DOI: 10.61782/fa.2023.0868

PREDICTION METHODS FOR THE ASSESSMENT OF BINAURAL SPEECH INTELLIGIBILITY IN CLASSROOMS

Giuseppina Emma Puglisi[1*] **Christopher Hauth**[2] **Greta Minelli**[1]
Arianna Astolfi[1] **Anna Warzybok**[2]
[1] Department of Energy, Politecnico di Torino, Italy
[2] Medizinische Physik and Cluster of Excellence Hearing4All,
Carl von Ossietzky University of Olden-burg, Germany

ABSTRACT

Learning environments deserve optimal acoustics to support speech communication, especially at the first stages of schooling. To guarantee an adequate sound environment for educational facilities, the control of background noise and the reduction late reverberation constitute two of the main intervention strategies at the designing stage. Both in the case of new construction and renovation of schools, prediction models for speech intelligibility can be used as evaluation tools to assess the efficacy of acoustic treatments. This work presents the results of the application of two versions of the Binaural Speech Intelligibility Model (BSIM) to a real primary school classroom that underwent an acoustical renovation. The BSIM versions mainly differed in the pre-processing of the speech signal, where the energy of the late reflections was either considered as detrimental to speech intelligibility or not. The acoustic renovation complied with the UNI 11532-2 standard, as objective measures of reverberation time decreased while speech clarity and definition increased, as expected. The BSIM gave outcomes in terms of speech recognition thresholds, considering the effect of increasing talker-to-listener distance and of the binaural spatial release from masking due to the separation of noise- and speech-sources.

Keywords: *speech intelligibility, prediction model, classroom, binaural listening.*

[*]***Corresponding author***: giuseppina.puglisi@polito.it

1. INTRODUCTION

The speech communication process is comprehensive of the premises pertaining to the talker and to the listener, which are supported and determined both by personal and by environmental characteristics [1]. In this framework, speech intelligibility is a main issue to be ensured in the built environment in order to support the transmission of verbal messages towards a listener. When the speech communication process takes place in teaching and learning environments, i.e., in classrooms, ensuring a high level of speech intelligibility turns into higher academic performances and into adequate cognitive support also towards disadvantaged subjects (e.g., young pupils, students with cognitive- or hearing-impairment) [2-5].
The main strategies to guarantee high speech intelligibility levels are related to the reduction of reverberation time and noise, and to the increase of speech clarity and definition [6-9]. However, these well-established acoustic indices to be controlled are mainly monoaural, thus do not reflect consistently the auditory task that benefits from the binaural listening. Indeed, the listening task in everyday life environments happens binaurally and the effect of the spatial distribution of target- and noise-sources, as well as the acoustic characteristics and the finishes in the environment, determine its greater or smaller complexity. Cherry [10] conceptualized the "cocktail party phenomenon" as the ability of a listener to discriminate a target signal in a noisy environment when noise sources are spatially distributed around the listener. There is a point where a listener benefits more from the so-called spatial release from masking (SRM) as the target- and noise-sources come to be spatially separated (i.e., in a azimuth range that is between about 120° and 135°) rather than when they are spatially co-located (i.e., at 0° or 180°)

[11,12]. In normal-hearing listeners, the SRM can be quantified in a speech intelligibility increase up to 12 dB in speech recognition thresh-old (SRT), which is intended as the signal-to-noise ratio (SNR) needed to yield 50% of correct recognition [11,13]. Such SRM turns to be significantly lower in reverberant environments due to the detrimental effect of reflections on the target signal as well as on noise and also in the case of increasing target-to-receiver distance [14-16]. Therefore, studies thank account from complex acoustic scenarios in terms of a large variety reverberation levels and of noise typologies (i.e., energetic or informational masking signals) are extremely needed. So far, complex acoustic scenarios have been investigated in the scientific literature primarily at a laboratory level in which reverberation and noise are artificially added [17,18] or based on prediction models [19,20]. Beutelmann et al. [21] and Rennies et al. [14] have implemented and further extended a binaural speech intelligibility model (BSIM) that is accurate for speech intelligibility predictions in noisy and reverberant environments. However, the steps available in [21] are not accurate in estimating the effect of detrimental reflections at large distances of target and receiver, therefore a further implementation by [14] was proposed.

