ICBEN 18 – 22 JUNE ZURICH 2017

12th ICBEN Congress on
Noise as a Public Health Problem

# Listening efficiency in university classrooms: a comparison between native and non-native listeners

Chiara Visentin[1], Nicola Prodi[1], Simone Torresin[2], Francesca Cappelletti[3], Andrea Gasparella[2]

[1] University of Ferrara, Department of Engineering, Ferrara, Italy
[2] Free University of Bolzano/Bozen, Faculty of Science and Technology, Bolzano/Bozen, Italy
[3] University IUAV of Venezia, Department of Design and Planning in Complex Environments, Venezia, Italy

Corresponding author's e-mail address: chiara.visentin@unife.it

## ABSTRACT

When listening to speech in one's native language a higher intelligibility is expected than when listening in a second language: perceptual and linguistic cues readily available for native listeners may be only partly accessed by non-native ones. In this study, the effects of different types of background noises on speech reception performance are compared between native and non-native listeners. Diagnostic Rhyme Tests (DRT) in the Italian language were proposed inside a university classroom of 197 $m^3$, with a reverberation time in occupied conditions of 0.6 s, complying with the target value suggested by the German DIN18041 standard. A group of 26 normal-hearing young adults participated in the experiment: half of them native (Italian), the other half non-native (German) speakers. Listeners' performance was assessed in three acoustic conditions (ventilation system, stationary, and fluctuating maskers) collecting data on speech intelligibility and response time. The interplay of perceptual and cognitive process in the speech reception process was then described by using the combined metric of listening efficiency.

## INTRODUCTION

Non-native listeners' performance in speech recognition and speech comprehension have been studied since long (see [1] for a review) and standardized tools were provided for predicting non-native listeners' intelligibility in room acoustics [2]. The knowledge available on accuracy tasks refers that non-native listeners suffer sub-optimal acoustical conditions more than native listeners do, and that the disadvantage is modulated by several variables. Non–native language proficiency is clearly a dominant one, but even in the presence of a second language acquisition since childhood (6-8 years) and of a continued exposition thereafter, there might be differences with respect to the performance of natives. Indeed, proper bilinguals are only those people simultaneously learning two languages since infancy, so that languages can be interchanged. Then, given a population that satisfies a certain categorization of native and non-native proficiency, most attention is drawn by the study of the other relevant factors, and primarily of the acoustical conditions. Variables considered in this case are the amount (i.e., level), quality (i.e., spectro-temporal characteristics) and informative content of noise, together with the effect of reverberation. It was found that, even if the structure of the coding mechanism does not vary between native and non-native listeners, the latter suffer from limited phonological, syntactical and lexical competence that impede to build up an unambiguous image of the signal, thus spoiling the word recognition and the extraction of meaning from the acoustically degraded signal.

Despite the large amount of works that explored the multifaceted topic of task accuracy for non-native listeners under adverse acoustics, an evaluation of the dimension of listening effort was not introduced until recently. In particular, in [3] two separate groups of native and non-native university students were presented with speech-in-noise tests, and response time was taken as a comprehensive indicator of listening effort [4]. It was found that non-native listeners displayed longer response times even in almost fully intelligible conditions. In the present study, the former investigation will be extended by considering a broader set of acoustical conditions and by assessing the effect of fluctuating noise on the speech reception performance.

## METHOD

### Participants

Twenty-six young adults participated in the experiment, divided in two homogeneous groups: 13 native Italian speakers (6 female, 7 male; mean age: 24.7 years, σ: 1.8 years) and 13 native German speakers (7 female, 6 male; mean age: 25.6 years, σ: 7.5 years). In the following, the groups will be named NI and NG respectively. None of the participants reported hearing impairments. All of them were recruited among the students and the academic staff of the Free University of Bolzano-Bozen, based on their self-declared mother tongue. They were all Italian citizens born in the bilingual context of South Tyrol and thus living since birth in an Italian/German speaking environment.

