Journal Title
XX(X):1–22

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Assessment of speech intelligibility in scholar classrooms by measurements and prediction methods

Samantha Di Loreto[1], Michela Cantarini[2], Stefano Squartini [2], Valter Lori[1], Fabio Serpilli[1], Costanzo Di Perna[1]

Abstract

According to Italian regulation, the Ministerial Decree of 11 October 2017 about Environmental Criteria, reference values for acoustic indoor quality descriptors in public buildings are imposed. Regarding school environments, indoor acoustic quality targets refer to reverberation time, clarity, and speech intelligibility, whose representative acoustic descriptor is the speech transmission index (STI). This paper presents *pyeSTImate*, a Python-based tool for speech transmission index prediction in lecture rooms. The tool returns fully simulated results from the dimensions and material characteristics of classrooms with parallelepiped geometry and without limitations in size. Extensive experiments have been conducted with different simulation methods, evaluating the accuracy by comparison with in situ measurements selected from primary, secondary, and university classrooms in school buildings of the Marche Region in Italy. The combination of simulated speech transmission indexes with a prediction method based on an artificial neural network has also been evaluated. The analysis of the performance demonstrates the computational robustness of the tool that enables its use for the analysis of existing rooms, as well as for the renovation and design of new spaces.

Keywords

Speech intelligibility prediction, objective intelligibility measurement, classroom acoustics, room acoustic simulation and modeling

Introduction

Speech communication is a complex phenomenon that involves different modalities of speaker-listener interactions and conversational environments. The UNI EN ISO 9921[1] standard defines speech communication as *conveying or exchanging information using speech, speaking, hearing modalities, and understanding*, encompassing the aspects of speech quality in terms of *the amount of audible distortion of a speech signal*, vocal effort, delays related to reverberant environments, and speech intelligibility. The purpose of communication requires different levels of speech intelligibility, defined as the *rating of the proportion of speech that is understood* that assumes a key role in environments where the aim is speech understanding, such as lecture rooms. For this reason, acoustical standards and guidelines currently in use are designed to ensure good speech intelligibility in classrooms by providing acoustical indicators that take into account the distinct and joint effects of background noise and reverberation.

Compliance with minimum indoor acoustic comfort values assumes considerable importance under the environmental criteria according to the Italian law for the procurement of design and construction services for new construction, renovation and maintenance of public buildings[2]. The Criteri Ambientali Minimi (Ministerial Decree of 11 October 2017)[3] are indications aimed at directing public bodies towards rationalization of consumption and purchases and underline the importance of integrating environmental criteria in the different phases of the tender procedures (subject of the contract, technical specifications, related rewarding

[1] Department of Industrial Engineering and Mathematical Sciences, Università Politecnica delle Marche, via Brecce Bianche 12, 60131, Ancona, Italy
[2] Department of Information Engineering, Università Politecnica delle Marche, via Brecce Bianche 12, 60131 Ancona, Italy

Corresponding author:
Samantha Di Loreto, Department of Industrial Engineering and Mathematical Sciences, Università Politecnica delle Marche, via Brecce Bianche 12, 60131, Ancona, Italy

Email: s.diloreto@pm.univpm.it

technical characteristics the method of awarding the most economically advantageous tender, conditions for carrying out the contract). This Decree defines acoustic descriptors that must respect reference values to ensure occupant comfort and maximize the functional performance of the environment.

The present study focuses on one of the acoustic descriptors for school buildings, the speech transmission index (STI), a physical metric of speech intelligibility degraded by additive noise and reverberation[4]. A stand-alone prediction tool or in combination with an artificial neural network has been implemented to calculate STI values of lecture rooms, and the outcomes have been compared with the findings of measurements performed in classrooms of different grades, construction types and sizes.

Background

This subsection summarizes some relevant work on the evaluation of objective and subjective factors affecting speech intelligibility in lecture rooms.

The theme of acoustic comfort related to ambient noise, sound insulation, reverberation time, and speech intelligibility in classrooms has been the focus of research worldwide[5,6]. The correlation between the acoustic quality of a classroom and the development of student's cognitive abilities has been analyzed in classrooms of different grades, showing that speech intelligibility is the first requirement for a good learning environment in each school-aged group and is affected by many factors, such as reverberation time (RT) and signal-to-noise ratio (SNR).

Several studies have addressed the influence of acoustic conditions on children's attention and memory during primary school[7–9] as, at the early stage of education, perceptual speech abilities are still developing, and the presence of noise and reverberation prevent maximization of teaching comprehension. In Prodi et al. (2013)[10], the speech perception performance of younger pupils in noisy classrooms has been investigated through listening tests in real classrooms and supplementary tests in quiet. The statistical analysis is based on stochastic ordering and is able to clarify the class behavior and the different impacts of noises on performance.

Although the ability to discern noise from speech increases with age, poor classroom acoustic quality also causes discomfort and annoyance in secondary school students[11,12], affecting skills such as memory for spoken lectures[13] and reading speed[14]. In Prodi et al. (2019)[15], the effects of type of noise, age, and gender on children's speech intelligibility and sentence comprehension have been investigated. The study was conducted on children aged 11-13 in ecologically-valid conditions (collective presentation in real, reverberating classrooms), showing that performance is influenced by the sound environment and the listener's characteristics.

The same issues are found for young adults in the case of university classrooms, where background noise influences attention processes, speech intelligibility and listening effort[16], especially for non-native listeners[17]. In Choi[18], speech intelligibility tests have been conducted in university lecture rooms, demonstrating that young adult listeners could achieve correct speech intelligibility scores at lower SNR values than younger primary school students.

Some authors addressed the topic of the effects of noise and reverberation on the listening effort in adults. Picou et al.[19] have shown that speech intelligibility results from the reduction of background noise more than from acoustic correction of the environment, in contrast with current models of listening effort. The degree to which noise interferes with speech depends on several factors, including the intensity of noise compared to speech, fluctuations in noise level over time relative to speech and the spectral characteristics of the signal. Puglisi et al.[20] explored the mechanisms underlying speech intelligibility for adults in primary school classrooms, representing real-world and complex acoustic scenarios in which speech communication occurs. Moreover, excessive noise, too much reverberation, or their combined effects impact not only the performance of listeners but also the speaking effort of the teacher[21–23].

While previous work has involved normal-hearing people, acoustic conditions of spaces assume greater importance for people with hearing impairments or other disabilities, for which recent studies have been conducted to assess the impact of acoustic discomfort in order to design comfortable and suitable living environments. Bettarello et al.[24] have investigated the combination of the needs of autistic people with hearing impairment or hypersensitivity to sound using assistive sensors and indoor acoustic requirements. Caniato et al.[25] have developed an approach to analyze stress induced on autistic people in uncomfortable domains, including the acoustic one.

Although the research involves different acoustic contexts, the findings suggest that factors reducing speech intelligibility, such as background noise and reverberation, must be constantly checked and mitigated. The requirement to maximize acoustic comfort in environments where the speech message is paramount has increased the interest in developing effective objective indicators of speech intelligibility correlated with measured parameters such as reverberation time, early decay time, energy ratio, and speech transmission index. Nowadays, scientists consider the STI the parameter that best reflects speech intelligibility in a sound transmission system[26], and consequently, its measurement correlates well with subjective intelligibility scores for stimuli distorted by linear filtering, reverberation and additive noise.

In Schwerin et al.[27], the STI approach has been revisited to improve its correlation to subjective intelligibility scores. A modification that processes the modulation envelope in short time segments has been proposed, requiring only a quasi-stationarity assumption (rather than the STI stationarity assumption) of the modulation signal. In Yang et al.[28], subjective speech intelligibility scores and speech transmission index in secondary school and university classrooms have been correlated by regression models. Based on the tests described by Liu et al.[29], the relationship between STI and speech intelligibility in large spaces has been modified, and a new threshold for the STI assessment has been proposed. Peters[30] studied the potential binaural effect of reduced reflection and reverberation, finding that these conditions decrease intelligibility due to echoes and strong discrete delayed reflections and lead to incorrect STI evaluation.

Predictive methods have been developed in the literature to determine the speech transmission index of a room from reverberation time and signal-to-noise ratio in reverberant and absorbent environments, comparing the results with measured STI values[31] or simulated using acoustic modeling software[32]. Leccese et al.[33] compared various experimental equations for fast estimation of the speech transmission index as a function of reverberation time with values obtained from a campaign of STI measurements in university classrooms. Analysis of various types of classrooms determined which equation produced the best prediction accuracy and new equations for fast estimation of STI have been presented.

Contribution of this work

This work originates from the assessment of speech intelligibility in 35 classrooms of several grades belonging to buildings with different structural characteristics and construction areas located in the Marche Region, Italy. Sorting the classrooms according to volumetry (small size with $V<200\,\mathrm{m}^3$, medium size with $200\,\mathrm{m}^3 \leq V<350\,\mathrm{m}^3$, and wide size with $V \geq 350\,\mathrm{m}^3$), we have a set composed of primary classrooms (7 small-sized), secondary classrooms (9 small-sized, 1 medium-sized, and 1 wide-sized), and university lecture rooms (1 small-sized, 5 medium-sized, and 11 wide-sized). In the analyzed classrooms, direct measurements of the RT have been performed by applying the assessment procedure in UNI 11532-2[34], and the STI values have been calculated by the indirect method described in the IEC 60268-16[35] (BS EN 60268-16)[36]. This standard specifies objective methods for rating the transmission quality of speech with respect to intelligibility, and the UNI 11532-1[37], providing reference values for descriptors representing the acoustic quality of an environment in relation to the destination of use (including schools), refers to it. In addition, we have assigned each classroom an intelligibility rating (IR), a discrete qualification according to a five-point scale of speech comprehension quality (“bad”, “poor”, “fair”, “good”, “excellent”) dependent on the range in which the computed STI falls as described in the UNI 11532-1, Table 1.

Direct reverberation time measurement may be difficult or even infeasible in some situations, such as for school buildings outside the region, classrooms during renovation works or at the design stage. According to the UNI 11532-1 standard, predictive methods are also allowed to compute the room impulse response in indoor environments in order to optimize the acoustic descriptor under consideration. Expensive simulators of reverberant rooms handling complex geometries via cad interface and providing extensive materials libraries are commercially available for this purpose, but lecture rooms do not always require sophisticated analyses. In most cases, the geometry is parallelepipedal, and the acoustic characteristics of the commonly used materials are tabulated. In light of the above, we have developed *pyeSTImate*[*], an open-source STI prediction tool based on the python[38] programming language and representing an extension of the room impulse response (RIR) simulator included in the pyroomacoustics[39] software package.

The main contributions of this work are summarized as follows.

1. The speech transmission index of the analyzed classrooms is calculated using the *pyeSTImate* tool, employing three different simulation methods to compute reverberation times.
2. For each simulation method, we compare STI values derived from measured and simulated reverberation times by monitoring the absolute error and absolute percentage error in single classrooms, lecture rooms belonging to schools of the same grade, and the entire set of classrooms.

[*] https://github.com/michelacantarini/pyeSTImate

3. The comparison of intelligibility ratings obtained through measurements and simulations is also conducted. In this regard, we introduce the intelligibility rating error (IRE) metrics, corresponding to the normalized absolute error in the range of actual values adapted to the discrete classification of speech comprehension quality.
4. We train an artificial neural network (ANN) with a synthetic dataset of STIs associated with classrooms with randomly chosen geometric and material characteristics, employing a significantly reduced number of input data compared to the full *pyeSTImate* tool. The model is then tested with the real dataset of classrooms by re-assessing the error between measured, simulated and predicted results.

The analysis of simulation results, compared with measurement findings also in terms of just noticeable difference (JND) units in STI values[40], illustrates in which settings and environments the tool exhibits greater robustness for its use in preliminary or design assessments of speech intelligibility in classrooms.

The paper is organized as follows: Section Materials and methodspresents the background on room acoustics metrics, the characteristics of the lecture rooms analyzed, the measurement equipment of the case studies, the acoustic assessment of the classrooms, and the simulation methodologies. The simulation tool is illustrated in detail in Section PyeSTImate: a Python-based tool for speech transmission index prediction, while the results of predictive analyses and discussion are reported in Section Results and discussion. Finally, Section Conclusions summarizes the work and comments on possible future extensions.

Materials and methods

This section first introduces the main descriptors that characterize the indoor acoustic quality of school environments, then illustrates the characteristics of the evaluated classrooms and the measurement equipment. The calculation method of the speech transmission index defined as "statistical" or "indirect" is explained in detail. Finally, an overview of the predictive STI methods using the implemented simulation tool, also in combination with an artificial neural network, is provided.

Room acoustic metrics

According to standards UNI 11532-1 and UNI 11532-2, the measurement session includes the reverberation time (RT), speech transmission index (STI) and clarity (C_{50}).

RT_{20} and RT_{30} are the reverberation times estimated by the slope of the Schroeder backward-integrated decay, respectively, in the [dB] ranges: [-5, -25] for RT_{20} and [-5, -35] for RT_{30}[41].

C_{50} is the ratio, in dB, between the "useful energy" received in the first 50 ms of the impulse response to the energy received in subsequent instants. The term "energy" represents the square of the instantaneous values of the pressure impulse response. The C_{50} is defined in the ISO 3382-1[42] through the following Equation (1).

$$C_{50} = 10 \log \frac{\int_0^{50\,ms} p^2(t)\,dt}{\int_{50\,ms}^{\infty} p^2(t)\,dt} dB \tag{1}$$

As mentioned in the introduction, STI is an objective measure to predict the intelligibility of speech transmitted from talker to listener by a transmission channel. The STI method applies a specific test signal to the transmission channel, and by analyzing the received test signal, the speech transmission quality of the channel is derived and expressed in a value between 0 and 1. Subjective and objective intelligibility indexes are correlated through the discrete qualification given by the intelligibility rating (IR), as expressed in UNI EN ISO 9921[1], Table F.1 and incorporated by UNI 11532-1 with regard to STI in confined spaces belonging to several uses, including the educational destination. In Table 1, the correlation between STI and IR is shown.

Table 1. Correlation between speech transmission index and intelligibility rating according to UNI 11532-1

STI values	IR
0.00<STI≤0.30	Bad
0.30<STI≤0.45	Poor
0.45<STI≤0.60	Fair
0.60<STI≤0.75	Good
0.75<STI≤1.00	Excellent

For the educational sector, the reference STI and C_{50} values are indicated in paragraphs 4.3–4.4 of the UNI 11532-2 standard and refer to a full environment with a maximum of two people (technicians). The categories of the environment concerning the destination use are shown in Table 2.

The C_{50} descriptor applies to categories A1, A2, A3 and A4 as an alternative to the STI for rooms with a volume less than 250 m^3, while the STI alone is allowed for rooms with a volume greater than 250 m^3. For both descriptors, the verification methods to be applied are provided in the standard and supplemented with the specifications needed for the case.

