inter•noise 2009
innovations in practical noise control
2009 August 23-26
Ottawa, Canada

Prediction of vocal effort and speakers' comfort in lecture rooms

David Pelegrín-García[a]
Jonas Brunskog[b]
Acoustic Technology Group
Technical University of Denmark
Kgs. Lyngby, DK-2800 Denmark

ABSTRACT

Vocal health is a major concern among teachers, and represents a significant cost for school authorities. The physical environment of the teaching room, and not only the background noise, plays a very important role in determining the average voice power levels at which the teachers do speak. When designing the acoustics of a classroom, it is necessary to predict quantitatively the vocal effort that the teacher will experience and the regulations should define an acceptable range for a measure that predicts it. Different measures, including a new magnitude called 'Room Gain', are proposed to predict the vocal effort while talking. The vocal effort is characterized here by the increase of voice power level experienced by a speaker talking in a classroom relative to a reference situation. The results of blind tests in an auditory virtual environment with different simulated classrooms show that speakers lower their voices when the room gain increases.

1. INTRODUCTION

During the last decades there has been growing awareness and effort toward improving working conditions. This has resulted in an improved understanding of which are the best ways of working that lead to maximum productivity and worker well-being. The fields of occupational health and safety (OHS) and ergonomics have acquired more importance in society and among professionals. An increasing number of studies have been carried out in many different areas where humans interact with the working environment.

Teachers are a group of professional voice users. Their voice is their primary working tool and it is exposed to a high number of risk factors.[1] In fact, the prevalence of vocal problems among teachers is higher than in other professions.[2] Some authors claim that, in average, each

[a] Email address: dpg@elektro.dtu.dk
[b] Email address: jbr@elektro.dtu.dk

teacher has several days of sick-leave every year due to voice problems, which can represent a significant cost for the state[3, 4].

According to the IEA (International Ergonomics Association), the term ergonomics refers to the "scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance".[5]

This definition is very convenient to illustrate the problem described in the present paper, where the interaction between the voice of a speaker and the physical environment is studied, in an attempt to quantify the magnitude of the interaction. Moreover, these kind of experiments must be seen under the context of the emerging field of voice ergonomics, term introduced by the Finnish phoniatrician Eeva Sala[6] to talk about the study and action on all the factors that enhance performance in speech communication, decrease risks for voice disorders and enable recovery from a voice disorder[7].

In the field of classroom acoustics, much work has been done to optimize the acoustic conditions from the point of view of the listener. Thus, there have been many studies reporting background noise levels, reverberation time[8-10] and also linking background noise and different metrics of speech intelligibility with actual performance[11, 12]. However, few studies have attempted to quantify the effect of room acoustics on the produced voice. Kob et al.[13] reported that good classroom acoustics have a beneficial effect on teachers with some kind of voice disorders.

A previous paper,[14] used as the basis for the present study, described a series of measurements of teaching sessions in real rooms, where the voice of the teachers was recorded, the objective acoustic parameters of the room were measured and the subjective impressions of the speakers were assessed through questionnaires. A significant relationship between the physical properties of a room and the voice power level used there was found. The change in sound power level (SWL) from a speaker was found to be dependent on the logarithm of the room volume and on the room gain (G_{RG}), i.e. the gain provided by the room to one's own voice. However, it is not possible to draw strong conclusions from the study, because it was performed with just 6 subjects in 6 different rooms.

For the experiments described in this paper, the interest was focused in studying the possible effect of auditory cues, independently of visual information associated to the room and to the distribution of the audience. Therefore, an auditory virtual environment was developed in order to simulate many different acoustic environments and study how the vocal behavior changes with modified auditory feedback, keeping all the other variables (e.g. visual impression, room temperature, air quality) steady.

2. EXPERIMENTAL METHOD

A. Setup Overview

The auditory virtual environment developed for these experiments consisted of 29 loudspeakers placed in a quasi-sphere around a subject in a highly damped room. The speech signal from the subject in the center is picked with a headworn microphone, convolved in real time with the impulse response (IR) of the environment, and recorded for analysis. Further details of the system and of its adaptation for real-time operation can be found in the literature[15, 16].