Aiming at a better understanding of the NG participants' language background, they were asked to fill out a questionnaire. Information were collected about the age of acquisition of German and Italian languages (*"When you first started speaking German/Italian?"*), their parents' mother tongue, the language most used at home and with friends, and the self-rated proficiency in the Italian listening. The responses analysis confirmed that all of NG participants learned German from birth; the 73.1% of them indicated German as the language most commonly used for communication, followed by Italian (19.2%) and Ladin (7.7%). As regards the skills of the NG listeners in the Italian language, based on the age of acquisition and the parents' mother tongue, two groups could be identified. The first group (NGI) was composed by four participants, declaring as age of acquisition of the Italian language "less than 3 years old"; at least one of their parents spoke Italian as mother tongue. The second group (NGG) was instead composed by 7 participants, who started the acquisition of the Italian language at the primary school (mean age: 6.4 yr, σ: 0.7 yr); their parents' mother tongue was German. The self-rated proficiency was assessed on a category scale, ranging from 1 to 7, with the latter extremity labeled as "mother tongue". All ratings were quite high (median: 5.0, iqr: 1.25), probably due to the prolonged exposure to the Italian language; no significant difference was found in the ratings of the NGI and NGG groups (Wilcoxon signed ranks test, $p$=0.243). No correlation was found between the self-rated proficiency and the results of the experiment.

### Room set up

The experiment took place in a university classroom, part of the Classroom Spaces Living Lab of the Free University of Bolzano. The room was box-shaped, with dimensions (7.3 x 7.6 x 3.6) m, resulting in a volume of 197 $m^3$. It was characterized by flat surfaces (ceiling: unpainted concrete, floor: linoleum, walls: painted plasterboard); the lateral partition with the adjacent corridor was acoustically treated with a Topakustik® 6/2 type finishing. The classroom was furnished with wooden desks and chairs; it was designed for a maximum of 25 students. Measurements in fully occupied conditions returned a mid-frequency reverberation time equal to 0.62 s. The value is close to the target value suggested by the DIN 18041 standard [5] for teaching classrooms of similar volume ($T_{soll}$= 0.56 s); the value refers to mother language, normal-hearing students.

For the experiment, the room was set up as shown in Fig. 1. A B&K type 4720 artificial mouth was placed close to the desk, at a height of 1.5 m, and oriented towards the audience; it was used to deliver the speech signal. Interfering background noises were played back with a B&K type 4292-L omnidirectional source located on the floor, exactly below the speech source. Two measurement positions (R1, R2) were defined within the room, located respectively 2.5 m and 7.1 m away from the loudspeakers. Omnidirectional microphones were positioned at a height of 1.25 m and used for the objective description of the listening conditions. The reverberation time of the classroom ($T_{30}$, averaged across 500-2000 Hz) derived from impulse responses measured with the sine-sweep technique during the experiment (occupancy: 50%), was 0.82 s in R1 and 0.85 s in R2.

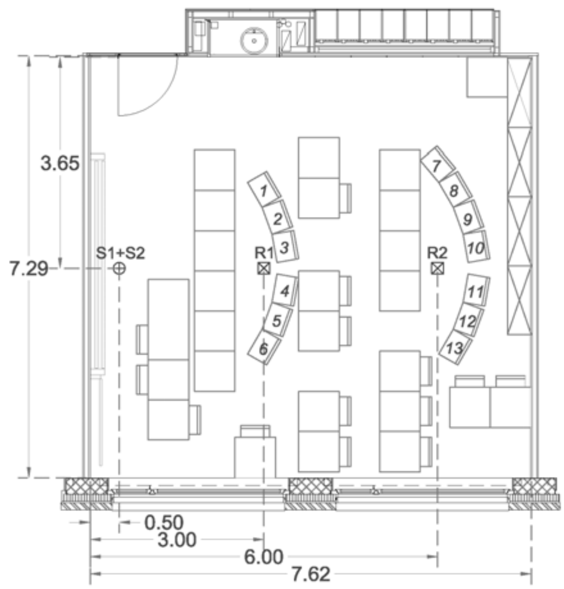


**Figure 1:** Set up of the classroom during the listening tests. The directional source delivering the speech signal and the dodecahedron playing back the masking noises were positioned in S1 and S2 respectively. Two measurement positions (R1 and R2) were set up, around which the participants were seated during the experiment.

### Test material

The Diagnostic Rhyme Test (DRT) [6] in the Italian language was used for the experiment. The test bases on a single, target word embedded in a carrier phrase; the target item is drawn from a corpus of 105 rhyming pairs, all of them meaningful, disyllabic words with a CVCV structure. Within each pair, the distinctive feature of the initial consonant varied, still keeping the consonant-vowel transition. The speech material is optimized as regards phonemic distribution of the Italian language and word familiarity. The test sequences were recorded by an adult, native Italian, female speaker; she was instructed to speak at conversational rate, maintaining a natural prosody and avoiding any emphasis on the final, target word. The recordings took place in a silent room, at a sampling frequency of 44.1 kHz. All of the sequences were filtered as to match the long-term spectrum of a female speaker indicated by the IEC60268-16 standard [2].

