



Effect of room acoustic properties and head orientation on practical speech-in-noise measurements for various spatial configurations

Larissa Warkentin^{1,*} , Florian Denk¹ , Alexandra Winkler² , Tobias Sankowsky-Rothe², Matthias Blau² , Inga Holube² , and Hendrik Husstedt¹ 

¹German Institute of Hearing Aids GmbH, Anschützstraße 1, 23562 Lübeck, Germany

²Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Ofener Str. 16/19, 26121 Oldenburg, Germany

Received 17 January 2024, Accepted 17 May 2024

Abstract – Speech recognition scores in noise can be affected by measurement conditions such as the spatial configuration, room acoustic properties, or the position and head orientation of a test person. This is critical when comparing repeated measurements, e.g., during hearing aid verification. While the basic effects of these factors are known, specific effect sizes for practically relevant configurations have not previously been reported. We investigated the effects of audiological test-room acoustics with low reverberation, and of head movements on speech-in-noise tests for typical spatial configurations. Speech-recognition thresholds (SRTs) and head movements were measured in 240 normal-hearing participants in six rooms for five loudspeaker configurations (S_0N_0 , $S_0N_{\pm 45}$, $S_0N_{\pm 90}$, S_0N_{180} and $S_{\pm 45}N_{\mp 45}$). Additionally, head positions were reproduced with an artificial head, and using a binaural speech intelligibility model, binaural room impulse responses were measured to estimate the influence of head movements on the SRT. The results show that the effects of room acoustic properties and head movements varies greatly between spatial configurations. Head rotations around the vertical axis can affect speech recognition scores by up to 8 dB, most critically for the configuration S_0N_{180} . For spatially separated sound sources, different room acoustic conditions caused differences in SRT of up to 5 dB.

Keywords: Speech intelligibility, Head orientation, Room acoustics, Speech-in-noise tests

1 Introduction

The results of speech-in-noise tests depend on multiple factors, including the temporal and spectral properties of the speech and noise signals, the spatial configuration of speech and noise sources [1, 2], the acoustic properties of the test room [3], and the participant's position and head movements [4–6]. Especially room acoustic properties and head movements are factors that are not well controllable in practical applications, and add variations to the data that may reduce the comparability between results obtained at different sites, and even repeated measurements in the same setting. However, comparability of such repeated measurements is assumed in practical applications such as audiological diagnostics and hearing aid verification. For example, demonstrating a hearing-aid benefit regarding speech recognition in noise by repeated measurements is a prerequisite for reimbursement for hearing-aid provision by health-insurance companies in Germany [7] and other countries. Still, the exact measurement conditions are not typically specified or documented, and potential systematic or random variations remain unknown.

While the basic effects of the factors mentioned above on speech intelligibility in noise are known, effect sizes in practically relevant conditions, and possible derived guidelines for procedures, have not previously been presented. The object of this work was therefore to quantify the influence of room acoustic properties and head orientations on Speech Recognition Thresholds (SRTs) in noise as obtained in a set of five common loudspeaker configurations with one speech and one noise source in six different rooms.

It is well known that a spatial separation between speech and noise leads to improved SRTs, most pronounced in the free field and in rooms within the critical distance [3, 8]. Different azimuths of speech and noise sources can lead to differences in phase, level, and SNR at the two ears, which allows the auditory system to separate the sound sources. This release from masking from a stationary noise source can be explained by two effects: better-ear listening and binaural unmasking. Better-ear listening relies on SNR differences and describes the effect that the auditory system primarily processes the information from the ear with the better SNR [9]. To improve signal detection and identification in noise in binaural unmasking, the auditory system extracts interaural time differences. These two effects can lead to an improvement in speech recognition,

*Corresponding author: l.warkentin@dhi-online.de

known as spatial release from masking (SRM) [10]. In anechoic environments, SRM can improve SRTs by up to 12 dB [1, 11].

In addition to different spatial configurations, reverberation and sound reflections that are governed by room size, wall absorption, and furnishings, are superimposed on the direct sound. This influence of room acoustic properties typically impedes auditory cues, and thus speech recognition, both in collocated and spatially separated conditions [12]. Lavendier and Culling [13] showed that the SRT is increased by up to 3 dB when decreasing the absorption coefficient from 1 to 0.2, making the room more reverberant. Duquesnoy and Plomp [14] found an SRT improvement of 10 dB by decreasing the reverberation time from 2.3 to 0 s. Further, Biberger and Ewert [3] demonstrated that rooms with the same reverberation time (T_{60}), but different volumes, could lead to different SRTs, but did not show significant differences in SRM. To the authors' best knowledge, a comprehensive evaluation of the influence of room acoustic properties on different common spatial configurations, especially of rooms with short reverberation times as used for clinical speech in noise tests, is still missing.

Even when room acoustic properties and the spatial configuration of sound sources are kept constant, speech recognition scores can be influenced rather randomly by the participants' head movements. Head movements can be divided into translation and rotation, related to the head position and head orientation. Conscious and unconscious head movements mainly help listeners to localize and distinguish sound sources [15]. It has been repeatedly shown that spontaneous, free head movements typically lead to improved SRTs [4, 6, 16]. For measuring speech recognition in noise, however, test persons are usually instructed to reduce head movements to a minimum [17]. However, there is not much available data on head orientation that can be used to estimate how much test persons move their heads with these kind of instructions. Some data is available for undesired head movements during measurements of Head-Related Transfer Functions (HRTFs) [18–20], showing typical movements of several degrees and centimeters over the course of minutes, in spite of instructions to keep the head still. However, it is unclear how well these data are transferable to speech tests in noise, where listeners have to repeat words typically under the supervision of an experimenter. Furthermore, the influence of head movements on SRT for various spatial configurations has not been sufficiently studied. HRTFs with varying head movements and their perceptual implications have been studied within virtual acoustics [21], but to the authors' best knowledge, the physical and perceptual influence of unconscious head movements on speech-in-noise tests is still missing.

The aim of the present work was to quantify the effects of room acoustic properties and unconscious head movements on measured SRTs in practically relevant settings. For this purpose, we measured SRTs in six different test rooms for five common spatial configurations of speech in noise, using a standardized German monosyllabic speech test [22] in a stationary noise. A total of 240 normal-hearing

listeners participated across two sites. Head movements of the participants during the whole experiment were recorded using a head tracker. In addition, Binaural Room Impulse Responses (BRIRs) of all test configurations were recorded in a KEMAR, including head positions and orientations that resembled those observed in the listeners. Based on these BRIRs, the influence of head movements on SRT in isolation could be assessed using a Binaural Speech Intelligibility Model (BSIM, [23]).

2 Methods

2.1 Setup

Both listening tests and technical measurements were conducted in six different test rooms located in Lübeck (LB) and Oldenburg (OL) in Germany. Two anechoic rooms at both locations served as references, and four audiological test rooms were chosen to represent a selection of realistic test rooms for speech audiometry. A detailed description of all rooms is listed in Table 1. Photos of the test rooms as well as 2D graphics are provided as [Supplementary data](#).

