

# Objective validation of alternative speech test sources for person-to-person speech intelligibility assessments

Building Acoustics

1–18

© The Author(s) 2024



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/1351010X241281539

[journals.sagepub.com/home/bua](https://journals.sagepub.com/home/bua)**Luis Gomez-Agustina**  and **Rusdy Hartungi**

## Abstract

A sound source that mimics the acoustics characteristics of a human talker is required in objective assessments of speech intelligibility in person-to-person speech communications (PP) scenarios. The few special speech test sources available that meet the strict specifications of the relevant standards are costly and can be inaccessible to practitioners. This can prevent investigations or can lead to non-special and more affordable loudspeakers to be used as alternative sources without knowledge of their suitability. This study evaluates the validity of non-special loudspeakers in PP assessments and supplements a previous preliminary examination by consolidating and expanding early findings. Speech Transmission Index Public Address (STIPA) assessments were conducted employing three representative non-special loudspeakers and one standardized reference speech test source under laboratory conditions. Absolute errors of STIPA measurements were analysed and evaluated against two just perceptible difference (JND) threshold levels. The mean absolute errors obtained for all non-special loudspeakers for almost all scenarios were within one JND of any of the two JND levels. The striking and consistent low level of error found validated the proposition that non-special and affordable loudspeakers may be used as suitable speech sources in pilot or survey-grade speech intelligibility assessments in place of a special standardized speech source when that source is not available, or in full investigations when the relevant standard to adhere to allows the use of those alternative speech sources.

## Keywords

Person-to-person speech communication, human talker, speech mouth simulator, speech intelligibility, direct speech communication, STIPA, speech test source

## Introduction

### Background

A satisfactory level of speech transmission quality is needed in many types of indoor spaces where speech intelligibility is of critical importance to support the spaces' main functions. Examples of these spaces include metro stations, stadia, airports lounges, train ticket halls, assembly halls,

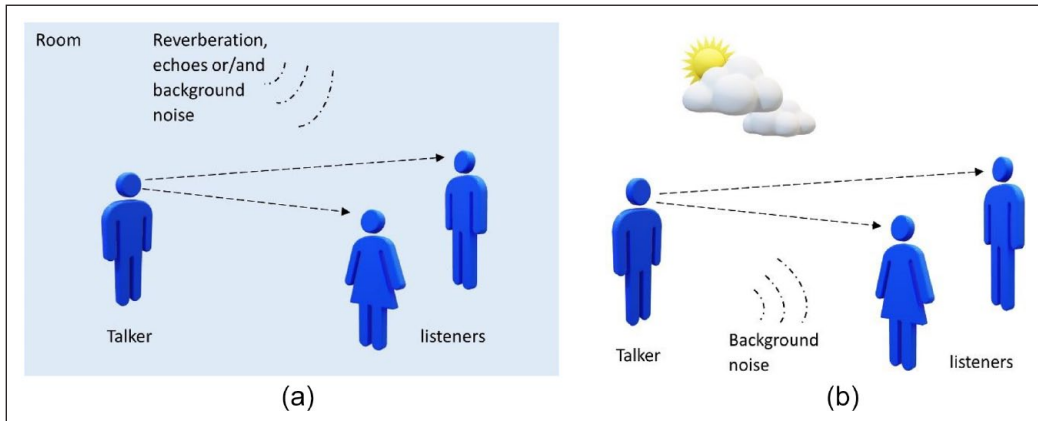
---

School of Built Environment and Architecture, London South Bank University, London, UK

### Corresponding author:

Luis Gomez-Agustina, London South Bank University, 103 Borough Road, London SE1 0AA, UK.

Email: [gomezagl@lsbu.ac.uk](mailto:gomezagl@lsbu.ac.uk)



**Figure 1.** (a) PP system in an indoor space or room. (b) PP system in an outdoor space.

foyers, classrooms, lecture theatres and workshops. Relevant research in the literature can be found dedicated to the improvement of speech transmission quality and intelligibility in those spaces.<sup>1–6</sup> Similarly, several international standards and codes of practice in the literature<sup>7–12</sup> offer their attention to the same topic in those venues. They provide guidance on test procedures and performance criteria for speech intelligibility in a range of buildings where Public Address (PA) or Voice Alarm systems (VA) are employed to reinforce and distribute voice messages.

In indoor spaces of smaller size, speech transmission quality and the resultant intelligibility are also crucial factors to enable the attainment of the intended purposes of specialist rooms. In these spaces, distances between human talkers and listeners tend to be of short ranges, hence amplified Speech Reinforcement Systems (SRS) are usually deemed unnecessary. Instead, verbal communications rely on Person-to-Person Speech Communications (PP) (also called direct communication or Natural Acoustics Communication (NAS)).

A PP system is comprised of the human talker (speech sound source), the room (transmission channel), and the listeners (receivers) and is characterized by the source and receiver being in the same environment in the absence of electroacoustic speech reinforcement devices such as microphones, amplifiers, or loudspeakers.<sup>11</sup> A PP system can also be employed in outdoor verbal communication scenarios in which the open space is the transmission channel.

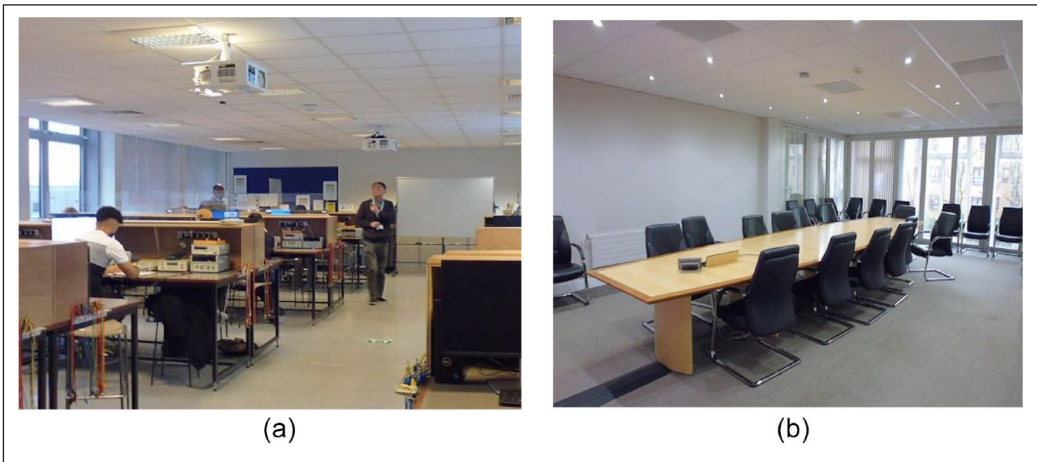
Figure 1(a) shows a simple model of human Person-to-Person(s) Speech Communication system (PP) in a room potentially affected by unfavourable factors in the transmission channel such as the presence of reverberation, echoes or/and background noise. Figure 1(b) shows the same PP model in an outdoor space potentially affected by the presence of background noise.

Table 1 shows the three components forming a PP communication system (column categories) and their corresponding main factor potentially affecting the speech intelligibility of the system. For the purposes of this investigation all the factors affecting the transmission channel are considered, as well as the directivity and vocal effort of the speech source and orientation of the receiver. Factors in Table 1 denoted by a dash (-) are not taken into account.