In paragraph 4.5 of the UNI 11532-2 standard, an optimal reverberation time RT_{ott} corresponding to a conventional

Table 2. Categories of the environment in relation to the destination use according to UNI 11532-2

Category	Activities in the environment	Methods of intervention
A1	Music	Objective achieved with integrated design of geometry, furniture, residual noise control
A2	Spoken/Conference	
A3	Lesson/communication as speech and lecture	
A4	Lesson/communication, special classroom lecture	
A5	Sport	
A6	Areas and spaces not intended for learning and libraries	Objective achieved with sound absorption and residual noise control

occupation of the environment equal to 80% for categories A1–A4 and the unoccupied environment for category A5 is defined.

For categories A1–A4, if the measurement is performed in a furnished but unoccupied environment, the measured values must be corrected with Equation (2) to compare them with the reference limits.

$$RT_{inocc} = \frac{RT_{occ}}{[1 - RT_{occ} * \frac{\Delta A_{pers}}{0.16V}]} \tag{2}$$

where:

- RT_{occ} is the optimal reverberation time for the room occupied at 80%, in seconds;
- RT_{inocc} is the optimal reverberation time when the room is not occupied (measurement result), in seconds;
- V is the volume of the room, in cubic meters;
- ΔA_{pers} is the equivalent additional surface area of acoustic absorption of people, in square meters.

The reference values of the optimal reverberation time for A1–A4 categories are reported in Table 3.

Table 3. Categories of the occupied environment in relation to the destination use according to UNI 11532-2

Category	Occupied environment 80%	
A1	RT_{ott}=(0.45 logV + 0.07)	$30\,m^3 \leq V < 1000\,m^3$
A2	RT_{ott}=(0.37 logV - 0.14)	$50\,m^3 \leq V < 5000\,m^3$
A3	RT_{ott}=(0.32 logV - 0.17)	$30\,m^3 \leq V < 5000\,m^3$
A4	RT_{ott}=(0.26 logV - 0.14)	$30\,m^3 \leq V < 500\,m^3$

Characteristics of lecture rooms and measurement equipment

The buildings selected for the measurement campaign are representative of Italian mild-climate schools. They comprise a heterogeneous sample of the diversity of school buildings in the Marche Region in terms of structure type, year of construction, educational stages, activities performed, built area, and materials and construction techniques used. The selected schools have no particular architectural qualities and pay no attention to bioclimatic issues and form factors. For each school building, the collection of design documentation, geometric survey and visual analysis of finishing materials of the classrooms have been performed. The main finishing materials in the classrooms are: stoneware tiles and pvc for floors; painted plasterboard, gypsumboard, and smooth plaster for walls; plasterboard, plastic material, wood planking, and smooth plaster for ceilings. In addition, surfaces with heavy glass for windows and light glass for the glazed parts of wall panels, metal material for fire doors, plastic or graphite for blackboards, wooden or unpadded plastic seats, and wooden desks and chairs are included. Table 4 summarizes the main characteristics of the lecture rooms: type of school (school number and grade), name of the room, year of construction, location characteristics and main finishing materials.

The list of geometric dimensions of each classroom is provided in Table 5. The acoustic characterization of the classrooms has been carried out in compliance with the UNI 11532-2 standard. Figure 1 illustrates the measurement positions according to the standard: four positions have been selected, three along the longitudinal axis of the classroom and one representative of the most unfavorable condition in terms of distance from the speaker and proximity to the noise produced by the indoor plant. Figure 2 and Figure 3 show plans and photographs of some classrooms selected for the case studies.

The measurement of RT_{30} has been performed according to the ISO 3382-2[41] standard, which requires measurements for at least two source positions and three microphone positions. All investigations have been carried out using the Dirac room acoustics commercial software, combined with an Edirol FA-101 external firewire soundcard. In the case of intelligibility measurements, the sound field is excited using a directional sound source. We employed the Echo Speech Source (Type 4720), a small active loudspeaker box providing calibrated acoustic signals for speech intelligibility measurements using the Dirac room acoustics software and is typically placed at a human speaker position (1.50 m from the floor). The acquisition of impulse responses has been performed by taking the output signal of a B&K 2250 sound level meter. The signals contained in the Echo source are real speech fragments that are used to set the volume to a "normal" level. The voice signal has a standard level of 60 dB(A) and is measured at a distance of 1 m from the speaker.

Table 4. List of classrooms indicating school type (school number and grade), name of the room, year of construction, location, and main finishing materials

ID	Type		Name of room	Year	Location characteristics	Floor	Walls	Ceiling
1	SCH 1	University	140/1	1970	Urban outskirts	Artificial stone tiles, PVC	Plasterboard	Suspended ceiling
2	SCH 1	University	140/2	1970	Urban outskirts	Artificial stone tiles, PVC	Plasterboard	Suspended ceiling
3	SCH 1	University	140/3	1970	Urban outskirts	Artificial stone tiles, PVC	Plasterboard	Suspended ceiling
4	SCH 1	University	155/D1	1970	Urban outskirts	PVC	Plasterboard	Suspended ceiling
5	SCH 1	University	155/D2	1970	Urban outskirts	PVC	Plasterboard	Suspended ceiling
6	SCH 1	University	155/D3	1970	Urban outskirts	PVC	Plasterboard	Suspended ceiling
7	SCH 1	University	155/D4	1970	Urban outskirts	PVC	Plasterboard	Suspended ceiling
8	SCH 1	University	160/1	1970	Urban outskirts	Artificial stone tiles, PVC	Plasterboard	Suspended ceiling
9	SCH 1	University	160/2	1970	Urban outskirts	Artificial stone tiles, PVC	Plasterboard	Suspended ceiling
10	SCH 1	University	AT1	1970	Urban outskirts	PVC	Plasterboard	Suspended ceiling
11	SCH 1	University	AT2	1970	Urban outskirts	PVC	Plasterboard	Suspended ceiling
12	SCH 1	University	AT3	1970	Urban outskirts	PVC	Plasterboard	Suspended ceiling
13	SCH 1	University	EN1	1970	Urban outskirts	Artificial stone tiles	Plasterboard	Suspended ceiling
14	SCH 1	University	EN3	1970	Urban outskirts	Artificial stone tiles	Plasterboard	Suspended ceiling
15	SCH 1	University	S1	1970	Urban outskirts	Artificial stone tiles	Gypsum board	Non-suspended ceiling
16	SCH 1	University	S2	1970	Urban outskirts	Artificial stone tiles	Gypsum board	Non-suspended ceiling
17	SCH 1	University	S3	1970	Urban outskirts	Artificial stone tiles	Gypsum board	Non-suspended ceiling
18	SCH 2	Primary	1B	2002	Urban center	Artificial stone tiles	Plaster	Non-suspended ceiling
19	SCH 2	Primary	2A	2002	Urban center	Artificial stone tiles	Plaster	Suspended ceiling
20	SCH 2	Primary	3A	2002	Urban center	Artificial stone tiles	Plaster	Suspended ceiling
21	SCH 3	Primary	2C	2014	Rural urban center	Artificial stone tiles	Plaster	Suspended ceiling
22	SCH 3	Primary	3C	2014	Rural urban center	Artificial stone tiles	Plaster	Suspended ceiling
23	SCH 4	Primary	4B	2008	Rural urban center	Artificial stone tiles	Plaster	Suspended ceiling
24	SCH 4	Primary	4A	2008	Rural urban center	Artificial stone tiles	Plaster	Suspended ceiling
25	SCH 5	Secondary	1	2013	Urban center	Artificial stone tiles	Plaster	Non-suspended ceiling
26	SCH 5	Secondary	2	2013	Urban center	Artificial stone tiles	Plaster	Non-suspended ceiling
27	SCH 5	Secondary	3	2013	Urban center	Artificial stone tiles	Plaster	Non-suspended ceiling
28	SCH 6	Secondary	4	2012	Urban center	Artificial stone tiles	Plaster	Non-suspended ceiling
29	SCH 6	Secondary	5	2012	Urban center	PVC	Plaster	Non-suspended ceiling
30	SCH 7	Secondary	6	2012	Urban center	PVC	Plaster	Non-suspended ceiling
31	SCH 7	Secondary	7	2016	Urban center	PVC	Plaster	Non-suspended ceiling
32	SCH 7	Secondary	8	2016	Urban center	PVC	Plaster	Non-suspended ceiling
33	SCH 8	Secondary	9	2016	Urban center	PVC	Plaster	Suspended ceiling
34	SCH 8	Secondary	10	2004	Urban center	PVC	Plaster	Non-suspended ceiling
35	SCH 8	Secondary	11	2004	Urban center	PVC	Plaster	Non-suspended ceiling

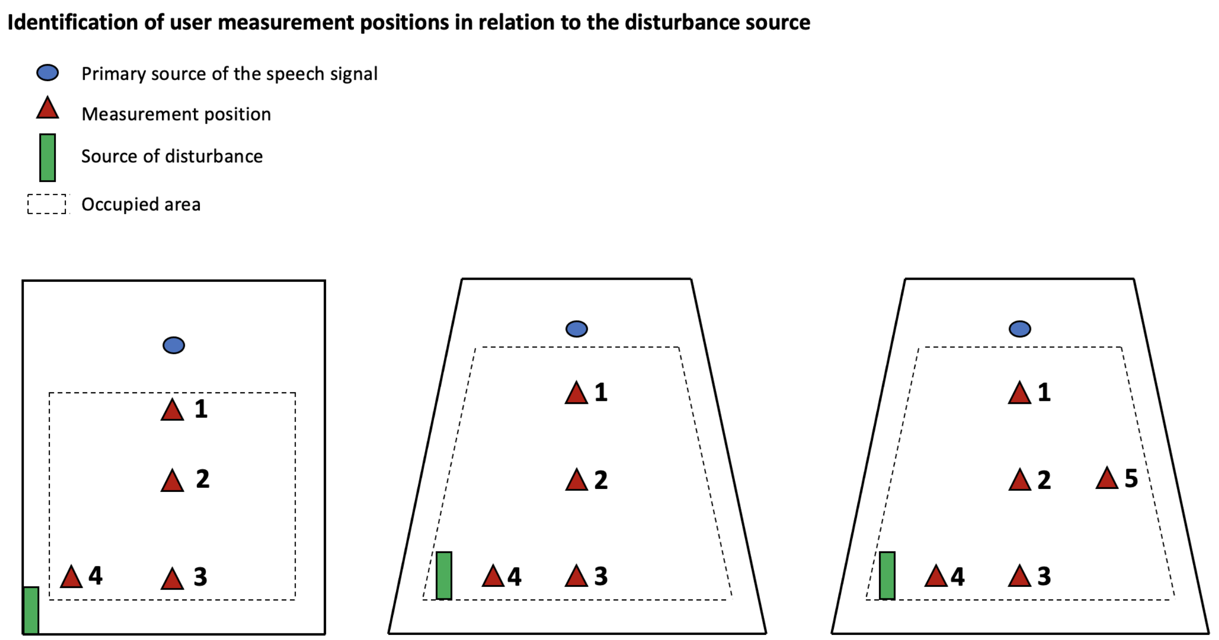

Figure 1. Measurement positions in classrooms according to the UNI 11532-2 standard

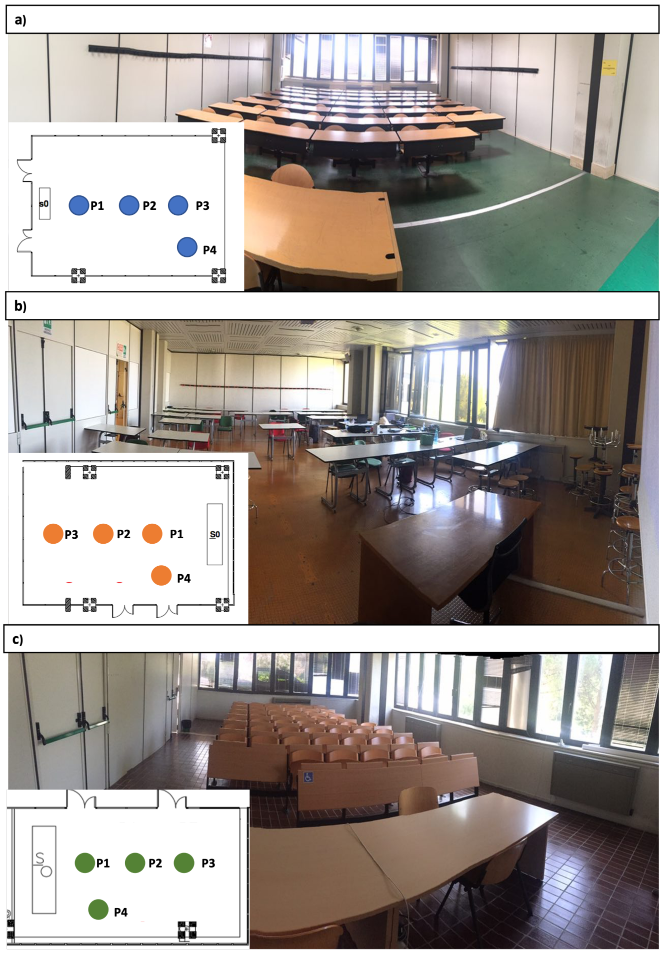

Figure 2. Plans and photographs of some classrooms of the School of Engineering of Università Politecnica delle Marche

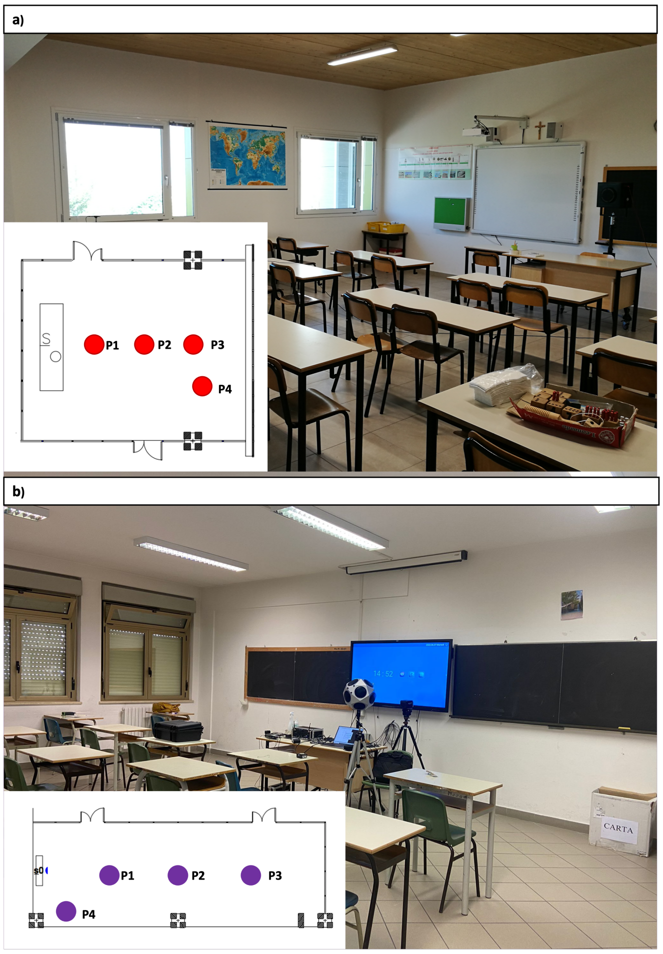

Figure 3. Plans and photographs of two classrooms of primary school (a) and secondary school (b)

Values for the RT and STI parameters for each classroom are presented in Table 6.

