Predicting speech intelligibility in university classrooms using geometrical acoustic simulations

Giulia Fratoni¹, Dario D’Orazio¹, Domenico De Salvio¹, Massimo Garai¹
¹ Department of Industrial Engineering, University of Bologna, Bologna, Italy

Abstract
In university classrooms a suitable acoustic condition is necessary to enhance the productivity of students and the vocal comfort of lecturers. With this purpose international standards recently introduced new requirements on the quality of verbal communication. The goal of this work is to propose a procedure for predicting speech intelligibility in large learning spaces through geometrical acoustic simulations. The performance of the present approach is investigated using analytical prediction models as well as the measurements results in six university classrooms surveyed.

Introduction
The acoustic design of a classroom should aim at providing suitable conditions for a good verbal communication between lecturers and students (Ricciardi and Buratti (2018), Leccese et al. (2018)). A proper environmental comfort is necessary to minimise the vocal effort for speakers and to facilitate the speech reception for listeners (Visentin et al. (2018), Astolfi et al. (2015), Pelegrín-García et al. (2011), Puglisi et al. (2018)). Several European regulations on classroom acoustics have been recently updated including intelligibility criteria requirements, as summarised by Pelegrín-García et al. (2014). The recent version of the Italian standard UNI 11532 (UNI (2018)), taking target ranges from BB93 (Building Bulletin (1993)) and DIN 18041 (Deutsche Institut für Normung (2016)), is an example of international trends. A proper acoustic condition should give support to the speaker voice without compromising the intelligibility at the listener positions. The objective and measurable criteria usually considered as descriptors of the verbal communication quality are the reverberation time ($T_{20}$), the early-to-late index ($C_{50}$) and the Speech Transmission Index (STI). Due to the significant role of the early reflected energy for the intelligibility of spoken words, it is a common practice to assess $T_{20}$, i.e. the reverberation time taken from -5 dB to -25 dB below the initial level of the decay curve and then extrapolated to a decay time of 60 dB. Among the several theories and regulations stating the suitable range for the intelligibility parameters in a classroom, the methods provided by UNI 11532 (UNI (2018)) were taken as reference point in the present work. The suitable range of $T_{20}$ values depends on the volume of the hall and it is considered in occupied conditions (80% of total occupancy). Generally in large classrooms ($V > 500 m^3$) $T_{20}$ target range turns out to be between 0.70 and 0.90 seconds, according to the method proposed by DIN 18041 (Deutsche Institut für Normung (2016)) and then adopted by UNI 11532 (UNI (2018)). The early-to-late index is defined as (EN ISO (2008)):

$$C_{50} = 10 \log \frac{\int_0^{50} p(t)^2 dt}{\int_0^{50} p(t)^2 dt} (dB) \quad (1)$$

where 50 is assumed as the threshold between early and late sound energy arriving and $p(t)$ is the instantaneous sound pressure of the impulse response. In large lecture halls $C_{50}$ should assume a value greater than 0 dB to ensure a proper speech clarity (UNI (2018)). The Speech Transmission Index is a parameter describing how much the speech is deteriorated by the reverberation time and the background noise of the room considered (BS EN (2011)). It is the result of the weighted contributions of individual frequency bands and its range spans from 0 (worst condition) to 1 (best condition), both the limit values representing ideal and theoretical cases. The minimum value recommended for STI values is 0.60, that is the threshold of “good” intelligibility according to to the ranking provided by IEC 60268-16 (BS EN (2011)).

In order to predict the intelligibility parameters in the acoustic design phase of a classroom, analytical prediction models are provided (UNI (2018)). Nevertheless predicted STI values remain hard to be found out in advance because quite complex evaluations are required. Annex L in IEC 60268-16 (BS EN (2011)) provides an indirect method for analytically calculating STI values based on the use of measured impulse responses and on the assumption of an ideal diffuse reverberant field. The modulation transfer function
If the acoustics of the room is poor, as frequently is from the sound source. This effect is even greater speech intelligibility significantly varies moving away time is a quite constant parameter in a room while the back the prediction of STI to the prediction of simpler field in a classroom, several studies attempted to lead since the diffuse sound field may be an assumption measurement carried out by D’Orazio et al. (2018) and the quality of the speech communication depends mainly on SNR, as may be confirmed also by measurements. Consequently given a certain room acoustics the theory assumed. According to the classical semi-reverberant theory the sound strength may be expressed as:

$$m_k(f_m) = \frac{1}{10^{\text{SNR_k}/10}} - 1$$

where $h_k(t)$ is the impulse response of $k$-octave band, $t$ is the integration variable for time, $f_m$ is the modulation frequency and SNR is the signal-to-noise ratio in dB. Equation 2 shows how the modulation transfer function $m_k(f_m)$ is affected by two terms: the room acoustics (impulse response) and the signal-to-noise ratio (SNR), i.e. the difference between the sound pressure level at the receiver and the background noise. Consequently given a certain room acoustics the quality of the speech communication depends mainly on SNR, as may be confirmed also by measurements. Therefore the modulation transfer function $m(f_m)$ for each octave band may be written as:

