On the acoustical design of university classrooms: effects of room acoustics on behavioral and subjective indices of listening effort

Chiara Visentin, Nicola Prodi
Department of Engineering, University of Ferrara, Ferrara, Italy.

Francesca Cappelletti
Department of Design and Planning in Complex Environments, University Iuav of Venezia, Venezia, Italy.

Simone Torresin, Andrea Gasparella
Faculty of Science and Technology, Free University of Bolzano/Bozen, Bolzano/Bozen, Italy.

Summary
The acoustic design of university classrooms, besides targeting a high speech intelligibility, should also ensure a minimal effort in the speech reception task. An effortful listening, which involves an increased amount of cognitive resources, may compromise students' learning and academic achievements, especially so for non-native listeners. The aim of this study is investigating the effect of room acoustic properties on the speech reception performance of university students, with specific reference to the listening effort as measured by either behavioral or subjective metrics. To achieve this objective the acoustic simulation model of a university classroom with a volume of 198 m^3 and acoustical treatment on a lateral wall was calibrated with field measurements of T30 and C50 in order to obtain realistic auralization. The room acoustic conditions were virtually modified by either changing the room shape or removing the acoustic treatment. The simulated binaural impulse responses were used for laboratory experiments with headphones. Consonant confusion tests (Diagnostic Rhyme Tests) in the Italian language were proposed to 21 students, aged 23 to 39 years and self-reporting normal hearing: 10 native speakers and 11 non-native (German) speakers. A speech-shaped stationary noise was used to mask the speech signal, and listening conditions with Speech Transmission Index values ranging between 0.42 and 0.52 were obtained. During the experiment, data on the number of words correctly recognized (speech intelligibility, IS), response times (RT) and subjective ratings of listening effort (LE) were collected. No differences were found between the compared acoustic conditions in terms of IS results. The effects of room acoustic properties were instead discriminated when the listening effort was considered, with the greatest number of significant differences observed with the behavioral metric of RT. A decrease in IS results was found in all conditions for non-native listeners; no differences were observed in the corresponding listening effort results.

PACS no. 43.55.Hy, 43.71.Gv

1. Introduction

The normative approach to the acoustical design of rooms for speech bases on either objective acoustical indices (e.g. reverberation time, Speech Transmission Index–STI) or the accuracy in a speech reception task (i.e. amount of words correctly recognized or speech intelligibility score, IS). However, this approach does not warrant per se a comfortable speech communication as it disregards the amount of cognitive resources required to perform the task. This aspect, which is named "listening effort", was recently defined as "the level of processing resource allocation to overcome obstacles in goal pursuit when performing a listening task" [1]. Interestingly the listening effort does not just mirror the changes of speech reception accuracy but may also 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. Owing to the limited availability of personal cognitive capacity [2], 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, etc.). This aspect is especially important in classrooms: when high levels of effort have to be sustained for long periods (e.g. during lessons), fatigue may arise with negative consequences on learning and cognitive achievements of listeners [3].

Many factors affect the construct of listening effort [1] and, to date no single measure is available to capture this multifaceted experience. Over the years, several methods have been proposed for the scope [1, 3], 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 [4], 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 not only on performance accuracy (as traced by IS) but also on feasible estimates of the listening effort: the subjective rating of listening effort (LE) and the response time (RT) to the auditory stimulus measured in a single-task paradigm. This latter quantity proved to reflect the amount of resources required to interpret and respond to the incoming signal [5, 6]. Since RT and LE potentially carry complementary information to IS, they could be used to improve the means of evaluation of rooms for speech. In particular, in this work realistic interventions that could be implemented in university classrooms for optimization purposes (acoustical treatment of a wall, change of room size) were simulated in a virtual model and comparative speech-in-noise tests were performed using the auralized outputs. The IS, RT and LE results were then compared to assess the advantages of an acoustical design also based on listening effort metrics.

