Publications from

Sound Environment Center at Lund University

No. 10

Final report of the project

Speakers comfort and voice disorders in classrooms

Core project team:
Jonas Brunskog[1]
Viveka Lyberg Åhlander[2]
Anders Löfqvist[2]
David Pelegrín-García[1]
Roland Rydell[2]

[1] Acoustic Technology Group
Department of Electrical Engineering
Technical University of Denmark

[2] Voice Research Group
Department of Logopedics, Phoniatrics and Audiology
Lund University

Sound Environment Center at Lund University - Department of Logopediscs, Phoniatrics and Audiology, Lund University - Technical University of Denmark
2011

Publisher: Sound Environment Centre, Lund University

e-ISBN	978-91-89415-30-0
DOI	https://doi.org/10.37852/oblu.143
ISSN	1653-9354

Publications from the Sound Environment Centre at Lund University Report no. 10

Printed in Sweden by Media-Tryck, Lund University, Lund 2011

(Citation, e.g.: Brunskog, J., Lyberg Åhlander, V., Löfqvist, A., Garcia, D. P., & Rydell, R. (2011). Speakers Comfort and voice disorders in classrooms. (Publications from Sound Environment Center at Lund University; Vol. 10). Sound Environment Center at Lund university, DOI: https://doi.org/10.37852/oblu.143)

Information about the Sound Environment Centre, Lund University, can be found here: https://www.lmc.lu.se/

Final report of the project
'Speakers comfort and voice disorders in classrooms'

Core project team:
Jonas Brunskog[1], PI
Viveka Lyberg Åhlander[2]
Anders Löfqvist[2]
David Pelegrín-García[1]
Roland Rydell[2]

[1]Acoustic Technology Group
Department of Electrical Engineering
Technical University of Denmark

[2]Voice Research Group
Department of Logopedics, Phoniatrics and Audiology
Lund University

Contents

Appendix: Papers

1 Introduction

An overall aim of the project has been to investigate the voice use of teachers in relation to the acoustic properties of the classroom, and to study whether speakers take into account auditory cues to regulate their voice levels, even in the absence of background noise. The most common means of communication in a classroom is speaking and listening. The teacher's voice is thus the tool for communicating with the students. The room acoustics in the classroom is the communication channel from the speaker to the listener. It affects the quality of the speech signal and thus the ability to understand what the teacher says.

During the last decades, an increasing focus has been put on teachers' voice and the consequences of vocal problems. A study from the mid 90's on voice and occupations in Sweden identified teachers as the most common occupational group at voice clinics, based on the percentage of the total number of teachers in the population at that time. The prevalence of voice problems in Swedish teachers is, however, largely a substantial number of unrecorded cases since teachers rarely seem to seek help for their voice problems. Voice difficulties at work seem to be regarded as more of an individual problem – depending on the individual's innate capacities or voice use or "abuse" – than as an occupational hazard. It has been estimated that the yearly costs for sick-days and treatment in US teachers amount to US$2, 5 billion.

There have been many studies trying to optimize the acoustical conditions for the students, in terms of measures of the speech intelligibility, signal-to-noise ratios, or reverberation time. Most of these studies have focused on the listener, but it has also been pointed out that a low reverberation time may affect teachers' voice.

In a pre-study of the present project, Brunskog et al. (2009), studied the classroom acoustics from the point of view of the speaker, and thus tried to relate the voice production process with different measurable parameters of the classroom, including the size of the room, acoustical parameters, and background noise. It was shown that the voice power used is related to the volume of the room and to the support, or room gain, provided at the position of the speaker.

In the field of voice therapy and phoniatrics, teachers' voice health problems are of major concern, not only due to the required clinical assistance, but also due to the financial impact that the teachers' absence produces in the overall budget of the country. There is a consensus that voice load is an important factor for voice problems, resulting from higher fundamental frequency (F0) and higher sound pressure level (related to the voice power).

One of the core concepts in this project is "speakers' comfort" that is tied to the voice use and the speaker's subjective perception of the voice. It is defined as the subjective impression that talkers have when they feel that their vocal message reaches the listener effectively [with no or low vocal effort]. In this subjective impression, experienced while hearing and perceiving one's own voice, some attributes play important roles: the voice-support provided by the room and the speech intelligibility along with the sensory-motor feedback from the phonatory apparatus.

Although much is known today about teachers' voices and voice use, only a few studies have taken into account the teachers' ratings of their work-environment in relation to their voice. Even fewer have explored the teachers' voice us in the work environment. Further, the work environment, *i.e.*

the classroom's air-quality and acoustics, has often been discussed and acknowledged to contribute to the vocal load, but these factors have not been very much investigated with the teacher *in action*. One purpose of the present project was thus to investigate the voices and the voice use of teaching staffs in their teaching environment and to explore the prevalence of voice problems in Swedish teachers. A second purpose was to explore the teachers' ratings of aspects of their working environment that can be presumed to affect vocal behavior and voice and to measure the teachers' voice use in relation to some of those factors. One more purpose was to clinically assess the voice function in the teachers with self-rated voice problems and compare it to their vocally healthy colleagues. To be able to do comparisons between the teachers, one further objective was to develop and assess a self-rating instrument for the rating of throat-related problems in relation to voice. The purpose was also to develop room acoustic measures related to the voice regulation, and to understand the physical parameters influencing the voice regulation. Finally, the knowledge built up in the project should be used to set up recommendations and design criteria for good speaking environments.

The original subprojects have all been carried out. The studies and subprojects of the project can be summarized as follows (the papers are included in the appendix to this report):

Voice Handicap Index – throat (A1)

- Lyberg-Åhlander V, Rydell R, Eriksson J, Schalén L. (2010)[1], Throat related symptoms and voice: development of an instrument for self assessment of throat problems. *BMC Ear, Nose and Throat Disorders*, 2010, 10:5. DOI: 10.1186/1472-6815-10-5.

Prevalence of voice problems (A2)

- Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A., (2010), Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff. *Journal of Voice,* in press. Corrected proof, available online 26 March 2010. DOI: 10.1016/j.jvoice.2009.12.006

Etiology of voice problems (A3)

- Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A. (2011), How do teachers with self-reported voice problems differ from their peers with self-reported voice health? Manuscript submitted for publication.

Voice level and speaker comfort in real rooms (B1)

- Brunskog, J., Gade, A.C., Payà-Ballester, G.; Reig-Calbo, L. (2009) [1], Increase in voice level and speaker comfort in lecture rooms. *Journal of the Acoustical Society of America,* **125**, 2072-2083.
- Pelegrín-García, D. (2011), Comment on "Increase in voice level and speaker comfort in lecture room"'. *Journal of the Acoustical Socety of America,* **129**, 1161-1164.
- Pelegrín-García, D., Smits, B., Brunskog, J, Jeong, C.-H. (2011), Vocal effort with changing talker-to-listener distance in different acoustic environments. *Journal of the Acoustical Socety of America,* **129**, 1981-1990.

---

[1] These papers are pre-studies, but have been finished within the project period.

The virtual environment (B2)

- Pelegrín-Garcia, D, Brunskog, J. (2011), Loudspeaker-based system for real-time own-voice auralization. Manuscript.
- Pelegrín-Garcia, D, Brunskog, J. (2010), Natural variations of vocal effort and comfort in simulated acoustic environments. *Proceedings of EAA Euroregio 2010*, Ljubljana, Slovenia
- Bottalico, P., Pelegrín-Garcia, D., Astolfi, A., and Brunskog, J. (2010), Measurement of vocal doses in virtual classrooms. *Proceedings of Internoise 2010*, Lisbon, Portugal
- Brunskog, J., and Pelegrín García, D. (2010), Speaking comfort and voice use of teachers in classrooms. *Italian Journal of Acoustics*, 34, 51-56.

Loudness of one's own voice (B3)

- Pelegrín-García, D., Fuentes-Mendizabal, O., Brunskog, J, and Jeong, C.H. (2011), Equal autophonic level curves under different room acoustics conditions. Manuscript submitted for publication.

Field study of voice use (C)

- Pelegrín-García, D., Lyberg-Åhlander, V., Rydell, R., Löfqvist, A. & Brunskog, J., (2010), Influence of Classroom Acoustics on the Voice Levels of Teachers With and Without Voice Problems: A Field Study. *Proceedings of Meetings on Acoustics,* Vol. 11, ASA.
- Lyberg-Åhlander, V., Pelegrin-Garcia, D., Rydell, R. and Löfqvist, A. (2011), Teacher's Voice Use in Teaching Environments: A Field Study Using Ambulatory Phonation Monitor (APM). Manuscript submitted for publication.
- Pelegrin-Garcia, D., Brunskog, J., Lyberg-Åhlander, V. & Löfqvist, A. (2011), Measurement and prediction of acoustic conditions for a talker in school classrooms. Manuscript.