A block diagram of the system is shown in Figure 1. Here, the IR loaded into the convolution software *jconv* contains all the reflections that a room produces at the talker's ears, with information about their delay, incidence angle and spectral level distribution. This requires the computer modeling of the desired room and the calculation of the different transmission paths with a room acoustics simulation software (Odeon). This IR has to be convolved with the on-axis, far-field speech signal. However, the speech signal is picked at the middle of the cheek. This requires the use of a near-to-far field filter to correct the biased spectral distribution of the picked signal.[17]

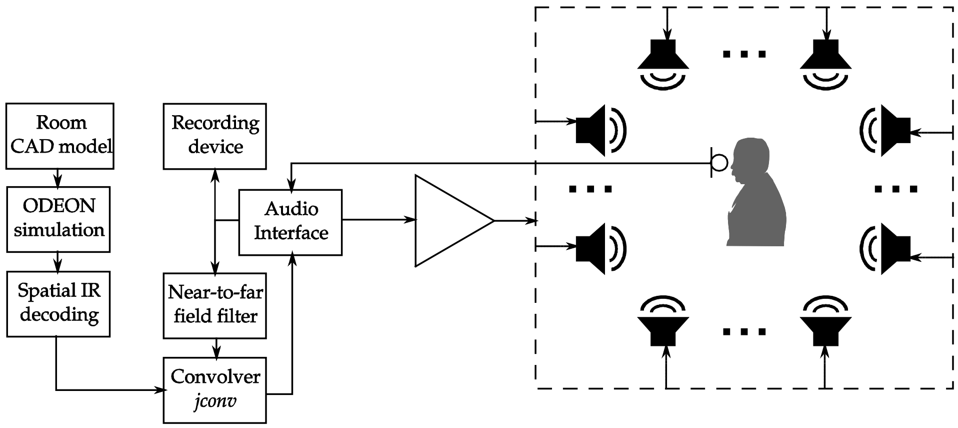


Figure 1: Block diagram of the auditory virtual environment used in the experiments.

During the experiments, a subject was seated at the center of the test room and was asked to wear a blindfold. Then, the simulation of a room started. The subject was instructed to give a lecture of 3 minutes to a group of 30 students, which they had to imagine.

B. Subjects

There were 5 subjects who participated in the experiments with age ranging 23-35. They were 3 fellow students and 2 teachers at the research group who had good hearing and vocal health at the time of the experiments.

C. Rooms

Each subject had to talk in 5 different situations for 3 minutes: four simulated rooms with different volume and acoustic features (see Figure 2) and the bare test room. The presentation order was random for each subject. R1 and R2 were a small meeting room, without and with some absorptive porous material (α_W=0.65) surrounding the speaker, respectively. R3 was a medium-sized lecture room with absorption in the back part and R4 was a large auditorium. These environments were chosen because they offered a big variation in a number of parameters (volume and acoustic measures).

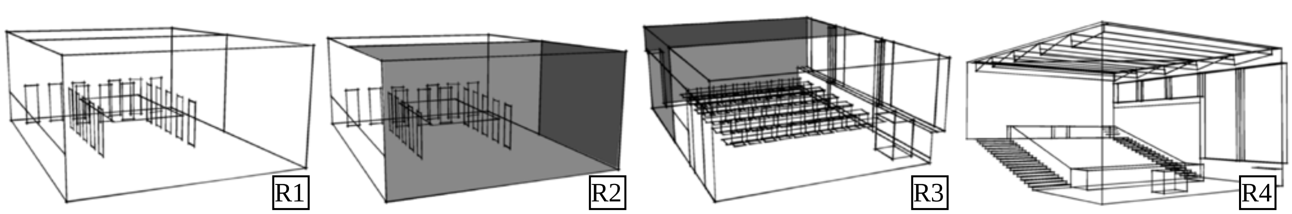


Figure 2: Four different simulated rooms.

E. Measurement of Objective Parameters

Objective parameters (support, room gain and reverberation time) were obtained from the IR measured in the test room, while the simulated environments were running, using the setup shown in Figure 3. A head and torso simulator B&K type 4128 with right ear simulator type 4158 and left ear simulator type 4159 was used. Additionally, a B&K NEXUS signal conditioner was used as the interface between the microphones and the internal sound card of the computer. The IR between the dummy mouth and the microphones at the ears was measured using room acoustic measurement software DIRAC. In all cases, results from the measured IR at the 2 ears were averaged.