2. Materials and methods

2.1. Participants

The experiment was proposed to 21 normal-hearing young adults, recruited among the students and the academic staff of the Free University of Bolzano/Bozen. Based on their self-declared mother tongue the participants were divided into two groups: 10 native Italian speakers (5 female, 5 male; mean age: 24.4 years) and 11 native German speakers (6 female, 5 male; mean age: 25.9 years.). In the following, the groups will be named NI and NG respectively. All the NG participants started the acquisition of the Italian language before the age of eight, and spoke Italian daily. They self-rated their proficiency in the Italian listening on a 7-point scale (7 corresponding to "mother tongue"); the resulting median value was 5.0 (interquartile range: 1.25).

2.2. Speech material

The speech material of the Diagnostic Rhyme Test (DRT) [7] in the Italian language was used in the experiment. The test bases on 105 pairs of rhymed words (e.g. *lupi*/*rupi*), one of which is presented embedded in a carrier phrase (*Ora diremo la parola lupi*, which is Italian for "Now we will say the word wolves").

The sentences were recorded by an adult, native Italian female speaker, within a silent room at a sampling rate of 44.1 kHz. All of the sequences were filtered as to match the long-term spectrum of a female speaker [8], and set to the same root mean square value. The recordings were then organized into five test lists of 18 words each, and a shorter training list of 15 words.

2.3. Classroom virtual models

An existing classroom, part of the Classroom Spaces Living Lab of the Free University of Bolzano/Bozen, was modeled using the room acoustics software Odeon® v14.01. The room has a rectangular plan (7.92 x 7.62) m and a height of 3.55 m, resulting in a volume of 197 m^3. It is characterized by flat surfaces (ceiling: unpainted concrete, floor: linoleum finishing, walls: painted plasterboard). The lateral partition with the adjacent corridor is acoustically treated with a Topakustik® 6/2 sound absorbing paneling. The classroom model included, besides surfaces, also desks and chairs for 25 students. The model was calibrated with reference to the T_{30} values measured in the room in occupied conditions. During the calibration, the absorbing and scattering properties initially assigned to the materials were adjusted, still keeping realistic values, in order to obtain differences between measured and simulated reverberation time values smaller than the Just Noticeable Difference (JND) of 5% defined in [9]. The model was further verified by checking that the differences between measured and simulated C_{50} values were smaller than the JND of 1 dB [9].
Starting from the calibrated virtual model (named C1 in the following), two other classroom models were created by changing either the room acoustic properties (model C2, without acoustical treatment on the lateral surface), or the room size (model C3, with a doubled volume along the longitudinal direction). Within the virtual models, two sources were defined: a speech source, having the directivity pattern of a human talker (*Tlknorm* in Odeon®) and located at the teacher's desk (height: 1.50 m) and a noise source, modeled as omnidirectional (*Omni* in Odeon®) and positioned on the floor directly below the speech source. Then, the virtual receivers were set, located in the front (R1) or in the back (R2, R3) of the classroom. The receivers R1 and R2 (defined for models C1 and C2 alone) were positioned at 2.50 and 5.50 m from the speech source, respectively. In model C3 only R3 was set, which was located, similarly to R2, 1.62 m from the end wall of the room (13.12 m from the sources). The same height (1.20 m) was set for all receivers. In order to obtain binaural impulse responses (BRIRs) at the receivers, the head-related-transfer-functions (HRTFs) of a B&K type 4100 head and torso simulator were used, which were already available from previous measures.

2.4. Listening conditions

Auralized listening conditions in the three virtual classrooms were created by convolving the anechoic speech material and noise (stationary noise, with the same spectrum of the speech signal) with the simulated BRIRs at the receiving positions for both speaker and noise sources.
The sound power levels of the sources were set in model C1 and maintained in the other two models. With reference to the speech source, the level was set to obtain 63 dB(A) at 1 m in front of the source, corresponding to a vocal effort of a speaker between "normal" and "raised" [10]. For the noise source, the sound power was fixed to measure a signal-to-noise ratio (SNR) of 0 dB at R1.
The auralized listening conditions are detailed in Table I. It is worth noticing that the STI values realized in the diverse listening conditions correspond to an intelligibility rated as "Poor" (STI <0.45) or "Fair" (0.45 < STI < 0.6) [10].