The PP system is widely employed in a variety of spaces of small to moderate sizes where speech transmission quality and the resulting speech intelligibility is of critical importance.<sup>7,11</sup> Classrooms, offices, disaster/air traffic control rooms, operation theatres, court rooms, laboratories, interview rooms, auditoria, theatres, and meeting rooms are examples of indoor specialist spaces where PP system are normally used (see examples of PP indoor applications in Figure 2(a)

**Table 1.** Person to person system components with their corresponding main factors.

	Person to person system components		
	Human talker (speech sound source)	Environment/room (transmission channel)	Listener ( receiver)
<b>Speech intelligibility main factors</b>	Gender, age, enunciation ability	Background Noise	Gender, age, hearing sensitivity
	Native/non-native	Reverberation	Language proficiency (native/non-native)
	Speech content complexity	Echoes	Self-hearing impairment
	Speech directivity	Talker–receiver distance	Orientation
	Self-hearing impairment	-	-
	Vocal effort	-	-

**Figure 2.** (a) Teaching and learning in a laboratory and (b) meeting/board room.

and (b)). Addressing a group of occupants outside a building in an evacuation, conveying information during a tour around outside facilities, outdoor theatre performances or outdoor soundscapes assessments are examples where PP is employed outdoors.

The research literature shows that in certain situations the use of SRS system in place of PP does not necessarily increase the estimated speech intelligibility score in rooms<sup>13</sup> or improve learning attainment in classrooms.<sup>14</sup> A theatrical performance is a clear example where PP is preferred over SRS systems.<sup>15</sup> Some standards<sup>11</sup> and guidance<sup>16</sup> recommend for certain applications the prioritization of the use of PP over SRS systems.

The topic of speech intelligibility in specialist rooms where PP is employed is considered to different degrees in relevant international/national standards and codes of practice<sup>7,10,16–22</sup> where some guidance and performance criteria is provided.

The assessment of the potential speech intelligibility and/or speech privacy in specialist room is crucial in the determination of their suitability for their intended purposes. The Speech Transmission Index Public Address (STIPA) is a globally accepted and standardized method<sup>7</sup> to objectively determine the potential speech intelligibility in PP applications. The STIPA metric is a subset

version of the parent full Speech Transmission Index (STI) method.<sup>7</sup> Both the full STI and subset STIPA method rate the estimated speech intelligibility of the transmission channel between 0 and 1, where '0' corresponds to total unintelligibility and '1' to a maximum or total intelligibility.

The STIPA method in PP scenarios requires the STIPA speech-like test signal to be reproduced acoustically by a sound source that simulates a human talker natural speech production.<sup>7</sup> International and national standards (see Table 2) recommend the use of a suitable test loudspeaker or a special electroacoustic sound source to emulate the speech acoustical characteristics of a human talker. To that purpose, the physical size, directivity, orientation, and frequency response of the speech sound source are the key parameters to consider in its application.

In assessments of speech intelligibility and measurements of STIPA from SRS systems, a speech sound source is also required to provide a human-like reference acoustical test signal into the microphone of the system.<sup>7,10,12,19</sup>

As it can be seen in Table 2 the requirements and level of specification for acceptable speech sound sources to be used in PP applications vary among relevant standards.

The most informative standards regarding speech sound sources<sup>7,8</sup> detail the electro-acoustics requirements for a 'special test loudspeaker' (e.g. artificial mouth<sup>24</sup> or talkbox) for STIPA testing in SRS and PP scenarios. It also provides strict specifications for other 'suitable transducers' as standardized speech sound sources in the absence of special test loudspeakers.

Some of the other standards in Table 2 provide limited or unspecific electro-acoustic requirements for a speech sound source<sup>10–12,20–23</sup> thus allowing a wide range of alternative non-special loudspeakers to be employed as conforming speech sound sources.

## Motivation

There are only a few special speech test sound sources available in the market that meet the electro-acoustics performance specifications of the most informative standard.<sup>7</sup> They are expensive devices and can be deemed unaffordable and inaccessible to researchers, consultants, and other practitioners.<sup>25</sup> The strict performance requirements for substitute speech sources indicated in that standard<sup>7</sup> can make it difficult for those users to find, test, or construct alternative sources that satisfies the standard requirements.<sup>7,25</sup>

Very limited research exists in the literature concerning the suitability of non-special or non-standardized loudspeakers as alternative speech sound sources in PP testing applications. A study<sup>26</sup> was found that employed various commercial loudspeakers as speech sources to investigate the influence of their directivity characteristics on STI in four simulated indoors spaces. The study showed that the largest differences in STI between loudspeakers occurred in the larger rooms. The author also preliminarily showed that loudspeaker drive units of around 60 mm diameter produced the smallest errors and suggested that this size is more appropriate to simulate the human speech directivity pattern. However, the results obtained were in noise-free conditions and based on room acoustic computer simulation methods involving limitations with the virtual characterization data of some of the loudspeakers employed. Non-special loudspeakers are widely used in the relevant industry and academia<sup>13,15,27–32</sup> to conduct speech intelligibility investigations in place of standardized special speech test loudspeakers (i.e. mouth simulator<sup>24</sup> or talkbox). Nevertheless, due to the absence of guidance and practical information in the literature, those investigations are conducted without knowledge on the alternative loudspeakers' practical suitability and validity of results.

To address that lack of information, a recent preliminary investigation<sup>25</sup> examined experimentally the suitability of a range of representative non-special loudspeakers as an alternative to standardized speech sources in PP speech intelligibility investigations. Frequency response and STIPA experimental results obtained from a reference standardized special speech source were compared

**Table 2.** Summary of international and national standards where artificial mouth simulators or alternative speech sound sources are referred to.

Standard number and citation number	Requires standardized mouth simulator or talkbox	Allows alternative non-special loudspeakers	Implicit use of unspecified non-special loudspeaker as speech sound source.	Intended use
<b>IEC 60268-16: 2020</b> <sup>7</sup>	Yes, Minimum electro acoustics specifications indicated	Yes, Minimum electro acoustics specifications indicated	No	SRS and PP speech intelligibility
<b>IEC 60268-16: 2011</b> <sup>8</sup>	Yes, Minimum electro acoustics specifications indicated	Yes, Minimum electro acoustics specifications indicated	No	SRS and PP speech intelligibility
BB93 <sup>16</sup>	Yes, refers to <sup>8</sup>	Yes, refers to <sup>8</sup>	No	Classroom speech intelligibility & privacy
ANSI Standard ANSI/ASA S12.60-2010 <sup>20</sup>	No	Yes, only requirement: (unspecified) 'directivity similar to human talker'	No	School speech intelligibility/privacy
ASTM E1179-13 (2019) <sup>21</sup>	No	Yes, Minimum electro acoustics specifications indicated	No	Open plan office speech intelligibility/privacy
<b>ASTM E1130-16 (2021)</b> <sup>22</sup>	No	Yes, refers to <sup>21</sup>	No	Open plan office speech intelligibility and privacy
<b>ANSI/ASA S3.5-1997</b> <sup>23</sup>	No	Yes, Minimum electro acoustics specifications indicated	No	Calculation of Speech Intelligibility Index
<b>BS 7827: 2019</b> <sup>10</sup>	Yes	Yes, only requirement (unspecified) 'calibrated loudspeaker'	Yes	SRS for sports grounds, large public buildings & venues
<b>BS EN 50849:2017</b> <sup>12</sup>	No	Yes (unspecified requirements)	Yes	SRS for emergency purposes
<b>BS 5839-8:2023</b> <sup>9</sup>	Yes, refers to <sup>7</sup>	Yes, refers to <sup>7</sup>	No	Fire alarm Voice Alarm systems
<b>BS EN ISO 9921:2003</b> <sup>11</sup>	No	No	Yes	SRS and PP systems

against results from various non-special loudspeakers measured under a range of real-world combinations of PP scenarios and influencing factors. The work found a close STIPA agreement between the reference and the alternative non-specialist speech sources.