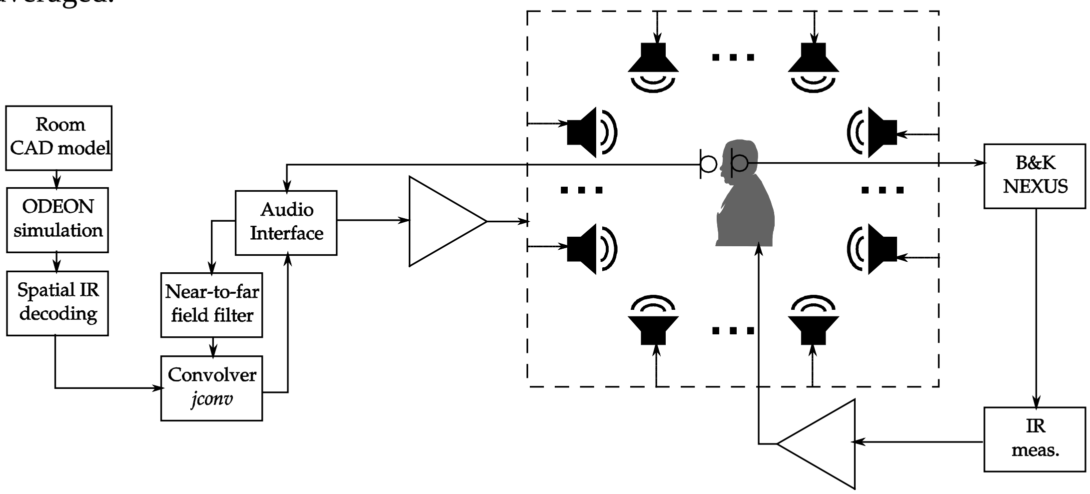


Figure 3: Setup used for measuring objective acoustic parameters.

The IR files $h(t)$ were analyzed with MATLAB. A window $w_d(t)$, of 5 ms duration, was applied to the signal in order to extract the component corresponding to the direct propagation path. Additional windows $w_r(t)$ of different duration (25, 45, 75, 95 ms and 1 s) were applied to the signal in order to extract the reflected component arriving to the ears at different time intervals. An illustration of the signal and the windows is shown in Figure 4. Only the window of 25 ms duration for the reflected component is shown.

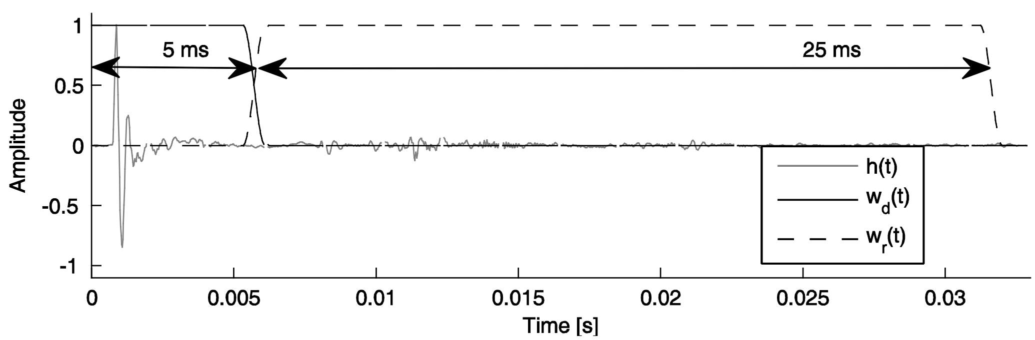


Figure 4: Example of measured IR and windowing applied to extract direct and reflected components.

The windowed signals $h(t)w_d(t)$ and $h(t)w_r(t)$ were filtered using one-octave bandpass filters with center frequencies between 125 Hz and 4 kHz, and also with two-octave bandpass filters in the same frequency region. These filters are generically called here $h_f(t)$. Thus, the energy levels $L_{E,dir}$ and $L_{E,ref}$, for the direct and the reflected components, respectively, are:

$$L_{E,dir} = 10\log\int_0^{\infty}\left[\left(h(t)w_d(t)\right)*h_f(t)\right]^2 \mathrm{d}t \tag{1}$$

$$L_{E,ref} = 10\log\int_0^{\infty}\left[\left(h(t)w_r(t)\right)*h_f(t)\right]^2 \mathrm{d}t \tag{2}$$

Furthermore, the total energy level L_E after filtering the IR is:

$$L_E = 10\log\int_0^{\infty}\left[h(t)*h_f(t)\right]^2 \mathrm{d}t \tag{3}$$

No reference value is used here, because the absolute value of these energy levels is not of concern, but only the difference between values of total, direct and reflected parts. It must be noted that the energy level of the direct sound is calculated properly when there are no reflections coming from surfaces closer than 1 m, approximately. If this condition is fulfilled, the energy level of the direct sound should be the same measured at any place.

