On The Enhanced Acoustic Design Of The Indoor Environments: Correspondence Of Perceptual Quantities Between Real And Simulated Sound Fields

Chiara Visentin
Department of Engineering, University of Ferrara, Ferrara, Italy, chiara.visentin@unife.it

Nicola Prodi
Department of Engineering, University of Ferrara, Ferrara, Italy, nicola.prodi@unife.it

Francesca Cappelletti
Department of Design and Planning in Complex Environments, University IUAV of Venezia, Venezia, Italy, francesca.cappelletti@iuav.it

Simone Torresin
Faculty of Science and Technology, Free University of Bolzano/Bozen, Bolzano/Bozen, Italy, Simone.Torresin@natec.unibz.it

Andrea Gasparella
Faculty of Science and Technology, Free University of Bolzano/Bozen, Bolzano/Bozen, Italy, andrea.gasparella@unibz.it

On The Enhanced Acoustic Design of the Indoor Environments: Correspondence of Perceptual Quantities Between Real And Simulated Sound Fields

Chiara VISENTIN[1]*, Nicola PRODI[2], Francesca CAPPELLETTI[3], Simone TORRESIN[4], Andrea GASPARELLA[4]

[1, 2]Department of Engineering, University of Ferrara,
Ferrara, Italy
[1]chiara.visentin@unife.it
[2]nicola.prodi@unife.it

[3]Department of Design and Planning in Complex Environments, University Iuav of Venezia,
Venezia, Italy
[3]francesca.cappelletti@iuav.it

[4,5]Faculty of Science and Technology, Free University of Bolzano/Bozen,
Bolzano/Bozen, Italy
[4]simone.torresin@natec.unibz.it
[5]andrea.gasparella@unibz.it

* Corresponding Author

ABSTRACT

In the design of indoor spaces where speech communication takes on a central role (e.g. classrooms, conference rooms, etc.), the influence of the sound environment on the occupants' performance needs to be addressed. In order to guarantee a comfortable communication experience, the acoustic design of such spaces has to ensure, beside a high percentage of correctly heard words (i.e. intelligibility), also a minimal effort in the speech reception process. An effortful listening, as produced for instance by the presence of background noise and/or reverberation, requires the involvement of an increased amount of cognitive resources. If sustained over a prolonged period, this additional cognitive burden may compromise occupants' learning and cognitive achievements.

The present study specifically addresses the issue of an improved acoustical design of the rooms for speech, based on both intelligibility and listening effort results. For the scope, the correspondence between the results of speech-in-noise tests presented within an existing university classroom and via headphones, using auralized signals obtained from acoustic simulations of the same environment, is investigated. In fact, whereas the reliability of acoustical simulations has been widely confirmed as regards the predicted objective acoustic parameters including speech intelligibility data, to date an ecological validation of the listening effort metrics is still lacking, and is needed to understand how well results obtained by virtual acoustics predict the everyday realistic communication situations.

Simulations of a university classroom with a volume of 198 m^3 and acoustical treatment on a lateral wall were carried out with an acoustical CAD software. In order to obtain realistic simulations, the model was calibrated with octave-band field measurements of reverberation and clarity parameters. Binaural impulse responses (BRIRs) were calculated in two listening positions within the classroom, and convolved with anechoic speech and stationary noise, to obtain the auralized stimuli for the speech-in-noise tests. Speech and noise sound pressure levels were calibrated with reference to the values measured during the *in situ* tests. Consonant confusion tests (Diagnostic Rhyme Tests) in the Italian language were proposed to normal-hearing young adults. The tests were firstly presented in the real classroom, and then in laboratory conditions via headphones. During the experiments, data on the number of words correctly recognized, auditory response times (RT, behavioral measure of listening effort) and subjective ratings of listening effort (LE) were collected.

The statistical analyses showed that both IS and RT data in auralized conditions matched the corresponding results obtained with *in situ* testing; the RT metric showed a greater sensitivity than IS, being able to discriminate

between the listening position within the classroom. As concern LE, the results were found to depend on the mode of presentation, suggesting that beside the auditory stimulus other factors (such as attention or experimental setup) affect the subjective response. Based on the comparison of the results in auralized and *in situ* conditions, it can be said that the auralization techniques allow recreating a perceptually equivalent environments as regards the IS and RT measures, and that the integration of the two metrics would be of benefit the acoustical design process.

1. INTRODUCTION

In the design of indoor spaces where speech communication takes on a central role (e.g. classrooms, conference halls, lecture rooms, etc.), the influence of the sound environment on the occupants' performance needs to be addressed. In particular, room acoustics should be carefully designed, aiming at a comfortable communication experience including both a high accuracy and a minimal effort in the speech reception process.