### *Aim and scope*

This study aims to further evaluate the performance validity of non-special loudspeakers as alternative to standardized special speech sound sources in person-to-person speech intelligibility assessment. It supplements the preliminary research undertaken previously<sup>25</sup> by providing additional evidence to consolidate earlier findings and to expand the range of their applicability. This is achieved by extending the scope of experimental testing and analysis of additional representative scenarios combinations of influencing factors. The incremental work respect the previous research, includes an expanded literature review and rationale, the incorporation of new test conditions including a higher interfering background noise level (55 dBA) and a wider off-axis source-receiver angle (45°), a comprehensive analysis of twenty scenarios, the addition of new discussion insights and a guidance for the selection of alternative speech test sources.

The complete findings and insights in this study will, for the first time, provide professional and research practitioners guidance and reassurance on the applicability and suitability of affordable non-special loudspeakers. This in turn will enable more researchers and practitioners to undertake PP speech intelligibility assessments and investigations when special standardized speech sources are not available or other relevant standards allowing non-special loudspeakers as alternative speech sources are adhered to.

### **Materials and methods**

The performance validity evaluation of non-special loudspeakers was based on the comparative analysis of STIPA results between alternative speech sources and the standardized reference speech test source. The materials and methods in this study followed those described in a previous preliminary investigation.<sup>25</sup>

Below is a summary of the speech test sources' description under test and the method procedure employed in this study. Additional technical specifications for each loudspeaker (such as directivity patterns) can be found via the corresponding citation reference indicated in the Model column of Table 3.

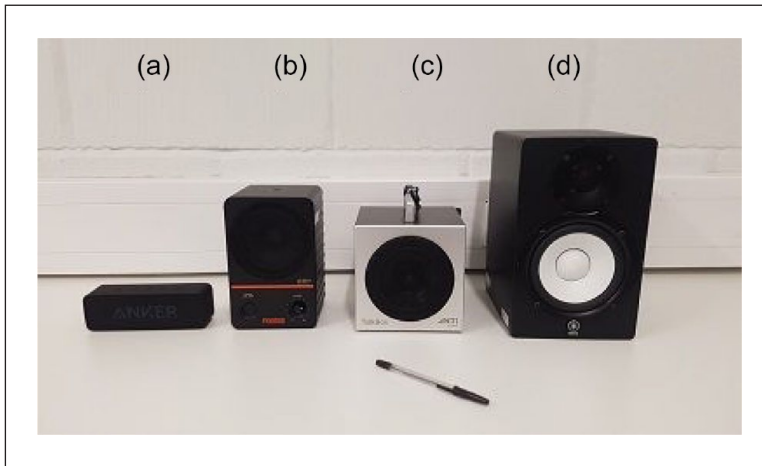
Figure 4 presents the frequency response in one-third octave bands to pink noise input signal reported in previous study<sup>25</sup> for the four speech test sources measured in anechoic conditions at 1 m on axis. The Anker green trace represent the average of the values from the three Anker units tested and the error bars indicate the standard deviation.

Measurements of the background noise sound pressure level (SPL) and STIPA were performed in turn on each of the speech sources in two different controlled physical acoustic environments (test rooms) which simulated a range of real-world PP acoustic conditions.

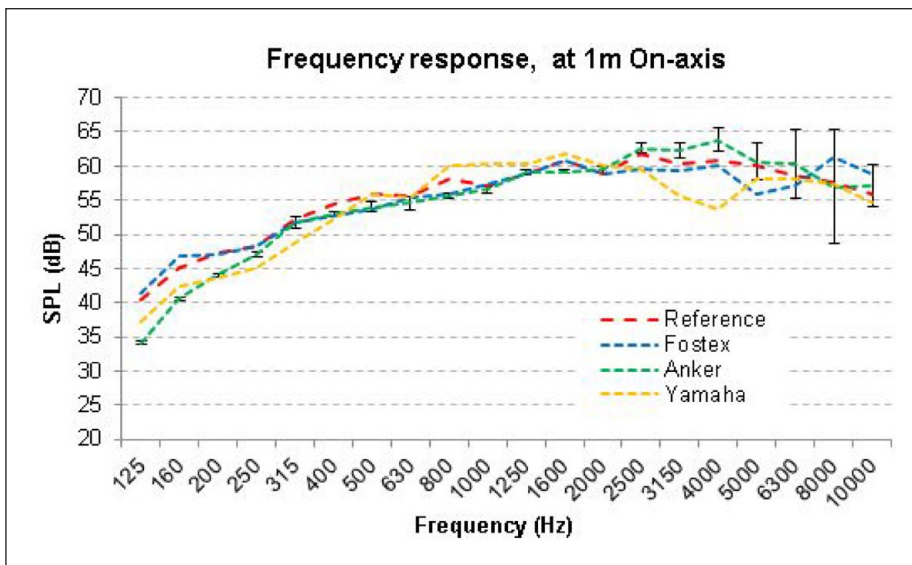
The first acoustic environment (semi-reverberant test room) consisted of the reverberation chamber at London South Bank University (LSBU) of 204 m<sup>3</sup> of volume, including 10 m<sup>2</sup> of highly sound-absorbing material (mineral wool) exposed on one of the chamber's walls (Figures 5(a) and 6(a)). The mid-frequencies average (500 Hz, 1 kHz, and 2 kHz) reverberation time RT30<sub>midfreq</sub> of the semi-reverberant test room measured to ISO 3382-1:2009<sup>37</sup> was 1.7 s. The second acoustic environment (anechoic test room) was the LSBU full anechoic chamber of 145 m<sup>3</sup> (excluding the volume occupied by wedges; Figures 5(b) and 6(b)).