In order to equalize intelligibility across all test sequences, preliminary evaluation measurements were done with a group of 18 normal-hearing, Italian speaking, young adults. The recorded sequences were presented in laboratory settings, energetically masked with a stationary noise, shaped as to match the frequency spectrum of the speaker. The Speech Transmission Index (STI) value associated to the listening condition was 0.48, which was expected to yield an Intelligibility Score (IS) close to 90% for the DRT [7]. Only the target words with an average IS higher than 75% were maintained and carefully organized in lists of 18 words each. The procedure ensured a deviation among the average IS of the lists smaller than 1.5 dB. Six test lists were selected for the present experiment.

### Stimuli

For the experiment, the speech level was calibrated at 63 dB(A) at 1 m in front of the source, corresponding to a vocal effort intermediate between “normal” and “raised” [8]. Three background noises were proposed to interfere with the speech reception. The first was the actual ambient noise of the room (A), consisting in the emissions from the classroom projector and the ventilation system. The noise levels measured at R1 and R2 during the experiment were 40.2 dB(A) and 40.7 dB(A) respectively; the corresponding STI values are reported in Tab. 1. The second masker was a steady-state noise (SSN), spectrally shaped to match the long-term spectrum of the speech [2]. Its level was set as to achieve a signal-to-noise ratio (SNR) equal to 0 dB at R1. The third masker was a single-speaker continuous fluctuating noise (ICRA). It was obtained by processing Italian phrases according to the established ICRA instructions [9], as to lose any informational content, and filtered as to match the frequency spectrum of the speaker [2]. Then, the two latter noises had the same spectral properties, but differed in their temporal envelope.

**Table 1:** Objective evaluation of the listening conditions presented during the experiment, for both position R1 and R2. The STI metric is reported for the ambient noise (A) and the stationary noise (SSN), whereas the short-term STIr is used for the fluctuating noise (ICRA). For comparison purposes, STIr values are also shown for SSN.

| Background noise | Listening position | STI | STIr |
|---|---|---|---|
| Ambient | R1 | 0.70 | - |
| | R2 | 0.64 | - |
| SSN | R1 | 0.55 | 0.47 |
| | R2 | 0.46 | 0.40 |
| ICRA | R1 | - | 0.50 |
| | R2 | - | 0.42 |

When masker fluctuations are present, the STI metric cannot provide a reliable description of the listening conditions, as signal and noise fluctuations are not properly disentangled [10].

Therefore, a short-time frame approach was used for the objective characterization of the listening conditions [11, 12]. Speech and noise signals were framed in segments of 186 ms duration (with the frame length reflecting the typical duration of a syllable in the Italian language [13]). For each frame, a STI value was calculated with the indirect method [2]; the average value over the entire recording is named STIr. The ICRA noise level was then set as to provide the same STIr as the SSN noise. The differences in the measured values reported in Tab. 1 are lower than the conventional *just noticeable difference* (JND) value of 0.04 [2]: supposing for STIr the same JND as the one assumed for STI, it is expected that no difference would be perceived between the listening conditions. The resulting SNR in position R1 with ICRA noise was -1 dB.

From the comparison of the STI values, it is observed that for all of the three noises the objective metric underwent a decrease between positions R1 and R2. The gap was comprised in the ΔSTI interval [-0.06;-0.09] for SSN and A noise, and it was equal to ΔSTIr =-0.08 for ICRA noise. Given the difference greater than the JND, it could be expected that a measurable reduction in IS should be obtained for all of the three noises. The STI data can also be compared with the informative Annex G of the Italian technical norm [14], which prescribes STI>0.6 for classrooms. It can be seen that for A noise both R1 and R2 satisfy the prescription on NI listeners. For the NG group, based on the above qualification of proficiency, the STI limit should be increased from 0.6 to 0.68, as indicated by Annex H of Ref. [2]. In this case only position R1 would be comprised in the appropriate range.

### Procedure

The experiment was presented separately to NI and NG listeners, in two subsequent one-hour sessions. During the test, participants sat around the two receiver positions; they were given a touchscreen handset to be used for responses selection. Participants listened to a target word embedded in the carrier phrase and selected one of the three options displayed on the touchscreen (the rhyming pair and the "*none of the two*" option). The *Intelligo* system was used to facilitate data collection and synchronization with the playback system [15]. A training session was firstly proposed, during which all background noises were presented; the aim was to familiarize the participants with the test procedure. During the experiment, the participants completed three test lists, each under a different background noise; they were then invited to change their sitting position (listeners sitting on the back of the classroom were asked to move frontward and vice versa). Then, three different test lists were presented, counterbalancing the order of presentation of the background noises. The procedure ensured that all participants experienced each noise condition at each receiver position, and was exactly replicated for the two listeners' groups.