The work of the project will also result in 2 PhD theses, of which one is published at the time of writing: Viveka Lyberg Åhlander, *Voice use in teaching environments Speakers' comfort* Lund university (2011).

The report has the following structure: Some of the methods being developed and used in the subprojects are first briefly described in chapter 2. The main results of the subprojects are then summarized in chapter 3. The findings within the project are then discussed in chapter 4. The major conclusions are given in chapter 5. Finally, the publications and other ways of spreading/implementing the findings of the project are described in chapter 6.

The project has been done in close cooperation between the Acoustic Technology group at the Department of Electrical Engineering, at the Technical University of Denmark (DTU), which is responsible for the technical-acoustic experiments and analyzing them, and the Voice Research Group at the Department of Logopedics, Phoniatrics and Audiology, Lund University (LU), Sweden. The project has been operating through the Sound Environmental Center (Ljudmiljöcentrum) at Lund University.

Informed, written consent was obtained from all subjects and all headmasters of the schools included. The protocols have been approved by the Institutional Review Board Lund University (No LU 366-01) and by the Regional Review Board (#248/2008).

2 Method overview

During the course of the project, several methodological issues have been considered. Some of them are briefly described here.

2.1 Prevalence of voice problems (A2)

In Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A., (2010), an epidemiology study, a screening questionnaire was developed to assess teachers' ratings of their working environment and also to estimate the prevalence of voice problems in teachers. The questionnaire covered fifty-two items in three main domains:1) background information; 2) room acoustics, perception of noise levels and other issues related to the environment: (items 1-13); and 3) voice problems, vocal behaviour and statements about skills in voice use: (items 14-32). Items in part 1 were answered by yes/no or description in free text. The items in part 2 were statements, e.g., "The air in the classroom is dry", which were rated on a scale from 0 to 4, where 0=completely disagrees and 4= completely agrees. The items in part 3 were statements, e.g., "I have to clear my throat", which were rated on a frequency-based scale from 0 to 4. Two statements were considered to be index-statements: #1:" The classroom acoustics help me talk comfortably" and #32:"I have voice problems". The questionnaire was tested in a pilot study of 63 teachers, all permanent staffs of one high school. A reference group attached to the project (experts in occupational and environmental medicine, voice, acoustics, and representatives of the teachers' unions, and building proprietors) also made comments. The validity of the questions was also discussed by a group of experienced teachers, representing the different teaching levels included in the study. Based on the pilot study and the feedback, the questionnaire was revised into its final form.

The questionnaire was distributed to 487 responders at their collegial meetings. The teachers were accessed via the headmasters of 53 randomly selected schools in the region. The choice of geographical area was based on a uniform distribution of air pollution, and on an equivalent population density. Participation was accepted by 22 schools. The teachers were informed about the study at regular, pre-scheduled, compulsory collegial meetings at each school. The questionnaire was distributed, completed, and collected during one and the same meeting. The teachers completed the questionnaire anonymously. If, however, a teacher was interested in continued participation in the project, contact information was obtained on a voluntary basis. All teachers participating at the conferences answered the questionnaire. Visits to distribute and collect the questionnaire were mainly made from January to April 2009. The questionnaire was completed by 73% of all the teachers of all the included schools. Nine of the questionnaires were excluded due to incomplete data. Further, eleven questionnaires were excluded since they had mistakenly been given to teacher-students who had participated in the collegial meetings where the questionnaire was distributed. Data from a total of 467 responders (336F:131 M, median age 47, range: 23-69) was thus finally evaluated. Teaching staff at all levels were included, except pre-school teachers at pre-schools and day-care-centres and teachers at specialised, vocational high schools, due to the large variety of teaching premises; see (Table 1) for the distribution of teaching levels.

Table 1. Teaching level of 467 teachers

Teaching level	N teachers
Junior+intermediate school	203
Secondary school	108
High school	156

Based on the ratings of statement #32 "I have voice problems", the participants were divided into two groups. *Group I*, (N=60) consisted of teachers suffering from voice problems sometimes, often, or always. *Group II* (N=407) included teachers having rated 0-1, *i.e.*, never or only occasionally experiencing voice problems. There were no significant differences between the groups for gender (*Group I* 80% F/20% M, *Group II* 71% F/29% M), age (*Group I* Md=49,5, *Group II* Md=46), smoking (*Group I* 10%, *Group II* 7%), or years of occupation (*Group I* Md=20, *Group II* Md=16), as shown by a chi square test.

2.2 Etiology of voice problems (A3)

The study by Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A. (2011) is prospective, has a case-control design and aimed at investigating the etiology of voice problems in teachers by exploring possible differences between 31 teachers with voice problems and their 31 age and gender matched voice healthy colleagues. All participants were recruited among the population of teachers from study A2. Planned continuation of the project was explained and 220 of the teachers were interested in further participation: n=41 who had rated themselves as suffering from voice problems and n=179 who had estimated no voice problems in study A2. The teachers with voice problems were matched for age and gender to voice healthy colleagues from the same schools. Ten subjects with voice problems were excluded: one due to lack of any control at his school; two smoking subjects since it was not possible to find a gender- and age matched smoking control at the school; one subject was not possible to reach and six subjects declined to participate due to lack of possibility or interest. Finally, two paired groups of teachers were formed: Group I (N=31, 26F/5M) included teachers with self-assessed voice problems, with a median age of 51 years (range 24-65) and a median time in occupation of 15 years (range 1-40); Group II (N=31, 26F/5M) included teachers without voice problems with a median age of 43 years (range 28-61) and median time in occupation of 14 years (range 2-39). The pairs came from 12 of the 22 schools in study II.

The teachers underwent examination of the larynx and vocal folds with a 70 degree rigid laryngoscope. A digital documentation system was used, HRES Endocam (Wolf, Germany). First, high resolution mode was used for evaluation of organic lesions, adduction and abduction. In high-speed mode 2000 frames/s were recorded for male subjects and 4000 frames/s for female subjects. These recordings were used to evaluate mode and symmetry of vibration at the glottal level. A recording of a read text was used for perceptual evaluation of the voice and for acoustic measurements. In addition, a standard Voice Range Profile was used to examine the range of intensities and fundamental frequencies that a participant could produce.

2.3 Voice level and speaker comfort in real rooms (B1)

For the laboratory experiments, real or simulated (auralizated) rooms can be used. The visual impression of the room cannot easily be included in the auralization. This might be a positive aspect in many cases, but sometimes it is important to get the visual size of the room and the distance to the audience right. Therefore, real rooms have been used for some of the laboratory experiments.

Within the project, new metrics describing the room acoustic conditions for a talker have been introduced. The room acoustic parameters for a talker are related to the possible ways in which his own voice reaches his ears. They require the measurement of the airborne acoustic path between the mouth and the ears, which is characterized by a room impulse response (RIR) $h(t)$. This airborne path has two components: the direct sound, transmitted directly from the mouth to the ears, and the indirect sound, coming from reflections at the boundaries. For this reason, the last component is also referred to as reflected sound. Two parameters are derived from the RIR measurement, using a head and torso simulator (HATS, dummy head), and the relation between the direct and the reflected sound, expressed in the quantities *room gain* and *voice support*. The background of the support measure comes from musical room acoustics, where the concept is used in connection to the stage, and is related to the possibility for the musicians to hear themselves when playing. Room gain was introduced in Brunskog, J., Gade, A.C., Payà-Ballester, G.; Reig-Calbo, L. (2009). The measurement principles were reconsidered and the voice support measure was introduced in Pelegrín-García, D. (2011).

The *room gain* G_{RG} was defined as the degree of amplification provided by the room to one's own voice, disregarding the contribution of the own voice which is transmitted directly through the body. This is the difference between the total energy level in a room and the direct energy level. Originally Brunskog et al. calculate the direct energy level with a RIR measurement in an anechoic environment.

$$G_{RG} = L_E - L_{E,d} \tag{1}$$

In Pelegrin-Garcia (2011), the *voice support* is chosen as an alternative measure for the degree of amplification of a room to one's own voice. In this case, the *voice support* compares the energy level of the reflections $L_{E,r}$ with the energy level of the direct sound, extracted from the impulse response corresponding to the path between the mouth and the ears.

$$ST_V = L_{E,r} - L_{E,d} \approx 10\log\left(10^{G_{RG}/10} - 1\right). \tag{2}$$

The *voice support* ranges from -18 dB to -5 dB in normal rooms, whereas the room gain is limited to a range between 0 dB and less than 2 dB.