Technical standards prescribe target values to be achieved, either in terms of objective parameters, for instance reverberation time and Speech Transmission Index (STI) (Deutsches Institute fur Normung, 2016; International Elechtrotecnical Commission, 2011) or in terms of speech intelligibility IS (i.e. number of words correctly heard). Objective and subjective quantities are related by task-specific psychometric curves (International Organization of Standardization, 2003). Despite this type of design ensures high accuracy, difficulties may be anyway experienced by occupants when listening to speech, due to the presence of background noise and/or reverberation. Indeed, whereas in ideal condition the speech is largely processed in an automatic way, mainly relying on perceptual cues, when speech is degraded as in case of suboptimal listening conditions, a mismatch arises between the external demands posed by the listening experience and the internal resources of the listener (Lemke and Besser, 2016). Then, a specific listening effort is experienced, indicating the listener is required to engage further cognitive resources to cope with the task. Listening effort was recently defined as "the level of processing resource allocation to overcome obstacles in goal pursuit when performing a listening task" (Pichora-Fuller et al., 2016). Interestingly, the construct was found to mirror changes of speech reception accuracy but also to vary independently as it happens in the most favorable listening conditions, when performance accuracy is maintained at the expenses of a more explicit cognitive processing (Suprenant, 1999). Owing to the limited availability of personal cognitive capacity (Kahneman, 1973), a practical consequence is that when increased resources are allocated to word reception, less capacity will be available for higher level processing of speech (e.g. recall of information, understanding of instructions, extraction of discourse meaning, etc.). So, when high levels of effort have to be sustained for long periods, such as during lessons, fatigue may arise with negative consequences on learning and cognitive achievements of listeners (Bess and Hornsby, 2014). At present, listening effort is not targeted by normative approaches, thus overlooking information of the deployment of cognitive resources that are instead relevant for the design of spaces where a comfortable communication may take place. Many factors affect this construct (Pichora-Fuller *et al.*, 2016) and to date no single measure is available to capture this multifaceted experience. Over the years, several methods have been proposed for the scope (Pichora-Fuller *et al.*, 2016, McGarrigle *et al.*, 2014), which can be divided into three categories: physiological, cognitive-behavioral and subjective ratings. The relationship between different metrics used as proxies of listening effort is still unclear: different measures can yield different results (van den Tillaart-Haverkate *et al.*, 2017), as supposedly reflecting underlying constructs that do not entirely match.

The present study specifically addresses the issue of the acoustical design of rooms for speech based on metrics that go beyond performance accuracy (IS) and are able to estimate the complex construct of listening effort. Following this research line, it will be possible to design environments where the occupants' performance is improved, as regards both perceptual and cognitive aspects. To the scope, a subjective and a behavioral metric that could be easily implemented in the context of *in situ* experiments have been selected: the subjective rating of listening effort (LE) and the response time (RT) to the auditory stimulus measured in a single-task paradigm. This latter proved to reflect the amount of resources required to interpret and respond to the incoming signal (Houben *et al.*, 2013; Pals *et al.*, 2015). Since RT and LE potentially carry complementary information to IS, they could be used to improve the means of evaluation of rooms for speech.

The first step toward this aim is to examine the correspondence between the metrics acquired from *in situ* and auralized speech-in-noise tests. Indeed, this is a preliminary requirement for the meaningful integration of the new quantities in the acoustic design process, which nowadays mainly relies on simulation techniques and virtual models that make it possible for objective and subjective metrics to be assessed before the room is built. During the past years, several studies have addressed the issue of the validity of acoustical simulations showing that, once the virtual models are carefully calibrated upon measures, the auralized sound field can almost be equivalent to the real one as concerns acoustical perceptual attributes (Potsma and Katz, 2016). A comparison of speech intelligibility data in the framework of auralization was also performed, indicating that a good agreement between real and virtual data can be obtained (Rychtáriková *et al.*, 2011) but consistency in speech intelligibility results seems to decrease for shorter reverberation times and too noisy sound fields (Yang and Hodgson, 2007). However, to date, no specific ecological validation of the RT and LE metrics has been carried out and a proof of correspondence between the values retrieved under natural and synthetized conditions is still lacking. In this work, auralization techniques based on calibrated acoustical simulations were used to playback sound fields via headphones in a laboratory setting, after the same listening conditions were presented ecologically in the real classroom. A university classroom was chosen as a case study, being a room typology for which good environmental comfort greatly influences the learning capacity of students.

2. METHODS

2.1 Participants and speech material

Ten normal hearing young adults participated in the experiment, including 5 female and 5 male. The ages ranged from 23 to 27 years (mean: 24.4 yr). Participants were recruited among the students of the University of Bozen-Bolzano and self-reported the absence of hearing impairments. All of them were native Italian speakers.

