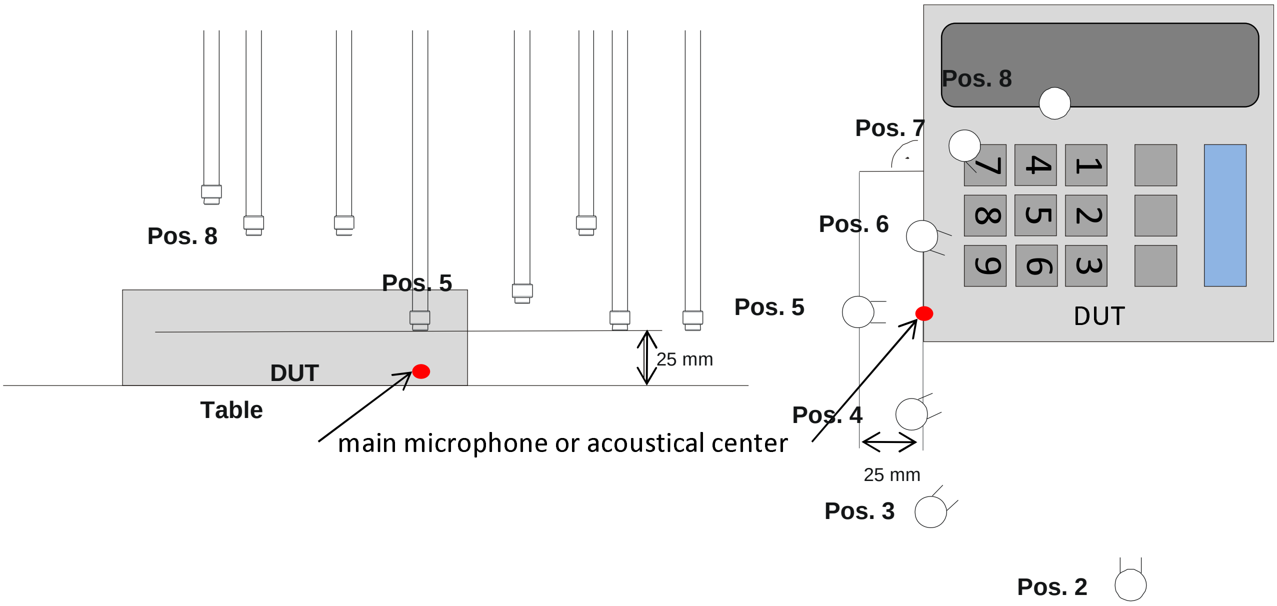

Figure 3: Positions of the recording microphones in a hands-free setup for desktop operated hands-free terminals

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.

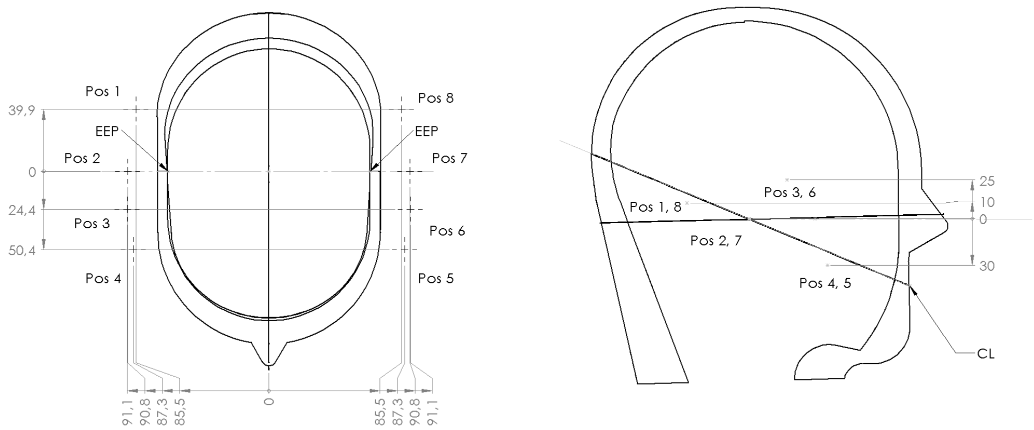

Figure 4: Positions of the recording microphones for binaural applications Vertical positions are related to the vertical position of the EEP

6 Loudspeaker setup for background noise simulation

6.0 General setup

It should be noted that the position height of the loudspeakers as well as the exact spacing between them in general is not crucial, since the equalization procedure described below accounts for the individual loudspeaker positions. The difference that might be observed between different loudspeaker positions is a different deviation from the original sound field at the intermediate positions of the microphone array. In order to allow better inter-lab accuracy of the sound field reproduction, the following positioning arrangement should be followed if possible.

Figure 5 shows the setup of the eight loudspeakers for the desired sound field reproduction. The vertical position of the loudspeakers is adjusted so that the centre of odd loudspeakers (i.e. 1, 3, 5 and 7) is about 15 cm above the HATS reference plane [1] and the centre of the remaining even loudspeakers (i.e 2, 4, 6 and 8) is about 15 cm below the HATS reference plane. The distance between the loudspeakers to the HATS as well as the horizontal distribution of the loudspeakers can be selected depending on the room, hence the spacing between the loudspeakers does not have to be exactly equal. The setup may be a square or a circle around the HATS or a setup in between depending on what fits best inside the room.