The RIR has to be measured with a dummy head that contains a loudspeaker at its mouth, used as source, and microphones at its ears, used as receivers. To ensure a correct separation of the direct and the reflected sound components, it is necessary to place the dummy head more than 1 m from reflecting or scattering surfaces, measured from the center of the mouth. In this way, there is a time gap free of reflections after the arrival of the direct sound. Then, the direct sound component is extracted by applying a window $w_d(t)$, of 5 ms duration, to $h(t)$. The complementary window, $w_r(t) = 1 - w_d(t)$ is applied to $h(t)$ in order to extract the reflected component arriving to the ears. An illustration of the signal and the windows is shown in Figure 1.

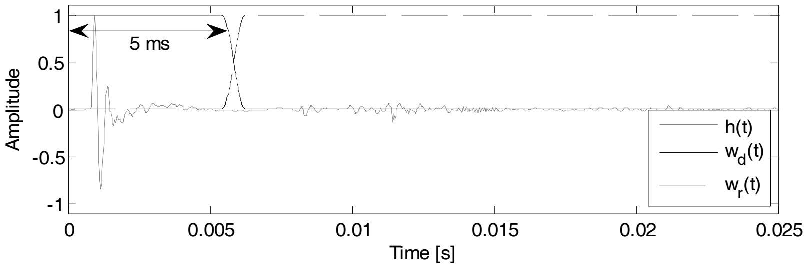

Figure 1: Example of an IR and windowing applied to extract direct and reflected components.

The windowed signals $h(t)w_d(t)$ and $h(t)w_r(t)$ can be filtered using one-octave bandpass filters with center frequencies between 125 Hz and 4 kHz to study the importance of directed and reflected sound in the octave bands of interest in room acoustics. These bandpass filters are here generically called $h_f(t)$. Thus, the energy levels $L_{E,dir}$ and $L_{E,ref}$, for the direct and the reflected components, respectively, are:

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

$$L_{E,ref} = 10\log\int_0^\infty \left[\left(h(t)w_r(t)\right) * h_f(t)\right]^2 \mathrm{d}t \tag{4}$$

The symbol $*$ denotes the mathematical operation ‘convolution’. Furthermore, the total energy level L_E after filtering the IR is:

$$L_E = 10\log\int_0^\infty \left[h(t) * h_f(t)\right]^2 \mathrm{d}t \tag{5}$$

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.

Talkers adjust their vocal effort to communicate at different distances, aiming to compensate for the sound propagation losses. In Pelegrín-García, D., Smits, B., Brunskog, J, Jeong, C.-H. (2011), the speech from thirteen talkers speaking to one listener at four different distances in four different rooms was recorded. The speech signals were processed to calculate measures of vocal intensity, F0, and the relative duration of the phonated segments. For each subject, the experiment was performed in a total of 16 different conditions, resulting from the combination of four distances

(1.5, 3, 6, and 12 m) and four different environments: an anechoic chamber, a lecture hall, a long, narrow corridor, and a reverberation room.

In Brunskog, J., Gade, A.C., Payà-Ballester, G.; Reig-Calbo, L. (2009), both subjective responses and objective measures of the room and of the voice level where collected. The range in the physical parameters of the six rooms of the study were wide, including small meeting and listening rooms; a medium size lecture room; two lager auditoria's, one with high reverberation time and one with low; and a large anechoic room.

Different instructions to the test subjects were used in different experiments. In Brunskog, J., Gade, A.C., Payà-Ballester, G.; Reig-Calbo, L. (2009) each of the speakers held a short lecture (about 5 minutes). A map test was used in Pelegrín-García, D., Smits, B., Brunskog, J, Jeong, C.-H. (2011). The talkers were given a map which contained roughly a dozen of labeled items (e.g. "diamond mine", "fast flowing river", and "desert"), starting and ending point marks, and a path connecting these two points. They were instructed to describe the route between the starting point and the finishing point, indicating the items along the path (e.g., "go to the west until you find the harbor"), while trying to maintain eye-contact with the talker. There were sixteen maps in total, and a different map was used at each condition. The order of the maps was randomized differently for each subject.

2.4 The virtual environment (B2)

A real-time self-voice auralization system has been developed within the project (Pelegrín-Garcia, D, Brunskog, J., 2011). The room, called SpaceLab, consists 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 room impulse response (RIR) of the environment, and recorded for analysis. As a result, the talker has the impression of being speaking in another room.

A block diagram of the system is shown in Figure 2 left, and in the right is shown a subject being in the room. Here, the RIR (stored in 29 WAV files, one for each loudspeaker) is loaded into the convolution software *jconvolver*. This requires the computer modeling of the desired room and the calculation of the different transmission paths with a room acoustics simulation software (Odeon). The output of Odeon is decoded and encoded in Ambisonics, adjusted to the requirements of the system. An equalizer filter is used to correct the biased spectral distribution of the speech signal at the head worn microphone. The system is implemented so that background noise can be added.

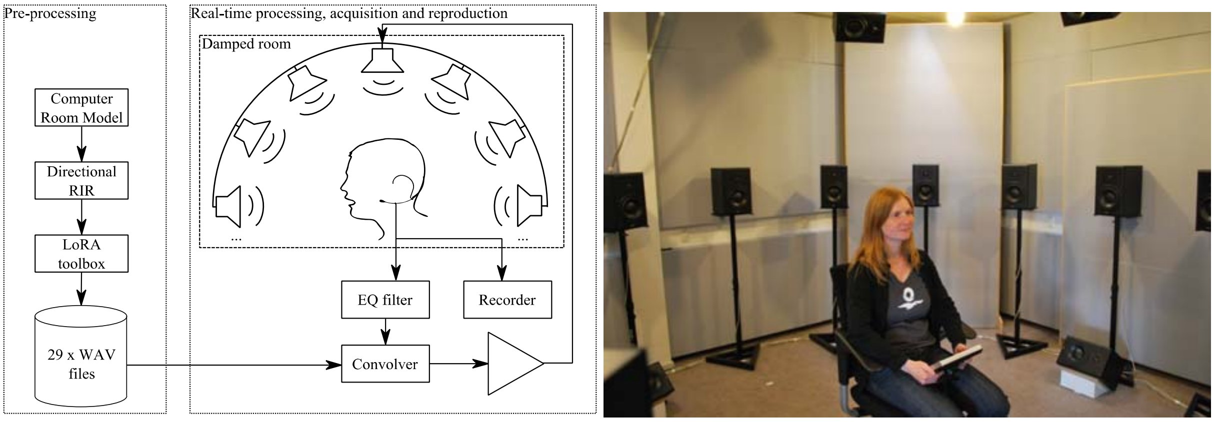

Figure 2: Left: Block diagram of the virtual acoustic system. Right: The auditory virtual environment.

2.5 Loudness of one's own voice (B3)

The loudness with which talkers perceive their own voice is called the autophonic rating. Pelegrín-García, D., Fuentes-Mendizabal, O., Brunskog, J, and Jeong, C.H. (2011) investigated the extent to which room acoustics can alter the autophonic rating and induce Lombard effect-related changes in voice. A reference sound at a constant sound pressure level (SPL) was presented, and the subjects were asked to produce a vocalization (either /a/, /i/, or /u/) with the same loudness as the reference. 14 subjects took part in the experiment. Each subject produced a total of 60 vocalizations that were stored and analyzed to extract the results.

The experimental setup is shown in Fig. 3., which is an alternative earphone implementation to the loudspeaker based auditory virtual environment in study B2. The experiment took place in an anechoic chamber in order to remove all reflections from the room. The indirect auditory feedback was generated by picking the voice from the talker, convolving it with a synthetic impulse response, and playing it back via earphones specially designed to minimize the blocking of direct sound and preserve the usual bone conduction path. The voice of the talker was picked with a microphone located on the cheek at a position 5 cm from the lips' edge in the line between the mouth and the right ear. This signal was sampled using an audio interface, which was connected to a computer running the convolution software *jconvolver* under Linux. The convolution system introduced an overall delay of 11.5 ms between the arrival of the direct sound at the ears and the indirect auditory feedback generated in the convolution process. The resulting signal was again converted into the analog domain and reproduced through the two channels (left and right) of the earphones. Figure 3 right shows the custom earphones used in the experiment.