STI calculation using the indirect method

Based on the indirect method, the speech transmission index is calculated thanks to the modulation transfer function (MTF) using the method of Houtgast et al.[26] STI is based on the measurement of the MTF[43,44], which quantifies the reduction in the modulation index of a test signal, depending on the modulation frequency. For each modulation frequency, the MTF is determined by the ratio between the modulation index of the signal at the listener (m_0) and the modulation index of the test signal (m_i). A family of MTF curves is determined, in which each curve is relative to each octave band of speech emission and is defined by the values that the modulation index reduction factor (m) assumes for each modulation frequency present in the envelope of natural speech signals. For the STI measurement, 7-octave bands from 125 Hz to 8 kHz, and 14 modulation frequencies between 0.63 Hz and 12.5 Hz in one-third octave intervals, are considered. The 98 (7×14) m-values are finally summarized in a single index, the STI, varying between 0 and 1, representing the effect of the transmission system on intelligibility with or without sound amplification system. The modulation transfer function of the transmission path is quantified by comparing the ratio of the modulation depth at the output and input of the test signal, which is written as Equation (3):

$$m(f_m) = \frac{|\int_0^\infty h(t)^2 e^{-j2\pi f_m t}\, dt|}{\int_0^\infty h(t)^2\, dt}[1 + 10^{-\frac{\mathrm{SNR}}{10}}]^{-1} \qquad (3)$$

where:

- $m(f_m)$ is the modulation transfer function of the transmission channel;
- h(t) is the impulse response of the transmission channel;
- SNR is the signal-to-noise ratio in dB.

Considering a diffuse reverberant field, the impulse response is written as Equation (4):

$$h(t) = \frac{Q}{r^2}\,\delta(t) + \frac{13.8\,Q}{r_c^2\,T}e^{-\frac{13.8\,t}{T}} \qquad (4)$$

Table 5. List of classrooms indicating school type (school number and grade), name of the room, and geometric dimensions

ID	Type		Name of room	length [m]	width [m]	height [m]	surface [m^2]	volume [m^3]
1	SCH 1	University	140/1	18.2	9.0	3.4	163.8	556.9
2	SCH 1	University	140/2	11.9	8.8	3.9	105.2	410.3
3	SCH 1	University	140/3	15.3	8.9	3.9	136.5	537.6
4	SCH 1	University	155/D1	12.3	8.9	3.4	109.4	371.9
5	SCH 1	University	155/D2	12.3	8.9	3.4	109.4	371.9
6	SCH 1	University	155/D3	12.3	8.9	3.4	109.4	371.9
7	SCH 1	University	155/D4	12.3	8.9	3.4	109.4	371.9
8	SCH 1	University	160/1	10.2	8.9	3.4	90.9	309.1
9	SCH 1	University	160/2	10.4	8.9	3.4	93.2	317.0
10	SCH 1	University	AT1	13.4	9.3	3.0	125.2	375.7
11	SCH 1	University	AT2	13.8	9.3	3.0	128.6	385.9
12	SCH 1	University	AT3	17.9	6.3	3.0	113.1	339.4
13	SCH 1	University	EN1	11.3	8.9	3.4	100.7	342.3
14	SCH 1	University	EN3	11.2	6.4	3.4	71.4	242.7
15	SCH 1	University	S1	7.5	7.4	3.0	55.5	166.5
16	SCH 1	University	S2	10.4	10.6	3.5	110.2	385.8
17	SCH 1	University	S3	14.3	16.2	3.0	231.7	695.0
18	SCH 2	Primary	1B	7.3	6.8	3.3	49.6	163.8
19	SCH 2	Primary	2A	7.8	7.4	3.3	57.7	191.2
20	SCH 2	Primary	3A	6.8	6.1	3.3	41.5	136.9
21	SCH 3	Primary	2C	7.4	6.8	3.0	50.3	151.0
22	SCH 3	Primary	3C	7.8	7.4	3.0	57.3	172.0
23	SCH 4	Primary	4B	7.9	7.5	3.0	59.3	177.8
24	SCH 4	Primary	4A	8.1	7.8	3.0	63.4	190.3
25	SCH 5	Secondary	1	7.9	7.3	3.1	57.3	177.6
26	SCH 5	Secondary	2	7.9	7.5	3.2	59.2	189.3
27	SCH 5	Secondary	3	8.0	7.6	3.1	60.1	186.3
28	SCH 6	Secondary	4	6.8	6.7	3.0	45.2	135.5
29	SCH 6	Secondary	5	7.3	6.4	3.0	46.7	140.0
30	SCH 7	Secondary	6	8.4	6.7	4.4	56.3	247.6
31	SCH 7	Secondary	7	10.7	7.5	6.5	80.3	521.6
32	SCH 7	Secondary	8	7.1	7.3	3.0	51.7	155.1
33	SCH 8	Secondary	9	7.2	6.2	3.0	44.6	133.9
34	SCH 8	Secondary	10	7.8	6.7	3.0	52.3	156.8
35	SCH 8	Secondary	11	6.0	9.0	3.0	53.8	161.3

where:

- Q is the directivity factor for the sound source (talker);
- r is the talker to listener distance;
- T is the reverberation time of the room space.

The impulse response of each classroom is determined at the four different positions in the room, and the reverberation time is calculated by the method described in UNI EN 12354-6[45], starting from the sound absorption of the room.

STI prediction using simulation tool

This paper proposes *pyeSTImate*, a predictive tool of speech transmission index in room acoustic environments. This application aims to compute the STI without recourse to direct measurements of room reverberation times. RIRs can be simulated with different techniques in the literature, given the geometry of the room, constituent materials, furniture and kind of occupants, and source and receivers' positions. After calculating reverberation times from the RIRs, STIs are determined using the indirect method. The extensive case history of classroom measurements has been used to fine-tune the simulator settings: comparing the STI values derived from the measured RTs and the results of different simulation methods allowed us to assess which settings most accurately reproduce the actual environment. The result is a tool for acoustics engineers suitable for the analysis of existing rooms, as well as for the renovation and design of new spaces.

Whereas *pyeSTImate* requires detailed input geometric data, especially regarding the surfaces to be associated with each material, in some cases, we only have some of this information, e.g., floor plans and photos of the room. To address this issue, we have adapted the algorithm to generate a synthetic dataset of classrooms characterized by randomly chosen sizes and materials, and in each of them, the average speech transmission index over four listening positions has been calculated. The dataset, composed of reduced input data compared to the full *pyeSTImate* tool, has been employed to train an artificial neural network capable of predicting STI with good approximation by providing only classroom size and material characteristics. This further application estimates the speech transmission index and speech comprehension quality with very limited details about the target room, demonstrating its usefulness in preliminary acoustic analyses.

In the next section, we present all parts of the *pyeSTimate* tool, from the required input data to the computational processes that provide the output STI descriptor. Then we describe the adaptation of the tool to the generation of a synthetic dataset of STIs obtained from classrooms with random geometric features and materials used for training an artificial neural network for STI prediction with a reduced number of data.

PyeSTImate: a Python-based tool for speech transmission index prediction

Overview

The core of the *pyeSTImate* tool is pyroomacoustics[39], a software package aimed at the rapid development and testing of audio array processing algorithms, properly adapted and extended to predict the STI just from dimensional room data, materials, and source/receiver positions. As illustrated in Figure 4, *pyeSTImate* is composed of two main blocks. The first, based on pyroomacoustics, returns the simulated room impulse responses and reverberation times in octave bands that input the second part of the algorithm, designed for STI computation using the indirect method. Based on the achieved STI value, the tool also provides the intelligibility rating, a qualification according to a five-point scale of speech comprehension quality.

Table 6. Measured RT_{30} (125 Hz-8 kHz), average RT_{30}, STI, average STI, and intelligibility rating for classrooms under acoustic speech intelligibility assessment

ID	Type		Name of room	RT_{30} [s]								STI					Intelligibility Rating
				125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	avg	P1	P2	P3	P4	avg	
1	SCH 1	University	140/1	0.90	0.93	0.76	0.71	0.60	0.75	0.72	0.77	0.57	0.52	0.46	0.43	0.50 ± 0.06	fair
2	SCH 1	University	140/2	0.51	0.86	0.97	0.72	0.50	0.77	0.57	0.70	0.52	0.40	0.37	0.34	0.41 ± 0.08	poor
3	SCH 1	University	140/3	0.55	0.57	0.50	0.60	0.69	0.93	0.83	0.67	0.60	0.58	0.35	0.33	0.47 ± 0.14	fair
4	SCH 1	University	155/D1	0.55	0.93	0.87	0.88	0.79	0.37	0.44	0.69	0.63	0.57	0.54	0.52	0.57 ± 0.05	fair
5	SCH 1	University	155/D2	0.53	0.88	0.87	0.86	0.75	0.32	0.37	0.65	0.62	0.57	0.54	0.52	0.56 ± 0.04	fair
6	SCH 1	University	155/D3	0.61	1.02	0.93	0.89	0.78	0.37	0.42	0.72	0.55	0.51	0.53	0.49	0.52 ± 0.03	fair
7	SCH 1	University	155/D4	0.61	0.86	0.88	0.87	0.75	0.35	0.41	0.68	0.60	0.55	0.54	0.52	0.55 ± 0.03	fair
8	SCH 1	University	160/1	0.67	1.19	0.81	0.81	0.71	0.65	0.61	0.78	0.59	0.42	0.51	0.39	0.48 ± 0.09	fair
9	SCH 1	University	160/2	0.62	0.76	0.99	0.99	0.87	0.75	0.61	0.80	0.58	0.50	0.45	0.40	0.48 ± 0.08	fair
10	SCH 1	University	AT1	0.42	0.85	0.74	0.72	0.82	0.95	0.26	0.68	0.73	0.61	0.59	0.45	0.60 ± 0.11	fair
11	SCH 1	University	AT2	0.55	0.35	0.45	0.74	0.78	0.82	0.21	0.56	0.70	0.60	0.60	0.56	0.62 ± 0.06	good
12	SCH 1	University	AT3	0.49	0.57	0.48	0.60	0.88	0.98	0.30	0.61	0.82	0.62	0.54	0.50	0.62 ± 0.14	good
13	SCH 1	University	EN1	0.72	1.16	0.85	0.78	0.64	0.71	0.41	0.75	0.61	0.56	0.55	0.41	0.53 ± 0.09	fair
14	SCH 1	University	EN3	0.71	0.99	0.88	0.77	0.63	0.68	0.38	0.72	0.56	0.52	0.49	0.43	0.50 ± 0.05	fair
15	SCH 1	University	S1	2.40	1.80	1.70	1.93	1.94	1.66	1.24	1.81	0.29	0.25	0.28	0.29	0.28 ± 0.02	bad
16	SCH 1	University	S2	2.30	1.90	1.72	1.94	1.87	1.66	1.30	1.81	0.28	0.26	0.28	0.27	0.27 ± 0.01	bad
17	SCH 1	University	S3	2.26	1.72	1.75	1.98	1.97	1.68	1.25	1.80	0.25	0.28	0.29	0.30	0.28 ± 0.02	bad
18	SCH 2	Primary	1B	0.63	0.73	0.83	0.88	0.90	0.81	0.71	0.78	0.66	0.64	0.62	0.57	0.62 ± 0.04	good
19	SCH 2	Primary	2A	1.19	1.24	1.33	1.39	1.22	1.13	0.89	1.20	0.54	0.57	0.54	0.53	0.55 ± 0.02	fair
20	SCH 2	Primary	3A	1.01	1.09	1.37	1.50	1.34	1.24	0.88	1.20	0.53	0.56	0.55	0.54	0.55 ± 0.01	fair
21	SCH 3	Primary	2C	0.89	1.22	1.61	1.75	1.71	1.47	1.03	1.38	0.48	0.50	0.52	0.48	0.50 ± 0.02	fair
22	SCH 3	Primary	3C	0.88	1.05	1.14	1.13	1.06	0.93	0.77	0.99	0.54	0.61	0.54	0.47	0.54 ± 0.06	fair
23	SCH 4	Primary	4B	0.70	0.84	0.87	0.93	0.90	0.81	0.68	0.82	0.57	0.62	0.63	0.59	0.60 ± 0.03	fair
24	SCH 4	Primary	4A	1.37	1.51	1.69	1.79	1.60	1.42	1.09	1.50	0.44	0.47	0.41	0.39	0.43 ± 0.04	poor
25	SCH 5	Secondary	1	0.66	0.81	0.94	0.98	0.88	0.84	0.69	0.83	0.62	0.60	0.59	0.54	0.59 ± 0.03	fair
26	SCH 5	Secondary	2	1.02	1.40	1.33	1.30	1.12	1.01	0.83	1.14	0.58	0.55	0.53	0.50	0.54 ± 0.03	fair
27	SCH 5	Secondary	3	0.76	0.89	0.89	0.95	0.89	0.86	0.71	0.85	0.57	0.58	0.57	0.56	0.57 ± 0.01	fair
28	SCH 6	Secondary	4	1.38	1.40	1.33	1.36	1.37	1.36	1.35	1.36	0.54	0.49	0.49	0.43	0.49 ± 0.05	fair
29	SCH 6	Secondary	5	0.81	0.62	0.96	1.07	1.22	1.08	0.99	0.96	0.58	0.56	0.49	0.44	0.52 ± 0.06	fair
30	SCH 7	Secondary	6	0.82	0.69	0.62	0.58	0.58	0.57	0.57	0.63	0.77	0.70	0.70	0.70	0.72 ± 0.04	good
31	SCH 7	Secondary	7	1.13	0.93	0.79	0.76	0.72	0.67	0.67	0.81	0.64	0.49	0.47	0.47	0.52 ± 0.08	fair
32	SCH 7	Secondary	8	1.03	0.70	0.64	0.52	0.56	0.58	0.59	0.66	0.65	0.53	0.51	0.51	0.55 ± 0.07	fair
33	SCH 8	Secondary	9	0.95	0.79	0.58	0.58	0.60	0.61	0.61	0.67	0.54	0.52	0.49	0.55	0.53 ± 0.03	fair
34	SCH 8	Secondary	10	2.43	2.14	2.31	2.46	2.30	1.95	1.49	2.15	0.38	0.31	0.33	0.34	0.34 ± 0.03	poor
35	SCH 8	Secondary	11	2.29	2.42	2.13	1.87	1.90	1.71	1.68	2.00	0.44	0.37	0.38	0.38	0.39 ± 0.03	poor

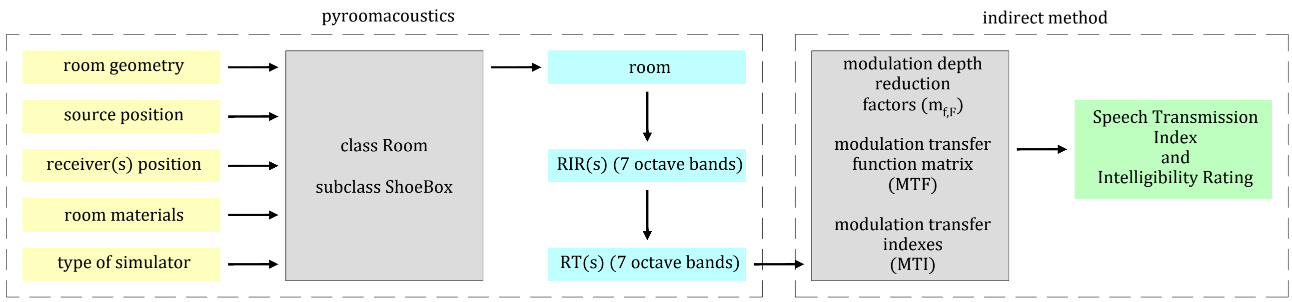

Figure 4. Block diagram of *pyeSTImate*, the proposed tool for predictive calculation of speech transmission index as an adaptation and extension of pyroomacoustics

Input data

The code has been implemented to handle input data with text-based files (e.g., json or csv), with the advantage of loading and processing multiple rooms simultaneously through an iterative process. The data required by the tool are room geometry, location of source and receivers, room materials, and type of simulator.