**Table 3.** Description of speech test sources.

Brand	Model	Driver/diameter (mm)	Powered by	Frequency range	Output power (W)	Application	Role in this study
Anker	Sound core <sup>33</sup>	Two way/30-50	Battery	170Hz–13kHz	6	All purpose	Alternative speech test source
Fostex	6301N <sup>34</sup>	Single/100	Mains	70Hz–15kHz	20	Studio monitor	Alternative speech test source
NTI-Audio	TalkBox <sup>35</sup>	Single/100	Battery	100Hz–10kHz	10	Reference precision speech test source	Reference speech test source
Yamaha	HS50M <sup>36</sup>	Two way/127-19	Mains	55 Hz–20kHz	70	Studio monitor	Alternative speech test source



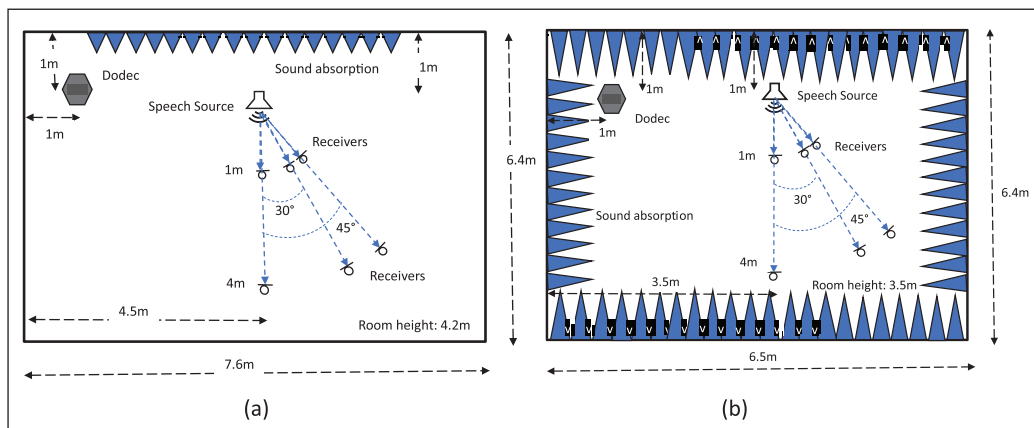
**Figure 3.** Photo of speech test sources (see Table 3) from left to right: (a) Anker, (b) Fostex, (c) Talkbox, and (d) Yamaha.



**Figure 4.** Frequency response in one-third octave bands for four speech test sources.

A fully in-calibration NTI-Audio XL2 class I acoustic analyser incorporating an NTI M2215 omni directional pre-polarized condenser microphone (frequency response: Class-1,  $\frac{1}{2}$ ", dynamic range: 25–153 dB) was employed to take background noise and STIPA signal sound pressure level (SPL) measurements (receiver SLM1). Another fully in-calibration XL2 class-I analyser incorporating an NTI M2211 omni-directional pre-polarized condenser microphone (frequency response : Class-1,  $\frac{1}{2}$ ", dynamic range: 21–144 dB) was used as the receiver to take STIPA measurements (receiver SLM2). Both analysers fully conformed with class-I specifications of sound level metres of relevant international standard.<sup>38</sup> Both acoustic analysers were field calibrated before and after





**Figure 5.** (a) Semi-reverberant room test layout and (b) anechoic room test layout (plan view, not to scale).

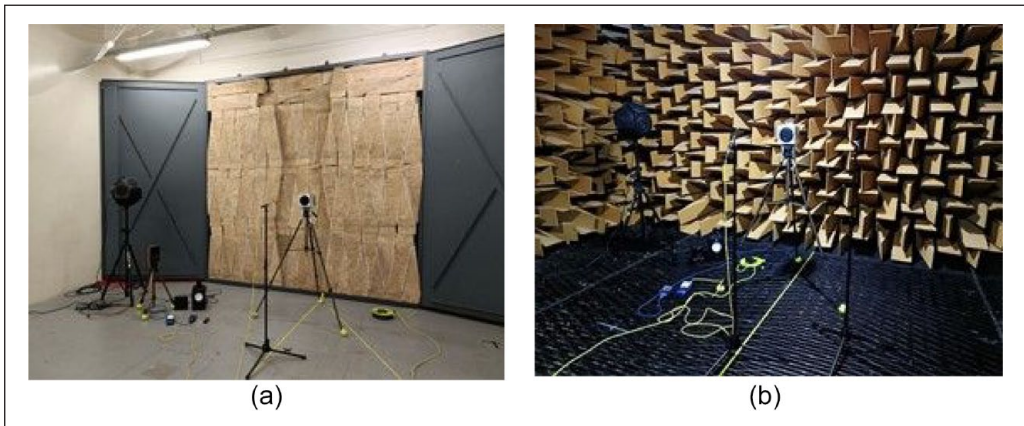
the measurement session by a Norsonics 1251 class-I calibrator, which itself was within traceable and valid laboratory calibration time window. No drift was observed in the field calibrations.

The adjustable background noise system consisted of a calibrated signal generator (NTI-Audio Minirator, MR-Pro) which fed pink noise signal into an audio amplifier driving an ANV dodecahedron sound source. Another signal generator (NTI-Audio Minirator, MR-Pro) provided the STIPA test signal via XLR cable connection into the line-in input of Yamaha and Fostex sources. STIPA signal was provided to the Anker source line-in input from a Toshiba Portege laptop via a mini-jack cable.

For both acoustic environments (or test rooms), the source position consisted of a reference mark point set at 1.6m height from the floor. This mark acted as a guide to situate with precision the approximate geometrical centre of each speech source. The receiver consisted of the SLM2 microphone set also at 1.6m height from the floor and situated at 1m on axis ( $0^\circ$ ) from the source position point (Figure 5(a) and (b)). The receiver SLM2 body was connected remotely to its microphone via an XLR extension cable to avoid contaminating reflections from the analyser's or operators' bodies.

The layouts for sources and receivers in both rooms (Figures 5(a) and (b) and 6(a) and (b)) were implemented to represent a range of potential and realistic PP scenarios and to examine the effects of source–receiver distance, angle, and acoustic conditions.

STIPA measurements were performed in both rooms following the test procedure specified in the latest version of relevant standard IEC 60268-16: 2020.<sup>7</sup> Each speech source under test was fed in turn with the STIPA test signal (fifth version) specified in the latest version of the relevant standard. The output level of the STIPA test signal was adjusted in the anechoic chamber for each source to measure 70 dBA at the SLM2 receiver with its microphone positioned on-axis at 1m from the speech source position. This calibration adjustment was performed to match the fixed signal output from the Talkbox (reference source) STIPA test signal Lombard level option. This selected output signal level corresponds to raised vocal effort exerted by talkers to overcome noisy backgrounds (Lombard effect). In line with the relevant standard IEC 60268-16:2020<sup>7</sup> test procedure, 70 dBA was chosen for this study as representative level of raised vocal effort expected to be exerted by a person addressing a group of people situated at different distances in an indoor or outdoor PP scenario. Once the speech sources' output levels were calibrated, they remained un-changed for the duration of the entire measurement session.



**Figure 6.** (a) Photo of the semi-reverberant room test layout and (b) anechoic room test layout.

Each source and receiver microphone height in both rooms was set at 1.6 m from the floor (i.e. adult average standing ear and mouth height).<sup>7</sup> During STIPA measurements in both test rooms, pink noise was emitted by the ANV dodecahedron sound source (Dodec) positioned at 4 m from the nearest receiver position at 1.6 m from the floor acting as a background noise source. The level of this adjustable background noise was set in both rooms to measure 55 dBA at each receiver position to represent interference background noise at a level typical of STIPA assessments in PP in indoor spaces<sup>9</sup> affected by external noise (e.g. traffic noise intruding into classrooms<sup>39</sup> or/and internal noise (e.g. occupancy noise or/and building services noise in workplaces<sup>40</sup>).

STIPA results from three non-special affordable loudspeakers were compared against the results from the reference standardized speech source (NTI Talkbox) and absolute errors were determined. Absolute error is defined in this study as the arithmetic difference between the STIPA value obtained from the reference source and the value from the non-special loudspeaker under test. Mean absolute error denotes the arithmetic average of a set of the absolute errors.

The validity evaluation was established through the analysis of absolute errors and standard deviation (std) of STIPA values as a function of four influencing factors: source-receiver distance, acoustic environment, source-receiver orientation and signal to background noise ratio (SNR).

