JANUARY 19 2017

Combined effect of noise and room acoustics on vocal effort in simulated classrooms

Marcella Cipriano; Arianna Astolfi; David Pelegrín-García

J. Acoust. Soc. Am. 141, EL51–EL56 (2017)
https://doi.org/10.1121/1.4973849

Cipriano *et al.*: JASA Express Letters [http://dx.doi.org/10.1121/1.4973849] Published Online 19 January 2017

Combined effect of noise and room acoustics on vocal effort in simulated classrooms

Marcella Cipriano[a] **and Arianna Astolfi**
Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi 24,
10129 Torino, Italy
marcella.cipriano@alice.it, arianna.astolfi@polito.it

David Pelegrín-García
ESAT-STADIUS, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10,
3001 Leuven, Belgium
david.pelegringarcia@kuleuven.be

Abstract: This work investigated the relationships between room acoustics, background noise level, and vocal effort of a speaker in simulated classrooms of various volumes. Under simulated acoustic environments, talkers adjusted their vocal effort linearly with the voice support, i.e., the degree of amplification offered by the room to the voice of a speaker, at his own ears. The slope of this relationship, called the room effect, of −0.24 dB/dB was significant only in the case of the highest noise levels of 62 dB. The vocal comfort for the speaker, however, was found to be more closely related to noise annoyance than to room reverberance.

[NX]
Date Received: November 15, 2016 **Date Accepted:** December 26, 2016

1. Introduction

Classroom acoustic design aims to improve the teacher-student speech communication.[1] Most guidelines for the classroom acoustic design have the goal of enhancing speech intelligibility[2,3] on account of its impact on the students' learning process. At the same time, teachers need to speak comfortably and without straining their voice, since around two-thirds of the teacher population suffer or have suffered from voice problems during their working career.[2] Bad acoustic design of classrooms can alter the auditory feedback of teachers, who as a consequence, increase the voice levels with potential risks for their own voice.[4–6]

In the literature,[2] two room acoustic speaker-oriented parameters have been found relevant to characterize the propagation of sound between the mouth and the ears of a speaker: the decay time $\mathrm{DT}_{40,\mathrm{ME}}$, linked to the perceived voice comfort in the absence of audible background noise, and the voice support ST_V, linked to the vocal effort. The vocal comfort in a room is a subjective attribute that is directly correlated to the positive evaluation of the room for speech production and to the perceived support. It is also negatively correlated to the feeling of having to raise the voice and to the tiredness after speaking long in the room.[2] The vocal effort can be defined[2,7] as a subjective physiological magnitude different from voice level, which accounts for the changes in voice production required for the communication at different distances, under different noise or room acoustics conditions. In a more pragmatic way, we define the vocal effort as the exertion of the speaker, quantified objectively by the A-weighted speech level at 1 m distance in front of the mouth.[3]

The decay time $\mathrm{DT}_{\mathrm{X,ME}}$ and the voice support ST_V are obtained from an oral-binaural room impulse response (OBRIR), i.e., an impulse response measured at a microphone located at the ears of a dummy head when a loudspeaker at its mouth acts as the source. The decay time is a measure of the rate at which the sound decays in the room and is defined as the time it would take for the reverse integrated energy curve of an IR to decay 60 dB after the arrival of the direct sound at t_0, calculated from the initial decay of X dB. This is,

$$\mathrm{DT}_{\mathrm{X,ME}} = \frac{60}{X}\left(t_{-X} - t_0\right), \tag{1}$$

---

[a] Author to whom correspondence should be addressed.

where t_{-X} is the time at which the reverse integrated energy curve of the IR is X dB lower than before the arrival of the direct sound.[5]

The voice support measures the degree of amplification offered by the room to the voice of a speaker at his own ears[2] and is defined as the difference between the reflected sound level (L_R) and the airborne direct sound level (L_D) of the voice of the speaker as found in an OBRIR,[5]

$$ST_V = L_R - L_D(dB). \qquad (2)$$

Pelegrín-García and Brunskog[5] carried out talking tests in simulated rooms with teachers, in absence of background noise, and found a quadratic relationship between $DT_{40,ME}$, averaged over the octave bands between 125 Hz and 4 kHz, and the perceived vocal comfort, with a maximum of preference for decay times between 0.4 and 0.5 s. In addition, they found that the talkers' voice level varies linearly with the voice support, averaged over the same octave bands, and they referred to this slope as "room effect." It depends on the type of talking instructions and on the individuals. Group-wise, the room effect ranges from −0.93 dB/dB, with free speech, to −0.1 dB/dB with a simple communication task such as reading a text.