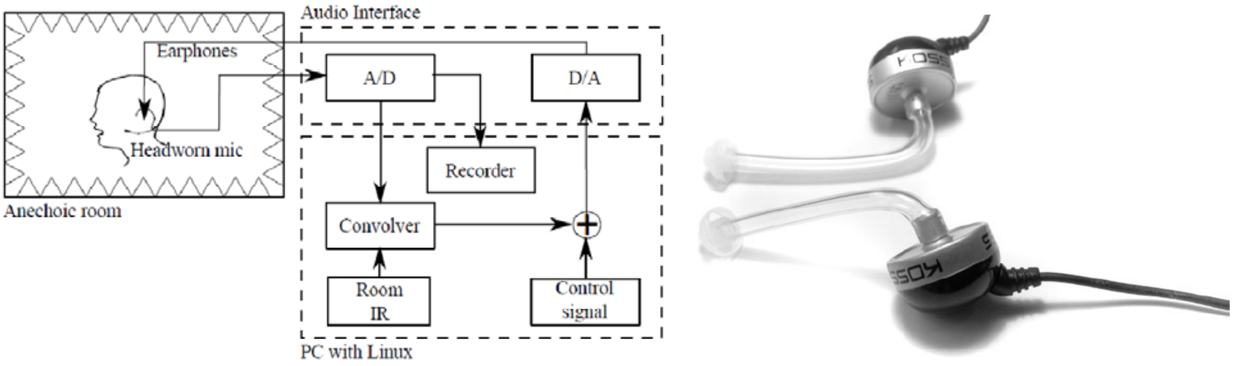

Figure 3: Left: the experimental setup for equal autophonic level experiments. Rigth: The custom earphones

The experiment was carried out using two different signals as the loudness reference. The first one is called “Voice Level Matching Test” (VLMT) which uses recordings from subjects’ own vocalizations as a reference, and the second one is called “Tone Level Matching Test” (TLMT). The reason for this decision was twofold. First, having a human vocalization as the reference could possibly lead to an imitation of the vocal effort, not a replication of loudness. Second, using a pure tone could have made the task more difficult because of the mismatch in the perceived sound quality of the reference and the vocalization.

2.6 Field study of voice use (C)

The field study is a prospective study with a case-control design, which investigated the voice use during a typical school day in teachers with voice problems and their voice healthy school colleagues, measured with a voice accumulator and a structured diary. For this study, n=28 teachers were recruited among the 62 participants in study A2. The pairs worked at the schools with the highest frequency of matched pairs, 3 schools, and they formed two groups: Group I: teachers with self-assessed voice problems (n=14, 12F:2M median age: 41, range: 24-62), and Group II: teachers without voice problems (n=14, 12F:2M median age: 43, range: 28-57). Median years in occupation: Group I: 13, range 2-40 and Group II: 18, range: 2-28. The groups did not differ for age or years in occupation as shown by a paired t-test.

In Lyberg-Åhlander, V., Pelegrin-Garcia, D., Rydell, R. and Löfqvist, A. (2011), ‘Teacher’s Voice Use in Teaching Environments: A Field Study Using Ambulatory Phonation Monitor (APM)’, the teachers were registered with the Ambulatory Phonation Monitor 3200 vers. 1.04 (APM)(APM, KayPentax New Jersey, USA). The APM uses an accelerometer to measure the skin vibrations of the neck that occurs during phonation. Based on the vibrations, the APM software estimates the phonation duration, fundamental frequency F0 (in Hz), sound pressure level SPL (in dB), and vocal doses. The APM does not record ambient noise, nor record the spoken message. Good accuracy has been shown for the APM’s estimation of F0 and phonation duration compared to recordings with traditional microphones. It also has a reasonably reliable estimation of the sound pressure level with an average error of 3.2 dB (SD 6 dB).

Simultaneously with the APM recordings, the noise and voice levels at the teacher's position were measured with a sound level meter Svantek, mod. SV-102. The signals were picked up by a lapel microphone at a distance of 15 cm from the teacher's mouth. The sound level meter was placed in the same waist-bag as the APM box (Pelegrín-García, D., Brunskog, J., Lyberg-Åhlander, V., Rydell, R., & Löfqvist, A., 2010), 'Influence of Classroom Acoustics on the Voice Levels of Teachers With and Without Voice Problems: A Field Study'. Moreover, the following acoustic properties of the classrooms were evaluated *background noise level*, *reverberation time*, *speech transmission index*, *sound strength* and *voice support* while the classrooms were empty, due to logistics. A head and torso simulator (HATS) was used for the *voice support* measurements, and an omnidirectional loudspeaker was used for the other room acoustic parameters. Additionally, the geometrical dimensions of the room were measured. The air humidity, room temperature, and the carbon dioxide (CO_2) contents of the air were simultaneously measured during the work-hours with an indoor air quality measuring device.

3 Results

The most important results and finding of the subprojects are summarized below. The complete results can be found in the papers in the appendix of the report.

3.1 Voice Handicap Index – throat (A1)

The aim of this study was to develop and evaluate an instrument that could simplify the patients' estimation of symptoms from the throat and to consider their relation to voice problems simultaneously. The Voice Handicap Index (VHI) had been in use at the voice clinic in Lund for a long period. A new subscale, named "throat scale" was constructed, using the same format, the same phrasing, and rating scale as in the VHI. The result, the VHI-Throat (VHI-T) was tested for validity, reliability, and test-retest stability. The test-retest reliability of the total VHI-T score was estimated with IntraClass coefficient (ICC), =0,968, proving a good reliability of the questionnaire. A paired samples t-test revealed no significant differences between the first and second occasion for neither the total VHI-T scores, nor the individual subscale in patients and controls. The VHI-T total score in all patients assigned to five different diagnose-groups was significantly higher than in the voice-healthy controls, thus indicating that the questionnaire separated persons with and without voice pathology. The difference in VHI-T scores between the patients and the controls was significant also for all subscales. Moreover, there was a good correlation of the test- retest occasions: the reliability testing of the entire questionnaire showed an alpha value of r = 0,90 which indicates a high degree of reliability, well in line with results reported by others. The Throat subscale separately reached an alpha value of r = 0,87, which is also considered a high reliability. The VHI-T thus proves to be a valid and reliable instrument for the estimation of self-perceived throat and voice problems. The throat subscale seems to reveal symptoms that are common in patients but that have not before been possible to uncover with the questionnaires designed for use in the voice clinic. The results show that symptoms from the throat are not uncommon in most voice diagnoses and that some scoring on the throat scale also occurs in completely voice-healthy individuals.

3.2 Prevalence of voice problems (A2)

This study examined how a group of Swedish teachers rate aspects of their working environment that can be presumed to have an impact on vocal behavior and voice problems. The secondary objective was to explore the prevalence of voice problems in Swedish teaching staff. A questionnaire was distributed to the teachers of 22 randomized schools. The results showed that 13% of the whole group reported voice problems occurring sometimes, often, or always.

The statements of the questionnaire were subjected to a principal component analysis (PCA). Prior to performing the PCA, the suitability of data for factor analysis was assessed. Inspection of the correlation matrix revealed the presence of many coefficients of ≥.3. The PCA revealed two components of eigenvalues exceeding 1 for the statements about room acoustics explaining 29.7% and 10.7% of the variance. There was a moderately strong correlation between the two factors (r=,542). For the statements about the voice, four components were found explaining 39,2%, 8.1%, 7,4%, and 5,7% of the variance. There was a weak positive correlation between components 1 and 2 (r=,338), 1 and 4 (r=,352) and 2 and 4 (r=,113) and a weak negative correlation between comp 1 and 3 (r=-,388), 2 and 3 (r =-,306) and 3 and 4(r=-,244). These findings indicate that the items listed under each component are highly loaded specifically onto one of these four independent underlying components. The loading of the acoustic and environmental statements on the two components of the PCA analysis were interpreted as follows:

- Component one includes the voice function and the interaction of the voice with the class room acoustics.

- Component two can be interpreted as covering external sources influencing the voice use.

The loading of the voice statements on the four components of the PCA analysis was interpreted as follows:

- Component 1 includes symptoms traditionally considered as early signs of voice problems and can most likely be interpreted as such also in this study, in particular due to the inclusion of statement 32 "I have voice problems" within this component.

- Component 2 can be viewed as "consequences of voice problems"

- Component 3 seems to reflect functional/emotional aspects of voice problems

- Component 4 includes symptoms from the throat.

Based on the ratings of statement 32 "I have voice problems", the participants were divided into two groups. Group I, (N=60) consisted of teachers having rated 2-4, i.e., suffering from voice problems sometimes, often, or always. Group II (N=407) included teachers having rated 0-1, i.e., never or only occasionally experiencing voice problems. There were no significant differences between the groups for gender or age computed by a chi-square test. There were no differences for smoking; years of occupation, voice training, possibility to rest, or for subject taught. Thus, we could not find teaching of any subject to be more hazardous to the voice.