In this work, the application of the BSIM to real complex acoustic scenarios is proposed. In field binaural room impulse responses (BRIRs) were measured and auralized in the model. Several challenging target- and noise-sources positions were designed in a primary school classroom that undergone an acoustical treatment, in order to evaluate by means of BSIM the effect of acoustics on binaural speech intelligibility before (ante-operam, AO) and after (post-operam, PO) the acoustic optimization. In terms of room acoustics parameters, reverberation time (T20), speech clarity (C50) and speech definition (D50) were derived from the monaural room impulse responses measured both in AO and PO conditions, and then were put in relation to the SRTs predicted using the BSIM. Particularly, the efficacy of the acoustic treatment was evaluated in terms of ΔSRT separately between the AO and the PO conditions, and using both the approaches proposed by Beutelmann et al. [21] and Rennies et al. [14] in order to establish which model could better reflect the modeling of real complex acoustic scenarios, thus could constitute a more robust tool for practitioners even at an early design stage.

2. METHODOLOGY

2.1 Case study

In a school building that dates back to the early XX century, one classroom undergone an acoustical treatment. From a geometrical point of view, it has a rectangular plan of about 56 m^2 and a constant height of 4.6 m. The classroom is placed at the ground floor of the building and presents three big windows that face a trafficked street. The floor is finished with Venetian tiles. Before the acoustical treatment, i.e., ante operam (AO), the lateral walls and the ceiling of the classroom are finished with plaster; after the acoustical treatment, i.e., post operam (PO), glassfiber absorptive panels with absorption coefficient at 0.5-1 kHz equal to 1.00 and to 0.95 were positioned on the lateral walls and on the ceiling, respectively.

2.2 Room acoustic measurements

Measurements in the classroom, both in AO and PO, were performed considering a simulated occupied condition. To do this, 100% polyester fiber panels were dimensioned to simulate the presence of 23 pupils seated, which results in 0.35 m^2 at 1 kHz, based on the findings of Astolfi et al. [22] and Puglisi et al. [1]. The measurement session in the classroom was organized to answer to two main questions. First, to characterize the room AO and PO by means of monoaural acoustic parameters in agreement with the standards. Second, to acquire the binaural room impulse responses (BRIRs) to be added in the binaural speech intelligibility model (BSIM) for speech intelligibility predictions under the different acoustic treatments of the classroom and spatial configurations.

Monoaural measurements were performed in agreement with the EN ISO 3382-2:2008 standard [23]. A directional sound source that embeds the directivity of human voice (TalkBox by NTi Audio) and a calibrated class-1 sound level meter (XL2 by NTi Audio) were used to this aim. Particularly, the sound level meter was placed at 1.2 m from the floor in several positions in the room, according to Minelli et al. [7], and was used to record the exponential sine sweep signals emitted by the TalkBox; then, room impulse responses were obtained after applying a deconvolution process. Room acoustics parameters were evaluated based on the thresholds and optimal ranges suggested in the Italian standard UNI 11532-2:2020 [24]. In particular, optimal values/ranges were calculated for reverberation time (T20, s) between 0.125 kHz and 4 kHz, for speech clarity (C50, dB) between 0.5 kHz and 2 kHz and for speech definition (D50, %) between 0.5 kHz and 1 kHz. The optimal values/ranges for the considered room acoustics parameters are shown in Table 2 together with the AO and PO characterization results.

Binaural measurements regarded the acquisition of BRIRs in challenging spatial configurations. A target source (T), a masking noise source (M) and a receiver (R) were mutually positioned in the classroom to consider both beneficial spatial configurations and detrimental ones, according to the available literature in the field. T consisted in the same TalkBox described above, which was placed at 1 m from the rear wall on the central axis of the classroom and at a height of 1.5 m from the floor in order to mimic the typical position of a teacher in the classroom. M consisted in a Larson Davis Omnidirectional sound source. It was positioned at 1.5 m from the floor in several even positions, which varied in azimuth and in distance with respect to the receiver's ears. R consisted in a Brüel & Kjær (B&K) Head and Torso Simulator (HaTS) placed at 1.5 m from the floor and at increasing distances from T (i.e., at 1.5 m, 4.0 m and 6.5 m). Table 1 summarizes the positioning of T, M and R in the classroom: these configurations were equally used in AO and PO conditions.