The reverberation time T_{30} is calculated from the reflected component of the IR (excluding the direct sound.

Room Gain

According to Brunskog et al.[14], the room gain G_{RG} is the degree of amplification provided by the room to the one's own voice, disregarding the bone conduction.

$$G_{RG} = L_E - L_{E,dir} \tag{4}$$

Support

One of the problems of the previous definition of Room Gain is the small dynamic range of the measured values, because the energy of the IR is mostly concentrated in the direct path. An alternative measure is the use of Support (*ST*) introduced by Gade[18] for the study of stage acoustics in concert halls. In this case, however, the measurement method is changed and the IR used to evaluate the parameter is measured between the mouth and the ears of a dummy head. The *ST* is defined as the difference between reflected and direct energy levels.

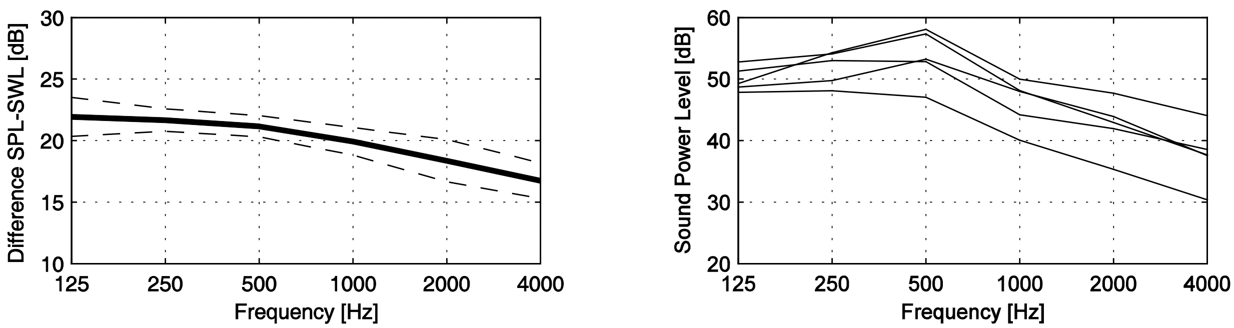
$$ST = L_{E,ref} - L_{E,dir} \approx 10\log\left(10^{G_{RG}/10} - 1\right) \tag{5}$$

F. Voice analysis

As stated above, the speech signals from all the subjects are recorded. The signal is first edited in MATLAB to remove automatically silences of long duration. Then, the mean sound pressure level (SPL) picked by the microphone is calculated for each subject and environment in the 6 octave frequency bands between 125 Hz and 4 kHz.

G. SPL to SWL

The sound power level (SWL) is used here to characterize the rate of energy radiated from a source. It is assumed that SWL is correlated with the vocal effort. Speech SWL is determined by performing sound power measurements in a reverberation room according to ISO 3741[19] and recording simultaneously the speech signal with the headworn microphone. The relationship between the mean SPL recorded close to the mouth and the SWL is shown in Figure 5a and Table 1. The individual values of speech SWL are shown in Figure 5b.



(a) (b)

Figure 5: (a) Difference between SPL close to the mouth and SWL. Bold line: mean value. Dashed lines: ±σ around the mean value. (b) Voice SWL measured for different individuals.

Table 1: Difference between SPL close to the mouth and SWL, and standard deviation.

	125	250	500	1000	2000	4000
Mean L_P-L_W (dB)	21.9	21.7	21.2	19.9	18.4	16.7
σ (dB)	1.58	0.90	0.86	1.11	1.71	1.42