On the other hand, voice levels increase under different noise levels at a rate of 0.5 to 0.7 dB/dB for moderate-high noise levels.[6] This increase, known as the "Lombard effect," is described as a natural response to account for loudness alterations of one's own voice or to maintain speech intelligibility in adverse conditions.

The present study aims to determine how the room effect is affected by changing in the noise level in simulated classrooms, without changing the visual appearance of the physical environment. The secondary goal is to find out the preferred acoustic conditions for talking in a classroom.

2. Experimental method

The acoustic environments were generated in an acoustic virtual reality (AVR) system[8] that convolved the speech signal produced by a user located in an anechoic room with the measured OBRIRs of five rooms, in order to give participants the auditory perception of speaking in a real environment. Under simulated acoustic conditions the participants were invited to perform a talking test, having no visual feedback of the measured room. At the end of each trial they were asked to answer three questions regarding the auditory speaking experience in order to find the correlations between the vocal comfort and the acoustic condition.

2.1 Conditions

The simulated room acoustic conditions were obtained loading the OBRIRs library measurements in the AVR system to give the participants the experience of auditory sensation measured in the real space.

OBRIRs were measured in five different rooms, ordered from lower to higher voice support: Room 1, an anechoic room of 293 m^3 ($ST_V = -12.2$ dB, $DT_{40,ME} = 0.04$ s);[9] room 2, a normal modern classroom of 302 m^3 with acoustical treatment ($ST_V = -9.7$ dB, $DT_{40,ME} = 0.50$ s); room 3, an old rather large classroom of 297 m^3 ($ST_V = -7.2$ dB, $DT_{40,ME} = 1.30$ s); room 4, an old fashioned narrow long classroom of 144 m^3 ($ST_V = -6.1$ dB, $DT_{40,ME} = 1.10$ s); and room 5, an empty reverberation room of 207 m^3 ($ST_V = -3.6$ dB, $DT_{40,ME} = 1.41$ s).

The auditory effect of the room was obtained in real time by streaming convolution between the speech signal of the participant and the selected OBRIR using the low-latency (4 ms) non-uniformly partitioned convolution engine described in Pelegrin-Garcia *et al.*[8] The speech signal was picked up with a headworn Microphone Sennheiser MKE 2P located at 2 cm from the edge of the lips in the line between the mouth and the right ear and the simulated sound reflections to the voice of a talker were delivered via Sennheiser HD650 headphones. The signals were also recorded for further post-processing.

In the AVR system, the OBRIRs were presented to the speaker in a random order, both without noise or combined with a speech shaped noise[3] at two different levels ($L_N = 50$ dB, $L_N = 62$ dB measured at the listener's ears), to allow the simulations of fifteen room acoustics conditions. "Speech shaped noise" is a stationary noise sequence whose spectrum follows the long term average speech spectrum, averaged for several male and female talkers, with a peak on the 1/3rd octave band of 250 Hz.

2.2 Participants and instructions

Twenty-two participants were involved in the experiment: ten were teachers (four females and six males) and twelve were Ph.D. or university students (two females and ten males) without self-reported hearing or vocal problems. Participants had not engaged into vocally demanding activities for a few hours before the experiments. They were asked to speak as if they were delivering a lecture in a classroom. For this, they were given 15 difference maps, one for each simulated room, with the instruction to describe as clearly as possible to the experimenter the route from a starting point to a finishing point, making reference to the labelled items along the path within a time limit of ninety seconds. English or Italian language was used throughout the experiments and participants were requested to maintain eye-contact with the experimenter seated at 6 m distance in front of her/him.

Each participant was invited to answer a questionnaire regarding the auditory/speaking experience and to write down any remarks about the experiment. Participants answered by drawing a vertical tick in a horizontal line of 100 mm length ranging from total disagreement (extreme left) to total agreement (right) with the following statements.

(1) The room helps me to speak comfortably.
(2) The room feels reverberant.
(3) I find it difficult to speak due to the noise.

The three statements have been linked below to voice comfort, room reverberance, and noise annoyance, respectively.

2.3 Post-processing of the speech signals

The voice recordings were post-processed with WAVESURFER 1.8.8p4 and Microsoft EXCEL. For each acoustic condition, the equivalent speech sound pressure level of the phonated segments ($L_{sp,eq}$) at the position of the headworn microphone was determined. The total average $L_{sp,eq}$ was assumed equal to 67 dB (Ref. 10) (unweighted $L_{p,eq}$ at 1 m from the speaker, averaged between males and females) to account for the lack of a reference measurement at the microphone and the arbitrary internal reference of WAVESURFER. The difference between 67 dB and the total average uncalibrated level was then added to each $L_{sp,eq}$ value.