Speech intelligibility scores (IS) and response times (RT) were calculated for each participant as the average across the 18 words composing a test list. The RT is defined as the time elapsed between the end of the audio reproduction (corresponding to the display of the alternatives on the touchscreen) and the selection of one of the three choices. The listening efficiency (DE) was calculated for each participant as the ratio between IS and RT results [16].

## ANALYSIS AND RESULTS

Prior to statistical analysis, individual RT data were examined as to remove outliers: as participants were not asked to provide a response as quickly as possible, some excessively long RT were observed. An absolute threshold of 4000 ms was set, yielding the rejection of 20 RTs (corresponding to 0.7% of the sample).

Shapiro-Wilk tests revealed that neither IS (undergoing a ceiling effect under the best listening conditions), nor RT data were normally distributed. Indeed, RT distributions are generally expected to deviate from normality and to be positively skewed [17]. Then, non-parametric tests were performed with the *nparcomp* package [18] of the R software, with the aim of assessing the effect of listening position, background noise type and mother language on the listeners' reception performance. Figures 2-4 report the results of the two groups of participants NI and NG as regards respectively IS, RT, and DE.

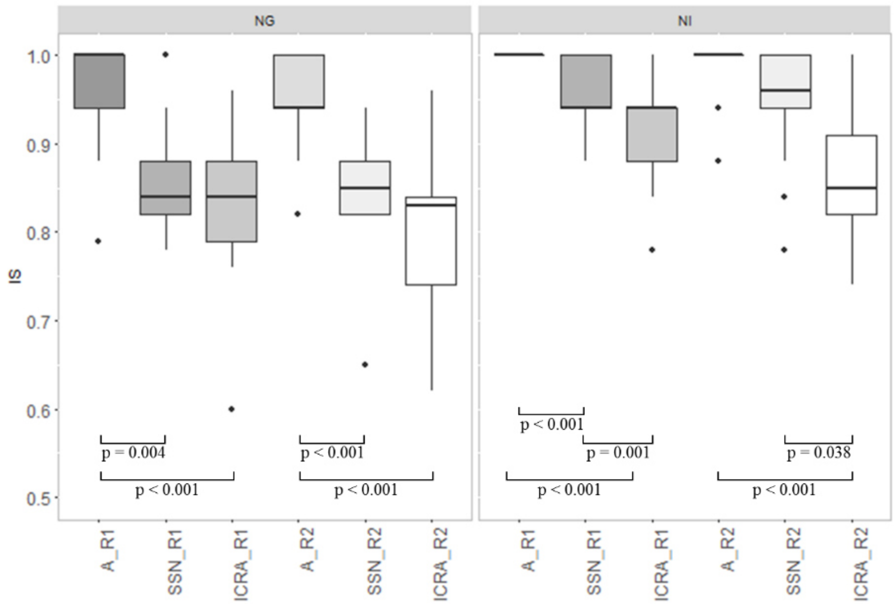


**Figure 2:** Speech intelligibility scores (IS) for native German (NG, on the left) and native Italian (NI, on the right) participants. Results refer to the three masking noises (A: ambient, SSN: stationary, ICRA: fluctuating), in the two listening positions (R1 and R2). The *p*-values associated to significant differences between background noises within the same position are also reported.

### Effects of listening position

The effect of listening position was separately assessed for each group of listeners, by means of 2-sample tests for paired data (function *npar.t.test.paired* of the R package) applied to the measured IS, RT and DE.

Firstly, the NG participants were considered. They displayed not-significant *p*-values in the comparisons between front and back positions. This applied to the three noises and the three metrics here considered. Concerning the results for NI listeners, it is noteworthy that IS in position R1 for noises A and SSN were fully comparable with the reference STI-IS curve for DRT [7], whereas IS higher than expected were found in position R2. Anyway, it should be pointed out that the reference curve does not specifically refer to the DRT in the Italian language, and that some small deviations could be expected according to the language-specificity of the test (e.g., number of syllables). Indeed, for A noise and SSN no significant changes were observed in IS when moving from R1 to R2. As concerns RT, no changes were found under A noise whereas a significant increase ($p$=0.029) was found in presence of SSN. The ICRA noise behaved differently. In fact, a significant worsening of IS was observed in position R2 versus R1 ($p$=0.049), together with a slowing down of RT results ($p$=0.01); their combined effect yielded a significantly lower DE ($p$=0.002).