3. RESULTS

A. Objective parameters

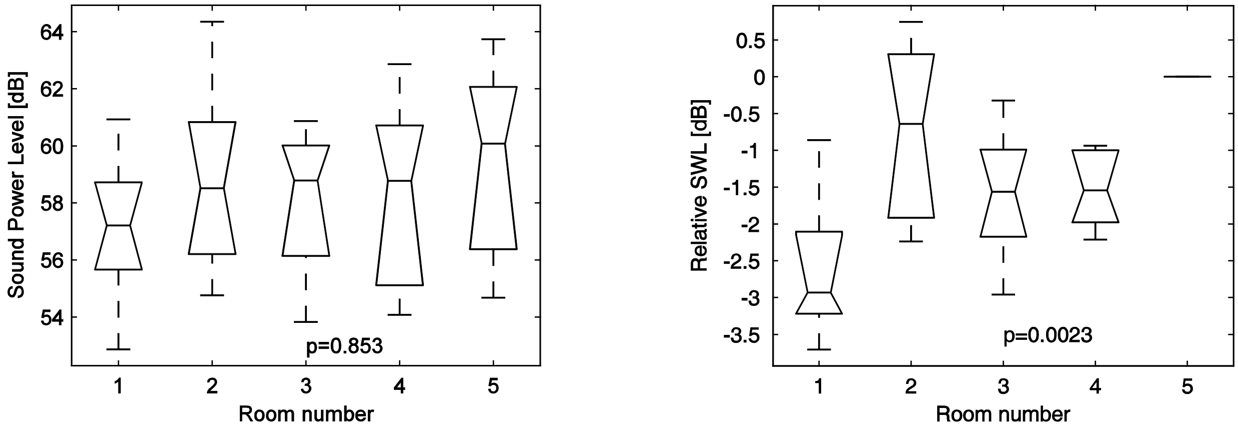
The measured parameters in the different environments, together with the volume of the simulated rooms (R1-R4) and the test room (R5) are presented in Table 2. The chosen frequency weightings (250-500 Hz for G_{RG} and *ST*, and 1 kHz for T_{30}) and the windowing of 30 ms for the *ST* are the ones that provide highest correlation with voice SWL changes.

Table 2: Room volumes and most important measured acoustic parameters

	V (m^3)	$ST_{30,\ 250\text{-}500}$ (dB)	$ST_{,\ 250\text{-}500}$ (dB)	$G_{RG,\ 250\text{-}500}$ (dB)	$T_{30,\ 1k}$ (s)
R1	130	-15.0	-12.0	0.26	1.03
R2	130	-15.4	-14.7	0.15	0.41
R3	344	-15.3	-14.2	0.16	0.71
R4	1174	-15.3	-14.2	0.16	1.24
R5	72	-15.5	-15.4	0.12	0.12

B. Speech levels

The measured global speech SWLs in each of the environments are shown in a box plot in Figure 6a and listed in Table 3. There is a big variability in the measured data, and thus the changes in mean values are not statistically significant at the 5% level (p=0.853). In order to reduce the variability among subjects, the SWL for each subject at each environment is subtracted from the SWL in environment 5 for that subject. The relative SWL (ΔL_W) values are shown in Figure 6b and in Table 3. With this modification, the mean values are statistically significant at the 5% level (p=0.0023). Thus, following relationships with other parameters are predictors of the mean value. The relative SWL spectra in each of the 5 environments, averaged for the 5 subjects, is shown in Figure 7a.

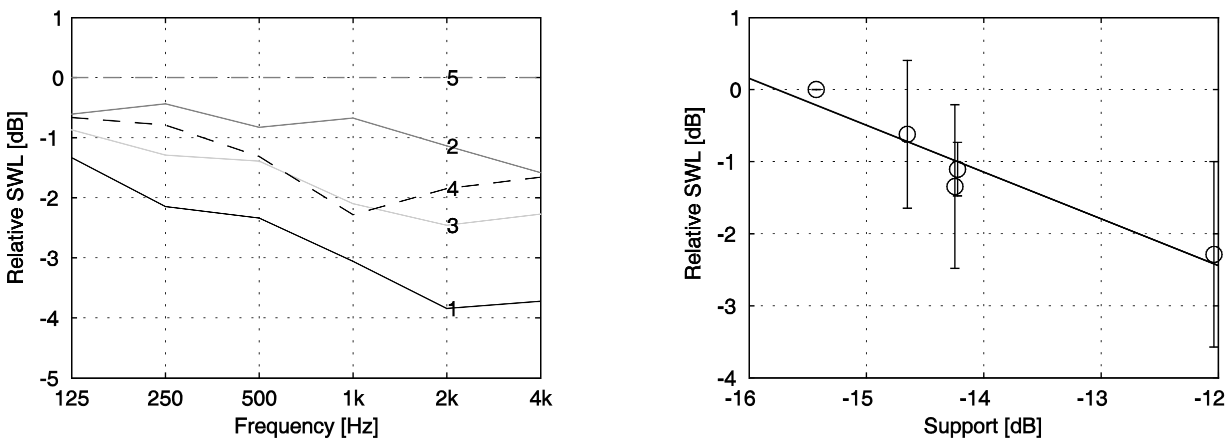


(a) (b)

Figure 6 (a) Box plot of mean speech SWL in different rooms. (b) Box plot of mean relative speech SWL in different rooms, where the speech SWL in room 5 for each subject is used as the reference. The extent of the boxes corresponds to the 25%-75% percentile values, its center line to the median value and the whisker lines to the total range of responses.