Technical measurements comprised both general room-acoustical quantities according to ISO 3382-2:2008 [24] and BRIRs at the listener's position. Regarding a general setup-agnostic characterization of the room acoustic properties in the respective rooms, the reverberation time T_{20} and clarity index $C50$ were measured for twelve source- and receiver positions, which were chosen according to ISO 3382-2:2008. To measure very low reverberation times, time-reversed filtering of room impulse responses into octave bands, according to Jacobsen [25], was employed. Single-number metrics were derived for T_{20} and $C50$ by averaging the respective values of the 500 Hz and 1 kHz octave bands. Both single-number metrics were averaged over three source- and four receiver combinations, in total twelve combinations (precision method according to ISO 3382-2:2008 [24]).

Regarding the loudspeakers, a setup comprising six GENELEC 8030 CP loudspeakers (Genelec Oy, Isalmi, Finland) as shown in Figure 1 was used in each room. The loudspeakers were arranged at 0° , 45° , 90° , 180° , -90° and -45° , in a horizontal ring with a radius of 1 m and a height of the acoustic axis of 1.3 m above the floor. The experimenters were located at different positions, depending on room size and furniture. In two rooms (OL anechoic, OL room 1), the experimenters sat outside the room, whereas in the other four test rooms they were positioned in front of and on the left or right of the test person (see Fig. 1). Measurements were made for eight configurations of speech and noise sources: S_0N_0 , S_0N_{45} , S_0N_{-45} , S_0N_{90} , S_0N_{-90} , S_0N_{180} , $S_{45}N_{-45}$ and $S_{-45}N_{45}$. For symmetrical loudspeaker configurations ($S_0N_{45/-45}$, $S_0N_{90/-90}$ and $S_{-45/45}N_{45/-45}$), measurements were split between both sides and the results were merged for evaluation, which resulted in five final speaker configurations: S_0N_0 , $S_0N_{\pm 45}$, $S_0N_{\pm 90}$, S_0N_{180} and $S_{\pm 45}N_{\mp 45}$. Measurements were controlled

Table 1. Description of the six test rooms. The reverberation time (T_{20}) and clarity indices for speech (C50) were averaged across the 500 Hz and 1 kHz octave bands. T_{20} and C50 values were averaged over twelve transducer-receiver combinations. For C50 the standard deviation (SD) across different measurement positions is given.

Room	Dimensions ($l \times w \times h$) (m)	Volume (m^3)	T_{20} (s)	C50 (dB)	
				Mean	SD
LB anechoic	$4.3 \times 4.6 \times 2.4$	47.47	<0.05		
OL anechoic	$3.5 \times 3.1 \times 2.1$	77.18	<0.05		
LB room 1	$2.6 \times 3.6 \times 2.5$	23.40	0.10	30.6	2.5
OL room 1	$2.4 \times 2.4 \times 2.2$	12.67	0.11	28.2	1.3
LB room 2	$4.8 \times 3.3 \times 2.4$	38.02	0.24	13.2	1.3
OL room 2	$3.9 \times 4.0 \times 2.5$	39.00	0.11	28.7	2.0

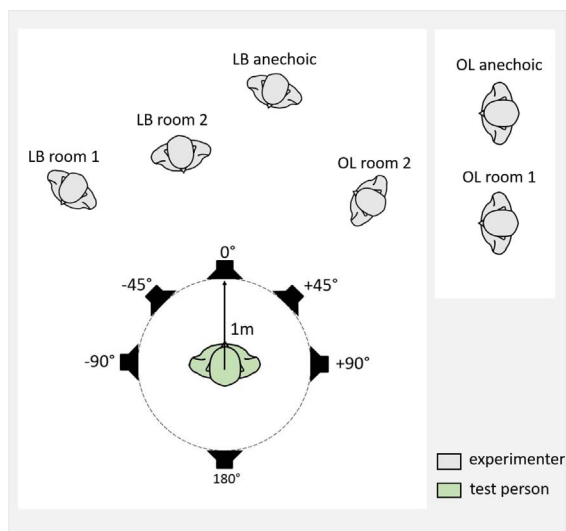


Figure 1. Experimental setup of six loudspeakers arranged in a circle with a radius of 1 m to present various configurations of speech and noise signals. Additionally the participants' and experimenter's positions are shown.

through custom routines implemented in MATLAB R2023a (MathWorks, Massachusetts, USA), and the transducers were connected to the computer using an RME Fireface UC soundcard (Audio AG, Haimhausen, Germany).

The loudspeakers were equalized using a calibrated microphone Type 4192 (Brüel & Kjaer, Nærum, Denmark). To achieve the same sensitivity for all incidence directions, the microphone was placed vertically in the center of the loudspeaker setup when the chair and the test person were absent. Note that for this pressure-field microphone type, the free-field frequency response is within ± 2 dB in the frequency range of interest for sound incidence at 90° , as used here. The frequency response was equalized by applying a minimum-phase filter that was computed by inverting the 1/12 octave smoothed, unequalized frequency response measured using exponential sweeps [26]. Furthermore, to match the capabilities of the loudspeaker, a bandpass between 80 Hz and 18 kHz was enforced.

The participants were seated on a chair in the center of the loudspeaker setup. An identical chair, which had a slightly flexible backrest and no headrest or armrests, was used in all rooms. The chair was not rotatable and had

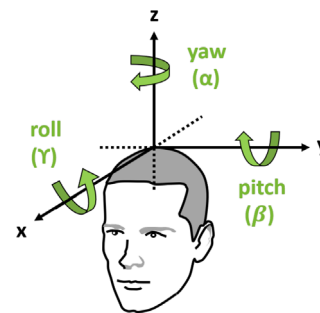


Figure 2. Illustration of three translational (x , y , z) and three rotational head movements (yaw, pitch, roll).

no wheels. The height of the chair was individually adjusted for each participant, such that the ears were at the height of the loudspeakers' acoustic axes at the mid-point of the bass and tweeter drivers. To measure head movements during the test, the participants were equipped with a headtracker PATRIOT (LB rooms) or Fastrack (OL rooms) (Polhemus, Colchester, USA). Three translational movements: front-back (x), right-left (y), up-down (z) and three rotational movements: yaw (α), pitch (β) and roll (γ) were monitored during the speech test (see Fig. 2). When referring to head movements in this article, the term head position is used in the translational plane (x , y , z) and head orientation is used in the rotational plane (yaw, pitch, roll). Head movements were tracked for all test rooms except for LB room 1, because the measurements in LB room 1 were already completed at the time of the decision to track the head movements of the participants.