3.3 Etiology of voice problems (A3)

This prospective, randomized case-control study compared pairs of teachers from study A2. Teachers with self-reported voice problems, n=31, were compared to age, gender and school-matched colleagues with self-reported voice health. The self-assessed voice function was related to factors known to influence the voice: laryngeal findings, voice quality, personality, hearing, psycho social and coping aspects, searching for objective manifestations of voice problems in teachers. Differences were found for all statements of all subscales of the VHI-T as shown by paired samples t-test and for time for recovery after voice problems computed by chi-square test: $\chi 2$, (7 n=60) = 17.608, *p*=0,014. Within the group of teachers with voice problems, 18% had considered change of work due to voice problems but none in the voice healthy group, as shown by Fisher's exact test (*p*=0,029). For the frequency of occurrence of voice problems, a chi-square test showed significant differences between the two groups: $\chi 2$, (5 n=60) = 20.138, *p*=0,01, Odds Ratio= 3.99, indicating that teachers with voice problems were close to four times as likely to rate a high frequency of voice problems. There were also significant differences between the groups for voice problems occurring without a concurrent upper-airway infection, $\chi 2$, (2 n=60) = 18,670 *p*=0.0008, OR=3.60.

Minor morphological abnormalities of the vocal folds were found in 13 subjects (5/31 in Group I (teachers with voice problems), 8/31 in Group II (voice healthy teachers)); some remarks on voice quality and hearing were made, and also some negative reports of psychosocial well being, but with no differences between the groups. The instrumental analyses of voice range (Voice Range Profile) and F0 in running speech did not show any differences between the groups. Further, there were no differences between the groups shown by the analysis of the Long Time Average Spectra. The ratios of the 0-1 kHz and 1-5 kHz frequency bands and the energy in the frequency band 5-8 kHz show that the voices should be considered to be modal to hyperfunctional.

3.4 Voice level and speaker comfort in real rooms (B1)

The pre-study by Brunskog, J., Gade, A.C., Payà-Ballester, G.; Reig-Calbo, L. (2009) showed a correlation between the physical characteristics of the rooms and the voice power, and with perceived quality, such that the room is perceived good or bad to talk in. The parameters in the room that primarily affect the voice power are the size of the room and the *room gain* provided by the room. In Pelegrín-García, D. (2011), a simplified and improved method for the calculation of room gain is proposed, in addition to a new magnitude called *voice support*. The new measurements are consistent with those of other studies. However, it turned out to be impossible to replicate the room gain measurements of Brunskog et al. in the original rooms of their study, probably due to a less stable measurement procedure, so the measurements were repeated.

The new room gain values differ considerably from the original ones. In order to enable a reliable comparison with future studies, the empirical model relating voice power level from the study of

Brunskog et al. to the room gain had to be recomputed. The relative voice power level (ΔL_W) is defined as the difference between the overall L_W in a certain room and the overall L_W measured in the anechoic room. A simplified linear model of only one explanatory variable is

$$\Delta L_W = 0.5 - 13.5 \cdot G_{RG} \, [\mathrm{dB}] . \qquad (6)$$

The model predicts a decrease in the expected voice power level with increasing room gain (R^2 = 0.83, p = 0.01).This can be interpreted as: rooms with low room gain demand higher vocal intensity from talkers.

Talkers adjust their vocal effort to communicate at different distances to compensate for the sound propagation losses. In Pelegrín-García, D., Smits, B., Brunskog, J, Jeong, C.-H. (2011), the speech from talkers speaking to a listener at four different distances in four different real rooms was recorded. The listener moved alternately at positions located at 1.5 m, 3 m, 6 m and 12 m away from the talker. This experiment was repeated in four rooms: an anechoic chamber, a reverberation room, a long narrow corridor and a big lecture room. The measurements show that speakers raise their vocal power when the distance to the listener increases, at a rate of 1.5~2.0 dB per double distance (see Figure 6, left). The voice power level produced in the anechoic room differed significantly from the other rooms.

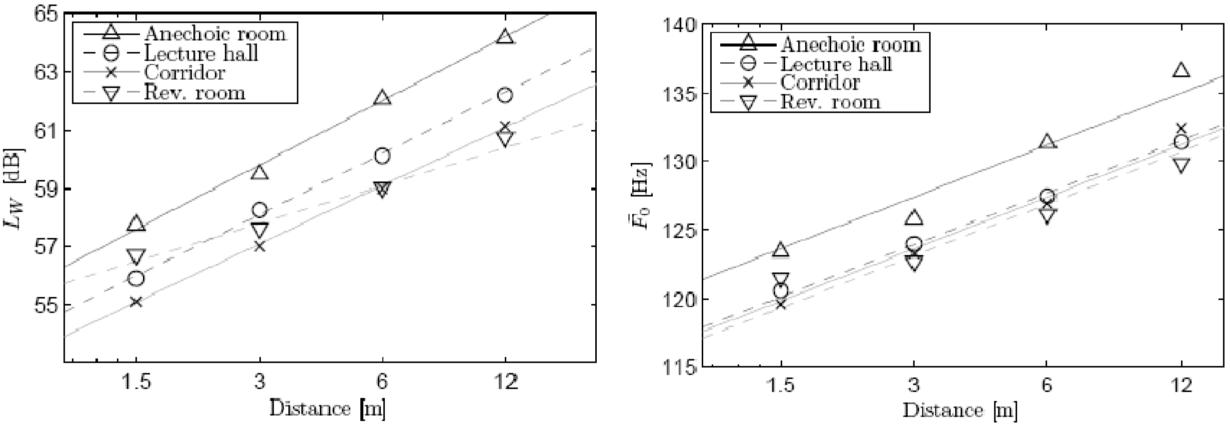

Figure 6: Left: Variations in voice power level versus distance. The lines show the predictions of the empirical model. Right: Phonation time ratio versus distance. The lines show the predictions of the empirical model.

The measured L_W, as a function of the distance and for each of the rooms, averaged across all subjects, is shown in Fig. 6 left. In the same figure, the lines show the fixed-effects part of the empirical model. L_W depends almost linearly on the logarithm of the distance (with slopes between 1.3 dB and 2.2 dB per doubling distance) and changed significantly among rooms (intercepts between 54.8 dB and 56.8 dB). At each distance, the highest L_W was always measured in the anechoic room. A significant interaction was found between the room and the logarithm of the distance, because the variation of L_W with distance in the reverberation room (1.3 dB per doubling distance) was lower than the variation in the other rooms (1.9 to 2.2 dB per doubling distance).

Figure 6 right shows the subject-averaged measured fundamental frequency F0 (data points) and the corresponding empirical model (lines) for the different distances and rooms. F0 changed significantly among rooms (intercepts between 119.3 Hz and 123.6 Hz) and had an almost linear dependence on the logarithm of the distance, with a slope of 3.8 Hz per doubling distance, identical for all the rooms. However, in the anechoic and reverberant rooms, there was less variation between the distances of 1.5 m and 3 m than at further distances. F0 in the anechoic room was about 4 Hz higher than in the other rooms for all distances. The standard deviation of the intersubject variation was estimated at 16.3 Hz, whereas the individual differences in the variation of F0 with distance had a standard deviation of 2.95 Hz per doubling distance.

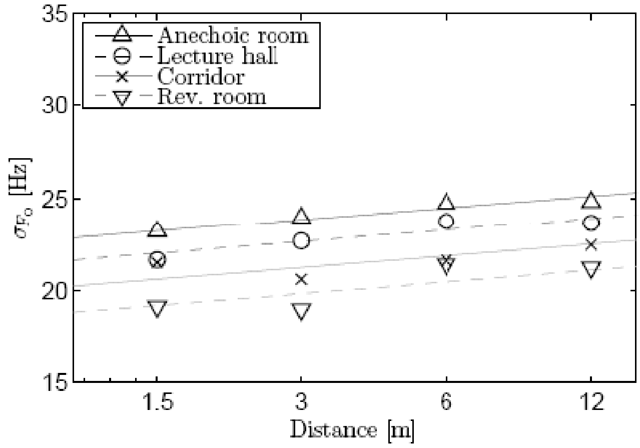

Figure 7: Average long-term standard deviation of the fundamental frequency used by talkers at different distances to the listener. The lines show the predictions of the empirical model.

The measured standard deviation of the fundamental frequency, σ_{F0} , as a function of the distance and for each of the rooms, averaged across all subjects, is shown in Fig. 7 The lines in the figure show the fixed effects part of the empirical model. σ_{F0} changed significantly between rooms (intercepts between 19.2 Hz and 23.2 Hz) and had a weak linear dependence on the logarithm of the distance, with a slope of 0.63 Hz per doubling distance, equal among the rooms.

As all of the measured parameters vary with distance and acoustic environment, they are potential indicators of vocal effort.

Furthermore, the subjects expressed their preference about vocal comfort, stating that the least comfortable environments were the anechoic room and the reverberation room. While the analysis of the voice levels cannot account for this preference, other parameters might be better suited. The phonation time ratio (ratio between duration of voiced segments and total duration of running speech) might be appropriated for this purpose. The subjects produce longer vowels in the anechoic room and the reverberation room, compared to the two other rooms, either to overcome the poorer speech intelligibility at the listener location (in the reverberation room) or due to the raised voice levels (in the anechoic room).