Then, despite the sensitive changes in the STI values between R1 and R2, the changes in IS and RT were never significant for NG listeners. This lack of dependence of the performance on the changes of acoustical conditions could be explained by the fact that both positions within the classroom were challenging enough for non-native listeners. Furthermore, when interpreting IS results, it should be considered that the speech recognition of non-native listeners differ from that of native ones not only on the mean of the psychometric function, but also on the slope [19]. Less steep curves are generally found for non-native listeners, and then, it could be hypothesized that for NG participants, STI differences greater than the JND are required to observe a decrease in the accuracy performance.

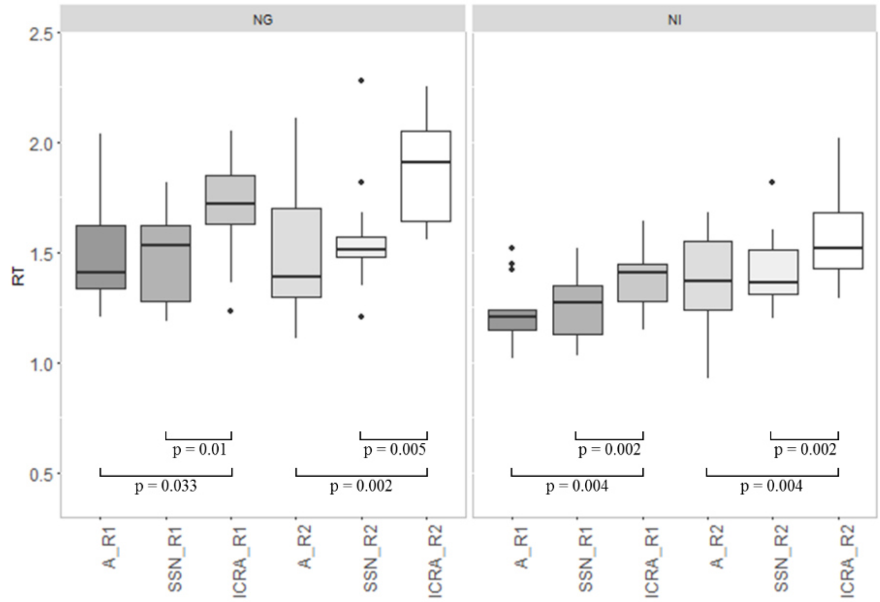


**Figure 3:** Response time (RT [s], defined as the time elapsed between the end of the waveform and the response selection) for native German (NG, on the left) and native Italian (NI, on the right) participants. Results refer to the three masking noises (A: ambient, SSN: stationary, ICRA: fluctuating), in the two listening positions (R1 and R2). The *p*-values associated to significant differences between background noises within the same position are also reported.

Even though the non-parametric tests used for the statistical analysis do not allow a proper investigation of the interaction effects, it could be said that the results of NI participants suggest the presence of a different recognition pattern depending on the noise type. For the A noise, no differences were detected between the two listening positions, probably due to the rather high STI values measured in both positions, already corresponding to the upper asymptote of the psychometric function. When SSN was introduced, whilst IS was found to be the same in both positions, a greater impairment on cognitive resources was observed in R2, pointed out by a slowing down of the RT results. Finally, the most unfavorable results in R2 were found under ICRA noise, stemming from the concurrent increase of energetic masking and on the necessity of involving more cognitive resources to distinguish the speech signal from the fluctuating masker.

### Effects of background noise

The effect of the typology of background noise was separately assessed for each position and for each listeners' group, by using the *mctp.rm* function [20] and by selecting the Tuckey contrast matrix to account for multiple comparisons. The resulting *p* values are reported in Figs. 2-4.

For both listeners' groups, significant differences were found between ICRA noise and the other two maskers, in both positions and for all metrics. The only exception was found in IS for NG listeners: in both positions, the same accuracy was found for SSN and ICRA noise, which was significantly lower than IS in A noise. In general, ICRA noise provided lower IS and higher RT compared to SSN and A noise. Thus, the first thing to observe is the absence of fluctuating masker release that, due to the reverberation, is completely ineffective [21]. Actually the reverse happens, that is the impact on the reception of the signal is further complicated by fluctuations both at sensory and at cognitive level, and this is reflected in both IS and RT results.