The participants were instructed to face the loudspeaker at 0° during the test and keep head movements to a minimum. Head movements were tracked relative to a reference position, which was measured for each test person individually each time they were seated and carefully positioned, i.e., once at the beginning, and repeated after mandatory breaks (see below). The head movements in all test rooms were evaluated during test-word presentation to measure head movements while listening and exclude undirected head movements during response times, e.g., head shaking. Figure 3 shows an example of a test person's head movements during the presentation time of one test list comprising 20 words, in which green shaded areas show

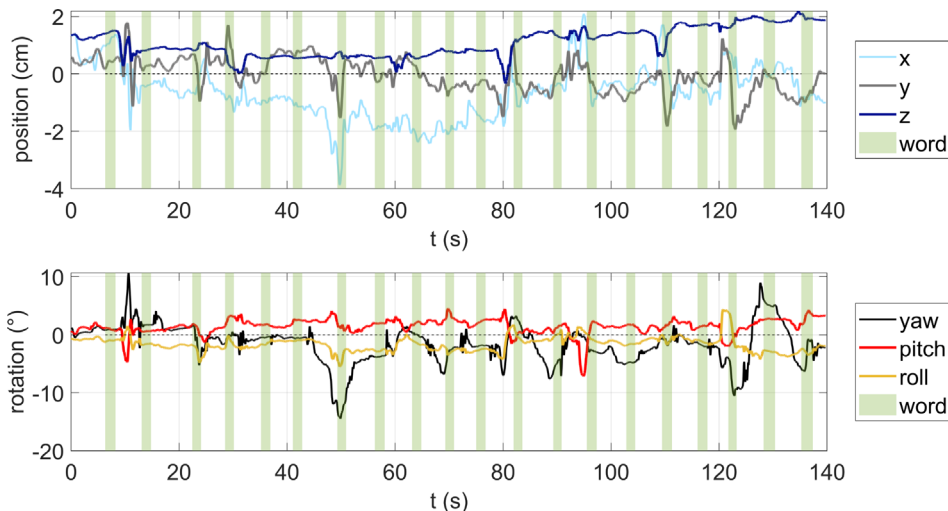


Figure 3. Example of time curves of three translational (x , y , z) and three rotational head movements (yaw, pitch, roll) of one participant during the presentation of one test list showing time periods of word presentation (green bars).

the times of word presentation that were used for further evaluations. The head movements were compared statistically using two-way ANOVA with Bonferroni correction to test the alternative hypothesis that there are significant differences in head movement across different loudspeaker configurations.

2.2 Test persons and speech test

In total, 240 test persons, 40 in each test room, participated in the listening experiment. All participants fulfilled the criteria of ISO 8253-3:2022 [17] for the measurement of speech recognition curves: Otologically normal, German native speakers, aged between 18 and 25 years, pure-tone hearing threshold of 10 dB or less between 250 Hz and 8 kHz and 15 dB for no more than two frequencies in this range. Additionally, the participants had no ear diseases, were not exposed to any noise in the 24 h prior to the measurement, and reported feeling well-rested and able to concentrate.

The speech material of the Freiburg monosyllabic speech test was used as a target stimulus, which is standardized and the most commonly used speech test in Germany [27]. The speech test consists of 20 test lists with 20 monosyllabic nouns each [22]. As common practice for this test, the continuous noise according to Comité Consultatif International Télégraphique et Téléphonique (CCITT) was presented as the masker [28]. The spectra of the speech and noise signals were published by Winkler and Holube [29] and are provided as [Supplementary material](#). The RMS levels of the speech and noise signals were adjusted according to DIN 45626-1 [22] and Winkler and Holube [29] using correction factors for C-weighted equivalent sound level ($\text{corrCeq} = 0.06$ dB) and impulse time weighting ($\text{corrI} = -0.5$ dB) to reproduce the common usage of the Freiburg monosyllabic speech test in clinical applications. The speech material was presented at 60 dB SPL, and noise levels were varied to produce the desired SNR

value. Based on the results of Winkler et al. [30] and pilot studies, four SNR values were chosen for each room and loudspeaker configuration to achieve speech-recognition scores below and above 50% to accurately sample the psychometric function. Speech-recognition scores were measured by presenting all 20 test lists of the Freiburg monosyllabic speech test to each participant. The test lists were assigned to twenty measurement conditions ($4 \text{ SNR} \times 5$ loudspeaker configurations), using Latin square balancing. The loudspeaker configurations $S_0N_{45/-45}$, $S_0N_{90/-90}$ and $S_{-45/45}N_{45/-45}$ were presented to half of the test persons from the positive, and for the other half of the test persons from the negative direction. The resulting SRTs were merged for evaluation and denoted as: $S_0N_{\pm 45}$, $S_0N_{\pm 90}$ and $S_{\pm 45}N_{\mp 45}$. After ten lists, to avoid fatigue effects a mandatory break was taken during which the test persons had the opportunity to leave the test room.

In all LB rooms, a wireless speech-transmission system, consisting of Roger Pen and Roger MyLink (Sonova Holding AG, Stäfa, Swiss) helped the experimenters assess the listener's responses as correct or false. In OL room 2, no speech transmission system was used. Due to limited room sizes in OL anechoic and OL room 1, the examiners were seated outside the rooms. In both rooms, the responses were transmitted using a wireless go microphone (Røde, Silverwater, Australia) inside the room to a 6301B loudspeaker (Fostex, Akishima, Japan) outside the room.

Using the speech-recognition scores in percent correct, speech-recognition curves were calculated for each test person per loudspeaker arrangement by fitting psychometric functions

$$p(\text{SNR}, \text{SRT}, s) = \frac{100}{1 + e^{s(\text{SRT} - \text{SNR})/25}}, \quad (1)$$

where p is the speech-recognition score in percent correct, SNR is the signal-to-noise ratio, SRT is the SNR that yields

a speech-recognition score of 50%, and s is the slope of the psychometric function at SRT.

The SRT of the individual speech-recognition curves were compared statistically. The Shapiro-Wilk test with Bonferroni correction revealed a normal distribution for 23 of the 30 groups. Since two-way Analysis of Variance (ANOVA) is robust against violations of the normality assumption, ANOVA was used with post-hoc t -tests and Bonferroni correction to evaluate differences between all six test rooms for each loudspeaker configuration.

2.3 BRIR measurements and speech intelligibility model

To capture the effect of the participants' head movements in the present setup, BRIR measurements were made for several positions of a head- and torso-simulator (HATS) with anthropometric pinnae type KEMAR 45BC-9 (G.R.A.S., Holte, Denmark). Using the exponential sweep method [26], BRIRs were measured at both ears of the HATS for all loudspeakers. Besides the reference position in the center of the loudspeaker setup, BRIRs were measured for translational displacements of ± 5 cm and ± 10 cm in x -, y - and z -directions and head rotations of $\pm 5^\circ$ and $\pm 10^\circ$ about the vertical axis (yaw). These head positions intentionally exceeded those of the participants recorded during the experiments, since it is assumed that in the current study we observed the lower end of possible head movements seen in practice (c.f. 4.2).

