Binaural Speech Intelligibility in a Real Elementary Classroom

Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Binaural Speech Intelligibility in a Real Elementary Classroom

To cite this article: G Minelli *et al* 2021 *J. Phys.: Conf. Ser.* **2069** 012165

Binaural Speech Intelligibility in a Real Elementary Classroom

G Minelli[1,3], G E Puglisi[1], A Astolfi[1], C Hauth[2] and A Warzybok[2]

[1]Energy Department, Polytechnic University of Turin, 24 Corso Duca degli Abruzzi, 10129 Turin, Italy

[2] Medizinische Physik and Cluster of Excellence Hearing4All, Carl von Ossietzky University of Oldenburg, D-26111 Oldenburg, Germany

[3]greta.minelli@polito.it

Abstract. Since the fundamental phases of the learning process take place in elementary classrooms, it is necessary to guarantee a proper acoustic environment for the listening activity to children immersed in them. In this framework, speech intelligibility is especially important. In order to better understand and objectively quantify the effect of background noise and reverberation on speech intelligibility various models have been developed. Here, a binaural speech intelligibility model (BSIM) is investigated for speech intelligibility predictions in a real classroom considering the effect of talker-to-listener distance and binaural unmasking due to the spatial separation of noise and speech source. BSIM predictions are compared to the well-established room acoustic measures as reverberation time (T30), clarity or definition. Objective acoustical measurements were carried out in one Italian primary school classroom before (T30= 1.43s±0.03 s) and after (T30= 0.45±0.02 s) the acoustical treatment. Speech reception thresholds (SRTs) corresponding to signal-to-noise ratio yielding 80% of speech intelligibility will be obtained through the BSIM simulations using the measured binaural room impulse responses (BRIRs). A focus on the effect of different speech and noise source spatial positions on the SRT values will aim to show the importance of a model able to deal with the binaural aspects of the auditory system. In particular, it will be observed how the position of the noise source influences speech intelligibility when the target speech source lies always in the same position.

1. Introduction

The development of speech, language and cognitive abilities in children is influenced by the acoustic quality of the classrooms. A good speech intelligibility in classrooms is of crucial importance for children's learning abilities, especially in lower school grades [1–7]. Speech intelligibility is defined as the percentage of correctly understood speech items [8]. Generally, it is lower than 100% when the competitive effect of the acoustic environment on speech is considered, i.e., when background noise and reverberation are present and do not comply with optimal acoustic conditions [2,9]. A number of studies showed that long reverberation times and excessive noise negatively impact not only speech intelligibility but also academic performance [10–12].

Although speech intelligibility in classroom is often corrupted by ambient noise and reverberation, listeners can focus on a specific speech signal even in challenging sound environments, which was described as the cocktail party effect [13]. In particular, when speech (target) is partially masked by noise (masker), intelligibility can be considerably improved if speech and noise sources are spatially separated. The ability to use the difference in location between a target and a masker to better understand the target is called spatial release from masking (SRM) [14]: up to 12 dB improvement in speech reception threshold (SRT) can be observed due to the spatial separation of the target and masker in anechoic conditions [14–16]. This ability is strongly reduced in reverberated conditions [17,18].

So far, many studies focused on the effect of reverberation and noise on speech intelligibility and SRM in laboratory [19,20], whereas there are only a few studies that have considered these aspects under ecologically-valid environments [15,18].

This work aimed at the objective assessment of binaural speech intelligibility and monaural room acoustic parameters in two real elementary classrooms. The two classrooms had the same size but differed considerably in their acoustics. One of them was acoustically treated and represents a good acoustic classroom in terms of reverberation and speech clarity, whereas the second one was an acoustically untreated classroom with rather poor acoustics. For the assessment of speech intelligibility, binaural room impulse responses (BRIRs) were measured at different target-to-receiver distances, with the noise source spatially co-located and separated from the listener's head and a fixed position of the target speech source corresponding to the typical teacher's position. With such a spatial design of experiments, the aim was to define the extent to which the masking noise position, in terms of distance and angle, influences the speech intelligibility at the receiver's position. The BRIRs were convolved with speech and noise anechoic stimuli, and then used as input in a binaural speech intelligibility model (BSIM, Beutelmann, 2010) used to predict Speech Reception Thresholds (SRTs). The predicted SRTs were used to compare speech intelligibility and SRM in the considered classrooms. Moreover, SRTs were compared to the well-established room acoustic measures as reverberation time (T30), early decay time (EDT), clarity (C50), and definition (D50) calculated from the monaural room impulse responses.