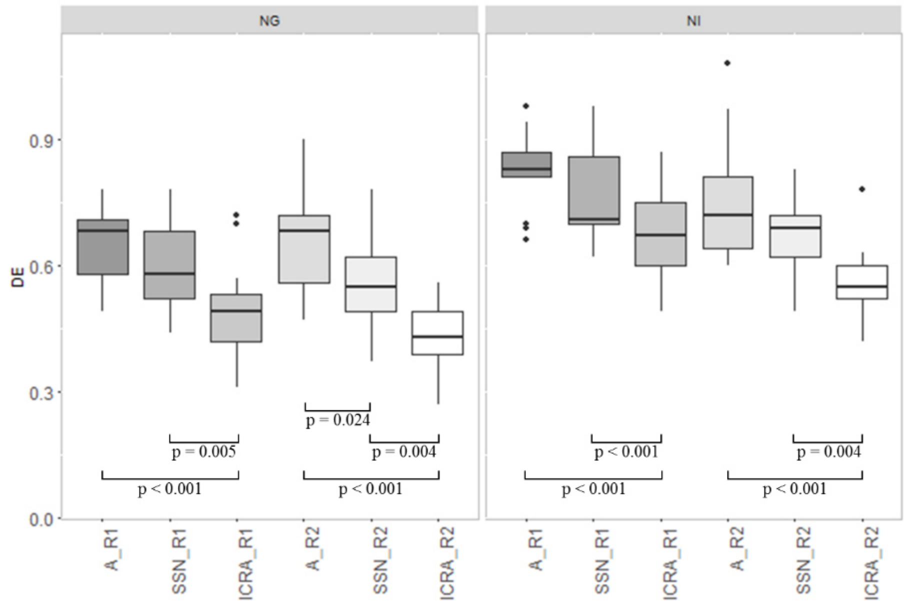


**Figure 4:** Listening efficiency (DE [$s^{-1}$]) for native German (NG, on the left) and native Italian (NI, on the right) participants. Results refer to the three masking noises (A: ambient, SSN: stationary, ICRA: fluctuating), in the two listening positions (R1 and R2). The *p*-values associated to significant differences between background noises within the same position are also reported.

As expected, differences were found when SSN and A noise were directly compared. A remarkable gap of STI was present between the two stationary noises A and SSN, and this would be sufficient to explain discrepancies. In particular in position R1, for both NI and NG, IS was significantly reduced by the presence of SSN. In position R2, a similar effect of the two noises was found for NI listeners, whereas a greater reduction of IS under SSN was experienced by NG participants. It seems that both groups suffer the presence of SSN compared to A noise, basically due to the increase of the noise level and thus of stationary energetic masking. Interestingly RT results were always similar between the two noises, whereas, accordingly to literature results [22] a slowing down was expected under more challenging listening conditions. This finding could be possibly explained by the large data dispersion, pointed out by the interquartile range of the RT results. Indeed, RT is a behavioral measure, which also reflects individual characteristics, such as susceptibility to noise and attention, and it is modulated by individual working memory capacity [4]. For example, the presence of SSN could elicit in some individuals a more focused attention, resulting in faster RT, similar to those under A noise.

### Differences between the listeners groups

Aiming at assessing the effect of participants' mother language, two samples tests (*npar.t.test*) were performed for each listening condition, comparing the results of NI and NG listeners; the resulting *p* values are reported in Tab. 2.

Data are best analyzed by position. In R1 all comparisons reached significance, that is NI always outperformed NG both in IS and RT, and finally in DE. When moving towards the back of the classroom, changes were found in the pattern of the comparisons. In particular A and SSN noises caused significantly different IS between the groups, whereas no difference was found in RT results. On the contrary, for ICRA noise, significantly different RT values were found, together with similar IS results. As to better understand the differences among the listeners groups, the position dependent results need to be considered, and specifically the absence of a positional effect for NG participants for both IS and RT metrics. Then, for SSN and A noise, no differences were found in RT results of the two groups of listeners, as NI participants reached in R2 the longest RTs that NG participants already showed in position R1. Similarly, under ICRA noise, NI listeners reached in R2 the lowest IS that NG listeners already scored in R1. On the whole, for all the three maskers, the outcome in terms of DE was significant, but the way it was realized differed, depending on the typology of background noise.

As expected, being non-native listeners implies a deficit in the coding of speech either in sub-optimal listening conditions (e.g., under A noise in position R1) or in more challenging acoustics scenarios (e.g., position R2). The analysis of RT data allows adding new information about the behavior of non-native listeners under fluctuating maskers. Indeed, whereas NG participants were able to keep the same RT of NI listeners under A noise and SSN, a greater involvement of cognitive resource was required under ICRA noise, reflected by a significantly higher processing time.