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

where $Q$ is the directivity factor of the sound source $(Q = 1$ for omnidirectional sound source), $r$ is the source–receiver distance in meters, $\delta(t)$ is the delta function, $r_c$ is the critical radius of the room in meters, $T$ is the predicted reverberation time of the room in seconds. Therefore the modulation transfer function $m(f_m)$ for each octave band may be simplified as:

$$m(f_m) = \frac{\sqrt{A^2 + B^2}}{C}$$

with the components $A$, $B$ and $C$ expressed as:

$$A = \frac{Q}{r^2} + \frac{1}{r_c^2} \left[ 1 + \left( \frac{2\pi f_m T}{13.8} \right)^2 \right]^{-1}$$

$$B = \frac{2\pi f_m T}{13.8 r_c^2} \left[ 1 + \left( \frac{2\pi f_m T}{13.8} \right)^2 \right]^{-1}$$

$$C = \frac{Q}{r^2} + \frac{1}{r^2} + Q \cdot 10^{-\text{SNR}/10}.$$
Figure 1: Inside views of the lecture halls surveyed (July 2017 - October 2018).

Table 1: Main data of the lecture halls surveyed ($V$ = volume, $S_A$ = students area, $N$ = occupancy). Measured $T_{20}$ values are averaged in the range 500–1000 Hz while measured $C_{50}$ in the range 125–4000 Hz. The occupied conditions (“occ”) are calculated with the method provided by UNI 11532 (UNI (2018)).

<table>
<thead>
<tr>
<th>Hall</th>
<th>$V$ ($m^3$)</th>
<th>Shape</th>
<th>Seats</th>
<th>$S_A$ ($m^2$)</th>
<th>$N$</th>
<th>$T_{20,occ}$ (s)</th>
<th>$C_{50}$ (dB)</th>
<th>STI</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1000</td>
<td>Amphitheatre</td>
<td>Wood</td>
<td>110</td>
<td>250</td>
<td>0.90</td>
<td>-2.5</td>
<td>0.48</td>
</tr>
<tr>
<td>b</td>
<td>900</td>
<td>Amphitheatre</td>
<td>Wood</td>
<td>95</td>
<td>200</td>
<td>0.99</td>
<td>-1.9</td>
<td>0.49</td>
</tr>
<tr>
<td>c</td>
<td>850</td>
<td>Shoe-box</td>
<td>Plastic</td>
<td>80</td>
<td>170</td>
<td>1.22</td>
<td>-4.1</td>
<td>0.43</td>
</tr>
<tr>
<td>d</td>
<td>910</td>
<td>Shoe-box</td>
<td>Wood</td>
<td>35</td>
<td>90</td>
<td>1.70</td>
<td>-3.3</td>
<td>0.44</td>
</tr>
<tr>
<td>e</td>
<td>780</td>
<td>Shoe-box</td>
<td>Wood</td>
<td>40</td>
<td>90</td>
<td>1.83</td>
<td>-4.4</td>
<td>0.41</td>
</tr>
<tr>
<td>f</td>
<td>800</td>
<td>Shoe-box</td>
<td>Wood</td>
<td>55</td>
<td>105</td>
<td>1.30</td>
<td>-2.1</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Method

The aim of the present study is to analyse the accuracy of speech intelligibility criteria predictions obtained with geometrical acoustic simulation techniques. A measurements campaign was performed in six university lecture halls to get a series of reference data during the work (see Figures 1 - 2). The calibration of 3D virtual models was carried out achieving the match between measured and simulated reverberation time (see Figures 3 - 4). The mutual match between measured, simulated and predicted $G$ values confirms that geometrical acoustic simulations could be able to give accurate information on normalized sound pressure level spatial distribution (see Figure 5). Measured STI and $G$ values displayed as functions of source-to-receiver distance show a significant relation between speech intelligibility and sound pressure level decays (see Figure 6). Therefore it could be suggested that the knowledge of $G$ spatial trend may allow to get reliable information on STI trend, overcoming the difficulties of STI prediction models’ calculations.