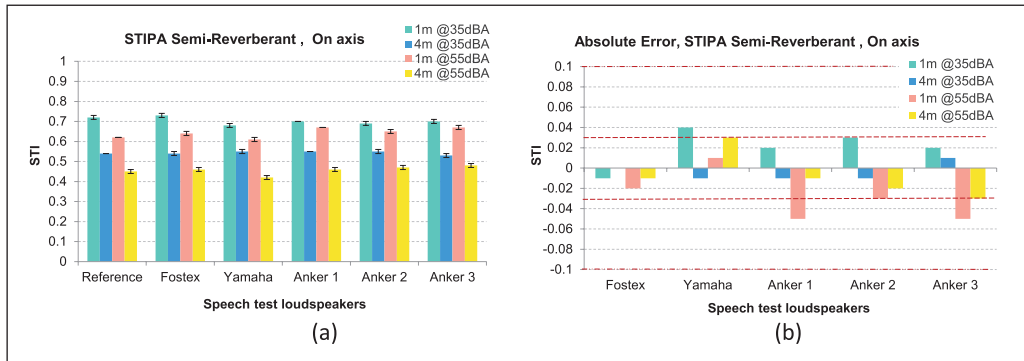
The new additional variables employed in this study with reference to the previous relevant investigation<sup>25</sup> and their purpose are instituted as follows.

To further assess the influence of off-axis source-receiver orientation, measurements were obtained at 45° off axis of the speech source under test, for two acoustics environments (semi-reverberant and anechoic) and for two source-receiver distances (1 and 4 m; see Figure 5(a) and (b)).

To examine the effect of lower signal to background noise ratio (SNR), a higher background noise level was artificially introduced to all measurements. Pink noise was employed as the controllable background noise signal. Background noise level at 55 dBA (up from 35 dBA in Reference<sup>25</sup>) was set at each microphone position for two acoustics environments, three source-receiver orientation angles (0° (on axis), 30° and 45°), and two source-receiver distances (see Figure 5(a) and (b)).

Reported STIPA values were calculated as the arithmetic average of five measurement readings (repeats) taken consecutively at the receiver position for each scenario or measurement condition.

To enable a comprehensive examination, STIPA data obtained from the previous investigation<sup>25</sup> at 35 dBA of background noise for 1 and 4 m in both test rooms was incorporated in the analysis of results.



**Figure 7.** (a) STIPA values measured in the semi-reverberant test room on axis at 1 and 4 m at 35 and 55 dBA of background noise. (b) STIPA mean absolute error values.

The combination of the influencing factors and variables considered for all speech sources under investigation, consisting of three source-receiver orientation angles ( $0^\circ$ ,  $30^\circ$ ,  $45^\circ$ ), two source-receiver distances (1 and 4 m), two background noise levels (35 and 55 dBA), and two acoustic environments (semi-reverberant and anechoic), would have generated twenty four measurement scenarios. However, measurements at  $45^\circ$  source-receiver angle with 35 dBA background noise level were not performed for the two source-receiver distances in any of the two environments. Hence the total measurement scenario combinations used for the overall data collection were twenty.

Absolute errors were evaluated in relation to the STI Just Noticeable Difference (JND). A value range of  $\pm 0.03$  STI is the uncertainty associated with the STIPA method.<sup>7</sup> This value is also accepted by some authors<sup>41–43</sup> as the estimated Just Noticeable Difference (JND) of the STI. However, researchers close to the inventors of the STI and successors of the development of the metric estimate the JND of the STI to be 0.1.<sup>44,45</sup> In the context of this study, absolute errors below one JND could be interpreted as non-perceivable and, therefore, negligible. In this investigation both JND threshold values are considered in the analysis of results.

Figures 5 and 6 show the test layout for measurements taken in both acoustic environments for combinations of two source-receiver distances for three source-receiver angle orientations.

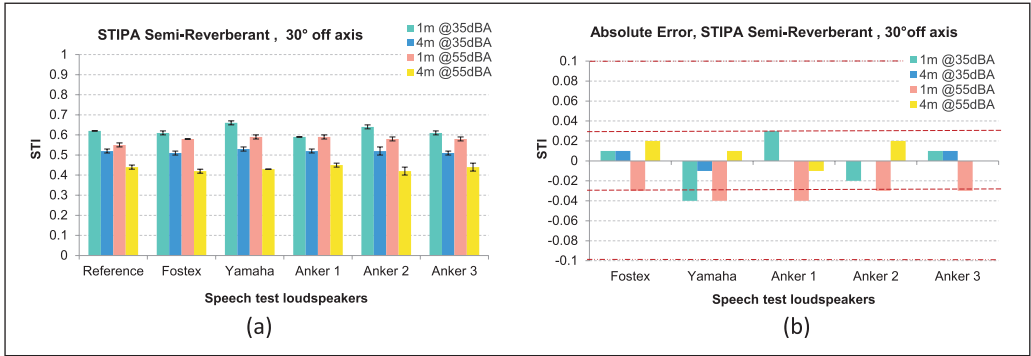
## Results and discussion

STIPA measurement results, measurement uncertainty and corresponding absolute error values obtained for each speech sound source at different distances, angles, background noise levels and test rooms are presented in Figures 7–12.

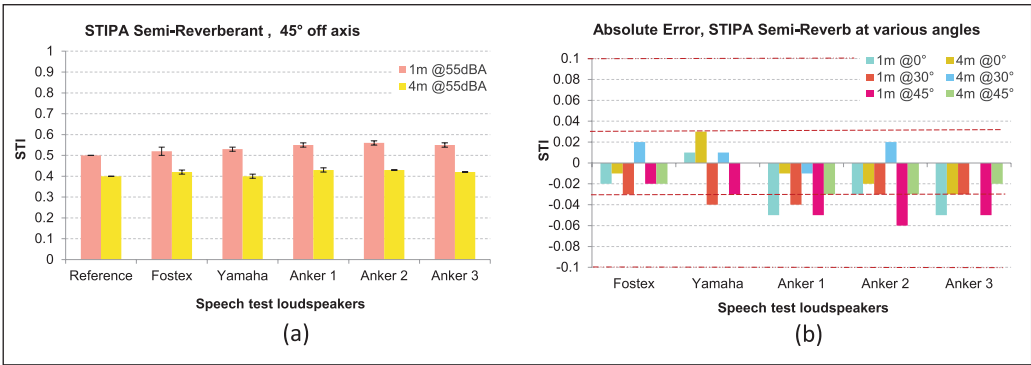
In all figures, each colour represents a scenario. In figures 7–12. (a), each column value is the average of five STIPA readings (repeats) taken of the same scenario. The error bars indicate the measurement uncertainty expressed by the estimated standard deviation of five STIPA readings taken per scenario.