To predict SRT changes as a function of HATS position, the stimuli were convolved with the BRIRs of each loudspeaker, and SRT values were calculated for each loudspeaker configuration using the binaural speech intelligibility model (BSIM) of Beutelmann et al. [23]. In short, BSIM mimics auditory binaural processing and outputs the SII of the resulting, binaurally enhanced signal from a binaural input, requiring separated speech and noise signals. Comparing the SII for various input SNRs then allows a prediction of the SRT difference between two conditions. The effect of head orientations on SRT was modeled by finding the SNR for a certain head orientation that led to the same SII as the reference head orientation, and doing this for each room separately. Efficiency in these calculations was increased by using a speech-simulating noise instead of the full test lists as the speech input [31]. SRTs were linearly interpolated between the measured head orientations.

3 Results

3.1 SRT results

Figure 4 shows the individual SRT of the study as boxplots for each loudspeaker configuration and test room. A comparison of different loudspeaker configurations shows the highest SRT values for the co-located loudspeaker configuration S_0N_0 and the lowest SRTs for $S_{\pm 45}N_{\mp 45}$, with the largest spatial separation in the visual field. When comparing SRTs in different test rooms, the anechoic rooms with the smallest reverberation times yielded the lowest SRT

values. The highest SRT values were obtained in non-anechoic rooms, depending on the loudspeaker configuration. For S_0N_0 , $S_0N_{\pm 90}$, and S_0N_{180} , OL room 1 showed the largest SRT values, and for the other two speaker configurations, LB room 2 revealed the highest SRTs. The differences between the test rooms were smallest for S_0N_0 and largest for $S_{\pm 45}N_{\mp 45}$.

A statistical evaluation of these effects using a two-way ANOVA showed a significant main effect of loudspeaker configuration ($F(4, 1199) = 1286.40$, $p < .001$) and test room ($F(5, 1199) = 169.39$, $p < .001$), as well as an interaction between loudspeaker configuration and test room ($F(20, 1199) = 17.84$, $p < .001$). Differences between the test rooms for each loudspeaker configuration were evaluated using post-hoc t -tests for independent samples with Bonferroni correction. Resulting p -values are given in Table 2.

The post-hoc analysis for t -tests showed no significant differences between test rooms for the loudspeaker configuration S_0N_0 . In this configuration, the smallest differences in median SRT were observed, with a maximum of 0.8 dB SNR between OL anechoic and OL room 1. The highest SRT differences of 5 dB SNR were observed for the loudspeaker configuration $S_{\pm 45}N_{\mp 45}$ between LB anechoic and LB room 2, which are the test rooms with the largest differences in reverberation times and clarity indices. Significant SRT differences were observed between all rooms except for the anechoic rooms, and for OL rooms 1 and 2 with similar reverberation times. For $S_0N_{\pm 45}$ and $S_0N_{\pm 90}$, the anechoic rooms and LB room 1 showed significant differences to all other rooms. For these loudspeaker configurations, no significant differences were seen between the anechoic rooms (LB anechoic and OL anechoic), and also between the test rooms OL room 1, LB room 2 and OL room 2. The loudspeaker configuration S_0N_{180} showed fewer significant SRT differences between different test rooms than the other spatially separated loudspeaker configurations, but ANOVA revealed a significant difference between the SRTs in the anechoic rooms ($p = 0.025$), with an SRT difference of 0.8 dB SNR.

3.2 Head-tracker measurements

To evaluate the influence of test room and experimenter position on the participants' head movements, the mean values and standard deviations in x (front-back), y (left-right), z (up-down), yaw (α), pitch (β) and roll (γ) direction are shown per test room in Figure 5. A statistical comparison of head positions and orientations in different rooms using pairwise t -tests for each movement dimension revealed that 56 of 60 of p -values (93.3%) indicated statistically significant differences due to the large amount of data per room. However, the main purpose here is to consider and evaluate the practical relevance of head-movement data by examining whether the participants oriented their heads in the direction of the experimenter. Therefore, yaw orientations are of particular interest, which is why the experimenter positions in yaw dimension are visualized with black markers (crosses) in Figure 5. Neither the yaw dimension nor the other dimensions showed an

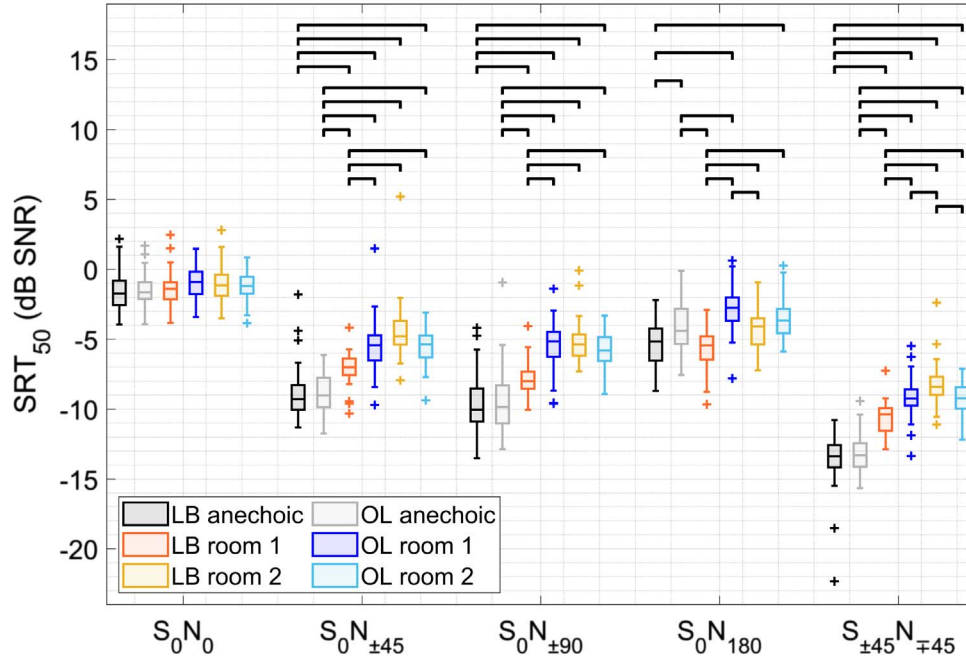


Figure 4. Boxplots of all individual SRT (dB SNR) measured in six test rooms per loudspeaker configuration. The boxplots show the medians (middle quartile), the inter-quartile range (IQR) between the first and third quartile (box length), the whiskers (1.5 times the IQR) and outliers (+). Significance bars indicate significant post-hoc differences between test rooms with Bonferroni correction. Corresponding p -values are listed in [Table 2](#).

Table 2. Results (p -values) of post-hoc t -tests with Bonferroni correction for a comparison of six test rooms for each loudspeaker configuration. Non-significant effects are shown in gray.