The speech material of the Diagnostic Rhyme Test (DRT) was used for the speech-in-noise tests, in the Italian version developed by Bonaventura *et al.* (1986). The DRT is a consonant confusion test, which bases on 105 pairs of rhymed, disyllabic words (e.g. */ˈtubo/* and */ˈkubo/*); the speech *corpus* is optimized as regards phonemic distribution and word familiarity. One item of each pair was recorded embedded in a carrier phrase (e.g. "La prossima parola è *tubo*", which is Italian for "The next word is *tube*") by an adult, native Italian, female speaker. The recordings took place in a silent room, at a sampling frequency of 44.1 kHz. The sentences were then filtered as to match the long-term spectrum of a female speaker (International Electrotechnical Commission, 2016), and set to the same root mean square value. The recordings were organized into five lists of 18 words each; the remaining 15 sentences were used in the training phase.

2.2 Tests in the real classroom: outline of the classroom and listening test setup

The listening tests were conducted in a university classroom, part of the Classroom Spaces Living Lab of the Free University of Bozen-Bolzano. The classroom has a rectangular plan (7.29 x 7.62) m and a height of 3.55 m; the resulting volume is 197 m^3. The boundary surfaces of the room have a flat finishing (ceiling: unpainted concrete, floor: linoleum finishing, walls: painted plasterboard); one wall is acoustically treated with Topakustik® 6/2 sound absorbing paneling. A plan of the classrooms is reported in Figure 1a.

During the tests, the classroom was furnished with wooden desks and chairs for a maximum of 25 students. A B&K type 4720 artificial mouth was placed close to the teacher desk, at the conventional height of a speaker's mouth (1.5 m); it was used to deliver the speech signal. A B&K type 4292-L omnidirectional source was located on the floor directly below the artificial mouth and used to playback the interfering noise. Two listening positions (R1 and R2) were defined within the classroom, located at the front and at the back of the audience, respectively 2.50 and 5.50 m from the sources. Two B&K type 4189 microphones were placed at a height of 1.25 m and used for the objective description of the listening conditions. Additionally, two head-and-torso simulators B&K type 4100 were positioned with the ears at a height of 1.15 m and used for the collection of binaural impulse responses (BRIRs). Around the microphones, the chairs for the participants were arranged.

The reverberation times (T_{30}) and the speech clarity values (C_{50}) were derived at R1 and R2 from monaural impulse responses, measured with the sine-sweep technique at the end of the experiment with the classroom still in occupied conditions. The mid frequency values (averaged over the 500–2000 Hz octave bands) were 0.82 and 0.84 s (T_{30}) and 3.6 and 0.5 dB (C_{50}).

2.3 Tests in the real classroom: listening conditions and procedures

For the experiment, the speech signal was played back with a fixed level of 63 dB(A) measured at 1 m in front of the source. It corresponds to a talker speaking with a vocal effort intermediate between "normal" and "raised" (International Organization of Standardization, 2003), and produced a speech level of 61.0 and 57.4 dB(A) respectively at R1 and R2. A stationary noise with the same long-term spectrum of the speech was also played back, setting its level to achieve a signal-to-noise ratio (SNR) of 0 dB at R1; the noise level measured in R2 was 58.7 dB(A). A comprehensive description of the tested listening conditions within the classroom was obtained by calculating the Speech Transmission Index (STI), which describes the combined effect of background noise and reverberation on the transmission quality of the speech signal. The STI values were 0.52 in R1 and 0.46 in R2, corresponding to an intelligibility rated as "Fair" (International Organization of Standardization, 2003).

A touchscreen handset was given to each participant to be used for response selection. The experiment was managed using a wireless test bench, with a server application controlling both audio playback and data retrieval (Prodi *et al.*, 2013). During the experiment, the participants sat around the two receiver positions and listened to a test sentence (carrier phrase + target word); at the audio playback offset three options were displayed on the touchscreen: the target word, its rhymed alternative at the "none of the two" option. After all participants have selected a response, the following test sentence was automatically presented.

A training session was firstly proposed to familiarize the participants with the test procedure; then they completed one test list of 18 words in each listening position. After each list, the participants were asked to rate their perceived listening effort (LE), answering to the following question: "*How much effort did it take to hear and understand the words?*" The responses were given on a 10-points scale, ranging from *minimum effort* (1) to *maximum effort* (10), which appeared on the handset touchscreen after the last pair of words from each list. The participants were instructed to pay attention, and asked to respond as accurately as possible but they were not urged to provide the quickest possible response.

The data retrieved for each participant during the experiment were word scores (correct/incorrect/none of the two), manual response times (time elapsed between the audio playback offset and the selection of a response) and subjective ratings of listening effort.