Table I. Listening conditions in the three virtual classrooms.

Virtual classroom	T30 [s]	Receiver	SNR (dB)	STI
C1	0.81	R1	0.1	0.52
		R2	-1.3	0.45
C2	1.21	R1	0.6	0.49
		R2	-0.1	0.42
C3	0.88	R3	0.5	0.48

2.5. Procedure

The experiment was performed via headphones, with groups of a maximum of four people at a time, in a quiet laboratory environment. It was managed using a wireless test bench [11], simultaneously controlling the audio playback and the data collection. The test set up was calibrated by placing the headphones over a B&K type 4100 head and torso simulator.
The participants were given a touchscreen handset for the response selection. They listened to a test sentence (carrier phrase + target word) and then selected one of the three alternatives appeared on the screen at the audio offset. The alternatives were the correct word, the rhymed word and the "none of the two" choice. Once all participants within a groups gave the response (or reached the 10 s timeout), the next sentence was played back. The participants were instructed to pay attention to the task, but were not urged to give a response as

quickly as possible. Firstly, the training session was proposed; afterwards the participants completed five lists of 18 words, each one proposed in a different listening condition. After each list, they were asked to rate their perceived listening effort (LE) on a 10-point scale, ranging from *minimum effort* (1) to *maximum effort* (10). Word lists and listening conditions were randomized across the groups of participants.

During the experiment the following data were collected: word scores (correct/incorrect/none of the two), response time (time elapsed between the audio offset and the response selection), and subjective listening effort.

2.6. Statistical analysis

Data analysis were performed using generalized mixed-effects models (GLMM), chosen on the account of the repeated measures design of the experiment and the not-normal distribution of the response variables.

The software R [12] was used for the analysis (packages *lme4*, *lsmeans*, *ordinal*); a significance level $\alpha=0.05$ 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 (2.9% of the data).

3. Results

For the analysis of the results, two separate GLMM models were set up, exploring diverse effects. The first model was devoted to analyze the effects of the acoustic treatment of the lateral wall on the response variables, and took into account the results of the listening tests in the virtual classrooms C1 and C2. In the model acoustical treatment, listening position, mother tongue and their interactions were entered as fixed factors; participants were considered a random factor. The second model explored instead the effects of room shape with reference to the rear listening position (R2 for C1 and C2 models, R3 for C3). Classroom type, mother tongue and their interaction were considered as fixed effects; participants were again considered as a random effect.

The descriptive statistics of the collected data, averaged across participants are depicted in Figure 1 and Figure 2, respectively for the two analysis. The results of the statistical analyses are instead summarized in Table II.

Table II. Results of the statistical analyses investigating for the three response variables: (a) the effects of the acoustical treatment of a lateral wall, (b) the effects of room shape. In the former case the main effects of acoustical treatment of the lateral wall (C1 – with, C2 – without treatment), mother tongue (NI *vs.*NG) listening position (R1 *vs.* R2), and their interactions were considered. In the latter case were instead taken into account the main effects of classroom type (C1, C2 and C3), mother tongue (NI *vs.*NG) and their interaction. The dashes indicate that the corresponding effect is not significant; only significant interactions are reported. **p*<0.05, ***p*<0.01, ****p*<0.001.

	Effect	IS	RT	LE
Effects of acoustical treatment	Acoustical treatment	-	$RT_{C1} < RT_{C2}$ ***	-
	Listening position	-	$RT_{R1} < RT_{R2}$ ***	-
	Mother tongue	$IS_{NI} > IS_{NG}$ *	-	$LE_{NI} > LE_{NG}$ *
	Treatment x position	C2: $IS_{R1} > IS_{R2}$ ***	-	$LE_{R1} < LE_{R2}$ *** R1: $LE_{C1} > LE_{C2}$ ***
Effects of room shape	Classroom type	-	$RT_{C2} > RT_{C1} > RT_{C3}$ **	$LE_{C1} < LE_{C3}$ *** $LE_{C2} < LE_{C3}$ **
	Mother tongue	$IS_{NI} > IS_{NG}$ *	-	-