		S_0N_0	$S_0N_{\pm 45}$	$S_0N_{\pm 90}$	S_0N_{180}	$S_{\pm 45}N_{\mp 45}$
LB anechoic	OL anechoic	1	1	1	.025	1
LB anechoic	LB room 1	1	<.001	<.001	1	<.001
LB anechoic	OL room 1	.257	<.001	<.001	<.001	<.001
LB anechoic	LB room 2	.578	<.001	<.001	.111	<.001
LB anechoic	OL room 2	1	<.001	<.001	<.001	<.001
OL anechoic	LB room 1	1	<.001	.002	<.001	<.001
OL anechoic	OL room 1	.536	<.001	<.001	.004	<.001
OL anechoic	LB room 2	1	<.001	<.001	1	<.001
OL anechoic	OL room 2	1	<.001	<.001	1	<.001
LB room 1	OL room 1	1	<.001	<.001	<.001	<.001
LB room 1	LB room 2	1	<.001	<.001	.003	<.001
LB room 1	OL room 2	1	<.001	<.001	<.001	<.001
OL room 1	LB room 2	1	.105	1	<.001	.042
OL room 1	OL room 2	1	1	1	.435	1
LB room 2	OL room 2	1	.060	1	.755	.013

orientation of the participants' heads towards the experimenter. Yaw orientations on average ranged from $\alpha = -1.4^\circ$ in LB anechoic, with an experimenter position at $\alpha = 10^\circ$, to $\alpha = 0.5^\circ$ in OL anechoic, where the experimenter sat outside the test room. This indicates that on average, the test persons' head orientations did not depend on the test room and the position of the experimenter. Therefore in the further evaluations, the head orientations of all test rooms were combined.

To evaluate whether the participants turned their heads depending on the directions of speech and noise signal,

[Figure 6](#) shows the mean values and standard deviations of the participants' head positions and orientations per loudspeaker configuration.

A comparison of mean values showed small differences of up to 0.17 cm (x) and 0.16° (pitch) between different loudspeaker configurations. A statistical evaluation with two-way ANOVA showed no significant main effect of loudspeaker configuration ($F(7,4320) = 0.516$, $p = 0.823$), but a significant influence of the axis of movement ($F(2,4320) = 108.15$, $p < .001$). Loudspeaker configuration and translation direction showed no interaction ($F(14,4320) = 0.28$,

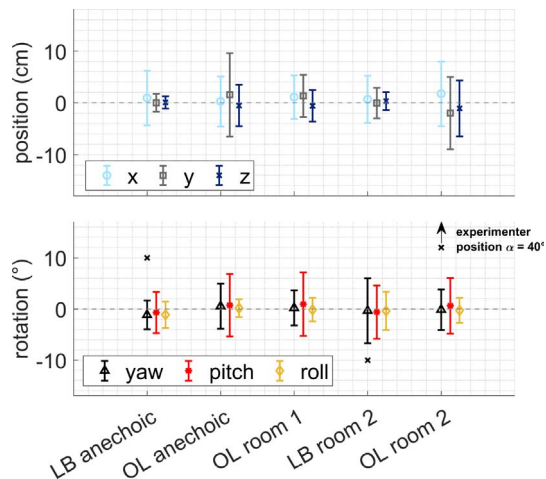


Figure 5. Mean values and standard deviation of the participants' head positions in x (front-back), y (left-right), z (up-down) direction and head orientations in yaw, pitch and roll direction per test room. Additionally, the experimenter positions in the yaw dimension are shown (black crosses). For OL anechoic and OL room 1, no yaw orientations are shown, because the experimenters sat outside the room. In LB room 1, no head movement data was measured.

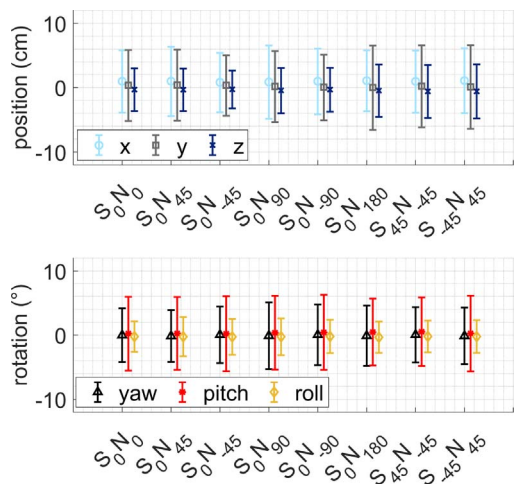


Figure 6. Mean values and standard deviation of the participants' head positions in x (front-back), y (left-right), z (up-down) direction, and head orientations in yaw, pitch and roll directions per loudspeaker configuration.

$p = 0.996$). Head movements in the x -direction significantly differed from movements in the y - and z -directions ($p < .001$). For rotational orientations, ANOVA also revealed no significant influence of loudspeaker configuration ($F(7,4320) = 0.26$, $p = 0.969$), but a significant effect of the axis of rotation ($F(2,4320) = 11.87$, $p < .001$). Post-hoc analysis for rotation direction showed that pitch orientations significantly differed from roll orientations ($p < .001$). Rotation direction and loudspeaker configuration showed no interaction ($F(14,4320) = 0.50$, $p = 0.933$).

To evaluate head movements over time, Figure 7 shows mean values of head movements in each room from the first

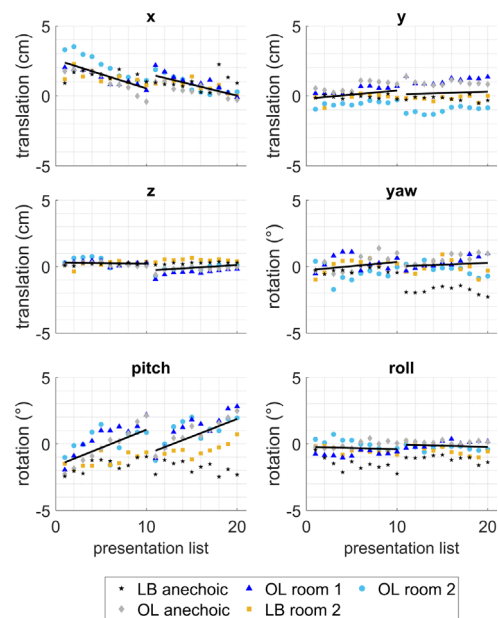


Figure 7. Scatter plot showing the mean values of the participants' head movements during test list 1-20 per test room. The subplots show head positions in x (front-back), y (left-right), z (up-down) directions, and head orientations in yaw, pitch, and roll directions. Additionally, linear regression lines for the first and second half of the test lists are shown.

to the last test list presented. To visualize drifts in head orientation over time, linear regression lines were fitted separately to the first 10 test lists before the break, and to the second 10 test lists after the break (during which the participants left the room). A comparison of all translational and rotational head movements demonstrated that the participants moved their heads about $\Delta x = -1.5$ cm back and $\Delta \beta = 2.5^\circ$ up before and after the break. Head positions for y and z and head orientations for yaw and roll show no drift in any particular direction over time. The largest deviations of mean head movements between all rooms can be seen for pitch orientations ranging from $\beta = -2.5^\circ$ to $\beta = 2.8^\circ$. The largest differences in translational head positions between the test rooms can be seen for the z -direction, with head positions between $x = -0.4$ cm and $x = 3.5$ cm.