3.5 The virtual environment (B2)

The experiment in the virtual environment aimed to investigate the voice used by a teacher to address a group of imaginary students under different simulated acoustics.

In a pre-experiment, Pelegrín-Garcia and Brunskog (2009b) (not included in the appendix, but also reported in Brunskog, J., and Pelegrín García, D., 2010), five subjects, aging 23-35 with normal hearing and voice status, talked freely in 5 different simulated acoustic environments during 3 minutes in each of them. The goal was to give a lecture of a familiar topic to an imaginary group of 30 students located in front of them. In addition, they had to answer a small questionnaire after speaking in each simulated room. The results in Figure 8 left show a significant linear dependence (R^2=0.92) between the changes in voice power level used by the speaker and the voice support provided by the room to the talker's voice, with a slope of -0.65 dB/dB, although the absolute mean variations were between 2 and 3 dB. The fundamental frequency used by the talkers changed significantly between environments, although it did not follow a linear trend.

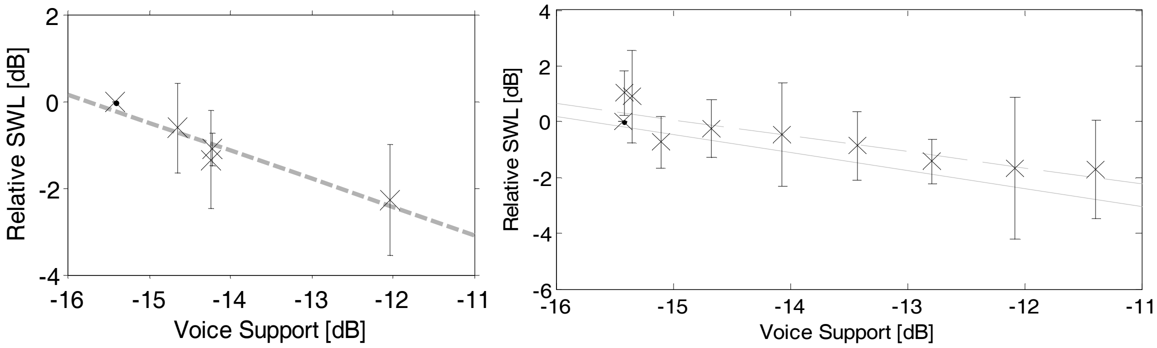

Figure 8: Measured relative voice power level versus support, using free speech. Left: Five subject (students) and five simulated room (Odeon), dashed line: regression. Right: Five subject (teachers) and ten simulated rooms (modified gain), dashed line: regression, solid line: regression from left figure.

The goal of the next experiment, reported in Brunskog, J., and Pelegrín García, D. (2010), was to measure the vocal output when the gain of the RIR was changed, and thereby also changing the voice support, but keeping the reverberation time fixed. Thus, the different stimuli did not correspond to actual simulated rooms, but to a single impulse response with 10 different gains. Five teachers talked freely in 10 different simulated acoustic conditions during 3 minutes in each of them. The goal was to give a free speech lecture of a familiar topic to an imaginary group of 30 students located in front of them. The measured variations in voice power level used by subjects are shown in Figure 8 right. The trend of the voice power level, indicated by the dashed line, lays very close to the voice power level measured in the first pre-experiment (solid line). The slope of the line is in this case -0.58 dB/dB. This indicates that the experiment is fairly repeatable, and that the acoustic environment can systematically change the vocal behavior.

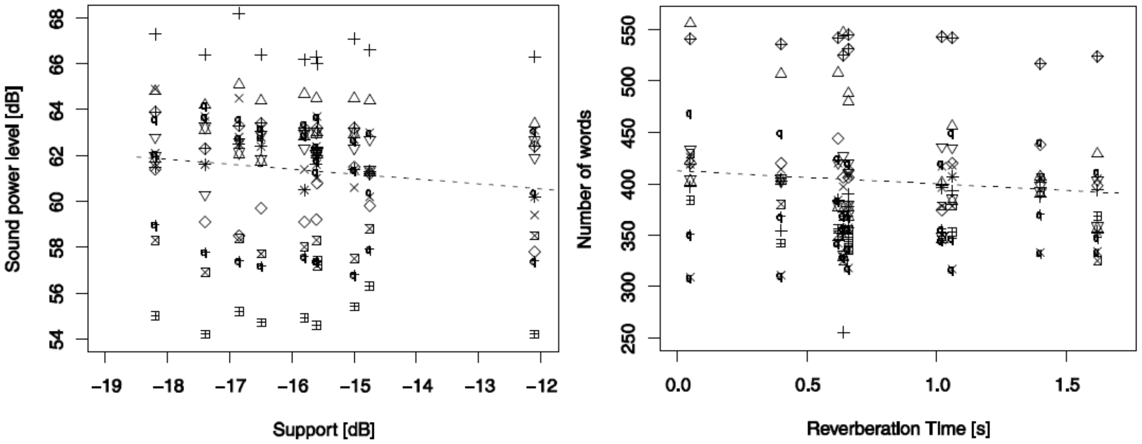

Figure 9: Left: L_W versus ST_V. Right: Number of read words versus T_{30}. Different symbols correspond to different subjects. The dashed lines correspond to regression lines calculated with linear mixed models

In the next experiment, thirteen teachers (4 females, 9 males) of secondary school, high school, and university, aging 30 to 67 years, participated in the experiment. The teachers did not have known voice problems (according to their statements) or hearing loss greater than 25 dB HL below 4 kHz. Once they were in the laboratory room, and for each condition, they were instructed to read a text during 2.5 minutes, addressing a listener located at a distance of 2 m. A dummy head was located at that position to provide the visual distance cue. There were ten experimental conditions, consisting of nine different simulated IR and the condition zero of no RIR simulated (and thus corresponding to the actual acoustic conditions of the laboratory room). The nine experimental conditions were the combination of three different classroom geometries and three different placements of absorptive materials in those rooms. Figure 9 shows the measured L_W against ST_V values (left), and the number of words versus the T_{30} (right). The figure shows a large spread among observations. Most of them are related to individual factors which only shift the absolute values, while keeping similar variations among conditions. The factor "subject" was considered a random effect, and a linear mixed model was used to evaluate the dependence of L_W with ST_V, finding a significant relationship ($p=0.004$). An identical procedure was followed to analyze the number of read words ($p=0.045$). The regression lines shown in Fig. 9 correspond to the output of the linear mixed models. The sound power level of the voice decreases with the ST_V, at a rate of -0.21 dB/dB. This rate is smaller (in absolute value) than reported in Fig. 8 or found in the pre-study. This deviation can be due to the different instructions given to the subjects: One reason for this might be that asking the talker to read a text aloud for a listener located at 2 m does not lead to the same voice adjustment as it would be required for addressing a group of people at further distances with spontaneous speech.

Another experiment was carried out at DTU in collaboration with the Politecnico di Torino, (Bottalico, P., Pelegrín-Garcia, D., Astolfi, A., and Brunskog, J., 2010), 'Measurement of vocal doses in virtual classrooms'. The goal was to measure vocal doses of speakers under different conditions of room acoustics and noise. Vocal doses are a set of measures derived from an estimation of the SPL and the fundamental frequency used by a talker during phonation. They are measured with an accelerometer attached to the talker's neck and an Ambulatory Phonation Monitor (APM).

In the SpaceLab, 22 untrained talkers (11 males, 11 females), without self-reported known problems with their hearing or their voice, had to read aloud a text passage from "Goldilocks" during two minutes under 13 different acoustical conditions. These conditions combined different kinds of background noise (traffic, ventilation, or babble noise), at levels ranging from 37 dB to 57 dB, and different room impulse responses, obtained by simulation of medium-sized classrooms with T30 in the range between 0.33 s to 1.47 s and ST_V in the range from -17.8 dB to -13.6 dB. There were significant differences in Vocal Load Index (VLI) between the conditions with low background noise and the conditions with higher background noise. Only when the background noise is sufficiently low (LN < 40 dB), there is an effect of different values of ST_V on the VLI. In this situation, conditions with high ST_V values result in lower Vocal Loading than in conditions with low ST_V.