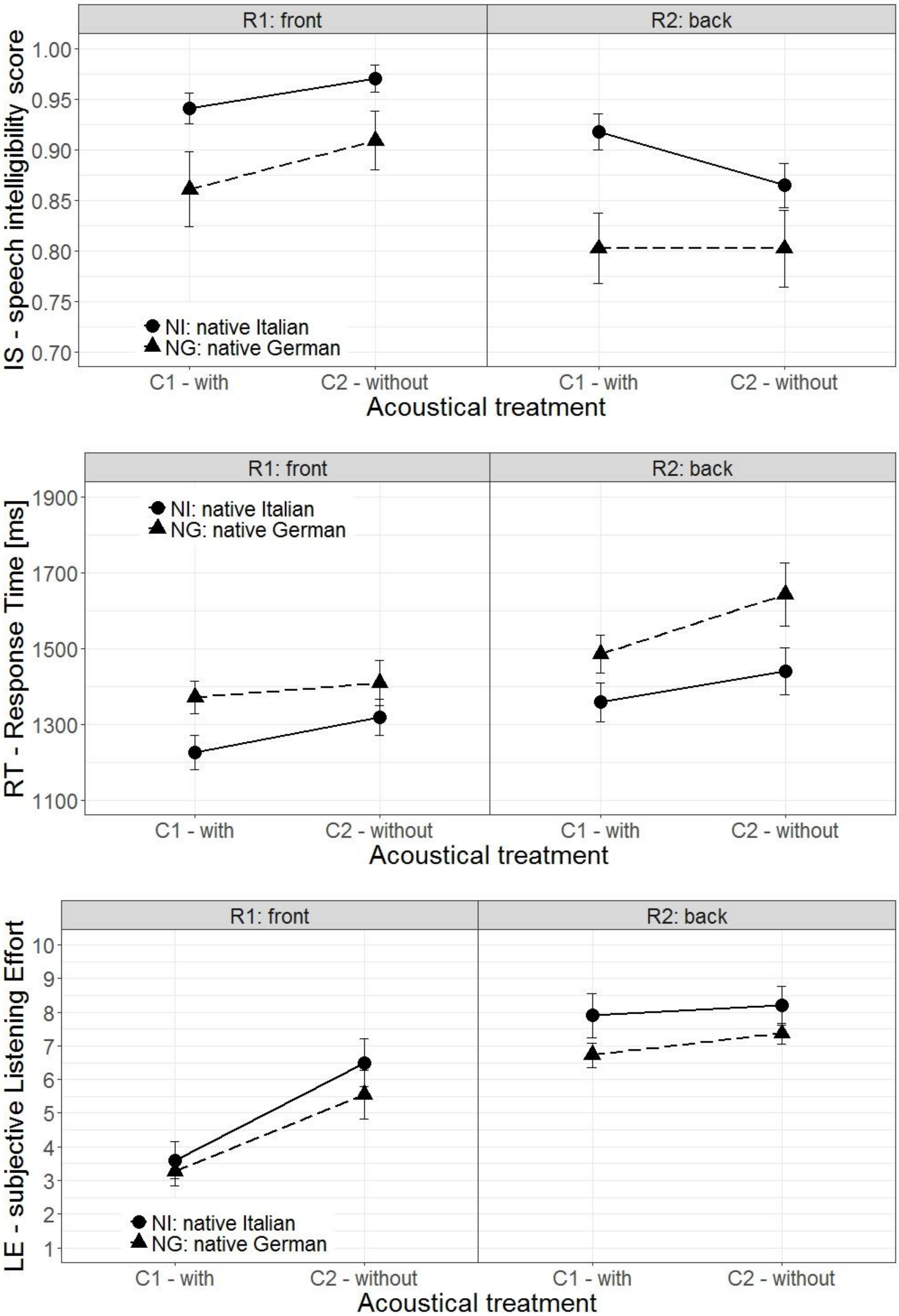

Figure 1. Mean results of the listening tests in the virtual classrooms C1 and C2 averaged across all subjects: speech intelligibility, response time and listening effort. The data are divided according to the participants' mother tongue (NI, NG), the listening position (R1, R2) and the finishing of the lateral wall (with or without acoustical treatment, corresponding to models C1 and C2). The error bars represent the 95% confidence intervals between participants.