3.3 Effect of head orientations on SRT

Figure 8 shows relative SRT changes calculated using BSIM for each loudspeaker configuration, and resulting from displacements of the HATS from the reference position. The SRTs were averaged across all test rooms. A comparison of all head movements shows that only yaw orientations have a notable effect on SRT in the S_0N_{180} condition. In that loudspeaker configuration, yaw orientations caused SRT improvements of up to 3.5 dB SNR in both directions. In the translational degrees of freedom, HATS displacements of ± 10 cm caused SRT changes of up to ± 0.8 dB SNR for the loudspeaker configurations $S_0N_{\pm 90}$ (y) and $S_{\pm 45}N_{\mp 45}$ (x, z), which is in the range of prediction error. To visualize the occurrence of these head

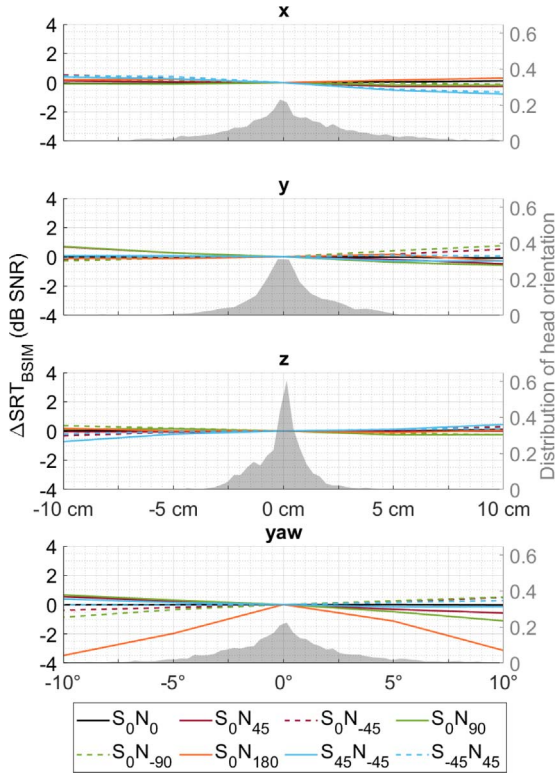


Figure 8. Relative SRT changes due to HATS displacements from the reference position, calculated with BSIM ($\Delta\text{SRT}_{\text{BSIM}}$) in dB SNR and averaged across all test rooms. The subfigures show HATS positions in x (front-back), y (left-right) and z (up-down) directions to ± 10 cm, and head rotations around the vertical axis (yaw) to $\pm 10^\circ$. Additionally, the distribution of head movements across all participants and test words is shown in gray as a histogram.

movements during speech-in-noise tests, the distribution of the participants' head movements is shown per movement dimension. All head movements had their peaks around $0^\circ/\text{cm}$, with no impact on SRT. Movements in x -, y -, and yaw-dimensions showed a larger spread of head positions and orientations, with standard deviations of about $\pm 5^\circ/\text{cm}$, whereas z -positions showed the smallest standard deviation of 3.7 cm.

To evaluate the effects of the participants' head movements on SRT for different loudspeaker configurations and test rooms, BSIM modeled SRTs are shown relative to the SRT at the reference position (see Fig. 9). For this evaluation, head movements per test word were pooled across all participants in all rooms to obtain one large data set of head movements measured in the listening experiment. These head movements in x -, y -, z - and yaw-dimensions were then used to predict the effect on SRT by associating them with the HATS measurements shown in Figure 8. Each head orientation was assigned to the corresponding interpolated $\Delta\text{SRT}_{\text{BSIM}}$ value obtained from a HATS displacement in each test room per speaker configuration. In Figure 9, mean values and the 95% confidence intervals across all SRT values are shown for each loudspeaker configuration and test room.

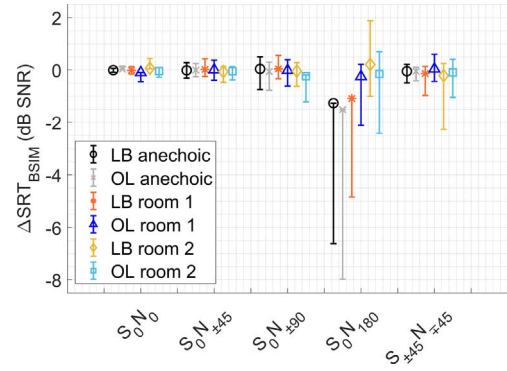


Figure 9. BSIM modeled SRT changes ($\Delta\text{SRT}_{\text{BSIM}}$) calculated from the participants' head movements per word for each loudspeaker configuration and test room. The symbols represent relative SRTs resulting from mean head orientations and the whiskers represent the 95% confidence intervals.

A comparison of the mean values and the 95% confidence interval of each loudspeaker configuration shows that the effects of head movements on SRT are largest for S_0N_{180} . The participants' head movements caused a mean ΔSRT of between -1.5 dB SNR in OL anechoic with the shortest reverberation time, and 0.2 dB SNR in LB room 2 with the longest reverberation time. Considering the 95% confidence interval, head movements caused ΔSRT between -8.0 dB SNR (OL anechoic) and 1.9 dB SNR (LB room 2). The second-largest effect of head movements on SRT was seen for $S_{\pm 45}N_{\mp 45}$, with ΔSRT s ranging from -2.3 dB SNR (LB room 2) to 0.6 dB SNR (OL room 1). S_0N_0 showed the smallest effect of head movements on SRT, with deviations of ± 0.4 dB SNR.

4 Discussion

4.1 SRTs in different test rooms

An evaluation of the participants' SRTs showed the highest values for S_0N_0 , and no significant differences between different test rooms, since the listeners could not make use of binaural information to separate speech and noise. For the loudspeaker configuration $S_{\pm 45}N_{\mp 45}$, the lowest SRTs and the largest differences between test rooms were observed. Rooms with different room-acoustical conditions showed significant SRT differences of up to 5 dB SNR for spatially separated sound sources, which confirms the findings of Lavendier and Culling [13] and leads to the conclusion that speech-recognition scores in different rooms should not be directly compared when speech and noise sources are spatially separated. Hence, the room used should be documented in clinical settings. According to the findings of Bradley et al. [12], Lavendier and Culling [13], and Duquesnoy and Plomp [13], better room acoustical conditions (reverberation time, absorption coefficient, clarity index) lead to better SRTs. This cannot be fully confirmed from the SRTs resulting from this study, because LB room 1 revealed better SRT values than OL rooms 1 and 2, despite similar T_{20} values (c.f. Tab. 1). The clarity index

explains the obtained SRT differences better, because LB room 1 with better SRT values had a larger C50 than OL rooms 1 and 2. Nonetheless, LB room 2 yielded similar, or in other spatial configurations, even lower SRTs than the OL rooms, but has a smaller C50. In conclusion, simple room acoustic parameters are apparently not sufficient to explain the room influence on SRTs in the present study, which included small rooms with low reverberation times as employed in standard audiological practice. This is because room acoustic parameters such as C50 are derived through quantitative calculations that consider early and late reflections, but not binaural speech perception cues.