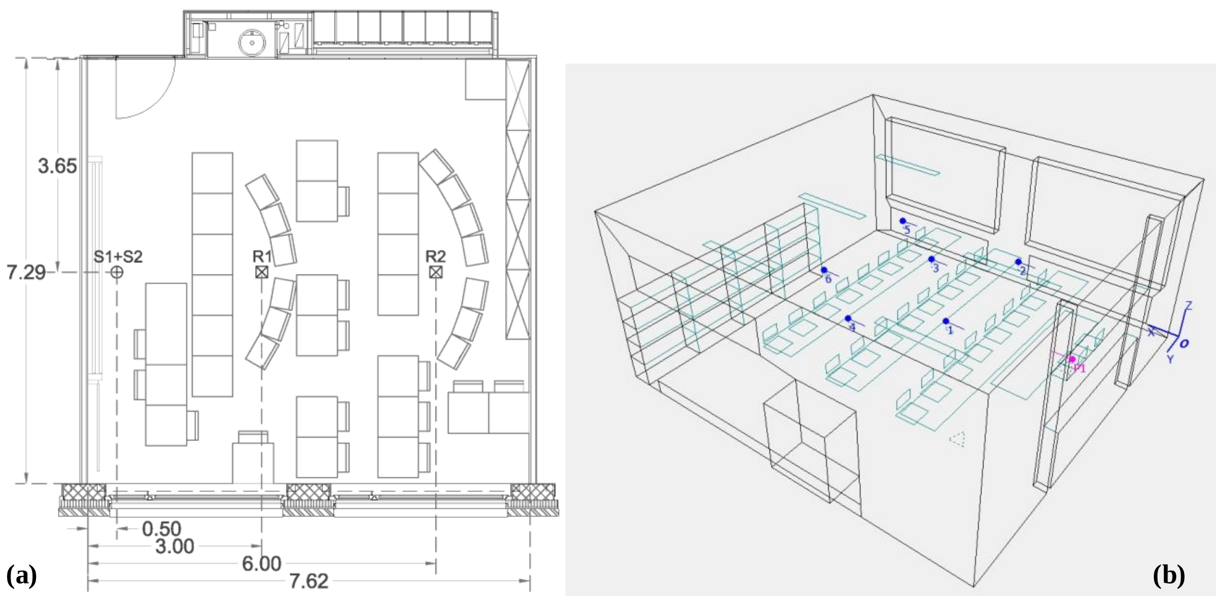

Figure 1: (a) Classroom plan with indicated the listening and the sources positions; (b) geometrical model of the classrooms imported in the CAD acoustic software.

2.4 Tests in auralized conditions: calibration of the model

A virtual model of the existing classroom was created using the room acoustics software Odeon® v14.01. The software employs a hybrid approach that combines, below a selected reflection transition order, a mixture of the image source method and ray-tracing and, above this transition order, a special ray tracing process that generates secondary sources radiating energy locally from the surfaces (ray-radiosity).

Firstly, a geometric model made of 261 surfaces was created in SketchUp® and then imported in the acoustic CAD software (Figure 1b). The geometric model included, besides boundary surfaces, also wooden desks, chairs and all the furnishing elements of the classroom that could be relevant for the acoustic simulation (e.g. lighting fixtures, radiators, shelves). Initial absorption coefficients were assigned to surfaces and objects based on the Odeon® material library and on data available from literature. A mid-frequency scattering coefficient of 0.05 was assigned to all boundary surfaces; for unoccupied desks and chairs a value of 0.5 was chosen (Astolfi *et al.*, 2008). A speech source with the directivity pattern of a human talker (*Tlknorm* in Odeon®) and emitting a signal spectrally shaped to match a female talker (International Electrotechnical Commission, 2011) was defined for the calculation of the room acoustics parameters. The virtual source was located at the same position as in the existing classroom.

A preliminary calibration of the virtual model in unoccupied conditions was initially performed. To the scope, six receiver positions were defined in the audience (Figure 1b) and measures in the real classroom with omnidirectional, B&K type 4189 1/2 inch microphones were achieved at the same locations (height of receivers: 1.25 m) with the speech source in the same position as during the *in situ* experiment. Simulations were performed with a transition order of two, with 2000 early rays and 16000 late rays. During the calibration process, the acoustical material properties were step-by-step adjusted, still keeping physically realistic values, until differences between measured and simulated values of the selected acoustical parameters smaller than the Just Noticeable Differences (JND) defined by the ISO 3382-1 standard (International Organization of Standardization, 2003) were obtained. The acoustical parameters reverberation time (EDT, T_{30}) and speech clarity (C_{50}) were selected as relevant indicators for the calibration, which was performed separately for each listening position. It was obtained that for all positions and acoustic parameters the difference between measured and simulated values was smaller than the corresponding JND (5% for reverberation time, 1 dB for clarity).