3. Statistical analysis

The statistical analysis was carried out with IBM SPSS STATISTICS 20. A linear regression model was applied to investigate both the room effect and the Lombard effect: the room effect was calculated as the slope of the linear relationship relating $L_{sp,eq}$ to ST_V for each subject, whereas the Lombard effect/slope was determined for each participant as the difference in $L_{sp,eq}$ for the conditions of $L_N = 50$ dB and $L_N = 62$ dB, averaged across rooms, divided by the difference in noise levels of 12 dB. Confidence intervals were extracted to estimate the reliability of results.[11]

Moreover, a bivariate correlation analysis was addressed to find out the relationships between objective acoustic data and subjective answers. In order to reduce the differences across people due to subjective scaling and to normalize the data, for each participant the answer to each question was substituted by the difference from the average value divided by the sample standard deviation.[11]

4. Results

4.1 Voice support and noise level

Table 1 shows the room effect averaged across participants, the related standard deviations and the confidence intervals for the different noise levels. The negative magnitude means that talkers tended to speak softer as the voice support increased. This slope increased with noise levels: the average room effect was −0.10 dB/dB on silence condition, −0.11 dB/dB for $L_N = 50$ dB and −0.24 dB/dB for $L_N = 62$ dB.

Nevertheless the reliability of the average results was given by a 90% confidence interval. In silence, the confidence range of −0.23–0.02 dB does not exclude the possibility of having an actual slope of 0 dB/dB (i.e., no room effect) on the overall population. In the case of $L_N = 50$ dB, where the confidence range was −0.23–0 dB, the same conclusion applies. On the opposite, in the noisiest condition of 62 dB, a clear tendency of raising the voice with increasing voice support was found by the confidence range of −0.33 to −0.15 dB, entirely shifted towards negative scores.

Figure 1 shows the graphical representation of the vocal effort dependency on ST_V (i.e., the room effect), on equal noise level. The large error bars indicate the high

Table 1. Average, standard deviation and 90% confidence intervals of room effect (dB/dB) per noise condition and Lombard slope (dB/dB).

	Room effect			
	No noise	$L_N = 50$ dB	$L_N = 62$ dB	Lombard effect
Average	−0.10	−0.11	−0.24	0.19
Std. dev	0.35	0.33	0.25	0.16
90% conf. int	−0.23–0.02	−0.23–0	−0.33 to −0.15	0.13–0.25

variability of the results, with standard deviations of about 4 dB. In the silence condition, the lowest measured $L_{sp,eq}$ corresponded to room 2, characterised by a low voice support of −9.7 dB. When $L_N = 50$ dB, the minimum vocal effort was measured in the room 5, with highest voice support (−3.6 dB). In the noisiest condition of $L_N = 62$ dB, room 4 ($ST_V = -6.1$ dB) was found to be the best room to speak in. The highest vocal effort was measured in room 1 (the anechoic room, $ST_V = -12.2$ dB) for two out of three noise conditions.

The Lombard effect results are summarised in Table 1. The positive confidence range shows the talkers' trend to raise their voice when the noise level increased.[2] The voice level averaged across participants and rooms ranged from 67.2 dB for $L_N = 50$–69.5 dB for $L_N = 62$ dB and the mean slope was 0.19 dB/dB. This slope is lower than the ones reported in literature,[3,6] which could mean that talkers lacked incentive for establishing a successful communication scenario. At the same time, the large standard deviation of the slope (0.16 dB/dB) is in good agreement with the common observation that the Lombard effect or reflex is highly variable from speaker to speaker.[12]

4.2 Subjective data

Table 2 displays the most significant Pearson correlations between input data of voice support (ST_v), decay time ($DT_{40,ME}$), noise level (L_N), and speech sound pressure levels ($L_{sp,eq}$), and the average normalised subjective scores on perceived voice comfort, room reverberance, and noise annoyance. Subjective data were obtained as the average over 22 participants for each acoustic environment (15 in total as the product of the three noise levels and the five rooms).

The voice comfort scores were negatively related to the noise annoyance scores, to the $L_{sp,eq}$ scores and to the L_N scores. The perceived room reverberance was positively related to the room acoustic parameters ST_v and $DT_{40,ME}$, and the noise annoyance was positively related to both L_N and $L_{sp,eq}$.