Contrary to expectations, the loudspeaker configuration S_0N_{180} showed significant differences between the two anechoic rooms, which could have several explanations. One reason for the SRT difference of 0.8 dB SNR may be a misalignment between the loudspeaker setups in the anechoic rooms. Considering the SRT calculations with BSIM in Figure 4, an offset of 2° between the loudspeaker setups can cause an SRT difference of 0.8 dB SNR, which may have occurred despite best efforts to accurately align the loudspeakers. Another factor influencing SRT for S_0N_{180} that needs to be considered is yaw orientations. HATS measurements showed that yaw orientations of about 2° in both directions can cause SRT improvements of 0.8 dB SNR. However, an evaluation of the effect of the participants' head movements on BSIM SRTs did not appropriately confirm the SRTs resulting from the speech-in-noise test. The BSIM calculations in OL anechoic revealed a better mean SRT of about 0.3 dB SNR than LB anechoic, whereas LB anechoic showed better SRTs. Nevertheless, the BSIM calculations showed that yaw orientations can significantly improve speech recognition in noise for S_0N_{180} . Hence, this needs to be considered when explaining SRT differences in rooms with similar room acoustic properties. However, the reason for the significant SRT difference was not finally identified.

4.2 Head movements during speech-in-noise tests

Despite many statistically significant differences between head movements in different test rooms, a comparison of mean head movements with the position of the experimenters showed no evidence that the presence and location of an experimenter influenced the measured head positions during word presentation. Apparently, the instruction was sufficient to avoid having the subjects look towards the experimenter, at least during periods of word presentation that are relevant for the speech test results.

Compared to Grange and Culling [6], the participants showed small translational and rotational head movements, which did not significantly depend on the speech- and noise-source directions. There are several explanations why small movements were detected and loudspeaker-configuration-dependent effects are not visible here. One reason is that the listeners were instructed to orient their head towards the loudspeaker at 0° during the whole test. The movements reported in the study of Grange and Culling [6] were larger, presumably because the listeners

were not instructed to keep their head straight forward, and undirected head movements were evaluated. Another possible explanation is that head movements in this study were evaluated during word presentation only, and were averaged over time and across test persons. Some individual participants may have shown substantially larger head movements, depending on the loudspeaker configuration. We assume that the head movements observed here are on the lower end of what is plausible in practice, since the participants were young and carefully instructed, and aware that their head position was recorded. Additionally, the chair used in this study may have caused smaller head movements, because it was not rotatable and had no wheels.

During the speech test, participants moved their heads slightly back (x) and up (pitch) over time, which indicates that they leaned back, and simultaneously tilted the head upwards. Pitch orientations additionally revealed the largest differences between the test rooms. Here the participants in Oldenburg had a greater pitch than the participants in Lübeck. It is likely that the presence of the experimenter in the Lübeck, but not the Oldenburg, rooms drove the participants to maintain a more constant posture during the experimental course. Another possible explanation could be that different sensor mountings on the participants' heads may have led to a slipping of the transmitter and caused these differences in LB and OL.

4.3 SRT benefit from head orientations

An evaluation of SRT changes due to HATS displacements showed that, compared to translational head positions, yaw orientations produced the larger SRT differences, even in anechoic environments. Especially for S_0N_{180} , the SRT is improved by yaw orientations to the left as well as to the right side, which confirms Grange and Culling's model predictions that the head-orientation benefit from yaw movements is largest for S_0N_{180} and similar for movements to both sides [6]. As already described by Stecker and Gallun [16], rotational head movements in both directions can modify interaural time- and level differences for the signals coming from 0° and 180° , which leads to a spatial release from masking and improves SRT with any deviation from the frontal head orientation. In contrast to head rotations (yaw), translational HATS positions showed small effects on SRT of only up to 0.8 dB SNR. For the conditions examined, changes of interaural cues due to head rotations were larger than any effect of level differences due to translational movements. Here, the loudspeaker configurations $S_{\pm 45}N_{\mp 45}$ and $S_0N_{\pm 90}$ were most affected by translational movements, since the speech and noise source had the largest spatial separation, and small movements towards one loudspeaker and away from the other could affect the SRT. Despite static displacements, head movements in daily practice occur in combination (e.g., movement to front-left), which is expected to induce larger SRT changes.

The SRT differences that were predicted from HATS measurements and the participants' head positions and orientations resulting from head tracker measurements

showed similar effects. These predictions showed the largest effect of head movements on SRT for the loudspeaker configuration S_0N_{180} , with improvements of up to 8.0 dB SNR. This again can be explained by the findings of Stecker and Gallun [16] that yaw-orientations resolve front-back confusions and improve the segregation of speech and noise sources. As described in section 4.1, these improvements due to head rotations could be an explanation for the differences in SRTs between similar rooms at S_0N_{180} . The predicted SRT improvements due to head movements in this study are smaller than the findings of Grange and Culling [6], but can be explained by differences in methodical procedures, such as in the instructions given to participants. For the other spatially separated loudspeaker configurations ($S_0N_{\pm 45}$, $S_0N_{\pm 90}$ and $S_{\pm 45}N_{\mp 45}$), head movements had less impact on SRT. This is due to the fact that yaw-orientations have a reduced influence on SRT for these specific speaker configurations. As expected, the smallest effects of head movements on SRT were seen for S_0N_0 , because head movements do not differently modify the perception of speech and noise. Thus, when making sequential speech recognition measurements in which room acoustic properties or head movements cannot be controlled, S_0N_0 is a good choice. No clear room-dependent effects of head movements on SRT were visible, except for head rotations at S_0N_{180} , where the improvements of SRT in the anechoic rooms were larger than in the audiology test rooms. This confirms the findings of Lavandier and Culling [13], that speech recognition can be decreased by reverberation, due to degraded interaural coherence.

5 Conclusion

The present study showed that room acoustic properties, loudspeaker configuration, and head movements have a significant effect on speech-recognition scores that were obtained in the Freiburg monosyllabic speech test in noise. This indicates that the results of speech-in-noise tests measured in different rooms cannot be directly compared. Individual room acoustic properties, even for rooms with low reverberation, led to significant changes in SRT, e.g., up to 5.0 dB SNR in our experiments. Hence, simple room acoustic parameters are not sufficient to explain the room influence on SRTs in the present study. Furthermore, not all loudspeaker configurations were equally affected by room acoustic properties. The loudspeaker configuration $S_{\pm 45}N_{\mp 45}$ was the most sensitive to room acoustical differences. In addition, head movements did not have the same influence on SRT for all loudspeaker configurations. Yaw movements could lead to SRT improvements for the loudspeaker configuration S_0N_{180} . For the other loudspeaker configurations, head movements, and thus the effect on SRT, could be sufficiently controlled by giving precise instructions. The loudspeaker setup should be chosen carefully for the specific application. While S_0N_0 is least influenced by room acoustic properties or head movements, S_0N_{180} is very likely to be influenced by uncontrolled head movements.