Table 1. Classification of target- and noise-sources in the classroom, with details on the mutual positioning and identification codes (ID).

ID	Noise-source distance	Noise-source azimuth	Spatial positioning of noise
R1 – target-to-receiver distance = 1.5m			
$M_{1m,120°}$	1 m	120°	separated
$M_{2.5m,120°}$	2.5 m	120°	separated
$M_{1m,180°}$	1 m	180°	co-located
$M_{2.5m,180°}$	2.5 m	180°	co-located
$M_{5m,180°}$	5 m	180°	co-located
R2 – target-to-receiver distance = 4.0m			
$M_{1m,120°}$	1 m	120°	separated
$M_{2.5m,120°}$	2.5 m	120°	separated
$M_{1m,180°}$	1 m	180°	co-located
$M_{2.5m,180°}$	2.5 m	180°	co-located
$M_{1m,0°}$	1 m	0°	co-located
$M_{2.5m,0°}$	2.5 m	0°	co-located
R3 – target-to-receiver distance = 6.5m			
$M_{1m,120°}$	1 m	120°	separated
$M_{1m, 0°}$	1 m	0°	co-located
$M_{2.5m,0°}$	2.5 m	0°	co-located
$M_{5m,0°}$	5 m	0°	co-located

2.3 The Binaural Speech Intelligibility Model

The BSIM was implemented to assess speech intelligibility in terms of SRTs in different types of environment, which can be characterized even by high reverberation time and noise. The work flow of this prediction model is accurately described in Beutelmann et al. [21], and its further updates in Rennies et al. [14].

Basically, the BSIM convolves BRIRs, which can be either measured in field or simulated, separately with anechoic speech and noise signals. As noise signal, it was used the speech-shaped noise of the Italian matrix sentence test [25,26]. What happens after a series of steps that manipulate the convolved signals (i.e., a gammatone filterbank, an Equalization-Cancellation, an equalization in frequency per interaural differences in level and time), as detailed in [14,21], the speech intelligibility index (SII) is used to weight the signal-to-noise ratio (SNR) with the aim of replicating the human perception of speech. Eventually, the SNRs are integrated in frequency and scaled to an index that ranges between 0 and 1, so that the SII values could be related to the SRTs that are then used as results. This latter step of mapping SII values to SRTs can be performed only as a consequence of an appropriate calibration of the model. Such calibration process is based on the selection of an empirical SRT measured in anechoic conditions and with co-located target and noise sources. Therefore, among those available in the set of AO and PO configurations, as a reference condition it was selected a PO condition (i.e., with very short reverberation time) with the close target-to-receiver distance and with the noise source co-located 1 m behind the listener. Further details on the calibration procedure can be found in [14,21,27].

The method and calibration process described above are the bases of both BSIM versions formalized by Beutelmann et al. [21] and Rennies et al. [14]. However, they exhibit a substantial difference in the pre-processing of the target speech signal. In Beutelmann et al. [21] the detrimental effect of the late reflections on the speech signal itself is not accurately accounted. Viceversa, Rennies et al. [14] proposes an approach that allows for the separation of the BRIR that is convolved with the target signal into an early part (≤100ms) and then into a late part (>100ms). After this separation, the target signal is convolved with both parts of the BRIR and the one convolved with the late part is added to the noise signal, so that the late reflections that still belong to the target but that are detrimental for speech intelligibility are thus considered not as useful signal.

3. RESULTS

3.1 Effect of the acoustic treatment

The acoustic treatment of the classroom resulted in room acoustics parameters that meet the optimal values/ranges proposed in UNI 11532-2:2020 [24]. Table 2 shows the compared outcomes of the monoaural measurements in terms of reverberation time, speech clarity and speech definition (each averaged in frequency as suggested in the standard).