**Table 2:** Statistical comparison between IS, RT, and DE results of the listeners groups: native Italian (NI) and native German (NG) speakers. The resulting *p* values are reported for each background noise (steady state: SSN, fluctuating: ICRA, ambient: A) and for each listening position (R1, R2). Bold italic figures for not-significant comparisons.

| Background noise | Listening position | IS | RT | DE |
|---|---|---|---|---|
| A | R1 | 0.040 | 0.001 | <0.001 |
| | R2 | 0.023 | ***0.188*** | 0.034 |
| SSN | R1 | <0.001 | 0.009 | <0.001 |
| | R2 | 0.001 | ***0.112*** | 0.006 |
| ICRA | R1 | 0.003 | <0.001 | <0.001 |
| | R2 | ***0.061†*** | <0.001 | <0.001 |

†For the statistical comparison of NI and NGG listeners: *p*=*0*.014

Finally, 2-samples tests (*npar.t.test*) were performed, as to compare the results of the sub-groups of participants NGI and NGG. The aim was assessing if the different age of acquisition would differently affect their speech reception performance in the Italian language. The two groups were found to behave significantly in a different way in R1, under A noise alone, as regards both IS ($p$=0.035) and RT ($p$=0.028). Specifically, NGI participants were slower to respond than NGG listeners but achieved a significantly higher IS. A new analysis was then performed to compare the speech reception performance of NI and NGG participants. It was found that all of the results reported in Tab. 2 were maintained, except for IS under ICRA noise in position R2, where the NGG group was found to score significantly lower than NI listeners did ($p$=0.014). This finding strengthen the hypothesis of the fluctuating noise greatly impairing the non-natives' performance: when a less proficient group of listeners is considered, ICRA noise worsen their performance not only as regards RT, but also on IS results.

## DISCUSSION

The evaluation of the acoustics for second language users is controlled by standardized parameters such as STI. In this experiment both higher, borderline values and lower values with respect to the relevant normative references were deliberately tested by considering two positions in the classroom under ambient noise, and by adding artificial maskers (SSN and ICRA noise) with a rather high level (SNR<0 dB). The choice of the SNR was intended to mimic conditions that may arise during group work, for instance in laboratory assignments or, most often, in the context of open plan group work. In these cases, the presence of fluctuating noise is due to the voices of other students, which are assumed irrelevant for the proper task; it can also be assumed that they do not cause informational masking on the target speech signal.

The results showed that second language students undergo a twofold disadvantage when attending lessons in a classroom primarily designed for native students. Firstly, the second language users hardly reach the same accuracy in intelligibility scores as the first language ones, even in sub-optimal conditions (ambient noise alone). A similar IS is only reached when conditions are the worst for both groups (e.g., in R2). Secondly, non-natives listeners systematically experience an increase in the amount of cognitive resource required for speech reception compared to their native peers. This is shown by slower response times in the front position and, in the back position for ICRA noise alone. It has to be remarked that, if the analysis were limited to IS, an unclear picture could be drawn on the effect of fluctuating masker. On the contrary, the RT results outline the detrimental effect of the fluctuating noise on the processing time required to accomplish the task. Unfortunately the primary constituents underpinning the listening effort construct are many [4], their relationships is still not entirely clear, and these open issues cannot be unveiled by this single, comprehensive behavioral measure of response time. Finally, more specific non-parametric statistical analysis are needed to assess the presence of interactions among the factors considered in the experiment; as an example, it would be of interest to understand if different masking noises would differently impair the RT results, depending on the mother tongue of the participants.

Despite limitations, the results add to the previous findings of [3] where even under optimal listening conditions a disadvantage of second language listeners was observed in terms of response time. In the present experiment, the disadvantage persists with the worsening of listening conditions, obtained with the energetic masking of the speech signal (stationary noise). A further detrimental factor is the presence of fluctuations in the masker, although not carrying informational content. It has to be noted that also for native listeners the impact of ICRA noise is quite relevant. First, the presence of reverberation cancels every possible benefit from listening in the gaps and second, whilst passing from A to SSN only causes a decrease in accuracy results, the presence of fluctuations in the masker implies a significant increase in RT too. This finding witnesses the increased request of cognitive resources needed for coping with fluctuations, as detailed in current models of working memory (ELU) [23].