3.6 Loudness of one's own voice (B3)

An experiment was conducted to obtain the relative voice levels that kept the autophonic level constant under different room acoustics conditions described by the parameters room gain and voice support. Fourteen subjects matched the loudness level of their own voice (the autophonic level) to that of a constant and external reference sound, under different synthesized room acoustics conditions. A four way ANOVA reveals that there is a significant effect of the acoustic condition ($F(8, 652) = 92.4$, $p < 0.0001$), responsible for almost the 90% of the explained variance. Gender has also a significant effect ($F(1, 652) = 43.2$, $p < 0.0001$), and is responsible for another 5% of the explained variance. The variables reference and vowel do not show significant effects. However, there are significant interactions between reference and vowel ($F(2, 652) = 5.55$, $p = 0.004$) and between vowel and gender ($F(2, 652) = 5.13$, $p = 0.006$), responsible however, for less than 3% of the explained variance. There are no significant interactions between the acoustic condition and any other variable. In the additive model, the average relative voice level ΔL_Z is -3.3 dB for females, whereas it is -2.2 dB for males. Analyzing the voice levels in one-octave bands and with different frequency weightings, a set of equal autophonic level curves was generated, Fig. 10.

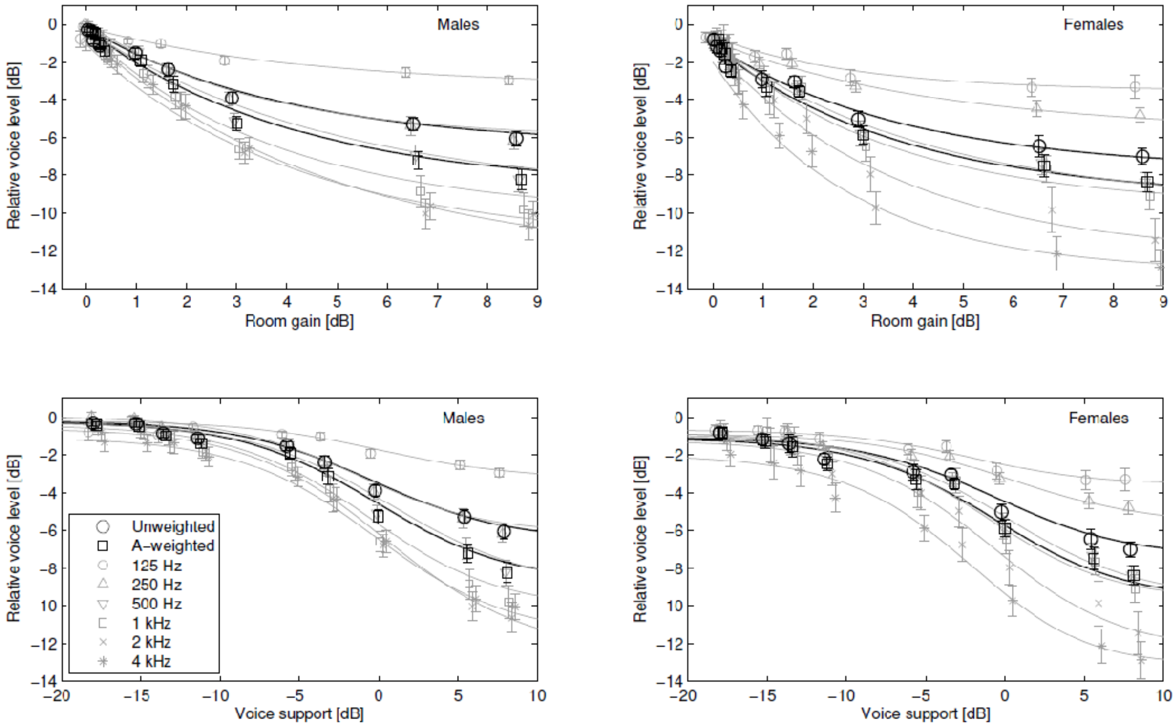

Figure 10: Relative voice levels as a function of the room gain (top row) and the voice support (bottom row), for male (left column) and female subjects (right column). The reference value for each subject is the voice level produced without simulated reflections. The curves are the best fitting models for each relative voice level descriptor. The bars around the points indicate ±1 standard error.

These curves allow to determine the expected voice level differences in different rooms which are purely related to the Lombard-effect or sidetone compensation. An average model for males and females together, for unweighted ΔL_Z and A-weighted ΔL_A

$$\begin{aligned} \Delta L_Z &= 8.4 \cdot e^{-0.24 G_{RG}} - 8.9 \; [\mathrm{dB}] \\ \Delta L_A &= 6.4 \cdot e^{-0.25 G_{RG}} - 6.9 \; [\mathrm{dB}] \end{aligned} \tag{7}$$

From the observation of the measured relative voice levels, it is possible to state that different acoustic environments alter the autophonic level for a talker. However, the reverberation time is not a good descriptor of the changes in voice level, since it is not directly related to the energy of the indirect auditory feedback. Figure 10 describes the changes in voice level that make the talker's voice sound equally loud at their ears when the indirect acoustic feedback is changed. The curves for ΔL_Z show a constant autophonic level under different room gain conditions (top row), or voice support conditions (bottom row). The A-weighted and the one-octave band values follow the same general trend of the non-linear model but with different model parameters. In normal rooms for speech without amplification ($G_{RG} < 1.0$ dB) the variations in voice level to keep a constant autophonic level are within 2.3 dB, according to model Eq. (7).

The main conclusions of the study are as follows: Voice level variations under different room acoustics conditions are related to the room gain or the voice support, and not to the reverberation time. Typical voice level variations in rooms for speech ($G_{RG} < 1.0$ dB) to keep a constant autophonic level are not higher than 2.3 dB. By comparison with other studies, talkers use other cues than loudness to adjust their voice level in rooms, resulting in larger voice variations than barely keeping the autophonic level constant.

3.7 Field study

The field study examined how classroom acoustics interacts with the voices of 14 teachers without voice problems and 14 teachers with voice problems. The assessment of the voice problems was made with a questionnaire and a laryngological examination. During teaching, the sound pressure level at the teacher's position was monitored. The teacher's voice level and the activity noise level were separated using mixed Gaussians. In addition, objective acoustic parameters of *Reverberation Time* and *Voice Support* were measured in the 30 empty classrooms of the study. An empirical model shows that the measured voice levels (see Figure 11) depend on the activity noise levels and the Voice Support. Teachers with and without voice problems were equally affected by the activity noise levels, raising their voice with increasing noise according to the Lombard effect, at an average rate of 0.6 dB/dB. Teachers with and without voice problems were differently affected by the Voice Support of the classroom. The results thus suggest that teachers with voice problems are more aware of classroom acoustic conditions than their healthy colleagues and make use of the more supportive rooms to lower their voice levels. This behavior may result from an adaptation process of the teachers with voice problems to preserve their voices.

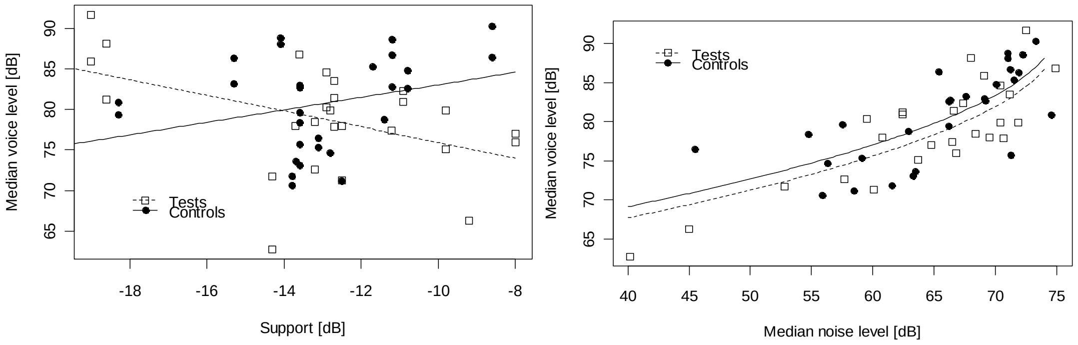

Figure 11: Comparison of the model and the measured values. Left: Median voice level vs Support. Right: Median voice level vs. Median noise level.

The study aimed at closer investigating the vocal behaviour and voice use in teachers with self-estimated voice problems and their age, gender and school matched colleagues without voice problems, using matched pairs. The teachers' fundamental frequency, Sound Pressure Level, and phonation-time were recorded with an Ambulatory Phonation Monitor (APM) during one workday and they also reported their activities in a structured diary. The main hypothesis was that teachers with and without voice problems act differently with respect to classroom acoustics and air-quality, and that the vocal doses obtained with a voice accumulator would separate the groups.

The analysis of the diaries confirms the results of the epidemiological and etiological studies. The group with voice problems rated their voice problems during the day significantly worse than their voice healthy colleagues, on the Visual Analogue Scale, according to a paired t-test3 (*p*=0.003). This group also rated their degree of vocal fatigue (*p*=0,007) and loss of air during speech (*p*=0,007) significantly higher than their voice-healthy matched peers.