Room geometry. It is possible to define rooms with parallelepiped geometry, called shoebox in the pyroomacoustics package and characterized by arrays composed of 'length', 'width' and 'height' in meters. The advantage of such geometry is the simplicity of definition and efficiency of the simulation, without incurring excessive approximations considering the geometric regularity of classrooms.

Source position. The primary source position (S) of the speech signal is defined by an array containing the coordinates in meters, denoted as 'source_coordinates'.

Receiver(s) position. Measurement user positions (P1–P4) are also defined by arrays containing the coordinates of each receiver in meters, referred to as 'receiver_1_coordinates', 'receiver_2_coordinates', 'receiver_3_coordinates', and 'receiver_4_coordinates', respectively. It is possible to configure up to four user positions for lecture rooms without amplification systems, as shown in Table 11 of UNI 11532-2.

Room materials. In pyroomacoustics, materials that constitute the shoebox faces, such as floor, ceiling and walls, are managed by the `pyroomacoustics.parameters.Material` object. A material is defined at least by an absorption coefficient, representing the ratio of sound energy absorbed by a wall upon reflection. In addition, one or more scattering coefficients corresponding to the ratio of energy dispersed upon reflection can be specified. The coefficients assigned to materials can be provided as scalars, representing uniform absorption or scattering at all frequencies, or as lists of coefficients, each associated with a specific octave band. Materials are set by directly providing the coefficients or choosing them from the pyroomacoustics library database. Different materials are assigned to each face of the shoebox, specified in a dictionary and named as 'ceiling', 'floor', 'east', 'west', 'north', and 'south'.

In *pyeSTImate*, the definition of materials has been extended and adapted to school environments. For this purpose, we have supplemented the pyroomacoustics dictionary with octave-band absorption coefficients of materials, furniture and occupants found in Tables C1–C3 of UNI 11532-2. We have created a dictionary for each table, in which the description of the entries follows the 'table–item_number' criterion. For example, the material listed at number 3 of Table C.2 (smooth plaster) is defined as 'C.2-3'. In most cases, classrooms are covered with several finishing materials, such as tile for the floors, plaster and glass for the walls, plasterboard for the ceiling, and sometimes even a single surface features more than one material. Moreover, lecture rooms contain furniture (e.g., chairs, desks, blackboards, projector screens, and lockers). To accurately represent the sound absorption of each k-surface in the room, we have defined a function that calculates the average absorption coefficient $\alpha_{\mathrm{k,m}}$, obtained as:

$$\alpha_{k,m} = \sum_{i=1}^{n} \frac{\alpha_{i,mat} S_{i,mat} + \alpha_{i,pers} S_{i,pers} + \alpha_{i,obj} S_{i,obj}}{S_i}, \tag{5}$$

where:

- $\alpha_{\mathrm{i,mat}}$ is the absorption coefficient of the i-th material (Table C.2 UNI 11532-2)
- $\mathrm{S_{i,mat}}$ is the area of the i-th material
- $\alpha_{\mathrm{i,pers}}$ is the absorption coefficient of the i-th category of occupant (person) (Table C.1 UNI 11532-2)
- $\mathrm{S_{i,pers}}$ is the area of the i-th occupant (person)
- $\alpha_{\mathrm{i,obj}}$ is the absorption coefficient of the i-th furniture (object) (Table C.3 UNI 11532-2)
- $\mathrm{S_{i,obj}}$ is the area of the i-th furniture (object)
- $\mathrm{S_i}$ is the total area of the k-surface

It is possible to define up to five types of materials ('floor_material', 'ceiling_material', 'wall_material'), occupant categories ('Apers'), and furniture ('Aobj') and assign them to each parallelepiped surface. The primary material (No. 1) is automatically associated with the corresponding surface area; if secondary materials are present, their surface area ('area_floor_material', 'area_ceiling_material', 'area_south_wall_material', 'area_east_wall_material', 'area_north_wall_material', 'area_west_wall_material') must be specified to compute the average absorption coefficient correctly. The same applies to surface areas of occupants ('area_Apers') and objects placed on the walls ('area_south_wall_Aobj', 'area_east_wall_Aobj', 'area_north_wall_Aobj', 'area_west_wall_Aobj').

Unlike absorption coefficients, the UNI 11532-2 standard does not give examples of scattering coefficients for materials that constitute or furnish classroom surfaces. At the same time, the pyroomacoustics library database reports a small number of scattering coefficients of commercial products, seating and audience in octave bands. For this reason, we have taken as reference the recommended scattering values in some common cases suggested in the ODEON Room Acoustics Software User's Manual[46], which are assignable as scalars to each shoebox surface and considered constant for all octave bands.

Type of simulator. In *pyeSTImate*, it is necessary to specify the simulation method to create artificial room impulse responses (RIRs) between the source and receivers. Taking advantage of the capabilities of pyroomacoustics, we can set up two types of simulators described as follows.

The first, based on the image source method (ISM), adopts Allen's implementation[47], funded on the principle that in a reflective environment, the reflected sound wave can be assumed to come from the virtual image of the actual source. This mechanism also applies to reflections after the first one, considering the virtual source of the previous reflection as the actual source. The simulation method assumes the walls as perfect reflectors and does not consider the effects of scattering. Moreover, the randomized image method has been introduced in pyroomacoustics: it adds a small random shift to the positions of the image sources so that they are no longer aligned over time, reducing sweeping echoes. In *pyeSTImate*, this simulation method is named 'ISM' and requires setting the maximum reflection order in the image source model ('max_order_reflections').

The second simulator is based on the ray tracing[48,49] theory: sound energy does not propagate on spherical wave fronts as in the case of virtual sources, but it is fractionated along rectilinear trajectories. The sound rays propagate from the source in all directions and gradually lose energy due to surface absorption, atmospheric attenuation, and scattering. Pyroomacoustics allows the use of a pure or hybrid ray tracing method for RIRs generation, in which the early reflections are simulated by the ISM and the diffuse tail by the ray tracing model. In *pyeSTImate*, this simulation method is set as 'Hybrid'; if pure ray tracing is to be used, 'max_order_reflections' $= -1$; otherwise, the combination of ISM and ray tracing requires 'max_order_reflections' ≥ 1.

Computational steps

The input data represent the arguments of the `Room` class, specifically the `ShoeBox(Room)` subclass of pyroomacoustics. According to the simulation method, the algorithm implements in C++ the image source and ray tracing models to generate the room impulse responses and simulate the propagation between sources and receivers. The output is thus an impulse response computed at the position of each receiver, with a sampling rate defined at the input stage and set equal to 16 kHz. The simulated RIRs are filtered in the 7-octave bands from 125 Hz to 8 kHz, from which the corresponding RT_{30} values are computed in order to be consistent with the measured reverberation times.

At this step, we have all the data for the STI calculation using the indirect method. First, the function for computing the modulation depth reduction factors has been implemented. It is applied to the 14 modulation frequencies for 7-octave bands and returns a modulation transfer function matrix of 98 values. For each octave band, the modulation transfer indexes are computed as the arithmetic mean of the transfer indexes for the 14 modulation frequencies. Finally, the STI is obtained from the weighted sum of the modulation transfer indexes for the 7-octave bands.

We have chosen to export to a textual dataframe the main results provided by the calculation code. Specifically, for each classroom, we have computed:

- the RT_{60} value averaged over all receivers' positions;
- the STI values for the individual receivers' positions;
- the average of the individual STI values computed at the previous step;
- the intelligibility rating (IR).

Since our work aims to develop a fully simulated tool for STI prediction, we tested its accuracy through comparison with the results of in situ measurements. In this regard, we have given as input the outcomes correlated with the measurements to obtain the error between measured and simulated scores, indicating the capability of the simulator to approximate the acoustic characteristics of an actual room.

For this purpose, we have monitored the following metrics:

- the absolute error (AE), absolute percentage error (APE) and intelligibility rating error (IRE) between average STI values from measurements and simulations in each classroom;
- the mean absolute error (MAE), mean absolute percentage error (MAPE) and mean intelligibility rating error (MIRE) over classrooms belonging to schools of the same grade and the entire test set of classrooms.

If the (mean) absolute error and (mean) absolute percentage error represent a regression loss between STI values, the metric (mean) intelligibility rating error quantifies the error associated with the discrete classification of the speech comprehension quality. The MIRE values fall in the range [0,1] and can also be expressed as a percentage. We assigned the five attributes of the speech comprehension quality scale (bad, poor, fair, good, excellent) a label from 1 to 5, where 1 indicates bad quality and 5 corresponds to excellent quality. The ratio of the absolute value of the difference between the measured and simulated IR labels and the range of the actual values averaged over the number of the analyzed classrooms provides the mean intelligibility rating error, defined in Equation (6) as:

$$MIRE(l,\hat{l}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \frac{|l_i - \hat{l}_i|}{max(l) - min(l)}, \tag{6}$$

where:

- l_i is the true intelligibility label of the i-th sample;
- $\hat{l}_i$ is the predicted intelligibility label of the i-th sample;
- max(l)-min(l) represents the range of the actual values;
- $n_{samples}$ is the number of samples in the test set.

In addition to the previously mentioned metrics, we also monitored the accuracy of the results in terms of just noticeable difference, the subjective limen representing the discernible difference of a room acoustic parameter. The study associated with the just noticeable difference in STI, i.e., the variation in STI values for which 50% of subjects can perceive the difference, determined a STI JND equal to 0.03 in simulated sound fields, but a STI JND of 0.1 is considered more realistic in everyday listening situations[40]. For this reason, we have calculated the number of JND units (STI JNDs) between STI values from measurements and simulations with both thresholds.

Application to artificial neural network

In this section, we describe the adaptation of *pyeSTImate* for generating a synthetic dataset to train an artificial neural network for speech transmission index prediction using a reduced number of input data. Then, we present the studies to yield a deep learning model with a high generalization capability. The procedures for generating the dataset, the neural architectures analyzed, the training settings and the performance metrics monitored are discussed.

Dataset and feature selection. Generally, traditional supervised deep learning models require large datasets in the training stage, but collecting a large amount of data through direct measurements and related labeling operations is often challenging. This issue can be overcome using public databases or synthetic data generated via algorithms. In this work, we have chosen to adapt the *pyeSTImate* tool to generate a synthetic dataset of classrooms with their corresponding average STIs. The code has been set to randomly define different geometries, source and receivers' positions, materials, types of occupants, and furniture with their associated surface areas, from which the average STI and the intelligibility rating over four listening positions have been calculated. The characteristics of the lecture rooms that compose the synthetic dataset are discussed as follows.

- Geometry: we set a range of minimum and maximum room length (l), depth (d), and height (h). We have defined $10\,\mathrm{m} \leq l \leq 30\,\mathrm{m}$, $5\,\mathrm{m} \leq d \leq 15\,\mathrm{m}$, and $3\,\mathrm{m} \leq h \leq 6.5\,\mathrm{m}$. Dimensions of each room have been randomly chosen in the respective ranges, with steps of 0.1 m for plan dimensions and 0.05 m for the height.
- Source and receivers' positions: the coordinates of the speaker and listeners have been placed according to Figure 1 (UNI 11532-2 standard). We set a margin of variability to the coordinates of the source and listening positions on the audience perimeter to increase the number of possible configurations. The source location has been assigned between 1–2.5 m from the west wall at the height of 1–1.7 m to consider the speaker sitting at the desk or standing. The sitting audience area is also not fixed, so point P1 has been allocated in the range of 1–2 m from the source and points P3–P4 at a distance of 1.5–2.5 m from east and south walls, respectively. Finally, point P2 has been placed in the center of the audience.
- Materials, occupants and furniture: we used the dictionary of *pyeSTImate* to define interior finishes, occupant categories, and classroom furniture. Materials have been subdivided so that the absorption and scattering coefficients of the appropriate finishes are assigned to floors, ceilings, walls, windows and doors.
- Surface areas: based on the characteristics of actual classrooms and to ensure the variability of simulated ones, we set up two types of materials (one main and one secondary with variable surface areas) for floors, walls and ceilings. Floors consist of the main material and optionally a secondary one for up to 20% of the surface area. Window areas have been assigned to two walls for up to 50% of the surface, while the room access door and lockers have been allocated on the remaining walls for up to 10% of the surface. Ceilings, like floors, consist of the main material and optionally a secondary one for up to 50% of the surface area. All lecture rooms have an area designated for the audience for up to 80% of the main surface area.

To calculate the speech transmission index, we have employed the hybrid ISM/ray tracing method, which has provided the best performance in previous experiments together with the pure ray tracing method. The parameter 'rooms_number' of the dataset generator code can handle the number of classrooms and the corresponding STI values to be computed.

We have generated a dataset of 1000 classrooms, exporting to a csv file all the parameters required by *pyeSTImate* for STI calculation. We have inspected the distribution of each feature, eliminating the ones with low variance and others that require accurate knowledge of the room, such as the surfaces to be attributed to each material, audience or furniture. Since our goal is to estimate the quality of speech transmission with basic information about the room of interest, e.g., with a floor plan and pictures, we kept as features only the geometry (length, depth, and height) and interior finishes from which octave-band absorption coefficients can be extrapolated using the *pyeSTImate* dictionary. Ultimately, we keep 73 features (3 dimensions and 10 finishes for 7 absorption coefficients in octave bands) and refer to STI as the label.

Deep Learning Model. The work aims to predict the output of a continuous value between 0 and 1, so we have devised an ANN based on a multilayer perceptron[50] (MLP) algorithm to solve a regression problem. An MLP is a mathematical model that performs non-linear mapping between input and output and is composed of an input layer that takes in the features of a set of examples (training set), one or more hidden layers and an output layer, each consisting of one or more computational nodes, also called neurons, connected among them in a network. In our case, the output layer comprises only one neuron computing the predicted STI value.

The iterative input-output mapping, or learning process, depends on the size of the training set and is controlled by a set of parameters chosen before the training starts, called network hyperparameters. Some common examples of hyperparameters are the number of hidden layers, the number of neurons in each layer, the activation function in a layer (the function that regulates and limits the output of each node), the number of iterations (epochs) in the training phase, train-validation split ratio, the learning rate and type of optimizer that control the speed of the learning process. To find the model that has the greatest generalization capability and provides the best results with data unseen during the training (test set), we have investigated the performance of the algorithm by varying the size of the training set, the number of hidden layers and neurons in each of them through a grid search[51]. We performed the experiments with training sets of 100 and 1000 examples and neural architectures composed of one, two and three hidden layers whose number of nodes has been chosen among a range between 8 and 1024.