Then, the virtual model of the classroom in occupied conditions was set up and further calibrated. The noise source was added to the model and set as omnidirectional (*Omni* in Odeon®), replicating the directivity pattern of the loudspeaker used in the real classroom. Two omnidirectional receiver points were created, corresponding to R1 and R2. The scattering coefficient of the chairs was modified and set to 0.7 (Astolfi *et al.*, 2008) to account for the presence of seated persons. Air temperature and relative humidity were set according to average values measured during the *in situ* tests (T=23°C, RH=23%). Simulations were performed with a transition order of two, with 2000 early rays and 16000 late rays. The virtual model in occupied conditions was calibrated with reference to the measured octave-band values of T_{30}, spatially averaged across the two monaural receivers. Then, the model calibration was further checked by considering, separately for each position, the comparison between the acoustical conditions in the real and the simulated classroom with reference to EDT and C_{50} (which, differently from T_{30}, are expected to vary with the listening position). Measured and simulated values (average value over 500–2000 Hz octave bands) are reported in Table 1. For both listening positions, and all acoustic parameters the differences were always smaller than one JND. Then, the virtual model was considered as properly calibrated, with an accuracy deemed appropriate for the scope of the work.

Table 1: Comparison between real and simulated parameters (T_{30}, EDT, C_{50} – average value over the 500–2000 Hz octave bands) at the two listening positions (R1 and R2).

Listening position	T_{30} [s]		EDT [s]		C_{50} (dB)	
	R1	R2	R1	R2	R1	R2
real classroom	0.82	0.84	0.81	0.86	3.6	0.4
virtual model	0.80	0.82	0.78	0.83	3.7	0.8

2.5 Tests in auralized conditions: listening conditions and procedures

After the calibration, auralized listening conditions were created in the virtual room by convolving the same anechoic speech signal and noise as used for the *in situ* experiment, with the simulated binaural impulse responses (BRIRs) for both speaker and noise sources.

Firstly, the sound power level of the virtual sources was defined. To the scope, it was required that the same sound pressure levels as measured *in situ*, 1 m away from the sources were also measured in the virtual model. Then, a virtual listener was defined in the acoustical CAD models having the head-related-transfer-functions (HRTFs) of the B&K type 4100 head and torso simulator, which were already available from previous measures. The auralization procedure involved creating separate BRIRs at each selected listening position within the virtual classroom, for both speech and noise sources; the BRIRs were then convolved with the corresponding anechoic material.

Real and auralized listening conditions are summarized in Table 2. It is relevant that, due to a proper calibrated virtual model, also the differences between measured and simulated STI values were smaller than the JND of 0.04 defined by the International Electrotechnical Commission (2011).

Table 2: Listening conditions within the real and the auralized classroom.

Classroom	Receiver	T_{30} [s]	Speech level dB(A)	Noise level dB(A)	SNR	STI
real classroom	R1	0.80	61.0	60.9	0.1	0.52
	R2	0.82	57.4	58.7	-1.3	0.45
virtual model	R1	0.82	61.0	61.0	0.0	0.52
	R2	0.84	57.4	58.7	-1.3	0.46

The same panel of testers taking part in the *in situ* tests also performed the auralized experiments in a quiet laboratory environment. The presentation of the stimuli and the data collection was controlled by the same wireless test bench as described in Sec. 2.3. The experimental set up was calibrated placing the headphones over a B&K type 4100 head and torso simulator.
The experimental session was held almost two months after the *in situ* listening tests, with groups of a maximum of four people at a time, following the same procedure as described in Sec. 2.3. Firstly, a training session was proposed; afterwards participants completed two lists of 18 words, each one proposed in a different listening condition. After the completion of each test list, the participants were asked to rate the subjective listening effort over a 10-points scale. Words lists and listening conditions were randomized across the groups of participants.

2.6 Statistical analysis

Data analysis was performed using generalized mixed-effects models (GLMM), chosen on the account of the repeated measures design of the experiment and the non-normal distribution of the response variables. The software *R* was used for the analysis (packages *lme4*, *lsmeans*, *ordinal*); a significance level α=0.05 was always set. In particular, a GLMM with a binomial distribution was used to analyze IS data, whereas RT results were analyzed using a Gamma distribution with a log-link function; the analysis of LE data was accomplished with a cumulative link mixed model. Model selection was based on a forward procedure using the likelihood ratio test, and the statistical assumptions of the final models were verified by checking the normality of the residuals. When appropriate, planned pairwise comparisons were performed, correcting for the test multiplicity using a Benjamini-Hochberg procedure.
Prior to data analyses, RT data greater than 5000 ms and corresponding to "none of the two" responses were removed from the dataset (2.8% of the data). For the analysis of IS results, the responses were coded using a binary score (0/1, corresponding to incorrect/correct); the selection of "none of the two" was considered an incorrect response.

3. RESULTS

In the setup of the statistical models, listening position (R1 *vs*. R2), mode of presentation (in situ *vs*. auralized), and their interaction were considered fixed factors. Participants were considered a random factor; a random slope was also specified, supposing that the effect of the mode of presentation might be different for each participant. The descriptive statistics of the measured IS, RT and LE data averaged over participants for the two experimental conditions are showed in Figure 2.