Given the above picture, it seems that a fragile speech recognition characterizes second language participants, even though they had known the target language since long and were exposed to it regularly. Indeed, the present experiment took place in the Bolzano area, which is multilingual and thus, NG participants had a very long practice and exposition to the Italian language. One may speculate that similar or even worse results would be obtained for a less specialized and proficient population, and that this could be the evidence of a serious problem with performance of second language students in classroom settings. Delays in response time cannot be directly correlated to academic achievements but, although the connection is mediated by personal factors and management of restoration pauses, it is realistic to foster an impact on learning performance.

## CONCLUDING REMARKS

This study confirmed previous results on the performance of non-native listeners, and extended the findings to a wider set of acoustical conditions. The experiment was restricted to speech reception, which is only partly matching the target of comprehension of assignments and of lesson explanations. On the other hand, this basic approach contributes to highlight more general trends pertaining to the communication channel. The results pinpoint the need for prescriptions considering both intelligibility scores and a measure of listening effort. Indeed systematically worse results were found for second language listeners not only in the intelligibility scores but also in the time required to process the task presented, especially so when a fluctuating masker was present. For this reason insight shall be put into the normative references to provide control over the fluctuating quality of noise to avoid excessive effort for first language users and even more so for second language ones.

## REFERENCES


[1] Lecumberri, M.L.G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech communication, 52,* 864-886.

[2] IEC 60268-16 (2011). Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index (International Electrotechnical Commission, Geneva, Switzerland).

[3] Lam, A., & Hodgson, M. (2016). *Intelligibility, response time and listening efficiency of native and non-native listeners in various acoustical conditions*. Paper presented at the Acoustics Week in Canada, Vancouver, Canada.

[4] 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'. *International journal of audiology, 53(7),* 433-440.

[5] DIN 18041 (2016). Acoustic quality in rooms – Specification and instructions for the room acoustic design (German Institute for Standardization, Berlin, Germany).

[6] 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 in the Italian language). International Technical Report No. 3C1286, Fondazione Ugo Bordoni, Rome.

[7] Steeneken, H.J.M. (2014). *Forty years of speech in intelligibility assessment (and some history)*. Keynote lecture at the IOA 40th Anniversary Conference, Birmingham (UK).

[8] ISO 9921 (2003). Ergonomics - Assessment of speech communication (International Organization of Standardization, Geneva, Switzerland).

[9] Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *Audiology, 40(3),* 148-157.

[10] Houtgast, T., Steeneken, H., Ahnert, W., Braida, L., Drullman, R., Festen, J., Jacob, K., Mapp, P., McManus, S., Payton, K., Plomp, R., Verhave, J., van Wijngaarden, S. (2002). Past, present and future of the Speech Transmission Index. TNO Human Factors, Soesterberg, the Netherlands.

[11] Payton, K.L., & Shrestha, M. (2013) Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data. *J. Acoust. Soc. Am., 134(5),* 3818-3827.

[12] Visentin, C., & Prodi, N. (2017). *Effects of the noise type on listening effort: relationship between subjective ratings and objective measurements*. Paper presented at the 12th ICBEN Congress, Zurich, Switzerland.

[13] Giordano, R. (2005). *Note sulla fonetica del ritmo dell'italiano*. Paper presented at the 2nd Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV), Salerno, Italy.

[14] UNI 11367 (2010). Acustica in edilizia - Classificazione acustica delle unità immobiliari - Procedura di valutazione e verifica in opera. (Ente Italiano di Normazione, Milano, Italia)

[15] Prodi, N., Visentin, C., & Bellettini, C. (2012). *Listening efficiency testing*. Paper presented at the AES International Conference, Helsinki, Finland.

[16] Prodi, N., Visentin, C., & Farnetani, A. (2010). Intelligibility, listening difficulty and listening efficiency in auralized classrooms. *J. Acoust. Soc. Am., 128(1),* 172-181.

[17] Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475-482.

[18] Konietschke, F., Placzek, M., Schaarschimidt, F., & Hothorn, L.A. (2015). nparcomp: An R software package for nonparametric multiple comparisons and simultaneous confidence intervals. *Journal of Statistical Software, 64(9),* 1-17.

[19] Van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *J. Acoust. Soc. Am., 111(4),* 1906-1916.

[20] Konietschke, F., Bathke, A. C., Hothorn, L.A., & Brunner, E. (2010). Testing and estimation of purely nonparametric effects in repeated measures designs. *Computational Statistics & Data Analysis*, *54*(8), 1895-1905.

[21] George, E. L., Festen, J. M., & Houtgast, T. (2008). The combined effects of reverberation and nonstationary noise on sentence intelligibility. *J. Acoust. Soc. Am., 124(2),* 1269-1277.

[22] McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2016). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*.

[23] Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health, 12(49),* 263-269.