### Results

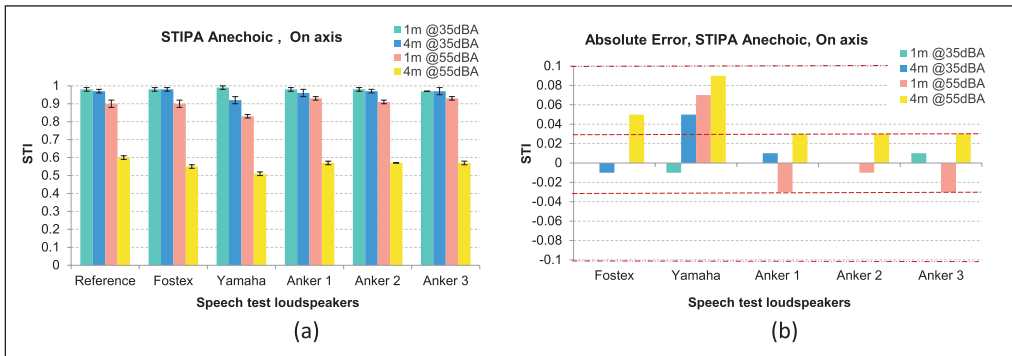
The red dotted lines in figures 7–12. (b) indicate the STI JND at 0.03 and 0.1. For readability STI JND of 0.03 and 0.1 will be referred as JNDa and JNDb respectively.



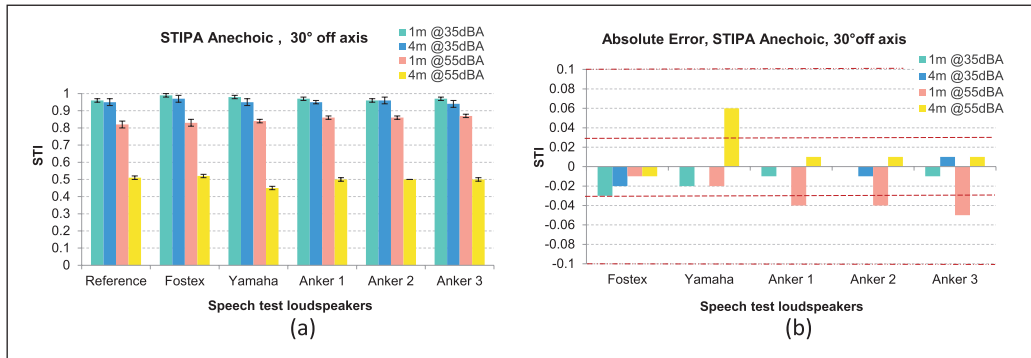
**Figure 8.** (a) STIPA values measured in the semi-reverberant test room 30° off axis at 1 and 4m at 35 and 55 dBA of background noise. (b) STIPA mean absolute error values.



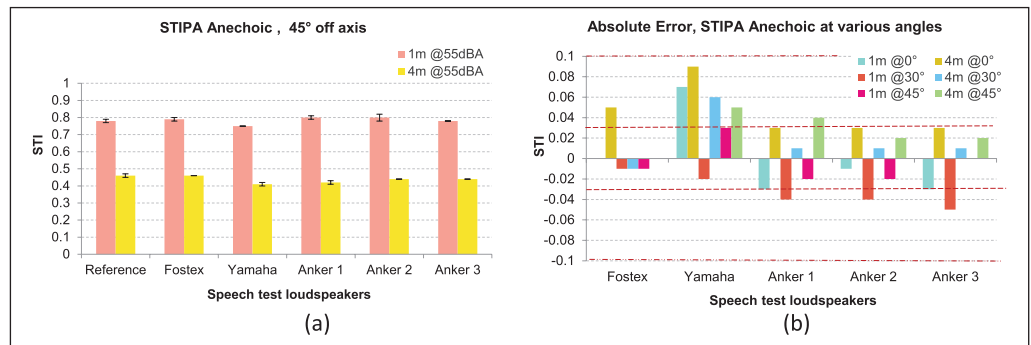
**Figure 9.** (a) STIPA values measured in the semi-reverberant test room 45° off axis at 1 and 4m at 55 dBA of background noise. (b) STIPA mean absolute error values at 1 and 4m at 0°, 30° and 45° at 55 dBA of background noise.



**Figure 10.** (a) STIPA values measured in the anechoic test room on axis at 1 and 4m at 35 and 55 dBA of background noise. (b) STIPA mean absolute error values.



**Figure 11.** (a) STIPA values measured in the anechoic test room 30° off axis at 1 and 4 m at 35 and 55 dBA of background noise. (b) STIPA mean absolute error values.



**Figure 12.** (a) STIPA values measured in the anechoic test room 45° off axis at 1 and 4 m at 55 dBA of background noise. (b) STIPA mean absolute error values at 1 and 4 m at 0°, 30° and 45° at 55 dBA of background noise.

### Measurement uncertainty and spread of results

The STIPA measurement uncertainty (shown as error bars in Figures 7(a)–12(a)) for the alternative speech sources (or no-special loudspeakers) for the all the scenario combinations was mostly within 0.01 (0.3 JNDA) with few cases within 0.02 (0.7 JNDA). The same values were obtained from the reference standardized speech source. These results showed the high repeatability of the alternative speech sources standing comparable in that characteristic to the reference speech source.

The spread of STIPA results (quantified by the standard deviation (std)) among the alternative sources was within 0.02 (0.7 JNDA), observed in all twenty scenarios combinations except for one which was within 0.04 (1.3 JNDA; Figures 7(a)–12(a)). That small spread indicates that the alternative speech sources perform almost equally irrespective of the four influencing factors.

The low measurement uncertainty and spread of results is consistent with those observed in the previous study.<sup>22</sup>

### Errors in the semi-reverberant environment

Examining absolute error results for all the alternative speech sources in all scenarios (Figures 7(b)–9(b)) it can be seen that errors were largely within 0.03 (or one JNDA), except for one Yamaha

scenario showing 0.04 (1.3 JNDa) and Anker1, Anker2 and Anker 3 showing few outlier results between 0.04 (1.3 JNDa) and 0.06 (2 JNDa). These slightly higher values were well within 1 JNDb and do not appear to be systematic or follow any systematic pattern or affected by the influencing factors at play.

### *Errors in the anechoic environment*

The absolute errors for all the alternative speech sources in all scenarios (Figures 10(b)–12(b)) were largely within 0.03 (1 JNDa), except Anker 1 and 2 that measured 0.04 (1.3 JNDa) and Anker3 that measured 0.05 (1.7JNDa) for only one scenario (Figure 11(b)). These Anker slightly higher errors (well within 1 JNDb) are not consistent with other related scenarios and do not appear to follow any pattern thus they can be deemed irrelevant. The Yamaha showed a slight pattern of decreasing errors with orientation angle (from 0.09 (3 JNDa) at  $0^\circ$  to 0.05 (1.7 JNDa) at  $45^\circ$ ) for the 4 m distance at 55 dBA background noise scenario (see Figures 10(b)–12(b)). This could suggest in principle that the Yamaha features a higher directivity than the rest of sources (including the reference) most likely caused its larger size. However, these errors were not observed at 35 dBA when tested also at 4 m at  $0^\circ$  and  $30^\circ$ . Those errors (within 1 JNDb) and the suspected pattern were not observed in the semi-reverberant environment.