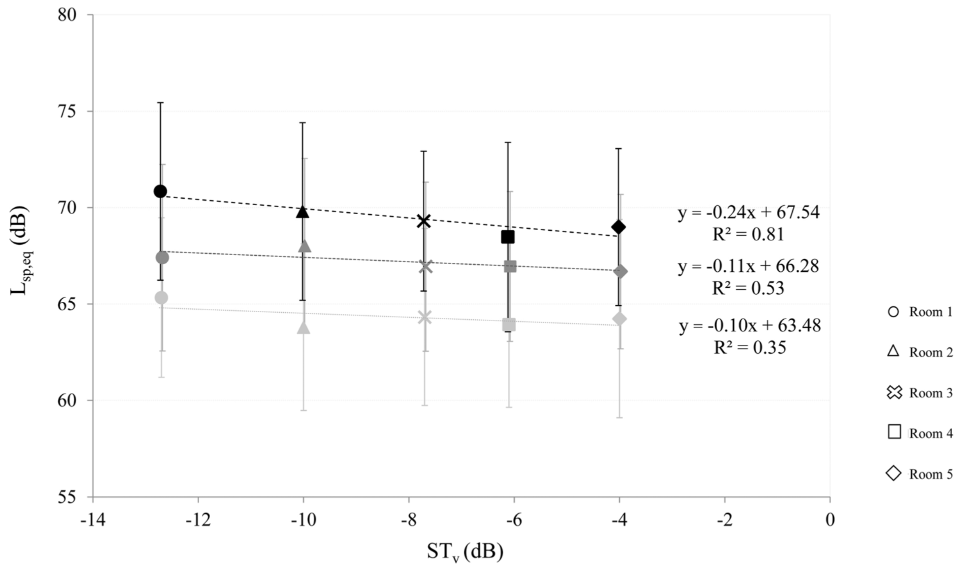

Fig. 1. Talker-averaged values and standard deviations of the equivalent speech sound pressure level ($L_{sp,eq}$) related to the voice support (ST_v) of the rooms, for the three noise level conditions. (No noise: light grey symbols, $L_N = 50$ Db: dark grey symbols, and $L_N = 62$ dB: black symbols.)

Table 2. Pearson's correlation coefficients r (p-value < 0.001) between voice support (ST_v), decay time ($DT_{40,ME}$), noise level (L_N), speech sound pressure levels ($L_{sp,eq}$), and subjective data of voice comfort (comfort), room reverberance (reverberance), and noise annoyance (noise).

	L_N	ST_V	$DT_{40,ME}$	$L_{sp,eq}$	Comfort
$DT_{40,ME}$		0.941			
$L_{sp,eq}$	0.419				
Comfort	−0.461			−0.254	
Reverberance		0.624	0.654		
Noise	0.729			0.342	−0.610

5. Discussion

5.1 Effect of noise on room effect

The rooms can be divided in two groups according to their voice support and decay time values: the dry rooms 1 (anechoic room) and 2, and the more reverberant rooms 3, 4, and 5 (reverberant room). Figure 1 shows that the worst rooms for speaking in both silence and noise conditions, with the highest measured voice levels, are the dry rooms 1 and 2. This result supports the findings by Brunskog *et al.*[13] and Pelegrín-García[2] and Pelegrín-García *et al.*[14]

Under simulated acoustic conditions, there is a room effect, i.e., the talkers' voice levels linearly increase with decreasing voice support, which increased with noise level. A larger spread in the results was found in silence and $L_N = 50$ dB, whereas a more reliable trend was identified for $L_N = 62$ dB. Pelegrín-García and Brunskog[5] found the same room effect as in the present silent condition (−0.10 dB/dB) in the case of describing a map in silence by a group of teachers in almost the same range of ST_V, which is approximately the value of the room effect that keeps the loudness of a talker's voice constant.[15]

They also highlighted the high sensitivity of the speakers toward changing acoustic conditions. In the present study, acquired knowledge in acoustics, in the case of students, or experience in explain something to somebody, in the case of teachers, could have brought to individual changing in the room effect and a large variability in the results. This variability reduced in the case of the highest noise level condition, and a possible explanation is that the presence of noise constituted a more demanding talking scenario which made talkers more aware about the acoustical conditions in order to profit from the voice support and lower their voice levels. A similar behavior of more remarkable room effect was observed in a group of teachers with voice problems compared to teachers without voice problems.[16]

5.2 Subjective outcomes

A significant correlation between objective and subjective variables has been found in the case of L_N with noise annoyance and the $DT_{40,ME}$ with the room reverberance, as expected.[5] Moreover, the significant correlation between noise annoyance and $L_{sp,eq}$ confirms the Lombard effect and points out noise as a main determinant of vocal effort. A third main outcome was the relationship between voice comfort and noise annoyance; however, no correlation was found between voice comfort and room reverberance. In a previous research,[1] conducted without background noise, the voice comfort for speaking was strongly related to decay time by a quadratic function with a maximum comfort at decay times between 0.4 and 0.5 s. From the present study, it appears that when noise is taken into account, its contribution on the perceived voice comfort overcomes the effect of reverberation but nevertheless leads to a stronger room effect (i.e., a decrease of vocal effort with increasing voice support).