2. Materials and Methods

2.1. Classrooms

Monoaural and binaural acoustical measurements were carried out in two classrooms belonging to the same school building. While the architectural features (i.e., geometry, finishes) of them were identical, they differed in acoustical properties due to an acoustical treatment applied to one of them. In the following, they will be considered as a one classroom before and after acoustic treatment, referred to as *ante operam* (AO) and *post operam* (PO) conditions, respectively. In particular, the classrooms present a rectangular plan of dimensions of 6.7 m x 8.4 m, and a volume of around 258 m^3. The environment is characterized by the presence of three windows facing the school courtyard adjacent to a road with medium volume of traffic; the floor's finishes are mainly made of venetian tiles, while the walls and the ceiling are covered by a plaster in the case of AO condition. The furniture consists of student's desks and chairs, bookshelves, and blackboards. In the PO classroom, the acoustical treatment involved the addition of absorbent panels in glass fiber ($\alpha_{0,5\text{-}1kHz}$ = 1) on the left longitudinal (9.7 m^2) and rear (16.2 m^2) walls, and of a countertop in glass fiber ($\alpha_{0,5\text{-}1kHz}$ = 0,95) on the ceiling (56.3 m^2).

Measurements were performed out of school time and 23 children sitting at their desks were simulated with the 100% polyester fiber panels (0.6 m x 0.6 m x 0.05 m each).

2.2. Acoustic measurements

2.2.1. Monoaural measurements

Monoaural measurements were carried out with the aim to characterize the acoustics of the room, according to EN ISO 3382-2. A calibrated NTi XL2 (NTi Audio, Schaan, Liechtenstein), sound level meter and a NTi Audio TalkBox (NTi Audio, Schaan, Liechtenstein) source were used for the

measurements. Recordings were performed for four receiver positions along the central axis at a distance of 1 m, 2.2 m, 3.6 m, and 6.2 m from the source and with a height of 1.2 m from the floor. The TalkBox source was always in the same position and was placed on the central axis of the classroom at 1 m from the frontal wall and at 1.5 m from the floor. Room impulse responses were acquired from exponential sine sweep signals emitted by the TalkBox and recorded by the sound level meter.
Reverberation time (T30, s) was averaged over a frequency range from 250 Hz to 2 kHz, while the early decay time (EDT, s), speech clarity (C50, dB) and definition (D50, dB) were averaged over a frequency range from 500 Hz to 1 kHz, all according to UNI 11532-2. Calculated parameters were compared to the recent recommendations for an optimal occupied classrooms. An optimal reverberation time should be around 0.5 s [21], EDT between 0.3 s and 0.7 s [22], C50 greater than approximately 2 dB [21], and D50 in a range of 0.86 to 1.0 for an excellent acoustics (Marshall, 1994).