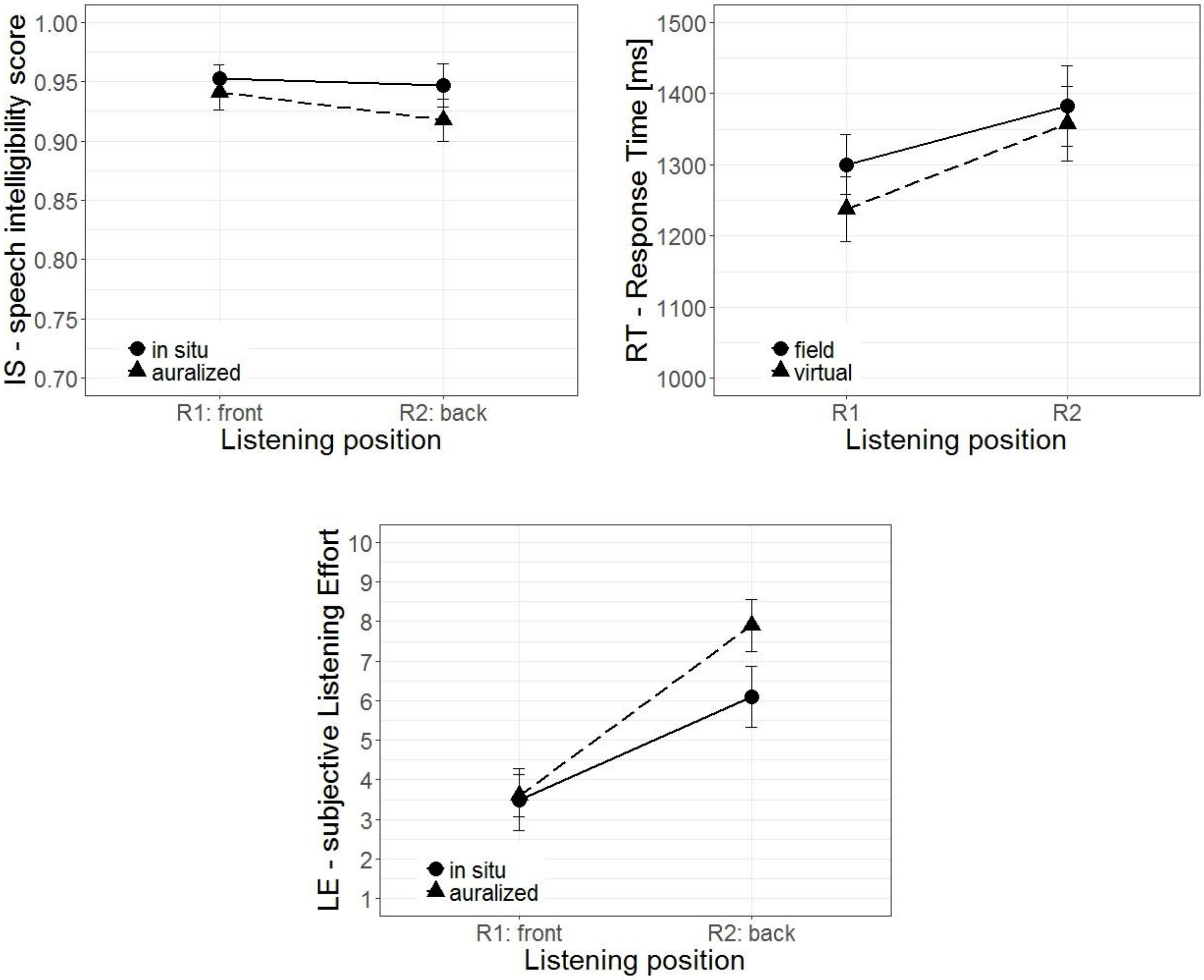

Figure 2: Mean results of the listening tests in the real and the virtual classrooms averaged across the subjects: speech intelligibility, response time and self-rated listening effort. The data are divided according to the listening position (R1, R2). The error bars represent the 95% confidence intervals between participants.

The statistical analysis of IS results revealed that the effects of mode of presentation ($\chi^2(1)=1.31$, $p=0.25$), listening position ($\chi^2(1)=0.67$, $p=0.41$) and the interaction ($\chi^2(1)=0.14$, $p=0.71$) were not significant.
In the analysis of RT data a significant effect of position ($\chi^2(1)=13.28$, $p<0.001$) was found, indicating that, when results were collapsed across mode of presentation, participants always showed significantly smaller RTs in position R1 versus position R2. The estimated mean difference in RT between the two listening positions was 92 ms. No effects of mode of presentation ($\chi^2(1)=1.58$, $p=0.21$) and interaction were found ($\chi^2(1)=1.29$, $p=0.26$). Indeed, the RT difference between the two modes of presentation was equal to 62 ms in R1 and 26 ms in R2, to be compared with a within-subjects standard deviation of 139 ms and 172 ms, respectively in the two positions.
Finally, the statistical analysis of the LE ratings showed a significant effect of listening position ($p<0.001$) and of the interaction between listening position and mode of presentation ($p=0.01$); the effect of mode of presentation was not significant ($p=0.14$). The presence of the significant interaction indicates that depending on the mode of presentation, the participants differently rated the perceived effort in position R1 and R2. The results of the *post hoc* tests indicated that, for position R2 alone, LE ratings in higher (more effortful) categories were more likely for the auralized tests than for the *in situ* tests ($z=2.08$, $p=0.037$). No significant difference in the LE ratings was found at position R1 ($z=0.04$, $p=0.70$). When examining the pairwise comparisons between positions within each mode of presentation it was found that, in both cases, the difference in the LE ratings was significant (*in situ*: $z=2.60$, $p=0.009$; auralized: $z=3.65$, $p<0.001$) pointing out that ratings of higher perceived effort were more likely in the back of the classroom.