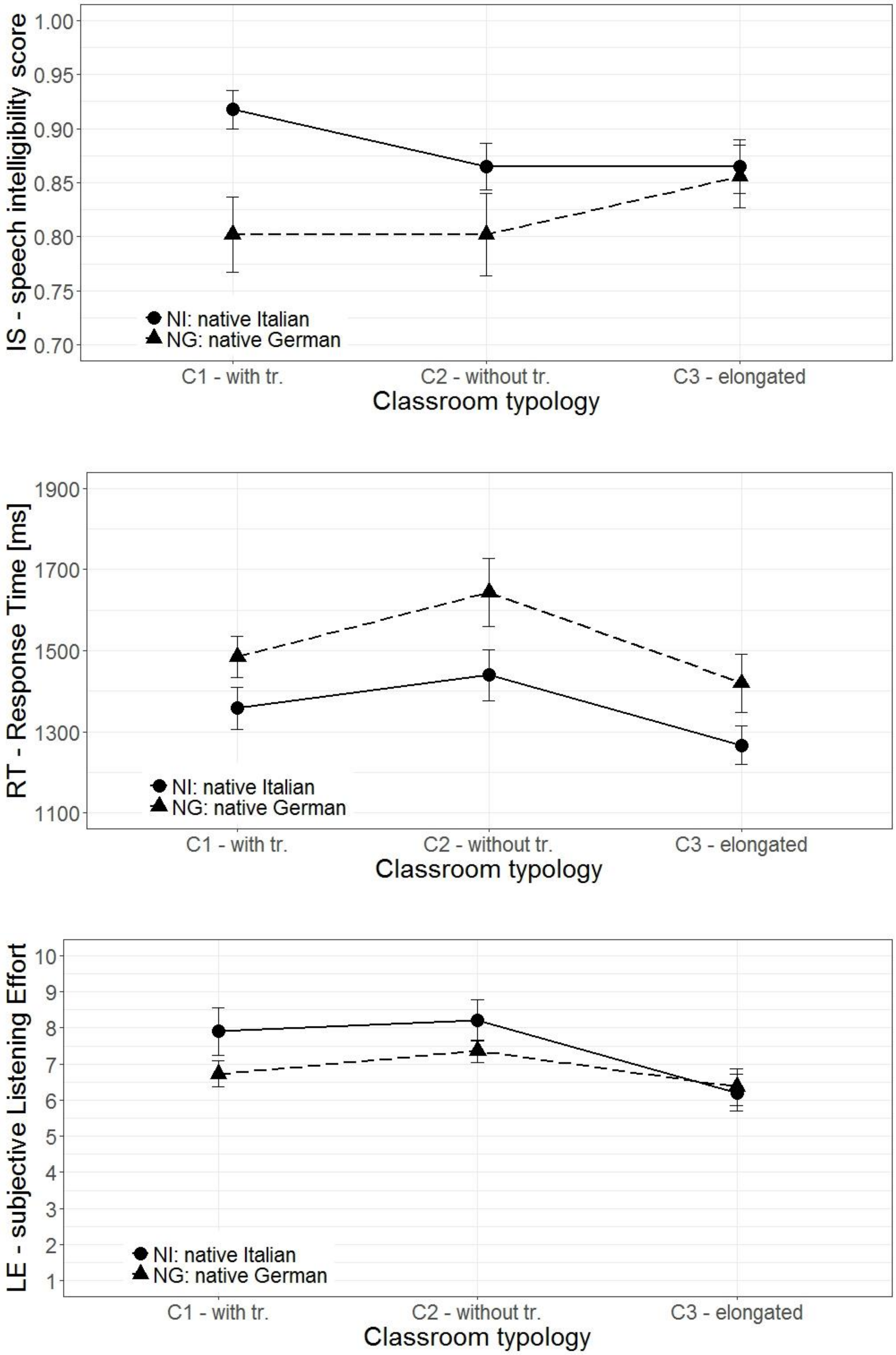

Figure 2. Mean results of the listening tests in the virtual classrooms C1, C2 and C3 in the rear listening position (R2 for C1 and C2, R3 for C3) averaged across all subjects: speech intelligibility, response time and listening effort. The data are divided according to the participants' mother tongue (NI, NG) and the classroom typology: rectangular shape with acoustical treatment (C1), rectangular shape without acoustical treatment (C2), elongated shape with acoustical treatment (C3). The error bars represent the 95% confidence intervals between participants.

4. Discussion

4.1. Effects of room acoustics

The analysis of the IS results pointed out that the metric was not sensitive to the modifications to the room acoustics and no statistically significant difference was found between the virtual models C1 and C2. Indeed, in both listening positions, the increase in the STI value achieved by inserting the acoustical treatment was smaller than the associated JND of 0.04 [8]. In the same way, as regards the STI difference between the listening positions (ΔSTI=0.07), it was found to be the same for both C1 and C2 and greater than the JND. Then, one would expect an almost equivalent measurable reduction in IS in both cases. Instead, a statistically significant difference between the front and rear position was found for the untreated room alone (10% decrease). The finding could be explained by considering the psychometric curve (i.e. the sigmoid curve relating STI and IS results [8]), which undergoes a ceiling effect at high STI values. Then, the same STI gap will result in fewer differences in the corresponding IS results when moving towards higher values of the objective metric.

On the contrary, a significant main effect of both acoustical treatment and listening position was found in the RT results. In particular, the RTs were significantly greater in the classroom without acoustic treatment and in the rear position of both classrooms, indicating that more time was spent to process the auditory information. Then, even though the number of correctly recognized words was not affected by the changes in acoustic conditions or listening position, the RT results imply that the task was cognitively more demanding.

When the self-rated LE was considered, it was found that the presence of acoustical treatment yielded lower (i.e. better) ratings for the anterior listening position alone. In R2, despite the significant increase in the RTs results, the two classroom configurations were rated as similarly effortful. The finding could be ascribed to a minor sensitivity of LE results in the range of unfavorable listening conditions [13].