2.2.2. Binaural measurements

Binaural room impulse response (BRIRs) were recorded with a Brüel & Kjær (B&K) Head and Torso Simulator (HaTS) using an NTi Audio TalkBox source with speech directivity pattern for talker-related BRIRs and a Larson Davis Omnidirectional (Dodecahedral) sound source for a masker-related BRIRs. In order to focus on such aspects as the influence of the acoustical treatment on speech recognition and the SRM as a measure of binaural speech processing, some conditions were kept constant: in particular, the sound pressure level and configuration of the target speaker (T), the head orientation and height of the receiver (R), the sound pressure level and typology of noise of the masking noise source (M) never changed during the measurement campaigns.
The speech source position (T) was fixed at 1.5 m height and at 1 m from the rear wall on the central axis of the classroom. The receiver's ears were at 1.5 m from the floor and it was placed at three distances from the T, namely at 1.5 m, 4.0 m and 6.5 m, referred in the following to as $T_{1.5}$, $T_{4.0}$ and $T_{6.5}$, respectively. Finally, the noise source was set at 1.5 m height in six positions that differed in angular position (0°, 180°, and 120°) and/or distance (1 m, 1.5 m. 2.5 m) with respect to the receiver.
Figure 1 shows the configurations of the binaural acoustical measurements, where:
i) the left side scheme shows the close target-to-receiver distance ($T_{1.5}$) corresponding to the first raw of student's desk. It is the most advantaged acoustic condition under evaluation. The positions of the noise source consider several distances (1.0, 2.5, and 5 m) and 2 azimuths(120° and 180°). The effect of increasing noise distance can also be evaluated, in agreement with literature studies such those of Westermann & Buchholz who studied the effect of noise distance on axis (from 1 m to 10 m). The two azimuths are considered in order to evaluate influence of reverberation and noise distance on SRM. It is expected that the SRM will be reduced in reverberated conditions compared to the anechoic conditions from the literature. Furthermore, it is expected that SRM will decrease with increasing distance between noise and receiver;
ii) the central scheme shows the conditions with the far target-to-receiver distance of 4 m ($T_{4.0}$) and corresponds to the central raw and is assumed to be the most disadvantaged condition for the students. Again, the noise source is placed at several distances (1, 2.5 and 4 m) and 3 azimuths (0°, 120° and 180°). These configurations allow evaluating differences in intelligibility when the noise source is located in front and in back of a listener which may introduce spectral cues resulting in speech intelligibility differences (14). Also, the SRM is examined and compared across the three target-to-receiver distances;
iii) the right-side scheme shows the very far target-receiver position ($T_{6.5}$). It corresponds to the last row of student's desk and it is assumed that, although the target-to-receiver distance increases, intelligibility can be improved compared to T4.0 due to the speech reflections of the rear wall. In line with conditions described above, the noise source is placed at several distances (1 m, 2.5 m, 5 m, and 6.5 m) and at 2 azimuths (120° and 0°) in order to examine the effect of increasing noise distance and SRM for a receiver located far away from the target speech.

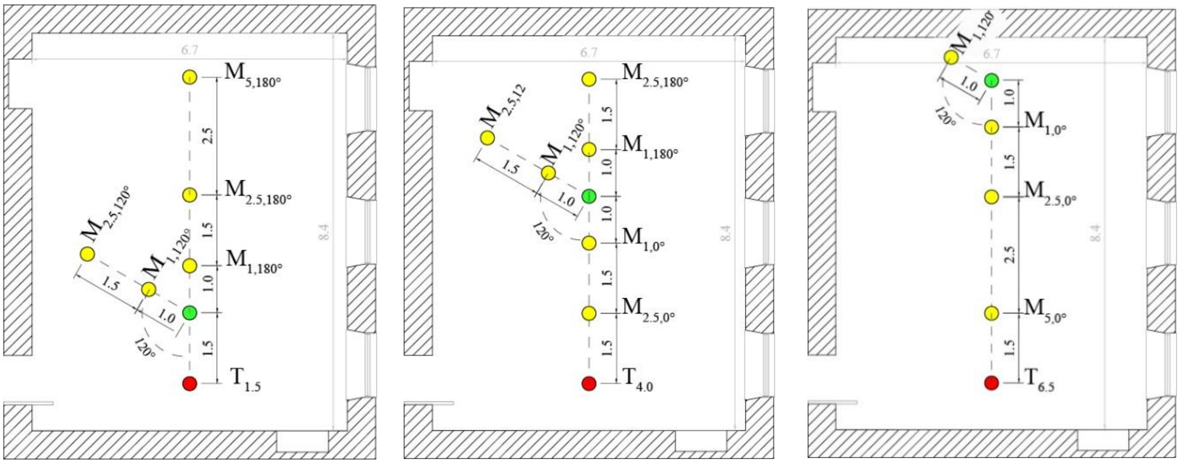

Figure 1. **S**chemes of the three different configurations. In each scheme the red circle represents the target position, the green circle identifies the receiver position, while the masker positions are indicated by yellow circles. Specification of linear and angular distance from the receiver are indicated in subscript for target and maskers: for example, $T_{1.5}$ indicates that the target is 1.5 m away from the receiver, while $M_{5,180°}$ indicates that the position of the masking is 5 m away from the receiver and with an angle of 180°.