Case studies

The present study is conducted in a sample of six lecture halls of the University of Bologna located in historical buildings of the city centre. They are characterised by large volumes (greater than 500 $m^3$) including up to 250 occupants. Of the six classrooms, two ($a$ and $b$) have an amphitheatre shape while the other four ($c, d, e, f$) have a quite regular shoe-box geometry except for the presence of small coupled volumes in the ceiling or in the side walls (see Figure 1). All the classrooms have similar surface finishes with sound reflecting materials such as smooth plastered walls, large windows placed on only one side of the halls and wooden (or plastic) seats, not counting a small part of the audience of hall $e$ with slightly upholstered seats. Table 1 shows the main data of the six lecture halls including the volume, the shape, the materials and the occupancy.

Measurements

Between 2017 and 2018, more acoustic measurements campaigns were performed in the lecture halls surveyed to collect objective speech intelligibility criteria. Measurements were conducted in full furnished
condition and in unoccupied states complying with ISO 3382-2 (EN ISO (2008)). Monaural impulse responses were acquired using the Exponential Sine Sweep technique, with a signal 512 K in length and sampled at 48 kHz (Guidorzi et al. (2015)). A high-SPL dodecahedron was used as omnidirectional sound source (D’Orazio et al. (2016)), rotated three times for each impulse response acquisition (Martellotta (2013)) in order to remove directivity factors of single loudspeakers as much as possible. The sound source was previously calibrated in a certified reverberation room according to ISO 3741 (UNI EN ISO (2010)). Two sound sources positions and a regular grid of receivers points were chosen for the measurements setup (see Figure 2). The extrapolation of data was conducted with B&K’s Dirac 6 software (Brüel and Kjær (2014)) and an accurate analysis was carried out on $T_{20}$, $C_{50}$, $G$ and STI, using for the latter the indirect method according to IEC 60268-16 (BS EN (2011)). In each receiver position and in each octave band the difference between the sound pressure level received and the background noise was greater than 20 dB, allowing to ignore the background noise level for focusing just on the signal received in each position (Hodgson and Wong (2009)).

**Numerical simulation**

The lecture halls surveyed represent a typology of rooms particularly suitable for numerical simulations due to their large volumes and quite regular shapes. The modelling process of approximation required in Geometrical Acoustic (GA) techniques is thus facilitated enhancing the reliability of all the simulations. The 3D models of the university classrooms were created using 3D SketchUp and then imported into ODEON Room Acoustics software (see Figure 3). The modelling process was carried out according to the state-of-the-art recommendations (Vorländer (2007)). Concerning the level of detail needed for the simulations, modelled surfaces were kept greater than 0.35 m following guidelines by Christensen (2011). Sound absorption coefficients were provided by materials databases present in previous scientific literature (Vorländer (2007), Cox and d’Antonio (2016)) and
Table 2: Absorption ($\alpha$) and scattering ($s$) coefficients for the main materials involved in simulation process (Cox and d’Antonio (2016), Vorl¨ander (2007)).

<table>
<thead>
<tr>
<th>Materials</th>
<th>125 Hz</th>
<th>250 Hz</th>
<th>500 Hz</th>
<th>1000 Hz</th>
<th>2000 Hz</th>
<th>4000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plaster-Floor</td>
<td>$\alpha$</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>$s$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>Seats</td>
<td>$\alpha$</td>
<td>0.22</td>
<td>0.10</td>
<td>0.07</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>$s$</td>
<td>0.01</td>
<td>0.10</td>
<td>0.45</td>
<td>0.65</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Figure 4: Calibration of geometrical acoustic models: measured and simulated mean $T_{20}$ values in unoccupied conditions as a function of the frequency. $T_{20}$ values are averaged over all the source-receivers couples. The error bars are referred to the Just Noticeable Difference (JND) defined by ISO 3382-1 EN ISO (2009).

Figure 5: Calibration of geometrical acoustic models: measured and simulated $G_{M}$ values as functions of source-to-receiver distance. “M” subscript indicates the values averaged over the central octave-bands 500 – 1000 Hz. Dashed and solid curves are respectively the Barron and Lee (1988) curve (see Equation 9) and the adjustment for university classrooms by Sato and Bradley (2008) (see Equation 10).
then adjusted through the calibration iterative process, according to a consolidated procedure (Astolfi et al. (2008)). Given that GA softwares describe the sound propagation along sound rays, the wave nature of sound had to be introduced by assigning scattering properties to each surface (Rindel (2000)). In particular, scattering coefficients took into account both the surface roughness and the occurrence of edges, compensating in this way the lack of details due to the simplification of the models. In order to simplify the whole procedure and reduce the uncertainty of input data for materials properties an attempt was made to reduce as much as possible the number of the layers used. During the workflow the layers were mainly divided in those ones characterised by higher absorption and scattering coefficients (seats) and all the other ones characterised by sound reflective and smooth surfaces (walls, floors, ceilings), as shown in Table 2. The numerical models were calibrated ensuring the match between measured and simulated values (see Figures 4 - 5). The calibration was considered to be achieved when the difference between measured and simulated values was within one Just Noticeable Difference (JND) according with ISO 3382-1 (EN ISO (2009)).