Similarly to the previous analysis, when the effects of room shape were considered no differences were found in the accuracy results, whereas both the RT and LE data indicated that less demanding listening was achieved in the treated long classroom, with smaller RTs and lower LE ratings in comparison with the treated normal-sized environment. Again, a change in the acoustic configuration of the room that yielded a STI difference lower than the JND between the tested sound fields, was not tracked by the accuracy results, while it was found to affect the listening experience as concerns the reported effort devoted to the task and the response time.

4.2. Effects of mother tongue

The issue of second-language listeners is especially relevant in school settings where the effect of the sound environment sums up with the incomplete linguistic knowledge. The statistical analysis showed that in all listening conditions the NI participants had a significantly greater accuracy than NG participant, with an average increase in IS results of 6%. The result was in line with previous literature studies [14, 15] indicating a disadvantage of the non-native listeners based on inaccurate perceptual processing of non-native words.

Conversely, using the RT metric, the difference between the two linguistic groups did not show up. Even though the descriptive statistics might suggest greater RTs for the group of NG participants, the trend was not confirmed by the statistical analysis. Indeed, an increase in RT results was expected for the non-native listeners, accounting for the interferences from their native language on the lexical or phonetic level [16, 17]. It has to be recalled that in the present study, the proficiency in the non-native listening was assessed through participants' self-ratings and judged on average as "high". Nevertheless, differences might still exist in the individual proficiency, which can only be disclosed by using an objective assessment (e.g. vocabulary testing) [18]. Controlling individual abilities within the statistical analysis using the test results as a covariate, or testing native and non-native listeners at the same IS level would help in better outlining the effects of room acoustics on the RT results. For instance, in [19], the performance of native and non-native listeners was compared in quiet and noisy acoustical conditions where 100% intelligibility was scored by both groups of participants. In this case, non-native listeners always showed greater RTs suggesting that when the same level of accuracy is reached, longer processing times are needed to cope with the task.