6. Conclusions

This research has been conducted in order to determine how the talkers adapt their voice depending on the combined effect of room acoustics and background noise level, as well as to find out the preferred acoustic environment for speaking.

Under simulated acoustic conditions, there is a room effect, i.e., talkers' increase voice levels linearly with decreasing voice support. This slope depends on the noise level and on individuals. In silence, it had an average value of −0.10 dB/dB, whereas on the noisier conditions of 50 and 62 dB, it had an average value of −0.11 dB/dB and −0.24 dB/dB, respectively. Lower variability and significance was found only for the highest noise level. In general, the dry rooms (anechoic or with the lowest reverberation) led to the highest vocal effort regardless of the noise levels.

A questionnaire investigation showed that the voice comfort is more closely related to the perceived noise annoyance than to the perceived room reverberance.

References and links

[1] A. Astolfi, P. Bottalico, and G. Barbato, "Subjective and objective speech intelligibility investigations in primary school classrooms," J. Acoust. Soc. Am. **131**, 247–257 (2012).

[2] D. Pelegrin-Garcia, J. Brunskog, and B. Rasmussen, "Speaker-oriented classroom acoustics design guidelines in the context of current regulations in European countries," Acta Acust. Acust. **100**, 1073–1089 (2014).

[3] ISO 9921, *Ergonomics—Assessment of Speech Communication* (International Organization for Standardization, Geneva, 2003).

[4] P. Bottalico and A. Astolfi, "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms," J. Acoust. Soc. Am. **131**, 2817–2827 (2012).

[5] D. Pelegrín-García and J. Brunskog, "Speakers' comfort and voice level variation in classrooms, Laboratory research," J. Acoust. Soc. Am. **132**, 249–260 (2012).

[6] A. Astolfi, P. Bottalico, A. Accornero, M. Garzaro, J. Nadalin, and C. Giordano, "Relationship between vocal doses and voice disorders on primary school teachers," in *Proceedings of the 9th European Conference on Noise Control* (2012).

[7] H. Traunmüller and A. Ericsson, "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**(6), 3438–3451 (2000).

[8] D. Pelegrin-Garcia, M. Rychtáriková, C. Glorieux, and B. F. G. Katz, "Interactive auralization of self-generated oral sounds in virtual acoustic environments for research in human echolocation," in *Proceedings of Forum Acusticum 2014* (2014).

[9] The quantities are averaged for the two ears and across the bands of 125 Hz to 4 kHz.

[10] I. R. Cushing, F. F. Li, T. J. Cox, K. Worrall, and T. Jackson, "Vocal effort levels in anechoic conditions," Appl. Acoust. **72**, 695–701 (2011).

[11] W. Navidi, *Statistics for Engineers and Scientists*, 3rd ed. (McGraw-Hill, New York, 2009), pp. 322–329.

[12] D. Pelegrin-Garcia, O. Fuentes-Mendizabal, J. Brunskog, and C.-H. Jeong, "Equal autophonic level curves under different room acoustic conditions," J. Acoust. Soc. Am. **130**, 228–238 (2011).

[13] J. Brunskog, A. Gade, G. P. Ballester, and L. R. Calbo, "Increase in voice level and speaker comfort in lecture rooms," J. Acoust. Soc. Am. **125**, 2072–2082 (2009).

[14] D. Pelegrín-García, "Comment on 'Increase in voice level and speaker comfort in lecture rooms' [J. Acoust. Soc. Am. **125**, 2072–2082 (2009)]," J. Acoust. Soc. Am. **129**, 1161–1164 (2011).

[15] D. Pelegrín-García, B. Smits, J. Brunskog, and C. Jeong, "Vocal effort with changing talker-to-listener distance in different acoustic environments," J. Acoust. Soc. Am. **129**(4), 1981–1990 (2011).

[16] D. Pelegrin-Garcia, V. Lyberg-Åhlander, R. Rydell, J. Brunskog, and A. Lofqvist, "Influence of classroom acoustics on the voice levels of teachers with and without voice problems: A field study," Proc. Mtgs. Acoust. **11**, 060001 (2010).