Table 2. Room acoustics parameters measured in before and after the acoustic treatment of the classroom. Optimal values/ranges are indicated; standard deviations are shown in parentheses and values that agree with the optimal ones proposed in the standards are written in italic.

	Optimal value/range	Ante-Operam	Post-Operam
$T20_{0.125\text{-}4kHz}$ (s)	~0.5	1.6 (0.1)	*0.6* (0.1)
$C50_{0.5\text{-}2kHz}$ (dB)	≥2.0	-0.1 (2.1)	*11.4* (3.7)
$D50_{0.5\text{-}1kHz}$ (%)	≥86	44 (12.8)	*90* (5.0)

3.2 Results of the Binaural Speech Intelligibility Model

As SRTs are expressed in terms of SNR, it means that the lower they are the better it is. To the aim of this study, the BSIM results were elaborated in terms of differences in SRTs between AO and PO conditions (ΔSRT). Therefore, values higher than 0 mean that the PO condition with shorter reverberation time and higher speech clarity and definition provide benefits to the speech recognition ability. Table 3 shows the ΔSRTs as results of the BSIM predictions performed both with the Beutelmann et al. [21] method and the Rennies et al. [14] one. As a general comment, it clearly appears that speech intelligibility benefits from the acoustical treatment of the classroom, as ΔSRTs are always positive values.

In the cases of those spatial configurations with a separated noise source and for all the target-to-receiver distances, the benefit of acoustical treatment in the classroom is the highest and ranged between ~9 dB SNR and ~17 dB SNR.

Comparing the findings of the two prediction models, it is evident that the ones obtained with the model of Rennies et al. [14] are higher on average, which means that a greater difference in SRTs was found between the AO and PO conditions. Prediction models do not allow for the obtainment of results variability. However, if an intrinsic variability of the model is considered in terms of just noticeable level differences (JND), as suggested in the EN ISO 3382-2 standard [23], the normalized error concept can be extended to the predicted data and then applied [28]. The normalized error (E_N) is used to compare data that have the same hierarchical level, and is calculated as the ratio between the absolute value of the difference between two mean values and the expanded uncertainty of such difference. When the result of E_N calculation gives values that are ≥ 1, it means that the difference between the compared means is not merely due to random effects, thus they are statistically different. On the other hand, when E_N is < 1 it means that the difference can be due to random effects and so there is no reason to refuse the hypothesis that they are equivalent. The application of the E_N has been formalized in past studies in the field of acoustics [29], so unless experimental data will be available to create a database with enough across-subjects variability, this method can be considered as reliable to provide at least an intrinsic variability.

Table 3. Differences in speech recognition thresholds (ΔSRT) before and after the acoustic treatment of the classroom. ΔSRTs are obtained based on the calculations using either the Beutelmann or the Rennies prediction model. The normalized error is applied and reported in bold when referred to a statistically significant difference.

Acoustic scenario	Beutelmann model	Rennies model	Normalized error
R1 – target-to-receiver distance = 1.5m			
$M_{1m,120°}$	14.3	15.2	0.34
$M_{2.5m,120°}$	10.1	12.3	0.76
$M_{1m,180°}$	8.5	9.2	0.24
$M_{2.5m,180°}$	8.5	9.1	0.22
$M_{5m,180°}$	8.8	9.4	0.19
R2 – target-to-receiver distance = 4.0m			
$M_{1m,120°}$	14.2	12.2	0.72
$M_{2.5m,120°}$	10.9	9.0	0.67
$M_{1m,180°}$	8.2	6.4	0.65
$M_{2.5m,180°}$	8.4	6.5	0.66
$M_{1m,0°}$	8.5	6.7	0.64
$M_{2.5m,0°}$	8.8	6.8	0.70
R3 – target-to-receiver distance = 6.5m			
$M_{1m,120°}$	14.8	17.3	0.88
$M_{1m, 0°}$	10.2	13.6	**1.19**
$M_{2.5m,0°}$	9.8	12.7	**1.01**
$M_{5m,0°}$	10.2	12.7	0.91