The results of the statistical analysis are summarized in Table 3.

Table 3: Results of the statistical analysis for the three metrics (IS, RT, LE). Dashes indicates that the corresponding effect was not significant. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Effect	IS	RT	LE
listening position (R1 *vs*. R2)	-	$RT_{R1} < RT_{R2}$***	$LE_{R1} < LE_{R2}$***
mode of presentation (real *vs*. auralized)	-	-	-
interaction: position X mode	-	-	R1: - R2: $LE_{real} < LE_{auralized}$**

4. DISCUSSION

The output of the statistical analyses allows drawing insights on the potential of the auralization techniques in recreating environments where, as regards a speech perception task, the same IS, RT and LE results as *in situ* can be obtained.
A virtual model of an existing classroom was created, and carefully calibrated in accordance to the state-of-the-art literature, with the aim of obtaining auralized sound fields that could be considered undistinguishable from the real ones as regards the IS output. Indeed, as expected, the IS results in auralized conditions matched the corresponding ones obtained with *in situ* testing, confirming that when properly calibrated virtual models are used, where the differences between auralized and measured relevant acoustic parameters are smaller than the JND, the same speech intelligibility as in real settings is obtained. No effect of listening position was obtained for IS. The participants performed the speech reception task with the same, close to the maximum accuracy in both R1 and R2, even though the STI gap between the two positions (ΔSTI=0.07) was greater than the JND of 0.04 and thus, a corresponding IS variation might be expected. However, it has to be considered where the absolute values of the objective metric locate on the psychometric curve (i.e. the sigmoid curve relating STI and IS results) (International Electrotechnical Commission, 2011; Steeneken, 2014). Indeed, given the sigmoid shape of the curve, when moving towards higher STI values fewer differences will be observed in the corresponding IS results, until for the highest STI values, it will undergo a ceiling effect.
As regards the RT metric, no effect of mode of presentation was found, indicating that, even for this quantity, upon proper calibration of the virtual model the same absolute values as measured *in situ* could be replicated in auralized conditions. Interestingly, the main effect of listening position was found to be significant for RT. The finding points out that, unlike IS and independently on the mode of presentation, relaying of the RT results it was possible to discriminate between front and rear position of the classroom. The RTs measured in the back position were always greater than the RTs measured in the front position, with a mean increase of 92 ms, indicating that whereas no difference was observed in the number of words correctly recognized, a greater cognitive load was experienced in the back of the classroom.
Finally, with reference to the LE results, the statistical analysis returned a significant interaction between mode of presentation and position. In particular, no difference was found in the LE ratings in R1 whereas the back position was rated as significantly more effortful in auralized than in real conditions (ΔLE=2). The result probably stemmed from the subjective nature of the LE rating, that beside the effect of the listening condition also reflects individual, extra-acoustic factors; the different experimental set up or the audio-visual interaction (for the *in situ* experiment) might indeed affect the ratings. In fact the tests in the lab did not incude a classroom-related visual feedback and the proximity of the other testers that was realized in the field tests was only loose in the lab case. Furthermore, recent studies pointed out that the listening effort depends not only on the task demands but also on an individual cost/benefit evaluation (Pichora-Fuller, 2016). In this sense, the LE rating will reflect the listening demands in relation to the participant auditory and cognitive abilities, but also the more subjective aspect of the participant appraisal of its capacity to meet the demands.