### *Errors as a function of orientation angle*

Figure 9(b) presents the STIPA errors from the speech sources in the semi reverberant environment as a function of distance and orientation angle with background noise fixed at 55 dBA. Examining the effect of the orientation angle, it can be seen that in general errors are within (0.03 (1 JNDa)) at both distances, are not affected by the orientation angle. Inspecting the relevance of the higher values, Anker2 gave an error of 0.06 (2 JNDa) at  $45^\circ$  at 1 m. Anker1 and 3 showed an error of 0.05 (1.7 JNDa) at  $0^\circ$  and  $45^\circ$  also at 1 m. However, these values were not observed at  $30^\circ$  or from Anker2 at  $0^\circ$  at the same distance. These errors potentially attributed to orientation angle are not observed for any of the Anker sources at 4 m. Moreover, the magnitude of those errors can be deemed even less noteworthy considering they are well within 1 JNDb.

For the anechoic environment, Figure 12(b) shows the STIPA errors as function of distance and orientation angle with background noise also fixed at 55 dBA. Analysing the effect of orientation angle, it can be observed that most of errors from of the sources are within 0.03 (1 JNDa) and they are not affected by orientation. The higher values were errors from the Yamaha for several angles and distances (errors ranged between 0.09 (3 JNDa) to 0.05(1.7 JNDa)) and less significant errors from three Anker sources ranging between 0.04 (1.3 JNDa) and 0.05 (1.7 JNDa) for the specific  $30^\circ$  at 1 m scenario. The analysis of those higher values is discussed above. The Fostex single error above 1JND of 0.05 (1.7 JNDa) at 4 m at  $0^\circ$  scenario, as is the case of other errors slightly above 1 JNDa, does not follow any pattern and cannot be attributed to any of the influencing factors and is therefore deemed irrelevant. In addition, the magnitude of all errors seen in the anechoic environment are well within 1 JNDb.

### *Errors as a function of increasing the background noise level*

In the anechoic environment (Figures 10(b) and 11(b)) the three Anker sources showed differences in error within 0.04 (1.3 JNDa) when the background was increased from 35 to 55 dBA. These differences are too small to indicate any influence due to the increase. However, the Yamaha showed

**Table 4.** Mean absolute error for each speech source in both environments.

	Semi reverberant			Anechoic			All
	1 m	4 m	mean	1 m	4 m	mean	mean
Fostex	0.02	0.01	0.02	0.01	0.02	0.01	0.01
Yamaha	0.03	0.01	0.02	0.03	0.05	0.04	0.03
Anker 1	0.04	0.01	0.03	0.02	0.02	0.02	0.02
Anker 2	0.03	0.02	0.03	0.01	0.01	0.01	0.02
Anker 3	0.03	0.01	0.02	0.02	0.01	0.02	0.02

error differences within 0.06 (2 JNDa) at 0° and 30° at both distances. Considering also the reduction of error with orientation angle seen in the anechoic condition earlier at 55 dBA but not at 35 dBA, it could be argued the Yamaha is more directional than the rest of sources (including the reference) and that feature on absolute errors is more affected by the higher background noise level. Nevertheless, these errors are relatively small and well within 1 JNDb so they could be considered not meaningful enough to be noteworthy.

In the semi-reverberant environment (Figures 7(b) and 8(b)) increasing the background noise level by 20 dBA did not influence the errors obtained for both distances at 0° and 30°. All error differences observed were too small (within 0.03 (1 JNDa)) to bear any relevance or be able to reveal any pattern.

### Summary

Table 4 shows the mean absolute errors calculated from the absolute errors obtained for each speech source at 1 and 4 m at 35 and 55 dBA in the semi reverberant and anechoic environments.

As it can be seen in Table 4 the mean absolute errors obtained for all alternative speech sources at both levels of background noise in both environments were within 1 JNDa; except for the Yamaha in anechoic condition which was 1.3 JNDa. The standard deviation of all mean absolute errors at both distances for both acoustic environments shown in Table 4 were below 1 JNDa.

All the errors measured in this study were well within 1 JNDb. STIPA errors within 1 JND can be interpreted as non-perceivable and therefore negligible.

The high measurement certainty and low spread of results consistently observed from all sources at all scenario combinations provides further confidence in the results and findings presented.

From the data analysis shown above, it follows that the non-special loudspeakers performance tested for the numerous scenarios have consistently shown close agreement with the standardized speech source. These findings validate the proposition that non-special and affordable loudspeakers may be used as suitable speech test sources in pilot or survey-grade speech intelligibility assessments in place of a standardized special speech source when that special source is not available, or in full investigations when the relevant standard to adhere to allows the use of those alternative speech sources.

From these conclusive findings, it could be implied that the STIPA metric when employed in PP situations, might allow for less restrictive tolerances in the speech test loudspeaker specifications than is currently specified in the most informative standard.<sup>7</sup>

Based on the literature and findings of this study the following guidance is suggested for the selection of suitable non-special loudspeakers for the applications described: To choose a quality commercial or professional grade loudspeaker for music or speech reproduction featuring an

enclosure of its largest dimension of around 170 mm incorporating a single-driver of wide frequency range (i.e. 100–10 kHz) of cone diameter between 60 and 100 mm.

## Conclusions

A study has been conducted to validate the suitability of non-special loudspeakers as speech sources employed in speech intelligibility assessments in Person-to-Person Speech Communications (PP). It has extended the scope of applicability and consolidated the findings obtained in a previous investigation.

Experimental Speech Transmission Index for Public Address (STIPA) tests were conducted on three representative non-special loudspeakers and a standardized reference speech source considering four influencing factors (source-receiver distance, acoustic environment, source-receiver orientation and signal to background noise ratio) affecting PP transmission channel. Twenty measurement scenario combinations generated a total of 600 measurement readings. Absolute errors of STIPA measurements obtained between the non-special loudspeakers and the reference source were analysed and their relevance evaluated against two threshold levels of Just Perceivable Difference (JND) for the STIPA parameter (0.03 and 0.1 of STI).

The mean absolute errors obtained for all non-special loudspeakers for almost all scenarios were within one JND of any of the two JND levels considered. The low measurement uncertainty and low spread of results consistently observed from all sources at all scenario combinations (standard deviation typically within 0.02 of STI) provided further confidence in the results obtained. These findings are in close accord with those from a previous investigation and serve here to corroborate the earlier conclusions.

The striking and consistent low level of errors found validate the proposition that non-special and affordable loudspeakers may be used as suitable speech test sources in pilot or survey-grade speech intelligibility assessments, in place of a special standardized speech source when that source is not available or in full investigations when the relevant standard to adhere to allows the use of those alternative speech sources.

The conclusive findings and insights of this study will inform and enable more researchers, consultants, and other practitioners to conduct speech intelligibility investigations in PP scenarios.

It is anticipated that the information presented in this work could influence future design, development and commercialization of loudspeakers for speech reproduction.

## Acknowledgements

For the purposes of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research has been supported by a REI Sabbatical Grant at London South Bank University.