Finally, as regards the self-reported measure of listening effort (LE) it is noteworthy that NG participants reported either no differences with NI participants or else a lower degree of perceived effort despite smaller IS results and no differences in the RTs. It can be speculated that the two groups either interpreted the concept of “listening effort” differently or scaled the judgments according to peculiar anchors differently (e.g. the same categorical value was associated with a different degree of effort, depending on the group of participants).

5. Concluding remarks

The results of the present study suggest that using a feasible metric to depict the complexity of listening effort to complement traditional intelligibility results is a valuable strategy, which allows the discrimination of listening conditions equivalent as regards speech intelligibility. In fact, realistic modifications to the room acoustics yielding similar accuracy in word identification were found to change the amount of processing resources involved in the speech reception task, which can be monitored by using the two proposed metrics.
Using the measure of response time it was possible to discriminate more listening conditions than using the self-reported effort. The finding suggests that RT might be a more sensitive metric than LE, providing information that positively contributes to the optimization of rooms for speech. Further details on this experiment are described in [20].

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.

References

[1] M. K. Pichora-Fuller, S. E. Kramer, M. A. Eckert, B. Edwards, B. W. Hornsby, L. E. Humes, et al.: Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). Ear Hear. 37 (2016) 5S-27S.

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

[3] R. McGarrigle, K. J. Munro, P. Dawes, A. J. Stewart, D. R. Moore, J. G. Barry, S. Amitay: 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. (2014).

[4] R. McGarrigle, P. Dawes, A. J. Stewart, S. E. Kuchinsky, K. J. Munro: Pupillometry reveals changes in physiological arousal during a sustained listening task. Psychophysiology 54 (2017) 193–203.

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

[6] C. Pals, A. Sarampalis, H. van Rijn, D. Başkent: Validation of a simple response-time measure of listening effort. J. Acoust. Soc. Am. 138 (2015) EL187-EL192.

[7] P. Bonaventura, F. Paoloni, F. Canavesio, P. Usai: 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, 1986.

[8] IEC 60268-16:2011. Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index.

[9] ISO 3382:2008, Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms.

[10] ISO 9921:2003, Ergonomics – Assessment of speech communication.

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

[12] R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2018).

[13] J. Rennies, H. Schepker, I. Holube, B. Kollmeier: Listening effort and speech intelligibility in listening situations affected by noise and reverberation. J. Acoust. Soc. Am. 136 (2014) 2642–2653.

[14] S. J. Van Wijngaarden, H. J. Steeneken, T. Houtgast: Quantifying the intelligibility of speech in noise for non-native listeners. J. Acoust. Soc. Am. 111 (2002) 1906–1916.

[15] M. L. G. Lecumberri, M. Cooke, A. Cutler: Non-native speech perception in adverse conditions: A review. Speech Commun. 52 (2010) 864–886.

[16] A. Wagner, M. Ernestus: Identification of phonemes: Differences between phoneme classes and the effect of class size. Phonetica 65 (2008) 106–127.

[17] M. Broersma, A. Cutler: Phantom word activation in L2. System 36 (2008) 22–34.

[18] L. Kilman, A. Zekveld, M. Hällgren, J. Rönnberg: The influence of non-native language proficiency on speech perception performance. Front. Psychol. 5 (2014), article 651.

[19] A. Lam, M. Hodgson, N. Prodi, C. Visentin: Effect of classroom acoustics on speech intelligibility and response time: a comparison between native and non-native listeners. Building Acoustics 25 (2018) 35–42.

[20] C. Visentin, N. Prodi, F. Cappelletti, A. Gasparella, S. Torresin: Using listening effort assessment to improve the acoustical design of rooms for speech. Build. Environ. (2018), under review.