5. CONCLUSIONS

In this study two metrics (RT and LE) have been introduced, which are supposed to be informative of aspects of the listening effort construct. Such measures beyond accuracy (as tracked by IS) in pursuing an acoustical design tailored on the occupants' needs. In particular, the correspondence between RT and LE results obtained with *in situ* speech-in-noise tests and with tests in auralized conditions was examined, with reference to a university classroom and native, normal hearing listeners. Relying on the study results, two main observation can be made:

- Upon proper calibration of the virtual model, IS and RT results obtained in the real setting could be replicated, without significant differences in auralized conditions. Conversely, a significant difference was found in the LE ratings between the two modes of presentation, depending on the listening position. The result is driven by the subjective nature of LE that, beside acoustic conditions, also reflect individual, extra-acoustic factors. An ecological validation was thus obtained for the first two metrics alone, suggesting that RT could be a reliable metric to be introduced in the framework of an enhanced acoustic design.
- The effect of listening position, which was not apparent relying on IS results, was found to be significant using both RT and LE, indicating a change in the amount of processing resources involved in the speech reception task, even though the same amount of words was correctly perceived. Introducing listening effort metrics in the field of the room acoustic design is then a valuable strategy that allow discerning between listening conditions equivalent as regards speech intelligibility and to detect the effects of room acoustic changes before the level where words cannot be identified (Visentin *et al.*, 2018).

REFERENCES

Astolfi, A., Corrado, V., & Griginis, A. (2008). Comparison between measured and calculated parameters for the acoustical characterization of small classrooms. *Appl. Acoust.*, *69*(11), 966–976.

Bess, F. H., & Hornsby, B. W. (2014). Commentary: Listening can be exhausting—Fatigue in children and adults with hearing loss. *Ear Hear.*, *35*(6), 592–599.

Bonaventura, P., Paoloni, F., Canavesio, F., & Usai, P. (1986). Realizzazione di un test diagnostico di intelligibilità per la lingua italiana (*Development of a diagnostic intelligibility test for the Italian language*). International Technical Report No. 3C1286, Fondazione Ugo Bordoni, Rome.

Deutsches Insistute fur Normung. (2011). *Hörsamkeit in Räumen – Anforderungen, Empfehlungen und Hinweise für die Planung (Acoustic quality in rooms – Specifications and instructions for the room acoustic design): DIN 18041:2016*. Berlin, Germany.

Houben, R., van Doorn-Bierman, M., Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *Int. J. Audiol.*, *52*(11), 753–761.

International Electrotechnical Commission. (2011). *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index: IEC 60286-16*. Geneva, Switzerland.

International Organization of Standardization. (2003). *Ergonomics–assessment of speech communication: ISO3382–1*. Geneva, Switzerland.

International Organization of Standardization. (2009). *Acoustics–Measurements of room acoustic parameters – Part 1: Performance spaces: ISO3382–1*. Geneva, Switzerland.

Kahneman, D. (1973). *Attention and Effort*. Vol. 1063. Englewood Cliffs, NJ: Prentice-Hall.

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: concepts and age-related considerations. *Ear Hear, 37*, 77S–84S.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int. J. Audiol.*, *53*(7), 433–440.

Pals, C., Sarampalis, A., van Rijn, H., & Başkent, D. (2015). Validation of a simple response-time measure of listening effort. *J. Acoust. Soc. Am.*, *138*(3), EL187–EL192.

Pichora-Fuller, M. K. (2016). How social psychological factors may modulate auditory and cognitive functioning during listening. *Ear Hear.*, *37*, 92S–100S.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... & Naylor, G. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear.*, *37*, 5S–27S.

Postma, B. N., Katz, B. F. (2016). Perceptive and objective evaluation of calibrated room acoustic simulation auralizations. *J. Acoust. Soc. Am.*, *140*(6), 4326–4337.

Prodi, N., Visentin, C., & Feletti, A. (2013). On the perception of speech in primary school classrooms: Ranking of noise interference and of age influence. *J. Acoust. Soc. Am.*, *133*(1), 255–268.

Rychtáriková, M., Van den Bogaert, T., Vermeir, G., & Wouters, J. (2011). Perceptual validation of virtual room acoustics: Sound localisation and speech understanding. *Appl. Acoust.*, *72*(4), 196-204.

Steeneken, H.J.M. (2014). Forty years of speech in intelligibility assessment (and some history). *Proceedings of IOA 40th Anniversary Conference*. Birmingham, UK.

Surprenant, A. M. (1999). The effect of noise on memory for spoken syllables. *Int. J Psychol.*, *34*, 328–333.

van den Tillaart-Haverkate, M., de Ronde-Brons, I., Dreschler, W. A., & Houben, R. (2017). The Influence of Noise Reduction on Speech Intelligibility, Response Times to Speech, and Perceived Listening Effort in Normal-Hearing Listeners. *Trends Hear.*, *21*, 1–13.

Visentin, C., Prodi, N., Cappelletti, F., Torresin, S., & Gasparella, A. (2018). Using listening effort assessment in the acoustical design of rooms for speech. *Build. Env.*, *136*, 38–53.

Yang, W., & Hodgson, M. (2007). Validation of the auralization technique: Comparative speech-intelligibility tests in real and virtual classrooms. *Acta Acust united Ac.*, *93*(6), 991–999.

ACKNOWLEDGEMENT

This study was funded by the internal project "Human-Centered Design of the Built Environment: definition of a methodology for the experimental assessment of the overall Indoor Environmental Quality" of the Free University of Bozen-Bolzano.