2.3. Binaural Speech Intelligibility Model

2.3.1. BSIM model

BSIM (Beutelmann et al., 2010) was used to predict SRTs for the different receiver and noise source positions in the two acoustical environments (AO and PO). The model requires the separate speech and noise signals for each ear. For that, the corresponding BRIRs were convolved with the speech and noise signal. As first, the signals are filtered with a gammatone filterbank (Hohmann, 2002) ranging from 146 Hz to 8300 Hz in 30 ERB spaced frequency bands simulating the frequency selectivity of the human auditory system. Then, binaural processing is considered by applying the Equalization-Cancellation (EC, Durlach, 1963) mechanism. In the EC mechanism, the interaural level differences (ILD) and the interaural time differences (ITD) are equalized in each frequency band, independently. In a next step, the equalized left ear channel is subtracted from the equalized right ear channel (or vice versa). If the target and noise signals differ in their interaural parameters, which is the case if they are spatially separated from each other, the SNR can be substantially improved due to constructive interferences in the target signal and more important, destructive interferences in the noise signal. The EC processing leads to SNR maximization. In the next step, the EC-maximized SNRs are compared to the monaural SNRs of the left and the right ear in each frequency band. The maximum SNR is chosen and is considered for further processing. In scenarios where speech and noise are spatially separated, EC mechanism can significantly increase the SNR at low frequencies (<=1500 Hz) due to the ITDs. For high frequencies (>1500 Hz) the ILDs due to the so-called head shadow effect result in SNR differences across the ears and better ear listening takes place. In a last step, the speech intelligibility index (SII, ANSI 1997) performs a weighting of the SNRs in each frequency band to mirror human speech perception. Afterwards, the speech-weighted SNRs are integrated over frequency and transformed to an index value between 0 and 1.

2.3.2. Calibration

In order to transfer an SII into an SRT, a reference condition needs to be defined and calibrated in the BSIM which requires an empirical SRT. Usually the reference condition corresponds to anechoic co-located target and masker sources placed in from of the listener at the same distance (18,23). Here, a different approach was used, because the acoustical sources for speech and noise were not the same

(Talkbox vs. dodecahedron). Instead, the BSIM was calibrated to the PO condition were the speech was coming from the target source position and the noise was located 1 m behind the listener (M3). This situation is most similar to an anechoic situation typically used for calibration since the binaural effects are almost neglectable and the target is not influenced by reverberation due to a very short target-receiver distance. The calibration obtained in this situation was kept constant across room acoustic scenarios and target/masker combinations.

The reference SII was set to 0.22, which corresponds to empirical SRT in anechoic and co-located speech/noise situations [18,24,25]. It is important to note that the relative differences across conditions which are of main interest here are not affected by choosing the reference SII.

3. Results

3.1. Classroom acoustics results

Table 1 shows the measured values of the acoustic parameters for both AO and PO conditions. After the acoustical treatment the parameters were improved: in fact, reverberation decreases parallel to the increase of speech clarity and definition. All the results in the PO condition are in agreement with the recommendations presented in the subsection 2.2.1. On the contrary, the obtained data in AO do not meet the aforementioned standards.

Table 1. Descriptive statistics of the acoustical parameters measured in the ante operam (AO) and post operam (PO) conditions. Standard deviations are indicated in parentheses. Values that meet the standards are highlighted in bold.

	Acoustical parameters			
Condition	$T30_{0.25\text{-}2\ kHz}$ [s]	$EDT_{0.5\text{-}1\ kHz}$ [s]	$C50_{0.5\text{-}1\ kHz}$ [dB]	$D50_{0.5\text{-}1\ kHz}$ [%]
AO	1.4 (0.1)	1.4 (0.1)	-1.1 (2.3)	44 (12.8)
PO	**0.4 (0.0)**	**0.3 (0.1)**	**10.6 (3.2)**	**90 (5.0)**