Table 3: Mean overall speech SWL and relative SWL in the different environments

	R1	R2	R3	R4	R5
L_W (dB)	57.1	58.8	58.0	58.2	59.4
ΔL_W (dB)	-2.2	-0.6	-1.3	-1.1	-



(a) (b)

Figure 7: (a) Relative speech SWL spectra in different environments. (b) Measured support vs relative speech SWL and least squares regression line.

C. Relationship between speech levels and objective parameters

The correlation between ΔL_W and the different objective parameters, with different frequency weightings and time windows, was studied. ΔL_W has not a significant correlation with the room volume or its logarithm ($R^2 < 0.25$), but it does with T_{30} (R^2=0.81) . It was found that the correlation between ΔL_W and ST (R^2=0.92) or G_{RG} (R^2=0.87) was strongest when a 2-octave frequency weighting (250-500 Hz) was applied to both the speech signal and the IR. Even strongest relationship was found when the time window applied to calculate the reflected energy in the support was until 30 ms after the direct sound (R^2=0.97). A single parameter linear regression model that predicts the mean ΔL_W in the frequency range 250-500 Hz is given by

$$\Delta L_W = -10.2 - 0.65ST\ . \quad (6)$$

Here, ST is used instead of G_{RG} and the 30 ms windowed version of the support, ST_{30}, due to the higher dynamic range of measured values (-15.4 dB ~ -12.0 dB).

4. DISCUSSION

As seen in Figure 6a, there is a big variation in measured SWL values and no value is significantly different from any other, as all the boxes overlap. By referring all measured values to the values measured at room 5 (see Figure 6b), it is possible to observe significant changes in SWL among different environments. One can discuss that is not totally correct to include the reference in the analysis, as can be seen from the zero-deviation in the relative SWL in room 5. This would reduce the significance (p-values) of the measurements. Therefore, the analysis exposed here must be seen as a description of observed trends, but not a clear relationship. However, there are still observable effects. For example, the relative SWL on room 1 is different from that on room 2, as the boxes do not overlap.

The magnitude of change in SWL (2.2 dB) is less than that found by Brunskog et al. (4.3 dB). There is a trend that indicates an average increase of SWL in the environments that provide less support to one's own voice. This is in agreement with Brunskog et al. who, however, used G_{RG} to describe this phenomenon.

Equation (6) indicates that the mean SWL decreases 0.65 dB when the support is increased 1 dB. Similar slopes have been found for sidetone compensation[20], i.e. the variation in mean SWL when the direct sound transmission path is artificially altered, without taking into account the room effect. However, the situation differs considerably from the present study. Here, the variations in total SPL at the ears are less than 0.5 dB, as suggested by the measured values of G_{RG}, whereas in the other studies, the SPL at ears was artificially modified by several dBs.

This observation leads to, at least, two different explanations. The first one is that, the perceived loudness of one's own voice is modified similarly here and in Lane and Tranel[20], and therefore the measured effects are similar. It should be necessary, however, to study more in detail in which way the bone conducted sound modifies the loudness of the perceived one's own voice. This component has been neglected in the present study. The second explanation is that, by listening oneself, it is possible to get some knowledge about the physical properties of the environment and adjust consequently the voice SWL, in order to provide the listeners a comfortable level.

In both cases, it seems possible to explain behavioral changes in the speakers' voice through the measurement of ST or G_{RG}. However, it is necessary to carry out more measurements in different environments with more subjects to draw stronger conclusions.

5. CONCLUSIONS

The present paper describes an attempt to reproduce in laboratory the measurements started by Brunskog et al[14]. Similar trends are observed, though the measured variation in SWL and G_{RG} differ substantially.