For training, the entire dataset has been randomly split into the training and validation sets with a percentage of 80% and 20% of the total number of samples, respectively. For testing, we considered the 35 classrooms already investigated with instrumental measurements and simulations using *pyeSTImate*. In previous experiments, we have found a discrepancy between STIs derived from measured and simulated reverberation times related to the uncertainty of the measurements and the computational simulator. Therefore, we have compared the test results with both the STI values from simulations and measurements.

Since STI does not allow negative values, we set the rectified linear unit (ReLU) activation function in the hidden layers, a piecewise linear function that outputs the input directly if it is positive; otherwise, it outputs zero. We also set the learning rate equal to 0.001, Adam[52] optimizer, and 5000 epochs with the early stopping, a form of regularization that stops the training when the monitored metric (in our case, the validation loss) has stopped improving. We set the MAE as regression loss and MAPE as monitored metrics.

Experimental setup. In the experiments, we used a NVIDIA DGX Station A100 with Dual 64-Core AMD EPYC 7742 @3.4 GHz, and 8 NVIDIA A100-SXM4-40 GB GPUs. The server was running Ubuntu 20.04.3 LTS, and we implemented the network designs with the Tensorflow[53] deep learning framework.

Results and discussion

In this section, we present and comment on experimental results. First, we use *pyeSTImate* to calculate the STI values of the classrooms described in Section Characteristics of lecture rooms and measurement equipment, providing the tool with all the required input data and comparing the outcomes with different impulse response modeling methods. Next, we illustrate and compare the results of the best of several deep learning models trained with the synthetic STI dataset generated by the *pyeSTImate* tool adapted for the purpose.

Assessment of speech intelligibility through pyeSTImate

Tables 7, 8, and 9 illustrate the comparison between the results predicted with the simulation tool and obtained from direct measurements of reverberation times calculated for each classroom with ISM, ray tracing and hybrid ISM/ray tracing methods, respectively. Specifically, the RT averaged over the four receiver positions (P1–P4), STI at individual listening points and the average value, absolute error (AE) associated with individual and average STI, absolute percentage error (APE) on the average STI, intelligibility rating, error on intelligibility rating (IRE), just noticeable difference units for the average STI with thresholds equal to 0.03 (STI JNDs(0.03)) and 0.1 (STI JNDs(0.1)) are presented. Table 10 summarizes the mean absolute error (MAE), mean absolute percentage error (MAPE), mean intelligibility rating error (MIRE), and mean just noticeable difference units with thresholds equal to 0.03 (mean STI JNDs(0.03)) and 0.1 (mean STI JNDs(0.1)) computed for classrooms of the same grade and overall the classrooms.

From the exploration of the summary results, we observe that the best performance are obtained from the RIR simulation methods of ray tracing and hybrid rather than the ISM modeling. The main issue of the ISM simulator is the choice of the maximum order of reflections to be assigned. This model assumes the walls as perfect reflectors, so later reflections due to scattering are not considered, and the activation of the reverberant tail is performed through a sufficiently large number of reflections handled by the

Table 7. Simulation results with the ISM method and errors between STI values from measurements and simulations: RT, individual and average STI values, absolute error (AE), absolute percentage error (APE), Intelligibility Rating, intelligibility rating error (IRE), and just noticeable difference units with thresholds equal to 0.03 (STI JNDs(0.03)) and 0.1 (STI JNDs(0.1))

ID	Type		Name of room	RT [s]	STI					AE					APE (%)	Intelligibility Rating	IRE	STI JNDs(0.03)	STI JNDs(0.1)
				avg	P1	P2	P3	P4	avg	P1	P2	P3	P4	avg*	avg*			avg*	avg*
1	SCH 1	University	140/1	2.60	0.83	0.28	0.26	0.26	0.41 ± 0.28	0.26	0.24	0.20	0.17	0.09	17.53	poor	0.2	2.9	0.9
2	SCH 1	University	140/2	1.73	0.79	0.41	0.38	0.38	0.49 ± 0.20	0.27	0.01	0.01	0.04	0.08	20.40	fair	0.2	2.8	0.8
3	SCH 1	University	140/3	2.24	0.82	0.33	0.30	0.30	0.44 ± 0.26	0.22	0.25	0.05	0.03	0.03	5.69	poor	0.2	0.9	0.3
4	SCH 1	University	155/D1	1.79	0.71	0.40	0.38	0.38	0.47 ± 0.16	0.08	0.17	0.16	0.14	0.10	17.21	fair	0	3.2	1.0
5	SCH 1	University	155/D2	1.81	0.71	0.40	0.38	0.38	0.47 ± 0.16	0.09	0.17	0.16	0.14	0.10	17.31	fair	0	3.2	1.0
6	SCH 1	University	155/D3	1.80	0.71	0.40	0.38	0.38	0.47 ± 0.16	0.16	0.11	0.15	0.11	0.05	10.43	fair	0	1.8	0.5
7	SCH 1	University	155/D4	1.77	0.71	0.40	0.38	0.38	0.47 ± 0.16	0.11	0.15	0.16	0.14	0.09	15.43	fair	0	2.8	0.9
8	SCH 1	University	160/1	1.49	0.87	0.46	0.44	0.44	0.55 ± 0.21	0.28	0.04	0.07	0.05	0.08	15.93	fair	0	2.5	0.8
9	SCH 1	University	160/2	1.49	0.87	0.46	0.44	0.43	0.55 ± 0.21	0.29	0.04	0.01	0.03	0.07	13.70	fair	0	2.2	0.7
10	SCH 1	University	AT1	1.59	0.78	0.41	0.40	0.40	0.50 ± 0.19	0.05	0.20	0.19	0.05	0.10	16.33	fair	0	3.2	1.0
11	SCH 1	University	AT2	1.54	0.77	0.42	0.41	0.41	0.50 ± 0.18	0.07	0.18	0.19	0.15	0.11	18.00	fair	0.2	3.7	1.1
12	SCH 1	University	AT3	2.13	0.78	0.34	0.33	0.33	0.45 ± 0.22	0.04	0.28	0.21	0.17	0.17	28.21	poor	0.4	5.8	1.7
13	SCH 1	University	EN1	1.65	0.76	0.44	0.41	0.41	0.51 ± 0.17	0.15	0.12	0.14	0.00	0.02	4.49	fair	0	0.8	0.2
14	SCH 1	University	EN3	1.63	0.55	0.28	0.26	0.26	0.34 ± 0.14	0.01	0.24	0.23	0.17	0.16	32.38	poor	0.2	5.4	1.6
15	SCH 1	University	S1	0.93	0.56	0.45	0.44	0.44	0.47 ± 0.06	0.27	0.20	0.16	0.15	0.20	70.44	fair	0.4	6.5	2.0
16	SCH 1	University	S2	1.36	0.77	0.48	0.45	0.46	0.54 ± 0.16	0.49	0.22	0.17	0.19	0.27	97.80	fair	0.4	8.9	2.7
17	SCH 1	University	S3	2.05	0.84	0.41	0.33	0.32	0.47 ± 0.25	0.59	0.13	0.04	0.02	0.19	69.25	fair	0.4	6.5	1.9
18	SCH 2	Primary	1B	0.86	0.77	0.48	0.47	0.47	0.54 ± 0.15	0.11	0.16	0.15	0.10	0.08	12.53	fair	0.2	2.6	0.8
19	SCH 2	Primary	2A	1.08	0.78	0.41	0.39	0.39	0.49 ± 0.19	0.24	0.16	0.15	0.14	0.05	9.80	fair	0	1.8	0.5
20	SCH 2	Primary	3A	0.92	0.76	0.48	0.47	0.47	0.54 ± 0.14	0.23	0.08	0.08	0.07	0.00	0.11	fair	0	0.0	0.0
21	SCH 3	Primary	2C	1.01	0.76	0.44	0.43	0.43	0.52 ± 0.16	0.28	0.06	0.09	0.05	0.02	4.40	fair	0	0.7	0.2
22	SCH 3	Primary	3C	1.07	0.77	0.41	0.40	0.40	0.50 ± 0.18	0.23	0.20	0.14	0.07	0.04	8.27	fair	0	1.5	0.4
23	SCH 4	Primary	4B	1.10	0.77	0.41	0.40	0.39	0.49 ± 0.19	0.20	0.21	0.23	0.20	0.11	18.13	fair	0	3.6	1.1
24	SCH 4	Primary	4A	1.13	0.78	0.40	0.37	0.37	0.48 ± 0.20	0.34	0.07	0.04	0.02	0.05	12.16	fair	0.2	1.7	0.5
25	SCH 5	Secondary	1	0.94	0.77	0.45	0.44	0.44	0.52 ± 0.17	0.15	0.15	0.15	0.10	0.06	10.74	fair	0	2.1	0.6
26	SCH 5	Secondary	2	0.93	0.78	0.44	0.43	0.43	0.52 ± 0.17	0.20	0.11	0.10	0.07	0.02	3.63	fair	0	0.7	0.2
27	SCH 5	Secondary	3	0.95	0.78	0.44	0.43	0.42	0.52 ± 0.17	0.21	0.14	0.14	0.14	0.05	9.30	fair	0	1.8	0.5
28	SCH 6	Secondary	4	0.82	0.75	0.51	0.50	0.50	0.57 ± 0.13	0.21	0.02	0.01	0.07	0.08	16.23	fair	0	2.6	0.8
29	SCH 6	Secondary	5	0.87	0.76	0.49	0.48	0.48	0.55 ± 0.14	0.18	0.07	0.01	0.04	0.03	6.74	fair	0	1.2	0.3
30	SCH 7	Secondary	6	1.04	0.79	0.43	0.39	0.38	0.50 ± 0.20	0.02	0.27	0.31	0.32	0.22	30.49	fair	0.2	7.3	2.2
31	SCH 7	Secondary	7	1.24	0.94	0.50	0.47	0.47	0.59 ± 0.23	0.30	0.01	0.00	0.00	0.08	14.87	fair	0	2.6	0.8
32	SCH 7	Secondary	8	0.91	0.76	0.47	0.46	0.46	0.54 ± 0.15	0.11	0.06	0.05	0.05	0.01	1.93	fair	0	0.4	0.1
33	SCH 8	Secondary	9	0.93	0.75	0.47	0.47	0.46	0.54 ± 0.14	0.21	0.05	0.02	0.09	0.01	2.31	fair	0	0.4	0.1
34	SCH 8	Secondary	10	0.96	0.58	0.45	0.44	0.44	0.48 ± 0.07	0.20	0.14	0.11	0.10	0.14	40.39	fair	0.2	4.6	1.4
35	SCH 8	Secondary	11	1.14	0.57	0.41	0.39	0.39	0.44 ± 0.08	0.13	0.04	0.01	0.01	0.05	12.18	poor	0	1.6	0.5

* Computed on average STI.

Table 8. Simulation results with the ray tracing method and errors between STI values from measurements and simulations: RT, individual and average STI values, absolute error (AE), absolute percentage error (APE), Intelligibility Rating, intelligibility rating error (IRE), and just noticeable difference units with thresholds equal to 0.03 (STI JNDs(0.03)) and 0.1 (STI JNDs(0.1))

ID	Type		Name of room	RT [s]	STI					AE					APE (%)	Intelligibility Rating	IRE	STI JNDs(0.03)	STI JNDs(0.1)
				avg	P1	P2	P3	P4	avg	P1	P2	P3	P4	avg*	avg*			avg*	avg*
1	SCH 1	University	140/1	1.56	0.80	0.38	0.36	0.37	0.48 ± 0.22	0.23	0.14	0.10	0.06	0.02	3.65	fair	0	0.6	0.2
2	SCH 1	University	140/2	2.20	0.79	0.37	0.32	0.29	0.44 ± 0.23	0.27	0.03	0.05	0.05	0.04	9.19	poor	0	1.2	0.4
3	SCH 1	University	140/3	1.45	0.80	0.42	0.38	0.39	0.50 ± 0.20	0.20	0.16	0.03	0.06	0.03	7.02	fair	0	1.1	0.3
4	SCH 1	University	155/D1	1.18	0.71	0.47	0.46	0.47	0.53 ± 0.12	0.08	0.10	0.08	0.05	0.03	6.15	fair	0	1.2	0.3
5	SCH 1	University	155/D2	1.21	0.71	0.47	0.46	0.47	0.53 ± 0.12	0.09	0.10	0.08	0.05	0.03	5.87	fair	0	1.1	0.3
6	SCH 1	University	155/D3	1.21	0.71	0.47	0.46	0.47	0.53 ± 0.12	0.16	0.04	0.07	0.02	0.01	1.44	fair	0	0.3	0.1
7	SCH 1	University	155/D4	1.21	0.71	0.47	0.46	0.47	0.53 ± 0.12	0.11	0.08	0.08	0.05	0.03	4.74	fair	0	0.9	0.3
8	SCH 1	University	160/1	1.35	0.86	0.47	0.43	0.45	0.55 ± 0.20	0.27	0.05	0.08	0.06	0.07	15.50	fair	0	2.5	0.7
9	SCH 1	University	160/2	1.28	0.86	0.47	0.45	0.46	0.56 ± 0.20	0.28	0.03	0.00	0.06	0.08	15.78	fair	0	2.5	0.8
10	SCH 1	University	AT1	0.99	0.79	0.54	0.50	0.46	0.57 ± 0.15	0.06	0.07	0.09	0.01	0.02	3.43	fair	0	0.7	0.2
11	SCH 1	University	AT2	0.81	0.80	0.60	0.50	0.55	0.61 ± 0.13	0.10	0.00	0.10	0.01	0.00	0.57	good	0	0.1	0.0
12	SCH 1	University	AT3	0.75	0.81	0.63	0.55	0.54	0.63 ± 0.12	0.01	0.01	0.01	0.04	0.01	2.30	good	0	0.5	0.1
13	SCH 1	University	EN1	1.14	0.76	0.52	0.46	0.47	0.55 ± 0.14	0.15	0.04	0.09	0.06	0.02	3.48	fair	0	0.6	0.2
14	SCH 1	University	EN3	1.03	0.59	0.35	0.31	0.37	0.40 ± 0.12	0.03	0.17	0.18	0.06	0.10	19.14	poor	0.2	3.2	1.0
15	SCH 1	University	S1	1.90	0.54	0.30	0.24	0.23	0.33 ± 0.15	0.25	0.05	0.04	0.06	0.05	18.02	poor	0.2	1.7	0.5
16	SCH 1	University	S2	2.56	0.80	0.39	0.29	0.30	0.44 ± 0.24	0.52	0.13	0.01	0.03	0.17	62.87	poor	0.2	5.7	1.7
17	SCH 1	University	S3	3.07	0.85	0.41	0.27	0.25	0.44 ± 0.28	0.60	0.13	0.02	0.05	0.16	58.28	poor	0.2	5.4	1.6
18	SCH 2	Primary	1B	0.51	0.80	0.59	0.56	0.57	0.63 ± 0.11	0.14	0.05	0.06	0.00	0.01	1.18	good	0	0.2	0.1
19	SCH 2	Primary	2A	0.87	0.78	0.42	0.40	0.40	0.50 ± 0.19	0.24	0.15	0.14	0.13	0.04	8.21	fair	0	1.5	0.4
20	SCH 2	Primary	3A	0.74	0.77	0.51	0.48	0.48	0.56 ± 0.14	0.24	0.05	0.07	0.06	0.01	2.14	fair	0	0.4	0.1
21	SCH 3	Primary	2C	0.82	0.77	0.47	0.44	0.44	0.53 ± 0.16	0.29	0.03	0.08	0.04	0.04	7.19	fair	0	1.2	0.4
22	SCH 3	Primary	3C	0.84	0.77	0.44	0.42	0.41	0.51 ± 0.18	0.23	0.17	0.12	0.06	0.03	5.51	fair	0	0.0	0.3
23	SCH 4	Primary	4B	0.84	0.78	0.44	0.41	0.40	0.51 ± 0.18	0.21	0.18	0.22	0.19	0.10	16.00	fair	0	3.2	1.0
24	SCH 4	Primary	4A	1.25	0.78	0.34	0.30	0.29	0.43 ± 0.24	0.34	0.13	0.11	0.10	0.00	0.28	poor	0	0.0	0.0
25	SCH 5	Secondary	1	0.89	0.78	0.43	0.41	0.40	0.50 ± 0.18	0.16	0.17	0.18	0.14	0.08	14.40	fair	0	2.8	0.8
26	SCH 5	Secondary	2	0.89	0.78	0.43	0.40	0.40	0.50 ± 0.19	0.20	0.12	0.13	0.10	0.04	7.34	fair	0	1.3	0.4
27	SCH 5	Secondary	3	0.89	0.78	0.43	0.40	0.40	0.50 ± 0.19	0.21	0.15	0.17	0.16	0.07	12.33	fair	0	2.3	0.7
28	SCH 6	Secondary	4	0.80	0.76	0.49	0.47	0.47	0.55 ± 0.14	0.22	0.00	0.02	0.04	0.06	12.32	fair	0	2.0	0.6
29	SCH 6	Secondary	5	0.81	0.76	0.49	0.46	0.45	0.54 ± 0.15	0.18	0.07	0.03	0.01	0.02	4.33	fair	0	0.7	0.2
30	SCH 7	Secondary	6	0.54	0.81	0.54	0.50	0.50	0.59 ± 0.15	0.04	0.16	0.20	0.20	0.13	18.41	fair	0.2	4.4	1.3
31	SCH 7	Secondary	7	1.06	0.93	0.50	0.47	0.45	0.59 ± 0.23	0.29	0.01	0.00	0.02	0.07	13.35	fair	0	2.3	0.7
32	SCH 7	Secondary	8	0.73	0.77	0.50	0.48	0.48	0.56 ± 0.14	0.12	0.03	0.03	0.03	0.01	1.32	fair	0	0.2	0.1
33	SCH 8	Secondary	9	0.70	0.77	0.51	0.50	0.49	0.57 ± 0.13	0.23	0.01	0.01	0.06	0.04	8.22	fair	0	1.4	0.4
34	SCH 8	Secondary	10	1.28	0.57	0.38	0.35	0.33	0.41 ± 0.11	0.19	0.07	0.02	0.01	0.07	19.44	poor	0	2.2	0.7
35	SCH 8	Secondary	11	1.40	0.57	0.35	0.32	0.31	0.39 ± 0.12	0.13	0.02	0.06	0.07	0.00	1.02	poor	0	0.1	0.0