Results of E_N exhibit values always lower that 1, except in the cases of far target-to-receiver distance (i.e., R3 at 6.5 m) when the noise source is co-located in front of the listener (i.e., 0°) and either at 1 m or 2.5 m of distance. This means that in those configurations the two BSIM models provide significantly different results and thus highlight that the

manipulation of the BRIRs with a split in early and late parts, as introduced in Rennies et al. [14], can account in a more accurate way for the degrading influence of reverberation on the speech signal with increasing distance. Furthermore, the model provided by Rennies et al. [14] can more effectively account for the binaural processing in real environments due to the physical and acoustic properties of the classroom, as the difference in the target level between the two ears in this model does not contain the late reflections that lead to the decorrelation of the signals across the ears.
Differences in SRTs across the different spatial configurations and with respect to the AO and PO conditions, were also evaluated to establish the extent to which binaural listening in complex acoustic scenarios could benefit from the spatial separation of the noise source. This benefit was evaluated in terms of spatial release from masking (SRM), that is, as the difference in SRTs between a co-located and a separated noise source position. In the AO condition, both models allow for the calculation of SRM values that range between -1 dB and 2 dB, therefore in the majority of cases they can be considered as negligible as a consequence of the long reverberation time and so of the detrimental effect of the reflections on speech intelligibility. In the PO condition, instead, moving the masker from 180° to 120° provides greater benefits compared to moving it from 0° to 120°, i.e., from ~3 dB to ~9 dB and from ~3 dB to ~6 dB maximum, respectively.

4. DISCUSSION

Overall, this study corroborates the findings from other studies and introduces research issues that deserve to be investigated in depth.
First, it is confirmed that SRTs are negatively affected by poor acoustic conditions in the environment, i.e., by long reverberation time and high noise levels as in the AO condition, rather than after the acoustic treatment that brought to short reverberation time and reduced noise in the PO condition. This, indeed, was also found in experimental studies such as that of Puglisi et al. [16] and D'Orazio and Garai [30], who found that the auditory ability of discriminating a target sound and assigning it to a specific source in the environment strongly depends on the reverberation characteristics of the sound field.
Second, the improvements in speech intelligibility can be considered as a consequence not only of the reduction of reverberation time in the classroom, but also of the increase of speech definition which was measured as significantly higher in the PO condition. Beutelmann et al., in two different studies [21,31], also consider D50 as a fundamental acoustic parameter to be considered in the understanding of the mechanisms that underlie speech intelligibility in complex acoustic scenarios.
Third, the strongly detrimental effect of reverberation on speech intelligibility has been underlined by the great differences in SRTs between AO and PO conditions that reach values up to ~14 dB SNR and ~17 dB SNR, respectively. Such negative effect was further degraded by the mutual spatial positioning of target- and noise-sources, which play a significant role in the listening experience when it happens in complex acoustic environments, as also underlined by Westermann and Buchholz [15].
Fourth, the findings on the spatial release from masking in the AO vs PO conditions reveal that the presence of reverberation significantly diminishes the efficacy of binaural cues, as the obtained SRM values in AO fall within the JND of ±1 dB and, only for the short target-to-receiver distance, reaches up to ~2 dB. The outcomes hereby presented in terms of SRM are consistent with other available studies. Particularly, Justine Hui et al. [32] and Kidd et al. [33] found that under similar conditions in presence of speech-shaped masking noise, SRM reaches up to 2 dB in reverberant environments.
Fifth, as far as the comparison between the results in the different acoustic conditions using the two BSIM models of Beutelmann et al. [21] and Rennies et al. [14] is concerned, statistically significant differences were only found applying the normalized error concept to the data in the case of far target-to-receiver distance (i.e., R3 at 6.5 m). This outcome puts in light that the two models both work accurately for a wide variety of cases, however in the option of using them as a prediction tool for early stages design, the extended model by Rennies et al. [14] is more effective as it better reflects the physical phenomenon that occurs in real complex acoustic scenarios due to the detrimental effect of reflections both on noise and on target signals.
Thanks to the outcomes of this study, and of those that similar research groups are leading at present, future investigations should also consider the use of the BSIM or of other prediction tools of binaural speech intelligibility for different design strategies of classrooms' acoustical treatment, including the use and the difference in quantity and positioning of acoustic surfaces with both absorbing and scattering properties. Furthermore, the prediction models' outcomes would benefit from a robust validation with empirical listening tests.