The distance between the surface of the artificial head and the loudspeaker fronts shall be at least 50 cm and should not exceed 2,5 m. Note that the maximum distance is also limited by the maximum sound pressure level that can be reproduced by the loudspeakers. The reproduction of a maximum sound pressure level of 105 dB SPL in the frequency range from 50 Hz (or down to 20 Hz, see clause 6.2.4) to 5 kHz is considered to be sufficient. Due to the typically much lower signal energy between 5 kHz and 20 kHz, the sound pressure level reproduced at such frequencies may be lower. In general it is advisable to select high quality loudspeakers providing flat free-field response characteristics and low distortion at maximum desired sound pressure level.

For certain applications, a sub-woofer may be used in addition. Either as an additional low-frequency support for the eight loudspeakers and/or for the reproduction of noise components of less than 50 Hz (see clause 6.2.4). The upper cut-off frequency of the subwoofer operating range shall be greater or equal 60 Hz to ensure an appropriate transition range between loudspeakers and subwoofer.

NOTE: Since wave lengths of lower frequencies are rather large compared to the distances between the microphones, the position of the subwoofer in the setup is almost arbitrary. The only constraints regarding positioning may be room modes that may have a negative impact on the linearity of the measured impulse responses. In such cases, modifying the positioning between subwoofer and microphone array may lead to an improved equalization result.

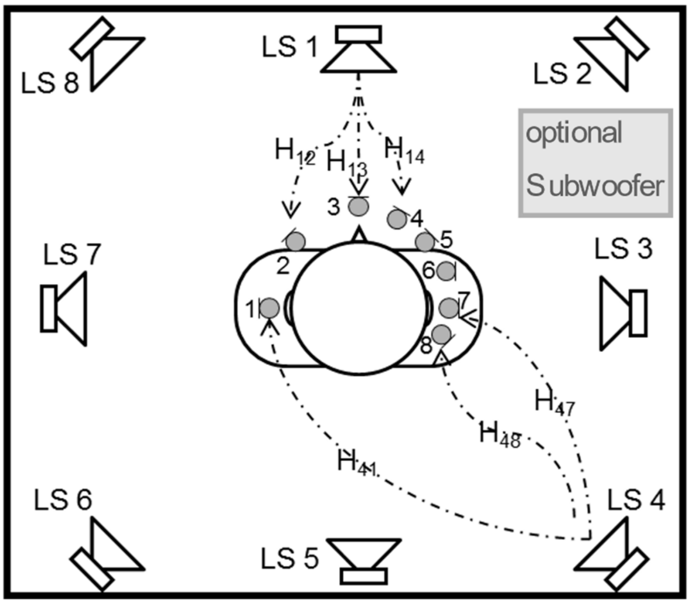

Figure 5: General loudspeaker setup and principle of the equalization paths for the handset and headset measurement setup (with optional subwoofer)

6.1 Test room requirements

The room required by the reproduction technique may vary from acoustically treated office rooms to anechoic rooms. The playback room should meet the following requirements:

- **Room size:**

  The room size should be in a range between 1,8 m × 2,4 m × 2,1 m to 8 m × 9 m × 4,5 m (L × W × H).

- **Room acoustic parameter clarity 80:**

  The most important criterion a room has to fulfil depends on the clarity 80 (C_{80}) [i.11]. This parameter is defined as the signal energy of the first *80 ms* of the Impulse Response (IR), $h(t)$, in relation to the remaining energy of the impulse response expressed in dB:

$$C_{80} = 10log\left(\frac{\int_{t_0}^{t_{80}} h^2(t)dt}{\int_{t_{80}}^{t_{end}} h^2(t)dt}\right) \quad (2)$$

  Where, t_0 is the arrival time of the impulse response direct sound wave, t_{80} is 80 *ms* after the arrival of the direct sound wave, and t_{end} is the effective length of the impulse response. For sound field reproduction systems, $t_{end} = 1\ 000\ ms$ shall be used. Impulse responses shorter than 1 000 *ms* shall be zero-padded to 1 000 *ms* prior to C_{80} calculation.

  For a noise field reproduction setup with l=1..L loudspeakers and i=1..N microphone positions, the system identification (see clause 6.2.2) provides $L \cdot N$ impulse responses, $h_{li}(k)$. For each impulse response, the clarity $C_{80,li}$ can be determined according to equation 2. The average clarity *C80* of the reproduction setup is determined as the arithmetic mean across all $L \cdot N$ *C80* values in dB.

  Suitable reproduction setups shall have an average *C80 > 20 dB*. [i.11]

- **Treatment of the room:**

  Office type rooms should be equipped with a carpet on the floor and some acoustical damping in the ceiling as typically found in office rooms. A curtain should cover one or two walls in order to avoid strong reflections by hard surfaces in the room. Additional damping materials may need to be applied in order to reach the C_{80} value given above.

  For anechoic or semi-anechoic chambers no additional treatment is needed.