* Computed on average STI.

'max_order_reflections' parameter. During the experiments, we observed that an insufficient number of reflections does not allow the entire reverberant tail to be simulated, providing underestimated reverberation times and, thus, better STI values. On the other hand, if this parameter is too high, the duration of late reflections is overestimated, and consequently, the results are lower speech transmission indexes.

The maximum order of reflections can be preliminarily estimated by calculating the RT value with Sabine's

Table 9. Simulation results with the hybrid ISM/ray tracing method and errors between STI values from measurements and simulations: RT, individual and average STI values, absolute error (AE), absolute percentage error (APE), Intelligibility Rating, intelligibility rating error (IRE), and just noticeable difference units with thresholds equal to 0.03 (STI JNDs(0.03)) and 0.1 (STI JNDs(0.1))

ID	Type		Name of room	RT [s]	STI					AE					APE (%)	Intelligibility Rating	IRE	STI JNDs(0.03)	STI JNDs(0.1)
				avg	P1	P2	P3	P4	avg	P1	P2	P3	P4	avg*	avg*			avg*	avg*
1	SCH 1	University	140/1	1.63	0.81	0.37	0.37	0.37	0.48 ± 0.22	0.24	0.15	0.09	0.06	0.02	3.33	fair	0	0.5	0.2
2	SCH 1	University	140/2	2.18	0.80	0.37	0.33	0.29	0.45 ± 0.24	0.28	0.03	0.04	0.05	0.04	9.67	poor	0	1.3	0.4
3	SCH 1	University	140/3	1.48	0.80	0.42	0.38	0.40	0.50 ± 0.20	0.20	0.16	0.03	0.07	0.03	6.99	fair	0	1.1	0.3
4	SCH 1	University	155/D1	1.25	0.71	0.47	0.46	0.48	0.53 ± 0.12	0.08	0.10	0.08	0.04	0.03	6.12	fair	0	1.2	0.3
5	SCH 1	University	155/D2	1.25	0.71	0.47	0.46	0.47	0.53 ± 0.12	0.09	0.10	0.08	0.05	0.03	6.11	fair	0	1.1	0.3
6	SCH 1	University	155/D3	1.25	0.71	0.47	0.46	0.47	0.53 ± 0.12	0.16	0.04	0.07	0.02	0.01	1.36	fair	0	0.2	0.1
7	SCH 1	University	155/D4	1.26	0.71	0.47	0.46	0.47	0.53 ± 0.12	0.11	0.08	0.08	0.05	0.03	4.79	fair	0	0.9	0.3
8	SCH 1	University	160/1	1.45	0.86	0.46	0.43	0.46	0.55 ± 0.21	0.27	0.04	0.08	0.07	0.08	15.91	fair	0	2.5	0.8
9	SCH 1	University	160/2	1.31	0.86	0.47	0.45	0.45	0.56 ± 0.20	0.28	0.03	0.00	0.05	0.08	15.91	fair	0	2.6	0.8
10	SCH 1	University	AT1	0.97	0.79	0.54	0.50	0.44	0.57 ± 0.15	0.06	0.07	0.09	0.01	0.03	4.45	fair	0	0.9	0.3
11	SCH 1	University	AT2	0.81	0.78	0.60	0.51	0.56	0.61 ± 0.12	0.08	0.00	0.09	0.00	0.00	0.36	good	0	0.1	0.0
12	SCH 1	University	AT3	0.75	0.80	0.64	0.57	0.55	0.64 ± 0.12	0.02	0.02	0.03	0.05	0.02	3.19	good	0	0.7	0.2
13	SCH 1	University	EN1	1.18	0.76	0.53	0.46	0.46	0.55 ± 0.14	0.15	0.03	0.09	0.05	0.02	3.47	fair	0	0.6	0.2
14	SCH 1	University	EN3	1.04	0.58	0.35	0.32	0.37	0.40 ± 0.12	0.02	0.17	0.17	0.06	0.10	19.01	poor	0.2	3.2	1.0
15	SCH 1	University	S1	1.92	0.54	0.30	0.24	0.23	0.33 ± 0.14	0.25	0.05	0.04	0.06	0.05	18.17	poor	0.2	1.7	0.5
16	SCH 1	University	S2	2.60	0.80	0.39	0.29	0.30	0.45 ± 0.24	0.52	0.13	0.01	0.03	0.17	64.02	poor	0.2	5.8	1.7
17	SCH 1	University	S3	3.13	0.85	0.41	0.26	0.24	0.44 ± 0.28	0.60	0.13	0.03	0.06	0.16	57.82	poor	0.2	5.4	1.6
18	SCH 2	Primary	1B	0.53	0.79	0.59	0.57	0.57	0.63 ± 0.11	0.13	0.05	0.05	0.00	0.01	0.84	good	0	0.2	0.1
19	SCH 2	Primary	2A	0.89	0.78	0.43	0.40	0.40	0.50 ± 0.14	0.24	0.14	0.14	0.13	0.04	7.75	fair	0	1.4	0.4
20	SCH 2	Primary	3A	0.76	0.76	0.51	0.48	0.48	0.56 ± 0.18	0.23	0.05	0.07	0.06	0.01	2.14	fair	0	0.4	0.1
21	SCH 3	Primary	2C	0.85	0.77	0.47	0.44	0.44	0.53 ± 0.16	0.29	0.03	0.08	0.04	0.03	7.00	fair	0	1.2	0.3
22	SCH 3	Primary	3C	0.87	0.77	0.44	0.42	0.41	0.51 ± 0.18	0.23	0.17	0.12	0.06	0.03	5.38	fair	0	1.0	0.3
23	SCH 4	Primary	4B	0.87	0.78	0.44	0.41	0.40	0.51 ± 0.18	0.21	0.18	0.22	0.19	0.09	15.73	fair	0	3.2	0.9
24	SCH 4	Primary	4A	1.29	0.78	0.34	0.30	0.29	0.43 ± 0.24	0.34	0.13	0.11	0.10	0.00	0.07	poor	0	0.0	0.0
25	SCH 5	Secondary	1	0.91	0.78	0.43	0.41	0.40	0.50 ± 0.18	0.16	0.17	0.18	0.14	0.08	14.28	fair	0	2.8	0.8
26	SCH 5	Secondary	2	0.92	0.78	0.43	0.40	0.40	0.50 ± 0.18	0.20	0.12	0.13	0.10	0.04	6.93	fair	0	1.2	0.4
27	SCH 5	Secondary	3	0.93	0.78	0.43	0.40	0.40	0.50 ± 0.18	0.21	0.15	0.17	0.16	0.07	12.08	fair	0	2.3	0.7
28	SCH 6	Secondary	4	0.82	0.76	0.49	0.47	0.47	0.55 ± 0.14	0.22	0.00	0.02	0.04	0.06	12.53	fair	0	2.0	0.6
29	SCH 6	Secondary	5	0.84	0.76	0.49	0.46	0.46	0.54 ± 0.15	0.18	0.07	0.03	0.02	0.02	4.59	fair	0	0.8	0.2
30	SCH 7	Secondary	6	0.55	0.80	0.54	0.50	0.50	0.59 ± 0.15	0.03	0.16	0.20	0.20	0.13	18.26	fair	0.2	4.4	1.3
31	SCH 7	Secondary	7	1.07	0.93	0.50	0.47	0.45	0.59 ± 0.23	0.29	0.01	0.00	0.02	0.07	13.51	fair	0	2.3	0.7
32	SCH 7	Secondary	8	0.75	0.77	0.50	0.49	0.48	0.56 ± 0.14	0.12	0.03	0.02	0.03	0.01	1.42	fair	0	0.3	0.1
33	SCH 8	Secondary	9	0.72	0.76	0.51	0.51	0.49	0.57 ± 0.13	0.22	0.01	0.02	0.06	0.04	8.03	fair	0	1.4	0.4
34	SCH 8	Secondary	10	1.30	0.57	0.37	0.35	0.32	0.40 ± 0.11	0.19	0.06	0.02	0.02	0.06	19.02	poor	0	2.2	0.6
35	SCH 8	Secondary	11	1.44	0.57	0.35	0.32	0.31	0.39 ± 0.12	0.13	0.02	0.06	0.07	0.00	0.81	poor	0	0.1	0.0

* Computed on average STI.

Table 10. Comparison of errors between average STIs from measurements and simulations for classrooms of the same grade and overall the classrooms: mean absolute error (MAE), mean absolute percentage error (MAPE), mean intelligibility rating error (MIRE), and mean just noticeable difference units with thresholds equal to 0.03 (mean STI JNDs(0.03)) and 0.1 (mean STI JNDs(0.1))

	Classrooms	MAE	MAPE (%)	MIRE (%)	mean STI JNDs(0.03)	mean STI JNDs(0.1)
ISM	University (IDs 1–17)	0.11	27.68	15.29	3.7	1.1
	Primary (IDs 18–24)	0.05	9.34	5.71	1.7	0.5
	Secondary (IDs 25–35)	0.07	13.53	3.64	2.3	0.7
	Overall	0.09	19.56	9.71	2.9	0.9
Ray tracing	University (IDs 1–17)	0.05	13.97	4.71	1.7	0.5
	Primary (IDs 18–24)	0.03	5.79	0.00	1.1	0.3
	Secondary (IDs 25–35)	0.05	10.23	1.82	1.8	0.5
	Overall	0.05	11.16	2.86	1.6	0.5
Hybrid	University (IDs 1–17)	0.05	14.16	4.71	1.7	0.5
	Primary (IDs 18–24)	0.03	5.56	0.00	1.0	0.3
	Secondary (IDs 25–35)	0.05	10.13	1.82	1.8	0.5
	Overall	0.05	11.17	2.86	1.6	0.5

formula[54]:

$$RT = 0.161 \frac{V}{\alpha S}, \tag{7}$$

where V, S, and α represent the volume, surface area and the average absorption factor of the room, and then applying the `pyroomacoustics.acoustics.inverse_sabine` module. Given the desired reverberation time, the dimensions of a rectangular room (shoebox), and the sound speed, it computes the energy absorption coefficient and maximum image source order needed. However, Sabine's theory assumes a perfectly diffuse sound field even during the transient regime, in which the uniform energy density in the room decreases exponentially over time. In the classrooms examined, the assumption of a perfectly diffuse field fails due to inhomogeneities in absorption coefficients and furniture distribution, as well as unbalanced geometric dimensions of some rooms. Sabine's approximation produces reverberation times lower than the measured values, making this procedure inadequate for our case studies of classrooms. In the experiments, we obtain the best fit when the linear interpolation of the power decay approximates the simulated curve. For this purpose, we have chosen a 'max_order_reflections' parameter equal to 33 that minimizes the mean errors over the entire test set of classrooms.

With regard to simulations using the pure ray tracing and the hybrid ISM/ray tracing methods, in the former case, we set the parameter 'max_order_reflections' equal to -1, and in the latter case, equal to 3, as recommended in the pyroomacoustics user's guide. Specifically, the hybrid simulator models direct and early reflections up to the third order with the ISM and late reflections with the ray tracing method. Since the ray tracing considers the scattering of sound from surfaces, the similarity of the performance yielded with the two simulators demonstrates that the diffusive components are predominant over the reflective ones in the analyzed classrooms, as the first reflections modeled with ISM do not provide appreciable variations compared with pure ray tracing.