5. CONCLUSIONS

Optimal acoustic conditions in classrooms are needed to support effectively the speech communication process, and particularly speech intelligibility toward the pupils listening. The optimal acoustic design of classrooms is thus mandatory at early stages and the use of accurate prediction tools can contribute significantly. The Binaural Speech Intelligibility Model (BSIM) has been developed and ameliorated to account for the detrimental effect of reflections in complex acoustic scenarios both on the target and on the noise signals.

This study highlights the potential of the application of the BSIM under significantly different room acoustic conditions, i.e., in a reverberant classroom without acoustic treatment (ante-operam, AO) and in the same classroom with optimal reverberation after an acoustic treatment (post-operam, PO). First, the predicted SRTs were lower (better) in the PO condition, as expected, as the differences found with both BSIM models were always positive. Second, the spatial separation of the noise source with respect to the receiver's ears (i.e., when it is at 120°) SRTs are lower and the SRM reaches values greater than ~3 dB both in AO and PO, and up to ~6 dB in AO and up to ~9 dB in PO. Third, the updated version of the BSIM, the one ameliorated in Rennies et al. [14], considers in a more accurate way the detrimental effect of late reflections, especially for speech intelligibility predictions with large distance between the target source and the receiver.

6. ACKNOWLEDGMENTS

This study is part of a research that has been partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 A5.

7. REFERENCES

[1] G.E. Puglisi, A. Astolfi, L.C. Cantor Cutiva, and A. Carullo: "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, 441–52, 2017.

[2] A. Astolfi, P. Bottalico, and G. Barbato: "Subjective and objective speech intelligibility investigations in primary school class-rooms," *J. Acoust. Soc. Am.*, 131, 247–57, 2012.

[3] N. Prodi, C. Visentin, and A. Feletti: "On the perception of speech in primary school classrooms: Ranking of noise interference and of age influence," *J. Acoust. Soc. Am.*, 133, 255–68, 2013.

[4] G.E. Puglisi, A. Prato, T. Sacco, and A. Astolfi: "Influence of classroom acoustics on the reading speed: A case study on Italian second-graders," *J. Acoust. Soc. Am.*, 144, 144-149, 2018.

[5] L.L. Mendel: "Current considerations in pediatric speech audiometry," *Int. J. Audiol.*, 47, 546–53, 2008.

[6] C. Visentin, N. Prodi, F. Cappelletti, S. Torresin, and A. Gasparella. "Using listening effort assessment in the acoustical design of rooms for speech," *Building and Environment*, 136, 38–53, 2018.

[7] G. Minelli, G.E. Puglisi, and A. Astolfi: "Acoustical parameters for learning in classroom: a review," *Building and Environment*, 208, 108582, 2021.

[8] D. D'Orazio, D. De Salvio, L. Anderlucci, and M. Garai: "Measuring the speech level and the student activity in lecture halls: Visual- vs blind-segmentation methods," *Applied Acoustics*, 169, 107448, 2020.

[9] D. De Salvio and D. D'Orazio: "Effectiveness of acoustic treatments and PA redesign by means of student activity and speech levels," *Applied Acoustics*, 194, 108783, 2022.

[10] E.C. Cherry: "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.*, 25, 975–979, 1953.

[11] A.W. Bronkhorst: "Cocktail party phenomenon," *Acta Acustica United With Acustica*, 86, 117–128, 2000.

[12] A. Warzybok, J. Rennies, T. Brand, S. Doclo, and B. Kollmeier: "Effects of spatial and temporal integration of a single early re-flection on speech intelligibility," *J. Acoust. Soc. Am.*, 133, 269–82, 2013.

[13] M.L. Hawley, R.Y. Litovsky, and J.F. Culling: "The benefit of binaural hearing in a cocktail party: effect of location and type of interferer," *J. Acoust. Soc. Am.*, 115, 833–843, 2004.

[14] J. Rennies, T. Brand, and B. Kollmeier: "Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet," *J. Acoust. Soc. Am.*, 130, 2999–3012, 2011.