Results

The last three columns of Table 1 show the overview on main measurements results where mean values are averaged over all source-receiver positions. Measured reverberation time (T20), sound clarity (C50) and speech intelligibility (STI) values indicate that each of the halls investigated present quite poor acoustic conditions for a suitable speech intelligibility. Nevertheless the aim of this work is to analyse the acoustics of typical university classrooms, often located in historical buildings that turn out to be unsuitable for the purpose. In order to assess the actual sound field behaviour of the lecture halls under study the 3D virtual models were calibrated basing on reverberation time in octave bands and the normalised sound pressure level as a function of source-to-receiver distance. Figures 4 and 5 provide the comparison between measured T20 and GM values and simulated ones. The error bars displayed in T20 graphs are referred to the JND defined by ISO 3382-1 (EN ISO (2009)). The dashed and solid curves displayed in GM graphs are referred to analytical prediction models, respectively the Barron and Lee (1988)’s curve (see Equation 9) and the adjustment for university classrooms by Sato and Bradley (2008) (see Equation 10). Results show a good agreement between measurements and simulated values allowing to consider reliable the calibration procedure used. Concerning the spatial distribution of sound pressure level the match with the specific prediction model developed for university classrooms confirms that the halls surveyed present the typical sound field behaviour of large learning spaces.

Discussions

As mentioned in the introduction the modulation transfer function, which is the basis of STI definition, mainly depends on two factors, the first concerning the room acoustics characteristics (impulse responses), and the second concerning the difference between the sound pressure level received and the background noise (SNR). Given a certain room acoustics, namely a certain impulse response, the deterioration of the modulation transfer functions is affected only by the sound pressure level at the receiver position considered. This is confirmed by the similar trend of linear regressions of measured STI and GM plotted as functions of source-to-receiver distance (see Figure 6). Nevertheless a good match between the regression lines is obtained neglecting receivers that are closer than 4 meters from the sound source, to assure that the reflected energy is prevalent rather than the direct one (see gray rectangles in Figure 6). It is possible to deduce that the critical radius of an actual room – defined as the distance at which the sound pressure level of the direct sound and the reverberant sound are equal – is greater than the traditional formula \( r_c = 0.057\sqrt[3]{VT} \) (m) (with \( V \) as the volume of the room in \( m^3 \) and \( T \) as the reverberation time in seconds) derived by classical semi-reverberant theory. It also means that the halls must be large enough for this kind of decay analysis. A reliable geometrical acoustical model, obtained with a simplified but rigorous procedure, may be considered an useful tool for validating the prediction of speech intelligibility by the analysis of sound pressure level spatial distribution. For this reason the calibration of geometrical acoustic models has been conducted taking into account the decay of sound pressure level with the increasing of source-to-receiver distance, as shown in Figure 5. A further analysis could also involve the comparison between STI and the early component of the total sound energy (Gearly) considering the integration time at 50 ms (De Cesari et al. (2015)), due to the relevance of this time interval for speech criteria.

Conclusions

The acoustic design of a classroom is an important issue because the reduction of the vocal effort for teachers and the increasing of the attention for students are quite significant aspects of the learning process. In order to predict the quality of verbal communication a preliminary approach based on geometrical acoustical simulations is here proposed. The data obtained by a measurements campaign conducted in six university lecture halls has been taken as reference point for the assessment of sound field behaviours and for the calibration process. The sample of university lecture halls considered in the present work represent a typology suitable for analysing the decay of the sound pressure level and consequently for connecting the
sound propagation losses to the variations of speech intelligibility throughout the space. The accuracy of simulated G spatial distributions and the similarity between G and STI trends suggest the possibility to get important information on speech intelligibility decays even in a preliminary acoustic design phase. A larger sample of lecture halls surveyed could allow to enhance the statistical significance of the model proposed.

References


Figure 6: Measured STI and $G_M$ values plotted as function of the source-to-receiver distance. “M” subscript indicates the values averaged over the central octave-bands 500 – 1000 Hz. Linear regressions of data are plotted excluding the values within 4 meters from the sound source (in rectangles).