The use of simulated environments can reproduce some of the effects observed in real rooms. Making use of the flexibility that the laboratory setup can provide, it is necessary to perform experiments with more subjects and more environments in order to elaborate a model that describes the average behavior of a representative population group.

If ST or G_{RG} can describe a change in vocal behavior, as presented here, its inclusion on design guidelines regarding classrooms acoustics should be considered. Furthermore, OHS authorities should support acoustic improvements in schools, considering that a good acoustic design in a classroom is a preventive actuation to care about teachers' primary working tool: their voice.

ACKNOWLEDGMENTS

The present paper describes a part of a PhD project which is funded by the Swedish company AFA Försäkring. The authors want to thank all the colleagues who have taken part in the experiment.

REFERENCES

[1] V. Jonsdottir, *The voice: An occupational tool. A study of teachers' classroom speech and the effects of amplification.* PhD thesis, University of Tampere and University of Oulu (2003).

[2] N. Roy, R.M. Merrill and S. Thibeault, "Prevalence of voice disorders in teachers and the general population," *Journal of Speech, Language & Hearing Research* **47**(2), 281-293 (2004).

[3] K. Verdolini and L. O'Ramig, "Review: Occupational risks for voice problems," *Logopedics Phoniatrics Vocology* **26**(1), 37-46 (2001).

[4] D. Lubman and L.C. Sutherland, "Good classroom acoustics is a good investment," *17th ICA Proceedings vol. V*, Rome, pp.138-139 (2001).

[5] IEA - The International Ergonomics Association. Available from: http://www.iea.cc/

[6] E. Sala, M. Sihvo and A. Laine, *Röstergonomi – rösten ett fungerande arbetsredskap* (Institutet för arbetshygien, Arbetarskyddscentralen, Helsinki, 2005).

[7] V. Jonsdottir and L. Rantala, "Nordic cooperation in the field of voice ergonomics," *Proceedings of the first nordic conference of voice ergonomics and treatment,* Helsinki, pp. 8-17 (2009).

[8] H.A. Knecht, P.B. Nelson and G.M. Whitelaw, "Background noise levels and reverberation times in unoccupied classrooms: Predictions and measurements," *American Journal of Audiology* **65**(2), 65-71 (2002).

[9] M. Hodgson, "Experimental investigation of the acoustical characteristics of university classrooms," *J. Acoust. Soc. Am.* **106**(4), 1810-1819 (1999).

[10] M.R. Serra and E.C. Biassoni, "Urban noise and classroom acoustical conditions in the teaching-learning process," *Int. J. Environ. Stud.* **56**(1), 41-59 (1998).

[11] S.R. Bistafa and J.S. Bradley, "Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics," *J. Acoust. Soc. Am.* **107**(2), 861-875 (2000).

[12] W. Yang and J.S. Bradley, "Effects of room acoustics on the intelligibility of speech in classrooms for young children," *J. Acoust. Soc. Am.* **125**(2), 922-933 (2009).

[13] M. Kob, G. Behler, A. Kamprolf, O. Goldschmidt and C. Neuschaefer-Rube, "Experimental investigations of the influence of room acoustics on the teacher's voice," *Acoustical Science and Technology* **29**(1), 86-94 (2008).

[14] J. Brunskog, A.C. Gade, G. Payà-Ballester and L. Reig-Calbo, "Increase in voice level and speaker comfort in lecture rooms," *J. Acoust. Soc. Am.* **125**(4), 2072-2082 (2009).

[15] S. Favrot and J. Bucholz, "LoRA – A loudspeaker-based room auralization system," *Acta Acustica*, submitted (2009).

[16] D. Pelegrín-García and J. Brunskog, "Development of an auditory virtual environment to measure the speakers comfort and increase of voice power levels in lecture rooms," *Proceedings of the first nordic conference of voice ergonomics and treatment*, Helsinki, pp. 38-49 (2009).

[17] C. Pörschmann, "One's own voice in auditory virtual environments," *Acustica – Acta Acustica* **87**(3), 378-388 (2001).

[18] A.C. Gade, "Investigations of musicians' room acoustic conditions in concert halls. I. methods and laboratory experiments," *Acustica* **69**(5), 193-203 (1989).

[19] International Standard Organization, *ISO-3741. Acoustics - Determination of sound power levels of noise sources using sound pressure: precision method for reverberation rooms.* 3rd edition (1999).

[20] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," *J. Speech Hear. Res.* **14**, 677-709 (1971).