[15] A. Westermann and J.M. Buchholz: "The effect of spatial separation in distance on the intelligibility speech in rooms," *J. Acoust. Soc. Am.*, 137, 757–767, 2015.

[16] G.E. Puglisi, A. Warzybok, A. Astolfi, and B. Kollmeier: "Effect of reverberation and noise type on speech intelligibility in real complex acoustic scenarios," *Building and Environment*, 204, 108137, 2021.

[17] A.W. Bronkhorst and R. Plomp: "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.*, 83, 1508–1516, 1988.

[18] D.D. Dirks and R.H. Wilson: "The effect of spatially separated sound sources on speech intelligibility," *J. Speech Hear. Res.*, 12, 5–38, 1969.

[19] H. Levitt and L.R. Rabiner: "Predicting binaural gain in intelligibility and release from masking for speech," *J. Acoust. Soc. Am.*, 42, 820–829, 1967.

[20] M. Lavandier, J.F. Culling, and S. Jelfs: "Prediction of reverberant speech intelligibility against multiple noise interferers in rooms: Binaural useful-to-detrimental ratios," *J. Acoust. Soc. Am.*, 128, 2361, 2010.

[21] R. Beutelmann, T. Brand, and B. Kollmeier: "Revision, extension, and evaluation of a binaural speech intelligibility model," *J. Acoust. Soc. Am.*, 127, 2479–97, 2010.

[22] A. Astolfi, V. Corrado, and A. Griginis: "Comparison between measured and calculated parameters for the acoustical characterization of small classrooms," *Applied Acoustics*, 69, 966-976, 2008.

[23] EN ISO 3382-2, Acoustics - Measurement of Room Acoustic Parameters – Part 2: Reverberation Time in Ordinary Rooms, International Organization for Standardization, Genève, 2008.

[24] UNI 11532-2. Caratteristiche acustiche interne di ambienti confinati - Metodi di progettazione e tecniche di valutazione - Parte 2: Settore scolastico (Acoustic characteristics of indoor environments – Design methods and evaluation techniques – Part 2: school sector), Ente Italiano di Normazione 2020.

[25] G.E. Puglisi, A. Warzybok, S. Hochmuth, C. Visentin, A. Astolfi, N. Prodi, and B. Kollmeier: "An Italian matrix sentence test for the evaluation of speech intelligibility in noise," *Int. J. Audiol.*, 54, 44-50, 2015.

[26] G.E. Puglisi, F. Di Berardino, C. Montuschi, F. Sellami, A. Albera, D. Zanetti, R. Albera, A. Astolfi, B. Kollmeier, and A. Warzybok: "Evaluation of Italian Simplified Matrix Test for speech-recognition measurements in noise," *Audiology Research*, 11, 73-88, 2021.

[27] L. Wagner, L. Geiling, C. Hauth, T. Hocke, S. Plontke, and T. Rahne: "Improved binaural speech reception thresholds through small symmetrical separation of speech and noise," *PLoS One*, 15, 1–7, 2020.

[28] ISO/IEC 17043: Conformity assessment – General requirements for proficiency testing, 2010.

[29] G.E. Puglisi, A. Astolfi, L.C. Cantor Cutiva, and A. Carullo: "Assessment of indoor ambient noise level in school classrooms," in *Proc. of the 10th European Congress and Exposition on Noise Control Engineering - Euronoise 2015*, (Maastricht, NL), pp. 715-720, 2015.

[30] D. D'Orazio and M. Garai: "The autocorrelation-based analysis as a tool of sound perception in a reverberant field," *Rivista di Estetica*, 66, 133-147, 2017.

[31] R. Beutelmann and T. Brand: "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.*, 120, 331–42, 2006.

[32] C.T. Justine Hui, Y. Hioka, H. Masuda, and C.I. Watson: "Differences between listeners with early and late immersion age in spatial release from masking in various acoustic environments," *Speech Communicaiton*, 139, 51–61, 2022.

[33] G.Jr. Kidd, C.R. Mason, A Brughera, and W.M. Hartmann: "The role of reverberation in release from masking due to spatial separation of sources for speech identification," *Acta Acustica United With Acustica*, 526–536, 2005.