A key aspect of the most accurate modeling of RIRs performed by ray-tracing-based methods is the contribution of scattering in the simulation of diffuse reflections. Studies of sound scattering in computer models show that sound rays follow a geometric pattern with specular reflections for surfaces with a scattering coefficient equal to zero, while a value of 0.2 is already sufficient to achieve more diffuse modeling[55]. Other works on the influence of scattered sound on objective room acoustical parameters have shown the correlation between the scattering coefficient and parameters such as early decay time, reverberation time, clarity, strength and definition[56,57], also in relation to the improvement of the acoustics in classrooms[58]. In our case, we observed that scattering coefficients equal to zero imply an overestimation of reverberation times (and an underestimation of STIs), while values in the range of the literature recommendations for the surfaces of interest return STIs closer to those obtained from measurements. Here, we have not yet conducted an analytical assessment of the tool's sensitivity in predicting RTs (and, consequently, STIs) as the scattering coefficient varies, but given the importance of the topic, we reserve further study for future work.

Now, we examine the results for groups of same-grade classrooms. All three simulation methods report more accurate STIs for primary classrooms, and in particular, the hybrid method yields the lowest MAE and MAPE values of 0.03 and 5.56%, respectively, correct intelligibility ratings in all classrooms, and mean STI JNDs within both thresholds. Next, the secondary classrooms also present the best achievements with the hybrid method, returning MAE equal to 0.05, MAPE equal to 10.13%, MIRE of 1.82% (1 incorrect prediction out of 11 classrooms), mean STI JNDs(0.03) equal to 1.8, and mean STI JNDs(0.1) within the threshold. Finally, for university classrooms, the ray tracing method gives MAE of 0.05, MAPE of 13.97%, MIRE of 4.71% (4 incorrect predictions out of 17), mean STI JNDs(0.03) of 1.7, and mean STI JNDs(0.1) within the threshold.

Referring to the ray tracing-based methods that provided results with the lowest error rates, we observe that the individual classrooms characterized by low accuracy in STI prediction, corresponding to the case studies where both thresholds of STI JNDs are exceeded, are no. 6 for the secondary, and S2-S3 for university classrooms. The reasons for the less accurate outcomes can be attributed to the approximations of the indirect method related to the source and ambient noise levels reported in UNI EN ISO 9921 that influence the experimental results. We observe that the no. 6 secondary classroom, as well as the EN3 university lecture room, are characterized by a volumetry of slightly less than 250 m^3 implying an average source level of 60 dB instead of 70 dB (UNI 11532-2, Table 11). As a result, in classrooms with volumes immediately below the regulation boundary, a decrease in source level of 10 dB penalizes the speech transmission index with underestimated performance. The opposite issue is found in university classrooms 160/1 and 160/2, with volumes of 309 m^3 and 317 m^3, respectively, where the simulated source level of 70 dB provides overestimated STIs, despite being within the same intelligibility rating range obtained from measurements.

In other lecture rooms, such as the S1, S2, and S3 university classrooms, the ambient noise spectrum from regulations does not adequately represent the noise level detected during the measurements, so the simulations provided overestimated STIs. The problem of inaccurate approximation of ambient noise can also be seen from the standard deviations on the average STI values calculated by all three simulation methods that assume higher values compared to STIs from measurements. Simulations generally return decreasing STIs as source-receiver distance increases, while measured reverberation times provide fairly consistent STI values at individual listening points, demonstrating that uniform background noise is a penalizing contribution to speech intelligibility even near the speaker's location.

In conclusion, the tool shows good accuracy in predicting speech transmission indices for small and wide classrooms, regardless of grade, year of construction, and finishing materials. The robustness of the simulated results is confirmed by STI JNDs(0.1) within one unit for 32 out of 35 classrooms, and at the same time, most classrooms report STI JNDs(0.03) within values slightly above one unit. The performance of the tool shows lower accuracy in medium-sized classrooms that are in between the different source levels specified by the regulations and in conditions of high ambient noise, which is not always adequately represented by the spectrum suggested by the standard specifications.

Assessment of speech intelligibility through deep learning models

Table 11 presents the performance (MAE, MAPE, MIRE) of the ANN trained with datasets of different sample sizes (specifically, 100 and 1000 examples) as the number of hidden layers (1–3) and corresponding hidden units (8–64) varies, and tested on the STI values calculated from both *pyeSTImate* in the best configuration (hybrid method) and from measurements over the 35 classrooms.

The first analysis concerns the results in testing as the size of the dataset varies. In all experiments, the synthetic dataset consisting of 100 classrooms provides better results in regression than that including 1000 examples. The reason can be attributed to the limited variety of materials, occupants and furniture in the dictionary, which implies low variability in sound absorption and scattering coefficients in the case of large datasets. Hence, an undiversified dataset leads the algorithm to overfit the training data, resulting in low testing generalization capability.

The second observation is related to the network hyperparameters because the error in regression on the test set decreases as the number of layers increases. In an artificial neural network, the depth, meant as the number of hidden layers, depends on the correlation between input and output. If their relationship is linear or approximable by an analytical function, up to one hidden layer is generally adequate. In contrast, if an analytical function cannot map this correlation, adding hidden layers allows for capturing complex representations. Although there is a functional link between material room coefficients and the speech transmission index, features necessary to accurately calculate the descriptor have been omitted, so a multilayer neural network allows for a better representation of the output.

The last analysis involves the results in classification, which sometimes do not follow the trend of errors in regression. This aspect shows that STI values that fall around the transition ranges of intelligibility ratings may be affected by low errors in regression but high errors in classification.

Table 12 compares the average STI predictions achieved with the ANN in the configuration that provides the best results (dataset composed of 100 samples, 3 hidden layers with 8, 64, and 16 hidden units, respectively) with the findings of *pyeSTImate* (hybrid method) and the outcomes of measurements. As in the previous assessment of speech intelligibility through *pyeSTImate*, we report in Table 13 the overall summary of the results calculated for the same grade and overall classrooms.

Although our goal is to obtain through the neural network findings as close as possible to those of in situ measurements, as expected, the lowest STI errors in individual, same-grade and overall classrooms are found between ANN predictions and tool outputs. The motivation is related to the dataset we used to train the ANN because the neural network learned the relationship between input data and STIs generated by the simulation tool, so the inaccuracy of the deep learning model must be added to the one associated with the simulator. In fact, the comparison with measurements returns in most case studies amplified errors with respect to the comparison with *pyeSTImate* outcomes.

The weaknesses of the synthetic data are confirmed by similar average STI values in classrooms with comparable geometric characteristics that result in a flattening of intelligibility ratings to the “fair” level of speech comprehension quality. This is an indication that the features represented by the geometric dimensions have the most significant weight in determining the STI and demonstrate that material characteristics have low variability in terms of absorption and scattering coefficients. An interesting investigation consists of evaluating the performance of the deep learning algorithm trained with data from measurements in real environments to be tested on our case studies of classrooms. This project involves collecting a large amount of reverberation time measurements in classrooms and is a further potential development of the present work.

In summary, despite the less accuracy of the results yielded by the ANN, STI JNDs(0.1) values with respect to STI from measurements are within the threshold in most of the case studies (26 out of 35 classrooms). For this reason, the ANN-based algorithm can be considered a fast method for preliminary assessments of speech intelligibility in classrooms, but the complete *pyeSTImate* tool is recommended for more reliable evaluations.

Conclusions

The topic of this work is the assessment of speech intelligibility in school environments through the prediction of the speech transmission index, a descriptor of indoor acoustic quality according to the Italian regulation for public buildings. In this regard, we presented *pyeSTImate*, a fully simulated tool developed in the Python programming language for speech transmission index prediction in lecture rooms with parallelepiped geometry and without limitations in size. The code has been implemented from the pyroomacoustics software package adapted to the

Table 11. Best ANN performance (MAE, MAPE, MIRE) with datasets of different sample sizes (100–1000), hidden layers (1–3), and hidden units (8–64) tested on the STI values calculated from both the simulated and measured RTs over 35 classrooms

dataset samples	hidden layers	hidden units	trainable parameters	Testing STIs simulated			Testing STIs from measurements		
				MAE	MAPE (%)	MIRE (%)	MAE	MAPE (%)	MIRE(%)
100	1	8	601	0.07	13.99	8.57	0.08	20.65	8.00
1000		8	601	0.11	23.78	19.43	0.12	30.06	19.43
100	2	8–64	1233	0.06	13.59	10.28	0.08	19.67	9.71
1000		8–64	1233	0.07	15.54	8.57	0.09	23.34	8.00
100	3	8–64–16	2225	0.06	11.98	8.57	0.08	19.36	8.57
1000		8–64–16	2225	0.07	15.18	8.57	0.09	22.05	8.00

Table 12. ANN results: average STI and IR computed by the deep learning model; absolute error (AE), absolute percentage error (APE), Intelligibility Rating, intelligibility rating error (IRE), and just noticeable difference units with thresholds equal to 0.03 (STI JNDs(0.03)) and 0.1 (STI JNDs(0.1)) from comparison with both *pyeSTImate* and measurements

ID	Type		Name of room	ANN		Comparison with *pyeSTImate*					Comparison with measurements				
				STI avg	IR	AE	APE (%)	IRE	STI JNDs(0.03)	STI JNDs(0.1)	AE	APE (%)	IRE	STI JNDs(0.03)	STI JNDs(0.1)
1	SCH 1	University	140/1	0.51	fair	0.03	6.58	0	1.0	0.3	0.02	3.03	0	0.5	0.2
2	SCH 1	University	140/2	0.56	fair	0.11	25.31	0.2	3.8	1.1	0.15	37.42	0.2	5.1	1.5
3	SCH 1	University	140/3	0.55	fair	0.05	10.56	0	1.8	0.5	0.09	18.28	0	2.8	0.9
4	SCH 1	University	155/D1	0.54	fair	0.01	1.80	0	0.3	0.1	0.02	4.42	0	0.8	0.2
5	SCH 1	University	155/D2	0.54	fair	0.01	2.25	0	0.4	0.1	0.02	4.00	0	0.7	0.2
6	SCH 1	University	155/D3	0.54	fair	0.01	2.45	0	0.4	0.1	0.02	3.85	0	0.7	0.2
7	SCH 1	University	155/D4	0.54	fair	0.01	2.65	0	0.5	0.1	0.01	2.26	0	0.4	0.1
8	SCH 1	University	160/1	0.55	fair	0.00	0.63	0	0.1	0.0	0.07	15.18	0	2.4	0.7
9	SCH 1	University	160/2	0.55	fair	0.01	1.66	0	0.3	0.1	0.07	13.99	0	2.3	0.7
10	SCH 1	University	AT1	0.47	fair	0.10	17.33	0	3.3	1.0	0.13	21.01	0	4.2	1.3
11	SCH 1	University	AT2	0.47	fair	0.14	23.30	0.2	4.8	1.4	0.15	23.58	0.2	4.8	1.5
12	SCH 1	University	AT3	0.42	poor	0.22	34.35	0.4	7.3	2.2	0.20	32.26	0.4	6.7	2.0
13	SCH 1	University	EN1	0.59	fair	0.04	7.08	0	1.3	0.4	0.06	10.80	0	1.9	0.6
14	SCH 1	University	EN3	0.53	fair	0.13	30.89	0.2	4.2	1.3	0.03	6.00	0	1.0	0.3
15	SCH 1	University	S1	0.52	fair	0.19	58.58	0.2	6.4	1.9	0.24	87.39	0.4	8.1	2.4
16	SCH 1	University	S2	0.53	fair	0.08	18.58	0.2	2.8	0.8	0.26	94.50	0.4	8.6	2.6
17	SCH 1	University	S3	0.52	fair	0.08	17.67	0.2	2.6	0.8	0.24	85.71	0.4	8.0	2.4
18	SCH 2	Primary	1B	0.52	fair	0.11	17.16	0.2	3.6	1.1	0.10	16.47	0.2	3.4	1.0
19	SCH 2	Primary	2A	0.54	fair	0.04	7.41	0	1.2	0.4	0.00	0.92	0	0.2	0.0
20	SCH 2	Primary	3A	0.54	fair	0.02	2.99	0	0.6	0.2	0.01	0.92	0	0.2	0.1
21	SCH 3	Primary	2C	0.53	fair	0.00	0.07	0	0.0	0.0	0.04	7.07	0	1.2	0.4
22	SCH 3	Primary	3C	0.53	fair	0.02	3.73	0	0.6	0.2	0.01	1.85	0	0.3	0.1
23	SCH 4	Primary	4B	0.53	fair	0.02	4.39	0	0.7	0.2	0.07	12.03	0	2.4	0.7
24	SCH 4	Primary	4A	0.53	fair	0.10	24.06	0.2	3.4	1.0	0.10	23.98	0.2	3.4	1.0
25	SCH 5	Secondary	1	0.51	fair	0.01	1.27	0	0.2	0.1	0.08	13.19	0	2.6	0.8
26	SCH 5	Secondary	2	0.51	fair	0.01	1.48	0	0.2	0.1	0.03	5.56	0	1.0	0.3
27	SCH 5	Secondary	3	0.51	fair	0.01	1.76	0	0.3	0.1	0.06	10.53	0	2.0	0.6
28	SCH 6	Secondary	4	0.51	fair	0.04	7.03	0	1.3	0.4	0.02	4.62	0	0.8	0.2
29	SCH 6	Secondary	5	0.50	fair	0.04	7.62	0	1.4	0.4	0.02	3.38	0	0.6	0.2
30	SCH 7	Secondary	6	0.54	fair	0.05	7.92	0	1.5	0.5	0.18	24.74	0.2	5.9	1.8
31	SCH 7	Secondary	7	0.57	fair	0.02	2.97	0	0.6	0.2	0.05	10.14	0	1.8	0.5
32	SCH 7	Secondary	8	0.51	fair	0.05	8.57	0	1.6	0.5	0.04	7.27	0	1.3	0.4
33	SCH 8	Secondary	9	0.53	fair	0.04	6.55	0	1.2	0.4	0.01	0.95	0	0.2	0.1
34	SCH 8	Secondary	10	0.49	fair	0.09	21.09	0.2	2.8	0.9	0.15	44.12	0.2	5.0	1.5
35	SCH 8	Secondary	11	0.48	fair	0.09	23.29	0.2	3.0	0.9	0.09	22.29	0.2	2.9	0.9

Table 13. Comparison of errors between average STIs from ANN and *pyeSTImate*, and from ANN and measurements for classrooms of the same grade and overall the classrooms: mean absolute error (MAE), mean absolute percentage error (MAPE), mean intelligibility rating error (MIRE), and mean just noticeable difference units with thresholds equal to 0.03 (mean STI JNDs(0.03)) and 0.1 (mean STI JNDs(0.1))

	Classrooms	MAE	MAPE (%)	MIRE (%)	mean STI JNDs(0.03)	mean STI JNDs(0.1)
Comparison with *pyeSTImate*	University (ID 1–17)	0.07	15.39	9.41	2.4	0.7
	Primary (ID 18–24)	0.04	8.54	5.71	1.5	0.4
	Secondary (ID 25–35)	0.04	8.14	3.64	1.3	0.4
	Overall	0.06	11.74	6.86	1.9	0.6
Comparison with measurements	University (ID 1–17)	0.10	27.28	11.76	3.5	1.0
	Primary (ID 18–24)	0.05	9.03	5.71	1.6	0.5
	Secondary (ID 25–35)	0.07	13.34	5.45	2.2	0.7
	Overall	0.08	19.25	8.57	2.7	0.8

scope; specifically, the data required by the tool are the room geometry, finishing materials, source and receiver positions, and the simulation method chosen among image-source, ray tracing and hybrid. To evaluate the accuracy of the predictive tool, we compared the results with the outcomes of speech intelligibility measurements conducted in 35 classrooms of several grades in the Marche Region in Italy. From the monitoring of the absolute error, absolute percentage error, intelligibility rating error, and just noticeable difference units associated with the average STI for individual classrooms and their mean values distinct for primary, secondary, university and overall lecture rooms, the best results have been achieved with the ray tracing and hybrid methods. The tool has returned a just noticeable difference in STI values within 0.1 for small ($V<200\,m^3$) and wide ($V\geq350\,m^3$) classrooms, regardless of grade, year of building construction, and finishing materials. In medium-sized ($200\,m^3\leq V<350\,m^3$) classrooms that are in between the different source levels specified by the regulations (60 dB and 70 dB) and in conditions of high ambient noise, the tool has shown slightly lower accuracy (between 1.3–1.7 STI JNDs with the threshold of 0.1).