Acknowledgments

The authors thank Patricia Fürstenberg, Patrick Scheumer, Kim Rullmann, Alina-Sophie Bockelmann, and Lauren Moschner for their assistance with data collection, and the motivated listeners for their participation. English language services were provided by stels-ol.de.

Funding

This research was funded by the German Federal Ministry of Economic Affairs and Climate Action [03TN0035B].

Conflict of interest

The authors declared no conflict of interests.

Data availability statement

Data are available on request from the authors.

Supplementary material

A supplementary material of this article is available at <https://acta-acustica.edpsciences.org/10.1051/aacus/2024018/olm>.

This article provides supplementary material (photos and 2D graphics) for the visualization of the six test rooms showing the sizes and positions of sound-absorbing surfaces.

Figure S1: 2D graphic and photo of the experimental setup in the anechoic room in Lübeck (LB anechoic).

Figure S2: 2D graphic and photo of the experimental setup in an audiometric test room in Lübeck (LB room 1).

Figure S3: 2D graphic and photo of the experimental setup in an audiometric test room in Lübeck (LB room 2).

Figure S4: 2D graphic and photo of the experimental setup in the anechoic room in Oldenburg (OL anechoic).

Figure S5: 2D graphic and photo of the experimental setup in an audiometric test room in Oldenburg (OL room 1).

Figure S6: 2D graphic and photo of the experimental setup in an audiometric test room in Oldenburg (OL room 2).

Figure S7: Third-octave spectra for CCITT noise (blue) and Freiburg monosyllabic speech test (FMST) (red).

References

1. R. Beutelmann, T. Brand: Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America* 120 (2006) 331–342.
2. A.W. Bronkhorst: The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica* 86 (2000) 117–128.
3. T. Biberger, S.D. Ewert: The effect of room acoustical parameters on speech reception thresholds and spatial release from masking. *Journal of the Acoustical Society of America* 146 (2019) 2188–2200.
4. R. Viveros, J. Fels: The benefit of head movements of normal listeners in a dynamic speech-in-noise task with virtual acoustics. *Fortschritte der Akustik – DAGA 2017*: 43. Deutsche Jahrestagung für Akustik: 06.-09. März 2017 in Kiel (2017) 1146–1149.
5. A.W. Bronkhorst, R. Plomp: The effect of head-induced interaural time and level differences on speech intelligibility in noise. *Journal of the Acoustical Society of America* 83 (1988) 1508–1516.

6. J.A. Grange, J.F. Culling: The benefit of head orientation to speech intelligibility in noise. *Journal of the Acoustical Society of America* 139 (2016) 703–712.
7. Hilfsmittel-Richtlinie des Gemeinsamen Bundesausschusses über die Verordnung von Hilfsmitteln in der vertragsärztlichen Versorgung (2020).
8. M. Rychtarikova, T. Van den Bogaert, G. Vermeir, J. Wouters: Perceptual validation of virtual room acoustics: Sound localisation and speech understanding. *Applied Acoustics* 72 (2011) 196–204.
9. O. Kokabi, F. Brinkmann, S. Weinzierl: Prediction of speech intelligibility using pseudo-binaural room impulse responses. *Journal of the Acoustical Society of America* 145 (2019) 329–333.
10. J.F. Culling, M. Levandier: *Binaural unmasking and spatial release from masking*. Springer International Publishing, 2021.
11. M.L. Hawley, R.Y. Litovsky, J.F. Culling: The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America* 115 (2004) 833–843.
12. J.S. Bradley, R.D. Reich, S.G. Norcross: On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. *Journal of the Acoustical Society of America* 106 (1999) 1820–1828.
13. M. Lavendier, J.F. Culling: Speech segregation in rooms: Effects of reverberation on both target and interferer. *Journal of the Acoustical Society of America* 122 (2007) 1713–1723.
14. A.J. Duquesnoy, R. Plomp: Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis. *Journal of the Acoustical Society of America* 68 (1980) 537–544.
15. F.L. Wightman, D.J. Kistler: Resolution of front–back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America* 105 (1999) 2841–2853.
16. G.C. Stecker, F. Gallun: *Binaural hearing, sound localization, and spatial hearing*. Plural Publishing, 2012.
17. ISO 8253-3:2022. *Acoustics – Audiometric test methods – Part 3: Speech audiometry*, 2022.
18. T. Hirahara, H. Sagara, I. Toshima, M. Otani: Head movement during head-related transfer function measurements. *Acoustical Science and Technology* 31 (2010) 165–171.
19. F. Denk, J. Heeren, S.D. Ewert, B. Kollmeier, S.M.A. Ernst: Controlling the head position during individual HRTF measurements and its effect on accuracy. *Fortschritte der Akustik – DAGA 2017 : 43. Deutsche Jahrestagung für Akustik: 06.-09. März 2017 in Kiel* (2017) 1085–1088.
20. J.G. Richter: *Fast measurement of individual head-related transfer functions*. RWTH Aachen University, PhD Thesis, 2019.
21. F. Brinkmann, R. Roden, A. Lindau, S. Weinzierl: Audibility and interpolation of head-above-torso orientation in binaural technology. *IEEE Journal of Selected Topics in Signal Processing* 9 (2015) 931–942.
22. DIN 45626-1. *Tonträger mit Sprache für Gehörprüfung*, 1995.
23. R. Beutelmann, T. Brand, B. Kollmeier: Revision, extension, and evaluation of a binaural speech intelligibility model. *Journal of the Acoustical Society of America* 127 (2010) 2479–2497.
24. ISO 3382-2:2008. *Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms*, 2008.
25. F. Jacobsen, J.H. Rindel: Time reversed decay measurements. *Journal of Sound and Vibration* 117 (1987) 187–190.
26. S. Müller, P. Massarani: Transfer-function measurement with sweeps. *Journal of the Audio Engineering Society* 49 (2001) 394–431.
27. K.-H. Hahlbrock: *Speech audiometry and new word-tests*. *European Archives of Oto-Rhino-Laryngology and Head Neck* 162 (1953) 394–431.
28. International Telecommunication Union. *Conventional telephone signal*. ITU-T Recommendation G.227, 1993.
29. A. Winkler, I. Holube: Freiburg monosyllabic speech test and EN ISO 8253-3: Technical analysis. *Zeitschrift für Audiologie* 55 (2016) 106–113.
30. A. Winkler, I. Holube, H. Husstedt: *The Freiburg monosyllabic speech test in noise*. *HNO* 68 (2020) 14–24.
31. C.F. Hauth, S.C. Berning, B. Kollmeier, T. Brand: Modeling binaural unmasking of speech using a blind binaural processing stage. *Trends in Hearing* 24 (2020) 1–16.

Cite this article as: Warkentin L. Denk F. Winkler A. Sankowsky-Rothe T. Blau M, et al. 2024. Effect of room acoustic properties and head orientation on practical speech-in-noise measurements for various spatial configurations. *Acta Acustica*, 8, 24.