**ORCID iD**

Luis Gomez-Agustina  <https://orcid.org/0000-0002-4460-7563>

**References**

1. Liu H, Ma H, Kang J, et al. The speech intelligibility and applicability of the speech transmission index in large spaces. *Appl Acoust* 2020; 167: 107–400.
2. Airport Cooperative Research Program National Academies of Sciences Engineering and Medicine (U.S.). *Improving intelligibility of airport terminal public address system*. Transportation Research Board and United States Federal Aviation Administration, 2017.
3. Mapp P and Hammond R. The effects of spectators on the speech intelligibility performance of sound systems in stadia and other large venues. *Audio Engineering Society Convention 147*. Audio Engineering Society, 2019.
4. Gomez-Agustina L, Dance S and Shield B. The effects of air temperature and humidity on the acoustic design of voice alarm systems on underground stations. *Appl Acoust* 2014; 76: 262–273.
5. Yang W and Bradley JS. Effects of room acoustics on the intelligibility of speech in classrooms for young children. *J Acoust Soc Am* 2009; 125(2): 922–933.
6. Astolfi A, Bottalico P and Barbato G. Subjective and objective speech intelligibility investigations in primary school classrooms. *J Acoust Soc Am* 2012; 131(1): 247–257.
7. IEC 60268-16:2020. Sound system equipment-part 16: objective rating of speech intelligibility by speech transmission index.
8. IEC 60268-16:2011. Sound system equipment-part 16: objective rating of speech intelligibility by speech transmission index.
9. BS 5839-8:2023. Fire detection and fire alarm systems for buildings. Design, installation, commissioning and maintenance of voice alarm systems. Code of practice.
10. BS 7827:2019. Designing, specifying, maintaining and operating emergency sound systems for sports grounds, large public buildings, and venues. Code of practice.
11. BS EN ISO 9921:2003. Ergonomics. Assessment of speech communication.
12. BS EN 50849:2017. Sound systems for emergency purposes.
13. Dance S, Backus B and Morales L. A methodology to objectively assess the performance of sound field amplification systems demonstrated using 50 physical simulations of classroom conditions. *Noise Health* 2018; 20(94): 77–82.
14. Dockrell JE and Shield B. The impact of sound-field systems on learning and attention in elementary school classrooms. *J Speech Lang Hear Res* 2012; 55(4): 1163–1176.
15. Luykx MPM and Vercammen MLS. Natural speech intelligibility in theatres in relation to their acoustics. *Build Acoust* 2011; 18(3-4): 293–311.
16. Department for Education. *Building bulletin 93. Acoustic design of schools: performance standards*. London: Education Funding Agency, 2015.
17. BS EN ISO 3382-3:2022. Acoustics. Measurement of room acoustics parameters. Part 3-Open plan offices.
18. Association of Noise Consultants and the Institute of Acoustics. *Acoustics of Schools: a design guide*, UK, November 2015. Acoustics of Schools: a design guide (ioa.org.uk)
19. The Association of Noise Consultants. *ANC good practice guide. Acoustics Testing of Schools*, UK, November 2015. ANC Schools GPG (association-of-noise-consultants.co.uk)
20. ANSI/ASA S12.60-2010/Part 1. American national standard acoustical performance criteria, design requirements, and guidelines for schools, part 1: permanent schools.
21. ASTM E1179-13:2019. Standard specification for sound sources used for testing open office components and systems.
22. ASTM E1130-16:2021. Standard test method for objective measurement of speech privacy in open plan spaces using articulation index.
23. ANSI/ASA. S3.5-1997 (R2020). Methods for calculation of the speech intelligibility index.

24. ITU-T Recommendation P.51 (08/96) Telephone transmission quality. Objective measuring apparatus: artificial mouth. [https://www.itu.int/rec/dologin\\_pub.asp?lang=e&id=T-REC-P.51-199608-I!!PDF-E&type=items](https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.51-199608-I!!PDF-E&type=items)
25. Gomez-Agustina L, Aygun H and Mohan LST. Non-special loudspeakers as speech test sources in natural acoustics speech intelligibility investigations. *Acoust* 2023; 5(3): 619–630.
26. Mapp P. Simulating talker directivity for speech intelligibility measurements. *Audio Engineering Society Convention 137*.
27. Scoczynski Ribeiro R, Moreira I, Marcelo Miscovicz L, et al. Assessing acoustical quality in university classrooms in Brazil: measurements and simulations. *Build Acoust* 2024; 31(2): 199–222.
28. Loreto SD, Lori V, Serpilli F, et al. ‘Great food, but the noise?’: relationship between perceived sound quality survey and non acoustical factors in one hotel restaurant in Italy. *Build Acoust* 2023; 30(4): 425–443.
29. Gover BN and Bradley JS. Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms. *J Acoust Soc Am* 2004; 116(6): 3480–3490.
30. Ahrens A, Marschall M and Dau T. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. *Hear Res* 2019; 377: 307–317.
31. McNeer RR, Bennett CL, Horn DB, et al. Factors affecting acoustics and speech intelligibility in the operating room: size matters. *Anesth Analg* 2017; 124(6): 1978–1985.
32. Peng J, Lau SK and Zhao Y. Comparative study of acoustical indices and speech perception of students in two primary school classrooms with an acoustical treatment. *Appl Acoust* 2020; 164: 107297.
33. Anker Soundcore A3102\_Soundcore\_Manual. Anker Soundcore manual. [https://ankertechnology-companyltd.my.salesforce.com/sfc/p/5g000004DkWQ/a/5g000000g2Ph/59RV\\_pJzzRkbKvQQDXVf-bQgXKdcI0tdbrsBp6QSiMvc](https://ankertechnology-companyltd.my.salesforce.com/sfc/p/5g000004DkWQ/a/5g000000g2Ph/59RV_pJzzRkbKvQQDXVf-bQgXKdcI0tdbrsBp6QSiMvc) (accessed 1 July 2024).
34. Fostex 6301N series. [https://fostexinternational.com/docs/products/6301N\\_Series.shtml](https://fostexinternational.com/docs/products/6301N_Series.shtml) (accessed 1 July 2024).
35. NTI Audio Talkbox. <https://www.nti-audio.com/en/products/noise-sources/talkbox> (accessed 1 July 2024).
36. Yamaha Manuals. [https://uk.yamaha.com/en/products/contents/music\\_production/downloads/manuals/index.html?l=en&c=music\\_production&k=HS50M](https://uk.yamaha.com/en/products/contents/music_production/downloads/manuals/index.html?l=en&c=music_production&k=HS50M) (accessed 19 May 2023).
37. BS EN ISO 3382-1:2009. Acoustics. Measurement of room acoustic parameters. Part-1 performance spaces.
38. BS EN 61672-1:2013. Electroacoustics. Sound level meters. Part 1-specifications.
39. Shield B, Conetta R, Dockrell J, et al. A survey of acoustic conditions and noise levels in secondary school classrooms in England. *J Acoust Soc Am* 2015; 137(1): 177–188.
40. Filus W, Lacerda AB and Albizu E. Ambient noise in emergency rooms and its health hazards. *Int Arch Otorhinolaryngol* 2015; 19(3): 205–209.
41. Bradley JS, Reich R and Norcross SG. A just noticeable difference in C 50 for speech. *Appl Acoust* 1999; 58(2): 99–108.
42. Duangpummet S, Karnjana J, Kongprawechnon W, et al. Blind estimation of speech transmission index and room acoustic parameters based on the extended model of room impulse response. *Appl Acoust* 2022; 185.
43. Morales N, Tang Z and Manocha D. Receiver placement for speech enhancement using sound propagation optimization. *Appl Acoust* 2019; 155: 53–62.
44. Wenmaekers R, Van Hout N, Hak C, et al. The effect of room acoustics on the measured speech privacy in two typical European open plan offices. In: *Proceedings of the inter-noise and noise-con*, Ottawa, Canada, 23–August 2009, vol. 2009, no. 5, pp. 2040–2045.
45. Houtgast T and Steeneken H. The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acta Acust United Acust* 1973; 28(1): 66–73.