In addition to the *pyeSTImate* tool, we implemented an artificial neural network able to yield a deep learning model with a good generalization capability and optimized for STI prediction in lecture rooms with a reduced number of input data. To calculate the speech transmission index, we generated a synthetic dataset of classrooms and the associated STI values employing the hybrid method and exporting a set of parameters considerably reduced with respect to data required by *pyeSTImate*. Regarding the ANN results, similar average STI values can be observed in classrooms with comparable geometric dimensions. This results in a flattening of intelligibility ratings to the "fair" level of speech comprehension quality, showing that the strong correlation between STI and room geometric dimensions implies a low variability in terms of absorption and scattering coefficients. Despite the less accurate results achieved by the ANN, the just noticeable difference is within 0.1 in most of the case studies.

In conclusion, the present work suggests an important role of prediction methods in speech intelligibility in the main acoustic feature dimensions. The results show that comparably measurements and calculations starting from real input data are surprisingly informative on the level of acoustic detail and the degree to which listeners are able to utilize it for speech comprehension. The predictive tool has demonstrated computational robustness that enables its use for preliminary assessments of speech intelligibility, to design the optimal type of scholar buildings and for sound amplification systems in classrooms in compliance with the Italian regulation.

This study represents a starting point for several future works. Some insights for further research consist of the sensitivity analysis of the tool results to varying absorption and scattering coefficients of materials, the implementation of complex geometries, and the design of an interface for data entry. The deep learning model can be improved with a training set obtained from real measurements or with innovative methods for generating synthetic data, such as generative adversarial networks.

References

1. UNI. Assessment of speech communication. Standard UNI EN ISO 9921, UNI, 2004. Available online: https://store.uni.com/uni-en-iso-9921-2004 (accessed on 08 December 2022).
2. Asdrubali F, Venanzi D, Evangelisti L et al. An evaluation of the environmental payback times and economic convenience in an energy requalification of a school. *Buildings* 2020; 11(1): 12. https://doi.org/10.3390/buildings11010012.
3. ministero dell'ambiente e della tutela del territorio e del mare. Criteri ambientali minimi per l'affidamento di servizi di progettazione e lavori per la nuova costruzione, ristrutturazione e manutenzione di edifici pubblici. Ministerial Decree 11 January, 2017. Available online: https://www.gazzettaufficiale.it/eli/id/2017/11/06/17A07439/sg (accessed on 08 December 2022).
4. Goldsworthy RL and Greenberg JE. Analysis of speech-based speech transmission index methods with implications for nonlinear operations. *The Journal of the Acoustical Society of America* 2004; 116(6): 3679–3689. https://doi.org/10.1121/1.1804628.
5. Mogas-Recalde J, Palau R and Márquez M. How classroom acoustics influence students and teachers: A systematic literature review. *Journal of Technology and Science Education* 2021; 11(2): 245–259. https://doi.org/10.3926/jotse.1098.
6. Minelli G, Puglisi GE and Astolfi A. Acoustical parameters for learning in classroom: A review. *Building and Environment* 2021; : 108582. https://doi.org/10.1016/j.buildenv.2021.108582.
7. Astolfi A, Puglisi GE, Murgia S et al. Influence of classroom acoustics on noise disturbance and well-being for first graders. *Frontiers in Psychology* 2019; 10: 2736. https://doi.org/10.3389/fpsyg.2019.02736.
8. Mealings K. Classroom acoustics and cognition: A review of the effects of noise and reverberation on primary school children's attention and memory. *Building Acoustics* 2022; 29(3): 401–431. https://doi.org/10.1177/1351010X221104892.
9. Murgia S, Webster J, Cutiva LCC et al. Systematic review of literature on speech intelligibility and classroom acoustics in elementary schools. *Language, Speech, and Hearing Services in Schools* 2022; : 1–14 https://doi.org/10.1044/2022_LSHSS-21-00181.
10. Prodi N, Visentin C and Feletti A. On the perception of speech in primary school classrooms: Ranking of noise interference

and of age influence. *The Journal of the Acoustical Society of America* 2013; 133(1): 255–268. https://doi.org/10.1121/1.4770259.

11. Astolfi A and Pellerey F. Subjective and objective assessment of acoustical and overall environmental quality in secondary school classrooms. *The Journal of the Acoustical Society of America* 2008; 123(1): 163–173. https://doi.org/10.1121/1.2816563.
12. Minichilli F, Gorini F, Ascari E et al. Annoyance judgment and measurements of environmental noise: A focus on italian secondary schools. *International journal of environmental research and public health* 2018; 15(2): 208. https://doi.org/10.3390/ijerph15020208.
13. Ljung R, Sörqvist P, Kjellberg A et al. Poor listening conditions impair memory for intelligible lectures: implications for acoustic classroom standards. *Building Acoustics* 2009; 16(3): 257–265. https://doi.org/10.1260/135101009789877031.
14. Puglisi GE, Prato A, Sacco T et al. Influence of classroom acoustics on the reading speed: A case study on italian second-graders. *The Journal of the Acoustical Society of America* 2018; 144(2): EL144–EL149. https://doi.org/10.1121/1.5051050.
15. Prodi N, Visentin C, Borella E et al. Noise, age, and gender effects on speech intelligibility and sentence comprehension for 11-to 13-year-old children in real classrooms. *Frontiers in Psychology* 2019; 10: 2166. https://doi.org/10.3389/fpsyg.2019.02166.
16. Castro-Martínez JA, Chavarría Roa J, Parra Benítez A et al. Effects of classroom-acoustic change on the attention level of university students. *Interdisciplinaria* 2016; 33(2): 201–214. Available online: http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1668-70272016000200001&lng=es&nrm=iso (accessed on 08 December 2022).
17. Visentin C, Prodi N, Cappelletti F et al. Speech intelligibility and listening effort in university classrooms for native and non-native italian listeners. *Building Acoustics* 2019; 26(4): 275–291. https://doi.org/10.1177/1351010X19882314.
18. Choi YJ. The intelligibility of speech in university classrooms during lectures. *Applied Acoustics* 2020; 162: 107211. https://doi.org/10.1016/j.apacoust.2020.107211.
19. Picou EM, Gordon J and Ricketts TA. The effects of noise and reverberation on listening effort for adults with normal hearing. *Ear and hearing* 2016; 37(1): 1. https://doi.org/10.1097/AUD.0000000000000222.
20. Puglisi GE, Warzybok A, Astolfi A et al. Effect of reverberation and noise type on speech intelligibility in real complex acoustic scenarios. *Building and Environment* 2021; 204: 108137. https://doi.org/10.1016/j.buildenv.2021.108137.
21. Bottalico P and Astolfi A. Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *The Journal of the Acoustical Society of America* 2012; 131(4): 2817–2827. https://doi.org/10.1121/1.3689549.
22. Bottalico P, Graetzer S and Hunter EJ. Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *The Journal of the Acoustical Society of America* 2016; 139(5): 2870–2879. https://doi.org/10.1121/1.4950812.
23. Puglisi GE, Astolfi A, Cantor Cutiva LC et al. Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. *The Journal of the Acoustical Society of America* 2017; 141(1): 441–452. https://doi.org/10.1121/1.4973805.
24. Bettarello F, Caniato M, Scavuzzo G et al. Indoor acoustic requirements for autism-friendly spaces. *Applied Sciences* 2021; 11(9): 3942. https://doi.org/10.3390/app11093942.
25. Caniato M, Zaniboni L, Marzi A et al. Evaluation of the main sensitivity drivers in relation to indoor comfort for individuals with autism spectrum disorder. part 1: Investigation methodology and general results. *Energy Reports* 2022; 8: 1907–1920. https://doi.org/10.1016/j.egyr.2022.01.009.
26. Steeneken HJ and Houtgast T. A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America* 1980; 67(1): 318–326. https://doi.org/10.1121/1.384464.
27. Schwerin B and Paliwal K. An improved speech transmission index for intelligibility prediction. *Speech Communication* 2014; 65: 9–19. https://doi.org/10.1016/j.specom.2014.05.003.
28. Yang D and Mak CM. An investigation of speech intelligibility for second language students in classrooms. *Applied Acoustics* 2018; 134: 54–59. https://doi.org/10.1016/j.apacoust.2018.01.003.
29. Liu H, Ma H, Kang J et al. The speech intelligibility and applicability of the speech transmission index in large spaces. *Applied Acoustics* 2020; 167: 107400. https://doi.org/10.1016/j.apacoust.2020.107400.
30. Peters R. *Uncertainty in Acoustics: Measurement, Prediction and Assessment*. CRC Press, 2020.

31. Galbrun L and Kitapci K. Accuracy of speech transmission index predictions based on the reverberation time and signal-to-noise ratio. *Applied Acoustics* 2014; 81: 1–14. https://doi.org/10.1016/j.apacoust.2014.02.001.
32. Nowoświat A and Olechowska M. Fast estimation of speech transmission index using the reverberation time. *Applied Acoustics* 2016; 102: 55–61. https://doi.org/10.1016/j.apacoust.2015.09.001.
33. Leccese F, Rocca M and Salvadori G. Fast estimation of speech transmission index using the reverberation time: Comparison between predictive equations for educational rooms of different sizes. *Applied Acoustics* 2018; 140: 143–149. https://doi.org/10.1016/j.apacoust.2018.05.019.
34. UNI. Caratteristiche acustiche interne di ambienti confinati - metodi di progettazione e tecniche di valutazione - parte 2: Settore scolastico. Standard UNI 11532-2, UNI, 2020. Available online: https://store.uni.com/uni-11532-2-2020 (accessed on 08 December 2022).
35. IEC. Sound system equipment - part 16: Objective rating of speech intelligibility by speech transmission index. International Electrotechnical Commission IEC 60268-16, IEC, 2011. Available online: https://webstore.iec.ch/publication/1214 (accessed on 08 December 2022).
36. BS. Sound system equipment. objective rating of speech intelligibility by speech transmission index. British Standards Document BS EN 60268-16, BS, 2011. https://doi.org/10.3403/30249993U.
37. UNI. Caratteristiche acustiche interne di ambienti confinati - metodi di progettazione e tecniche di valutazione - parte 1: Requisiti generali, 2018. Standard UNI 11532-1, UNI, 2018. Available online: https://store.uni.com/uni-11532-1-2018 (accessed on 08 December 2022).
38. Van Rossum G. Python reference manual. *Department of Computer Science [CS]* 1995; (R 9525).
39. Scheibler R, Bezzam E and Dokmanić I. Pyroomacoustics: A python package for audio room simulation and array processing algorithms. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, pp. 351–355. https://doi.org/10.1109/ICASSP.2018.8461310.
40. Bradley JS, Reich R and Norcross S. A just noticeable difference in c50 for speech. *Applied Acoustics* 1999; 58(2): 99–108. https://doi.org/10.1016/S0003-682X(98)00075-9.
41. International standard ISO. Acoustics - measurement of room acoustic parameters - part 2: Reverberation time in ordinary rooms. Standard ISO 3382-2, International Organization for Standardization, 2008. Available online: https://www.iso.org/standard/36201.html (accessed on 08 December 2022).
42. International standard ISO. Acoustics — measurement of room acoustic parameters — part 1: Performance spaces. Standard ISO 3382-1, International Organization for Standardization, 2009. Available online: https://www.iso.org/standard/40979.html (accessed on 08 December 2022).
43. Schroeder MR. Modulation transfer functions: Definition and measurement. *Acta Acustica united with Acustica* 1981; 49(3): 179–182.
44. Houtgast T and Steeneken HJ. The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acta Acustica united with Acustica* 1973; 28(1): 66–73.
45. UNI. Building acoustics - estimation of acoustic performance of building from the performance of elements - part 6: Sound absorption in closed environments. Standard UNI EN ISO 12354-6, UNI, 2006. Available online: https://store.uni.com/uni-en-12354-6-2006 (accessed on 08 December 2022).
46. Christensen CL. Odeon room acoustics program, version 17.0, user manual, 2021. Available online: https://odeon.dk/download/Version17/OdeonManual.pdf (accessed on 08 December 2022).
47. Allen JB and Berkley DA. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America* 1979; 65(4): 943–950. https://doi.org/10.1121/1.382599.
48. Schröder D. *Physically based real-time auralization of interactive virtual environments*, volume 11. Logos Verlag Berlin GmbH, 2011.
49. Vorländer M. *Auralization*. Springer, 2020. https://doi.org/10.1007/978-3-030-51202-6.
50. Noriega L. Multilayer perceptron tutorial. *School of Computing Staffordshire University* 2005; 4: 5. Available online: http://www.amno.moph.go.th/research/uploadfile/1365058846mlp.pdf (accessed on 08 December 2022).
51. Liashchynskyi P and Liashchynskyi P. Grid search, random search, genetic algorithm: a big comparison for nas. *arXiv preprint arXiv:191206059* 2019; https://doi.org/10.48550/arXiv.1912.06059.
52. Kingma DP and Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980* 2014; https://doi.org/10.48550/arXiv.1412.6980.
53. Abadi M, Barham P, Chen J et al. {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*. pp. 265–283. https://doi.org/10.48550/arXiv.

1412.6980.

54. Sabine WC and Egan MD. Collected papers on acoustics, 1994. https://asa.scitation.org/doi/pdf/10.1121/1.409944.
55. Rindel JH. The use of computer modeling in room acoustics. *Journal of vibroengineering* 2000; 3(4): 219–224. Available online: https://odeon.dk/pdf/Vilnius_2000-rindel.pdf (accessed on 08 December 2022).
56. Shtrepi L, Pelzer S, Rychtáriková M et al. Objective and subjective assessment of scattered sound in a virtual acoustical environment simulated with three different algorithms. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 2012. Institute of Noise Control Engineering, pp. 7394–7405.
57. Shtrepi L, Astolfi A, Pelzer S et al. Objective and perceptual assessment of the scattered sound field in a simulated concert hall. *The Journal of the Acoustical Society of America* 2015; 138(3): 1485–1497. https://doi.org/10.1121/1.4929743.
58. Choi YJ. Effects of periodic type diffusers on classroom acoustics. *Applied acoustics* 2013; 74(5): 694–707. https://doi.org/10.1016/j.apacoust.2012.11.010.