3.2. Speech Intelligibility results

3.2.1. Effect of classroom acoustics

Figure 2 shows the intelligibility benefits of the PO intervention for all target and masker combinations. An improvement of speech intelligibility after acoustical treatment ranges from predicted SRT improvements (ΔSRT) of 8.2 dB SNR to 14.8 dB SNR and is observed in all the target-masker configurations. Greater benefits of about 10 – 15 dB SNR are in the spatially separated noise position for all target-to-receiver distances. The benefit is larger when the masker is closer to the receiver (at 1 m = 14 dB SNR, at 2.5 m = 10 dB SNR). Similar SRT benefits are obtained for the co-located noise positions of 0° and 180°. This is not influenced by the masker distance, i.e., SRT benefit does not change with increasing the masker distance. The receiver positions at 1.5 and 4.0 m away from the target and co-located noise source show very similar SRT benefit of about 8.5 dB, this is increased by about 1.6 dB for the far-far receiver position at 6.5 m from the target. Larger benefit for the $T_{6.5}$ could be due by the stronger influence of the absorbing panels on the rear wall for the far-far target-to-receiver distance; on the contrary, in AO the noise signal is reflected by the rear wall and delayed to the listener, that receive simultaneously the covered speech signal, arriving later because of the greater distance.

Generally, the predicted results of this study confirmed the observation from the literature [17,18] where SRTs increased (got worse) with the degree of reverberation. Also for the monaural acoustic measures, substantial improvement for short reverberation time was observed (Table 1). In this study, D50 values were better in all the frequency bands for the PO condition (AO/PO 125 Hz: 0.33/0.58, 250 Hz: 0.44/0.80, 500 Hz: 0.44/0.87, 1000 Hz: 44/95, 2000 Hz: 59/95, 4000 Hz: 67/96, 8000 Hz: 87/97), which correctly predict better intelligibility in the PO condition in agreement with [15,18].

However, application of the objective acoustical measures is limited to co-located conditions since they cannot account for the binaural aspects. In contrast, BSIM can account for binaural processing and assess speech intelligibility not only for co-located conditions but also for different spatial configurations of target and masker which reflects better real listening conditions.

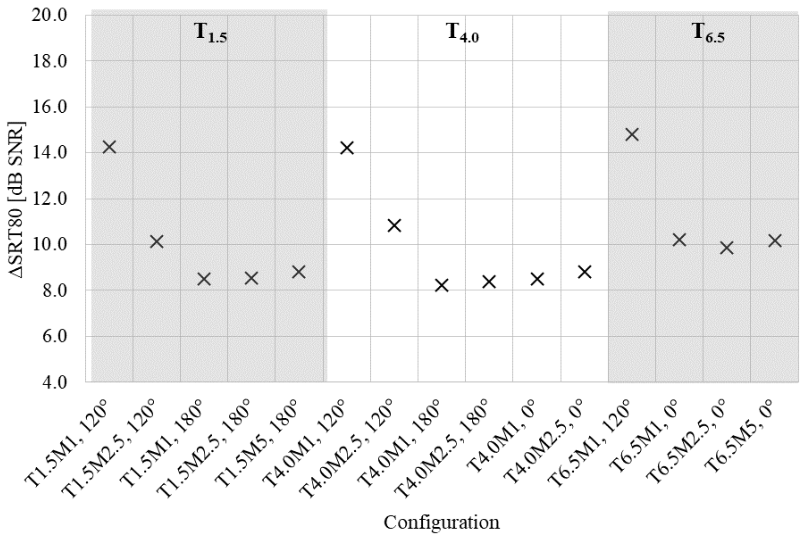

Figure 2. Difference between the obtained speech reception thresholds (SRT80s) in AO and PO conditions, considering the positions of target-to-receiver ($T_{1.5}$, $T_{4.0}$, $T_{6.5}$) and the masker-to receiver distance (N_1, $N_{2.5}$, N_5) and angle (0°, 120°, 180°).

3.2.2. Spatial release from masking in angle

Figure 3 shows that, in general, the benefit of SRM in angle is greater in PO than in AO and ranges from 2.8 to 7.3 dB SNR, with the greatest benefit in $T_{1.5}M_{1,180°\text{-}120°}$ in PO. Moving the masker from 0° to 120° results in lower benefits than moving it from 180° to 120°, in both AO and PO conditions. The SRMs for the AO range between -0.4 and 1.5 dB indicating that reverberation strongly degrades the binaural cues that are used to improve speech intelligibility in spatially separated conditions. Reverberation decorrelates the signals at the left and right ear and by that the human binaural system is not able to exploit the binaural cues. In a consequence, SRTs in co-located and spatially separated conditions are similar leading to no or strongly reduced SRM. Focusing only on the configurations with the masker at 1 m from the receiver (i.e., with the largest SRM), the binaural release from masking is in the range of 6 dB for the PO condition, while it is very much reduced (0.4 dB) for the AO. It demonstrates how important is reverberation for the binaural aspects and quantifies how big can be the benefit from acoustical treatment in terms of SRM. This has very important implications for the acoustical design and treatment of classrooms. The binaural aspects have a crucial role in source segregation, enabling focus on the speech target and at the same time supressing the interfering sources. It is, however, only possible in the classrooms with a good acoustics.

Finally, comparing the predicted SRM in this study to the data from literature obtained in the anechoic conditions with a speech source in front of the listener and varying azimuth of a noise source [14], several observations can be made. According to expectations, anechoic SRMs are higher (better) than those in real rooms. For a noise azimuth of 120°, up to 13 dB SRT improvement can be obtained in anechoic condition [14] whereas it is reduced to about 6 dB in rooms with a good acoustics as indicated by the results of PO condition, and is almost absent in rooms with high reverberation time (AO condition). It is in line with other studies examining SRM under reverberant conditions [15,18]. For example, as shown in [15], where room acoustic of an office and of a cafeteria were considered, and in [18], where a classroom, a listening room and a church were investigated, the SRM at 120° was in the range of about 2-8 dB, i.e., substantially lower (worse) than in an anechoic condition.

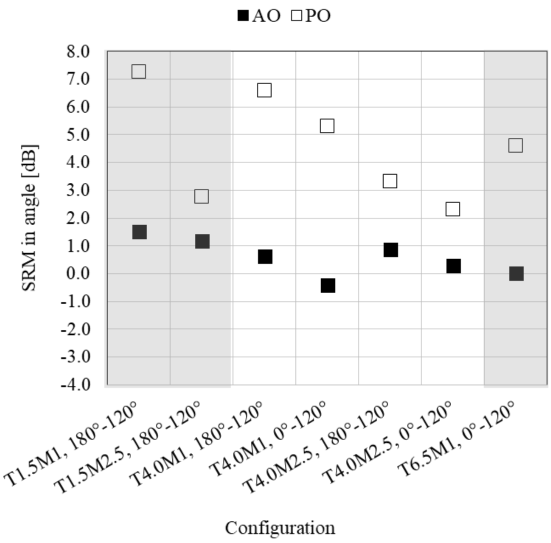

Figure 3. Spatial release form masking (SRM) of close target and masker ($T_{1.5}M_{1,180°\text{-}120°}$), close target and far masker ($T_{1.5}M_{2.5,\ 180°\text{-}120°}$), far target and close masker ($T_{4.0}M_{1,\ 180°\text{-}120°}$; $T_{4.0}M_{1,\ 0°\text{-}120°}$), far target and far masker ($T_{4.0}M_{2.5,\ 180°\text{-}120°}$; $T_{4.0}M_{2.5,\ 0°\text{-}120°}$), very far target and close masker ($T_{6.5}M_{1,\ 0°\text{-}120°}$). Black squares indicate the AO condition, while the white ones the PO condition.

4. Conclusions

The goal of the present work was to investigate the binaural aspects of the auditory system of normal-hearing listeners through predictions of binaural speech intelligibility in classrooms with and without acoustical treatment when an energetic masking noise is used. The results showed that:

- the predicted SRTs are lower (better) in the acoustically treated room; predictions were consistent also with D50 checked in each frequency band;
- when considering the noise source position spatially separated from the listener's head, SRTs are lower (better) than those with the noise source spatially co-located;
- SRM improves substantially as less reverberation directly influences the listeners ability to benefit from interaural differences between the target and the interfering signal.

As an overall conclusion, guaranteeing optimal acoustic condition in classrooms is mandatory for good speech intelligibility. This can be pursued also in existing classrooms with a proper retrofit through the addition of absorbing material on the walls and on the ceiling.

Acknowledgment

This research was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 A5.

References

[1] Cardon G, Campbell J and Sharma A, 2012. Plasticity in the developing auditory cortex: Evidence from children with sensorineural hearing loss and auditory neuropathy spectrum disorder. J Am Acad Audiol.;23(6):396–411.
[2] Astolfi A, Bottalico P and Barbato G, 2012. Subjective and objective speech intelligibility investigations in primary school classrooms. J Acoust Soc Am.;131(1):247–57.
[3] Prodi N, Visentin C and 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):255–68.
[4] Puglisi GE, Astolfi A, Cantor Cutiva LC and Carullo A, 2017. Four-day-follow-up study on the

voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. J Acoust Soc Am;141(1):441–52.
[5] Visentin C, Prodi N, Cappelletti F, Torresin S and Gasparella A, 2018. Using listening effort assessment in the acoustical design of rooms for speech. Build Environ;136(January):38–53.
[6] Puglisi GE, Prato A, Sacco T and Astolfi A, 2018. Influence of classroom acoustics on the reading speed: A case study on Italian second-graders. J Acoust Soc Am;144(2):EL144–9.
[7] Mendel LL, 2008. Current considerations in pediatric speech audiometry. Int J Audiol.;47(9):546–53.
[8] IEC 60268-16, 2011. Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. BSI Standards Publication.
[9] Prodi N, Visentin C and Farnetani A, 2010. Intelligibility, listening difficulty and listening efficiency in auralized classrooms. J Acoust Soc Am.;128(1):172–81.
[10] Shield BM and Dockrell JE, 2008. The effects of environmental and classroom noise on the academic attainments of primary school children. J Acoust Soc Am.;123(1):133–44.
[11] Picard M and Bradley JS, 2001. Revisiting Speech Interference in Classrooms:Revisando la interferencia en el habla dentro del salón de clases. Int J Audiol.;40(5):221–44.
[12] Bradley JS, 1986. Speech intelligibility studies in classrooms. J Acoust Soc Am.;80(3):846–54.
[13] Cherry EC, 1953. Experiments Described Herein Are Intended As a. J Acoust Soc Am;25:975–9
[14] Bronkhorst AW, 2000. Cocktail party phenomenon, Vol. 86, Acta Acustica United With Acustica, p. 117–128.
[15] Beutelmann R and Brand T, 2006. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. J Acoust Soc Am.;120(1):331–42.
[16] Warzybok A, Rennies J, Brand T, Doclo S and Kollmeier B, 2013. Effects of spatial and temporal integration of a single early reflection on speech intelligibility. J Acoust Soc Am.;133(1):269–82.
[17] Rennies J, Brand T and Kollmeier B, 2011. Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet. J Acoust Soc Am.;130(5):2999–3012.
[18] Beutelmann R, Brand T and Kollmeier B, 2010. Revision, extension, and evaluation of a binaural speech intelligibility model. J Acoust Soc Am.;127(4):2479–97.
[19] Bronkhorst AW and Plomp R, 1987. The effect of head-induced interaural time and level differences on speech intelligibility in noise. J Acoust Soc Am.;81(S1):S27–8.
[20] Peissig J and Kollmeier B, 1997. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. J Acoust Soc Am.;101(3):1660–70.
[21] UNI. UNI 11532-2, 2019. Acoustic characteristics of indoor environments – Design methods and evaluation techniques – Part 2: school sector, p. 1–42.
[22] Bradley JS, 2011. Review of objective room acoustics measures and future needs. Appl Acoust;72(10):713–20.
[23] Hauth CF and Brand T, 2018. Documentation of the binaural speech intelligibility model (BSIM) written by.
[24] Rennies J, Warzybok A, Brand T and Kollmeier B, 2014. Modeling the effects of a single reflection on binaural speech intelligibility. J Acoust Soc Am.;135(3):1556–67.
[25] Wagner L, Geiling L, Hauth C, Hocke T, Plontke S and Rahne T, 2020. Improved binaural speech reception thresholds through small symmetrical separation of speech and noise. PLoS One. 15:1–7.