Teachers with voice problems behaved vocally different from their voice healthy peers, in particular during teaching sessions. The *time dose* (percent of voicing) was significantly higher in the group with voice problems as shown by a paired t-test for the entire work-day and specifically for teaching. The phonation time for teachers in this material varied between 17-24%. Further, the *cycle dose* (number of cycles) during work-time differed significantly between the groups as shown by a paired t-test. The cycle dose varied between activities for both groups as shown by a one-way ANOVA and post-hoc comparisons with Tukey HSD test indicated that the mean score for ”teaching” differed significantly from “preparation/break” for both groups with the higher cycle dose for teaching.

Also the F0 pattern, related to voice-SPL differed between the groups. The group with voice problems did not raise their F0 with increasing SPL of the voice, whereas the voice healthy group raised the F0 with the SPL increase. The voice-problem group either kept the F0 stable or decreased it as shown by Figure 12. This is shown by the difference between the groups in the

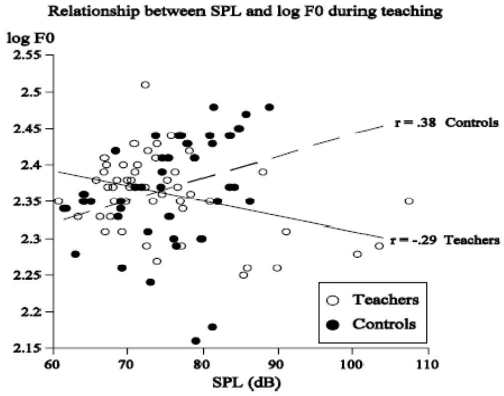

Figure 12: The sound pressure level and fundamental frequency during teaching.

In Pelegrin-Garcia, D., Brunskog, J., Lyberg-Åhlander, V, & Löfqvist, A. (2011), ‘Measurement and prediction of acoustic conditions for a talker in school classrooms’, data from the field measurement where used to validate a simplified prediction model of the *voice support,*

$$ST_V = 10\log\left(\frac{24.8T}{V} - \frac{4}{S} + \frac{Q^*}{4\pi(2d)^2}\right) + \Delta L_{HRTF} - K \qquad (8)$$

where T is the reverberation time, V is the volume of the room, S is the total surface area of the room, Q^* is the directivity factor of speech in the downward direction and d is the height from the ground for the head position of the talker. The two corrections are first the correction ΔL_{HRTF} due to the head related transfer function (HRTF) and secondly the correction K between sound power and sound pressure level at the receiver. Figure 13 compares the model with the measured values of the *voice support* in the class rooms.

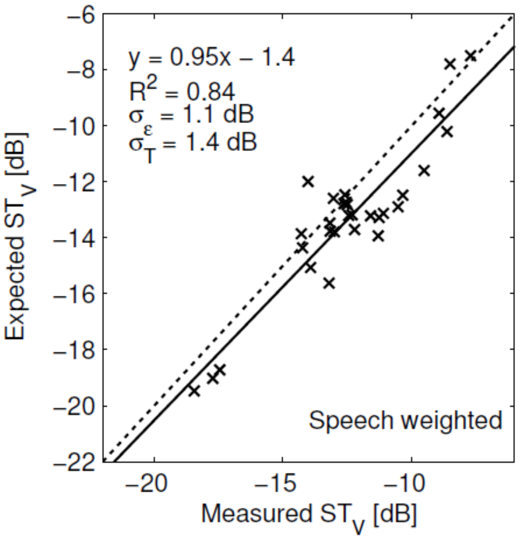

Figure 13: Expected versus measured speech-weighted overall values of voice support. The solid lines show the regression lines for the predictions and the dotted lines indicate the ideal and unbiased prediction lines.

4 Discussion

The basic findings within the project and their consequences are briefly summarized here.

4.1 The environmental factors of vocal load

The environmental factors affecting the vocal load can be summarized as: voice use, rest and recovery, background noise, room acoustics, air quality, and stress and psychological factors (Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A. 2010). Teachers' voice problems can be seen in the interaction with the environment and exist even if it is not possible to find any clinical evidence in teachers with voice problems. In addition, the ST_V is an important measure for understanding voice control. The teachers have something to gain from paying attention to the room acoustics and taking advantage of it for their voice use. Teachers with voice problems are more dependent on good working conditions and need to learn how to optimize their use of the voice and of the room acoustics. Discussions about the use of the acoustic properties of the classroom should be included in voice therapy and preventive voice care designed for teachers. Field measurements of the voice should be included when exploring occupational voice problems, since it is apparent that voice problems arise out of the interplay between the individual and the work environment.

Figure 14 shows how the teachers perceived the room acoustics of the classrooms. Most of these statements can be related to a too low voice support of the room.

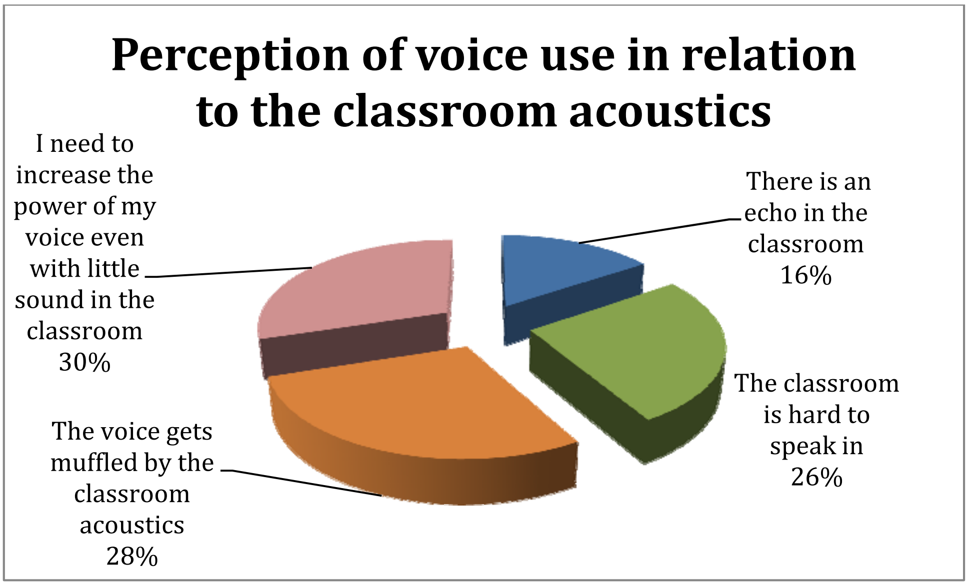

Figure 14: Perception of voice use in relation to the classroom acoustics

Figure 15 shows how the teachers perceived the importance of different noise sources in the class rooms. The noise cause by the pupils is the most important one.

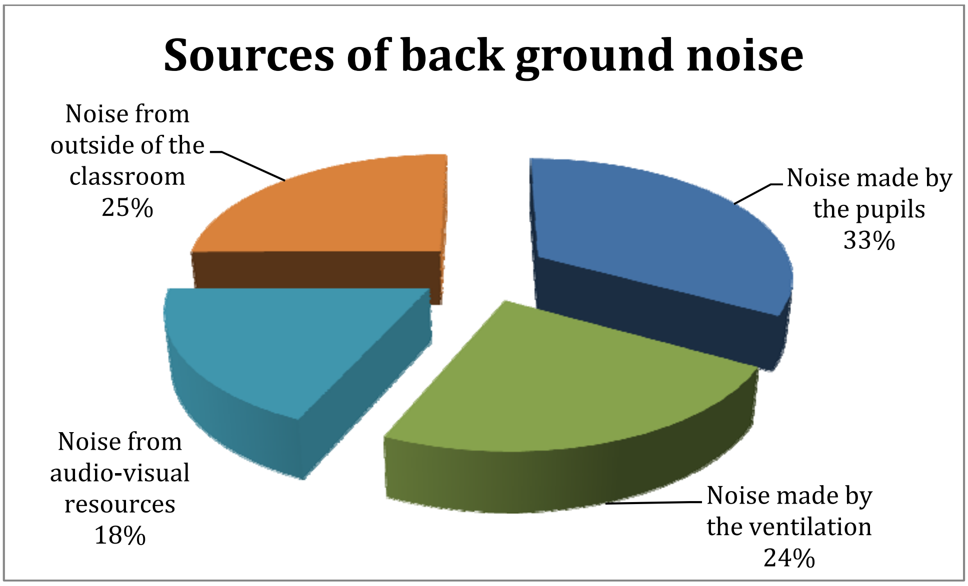

Figure 15: Perceived sources of background noise

Figure 16 shows how the teachers express their voice problems. Several typical voice symptoms are used.

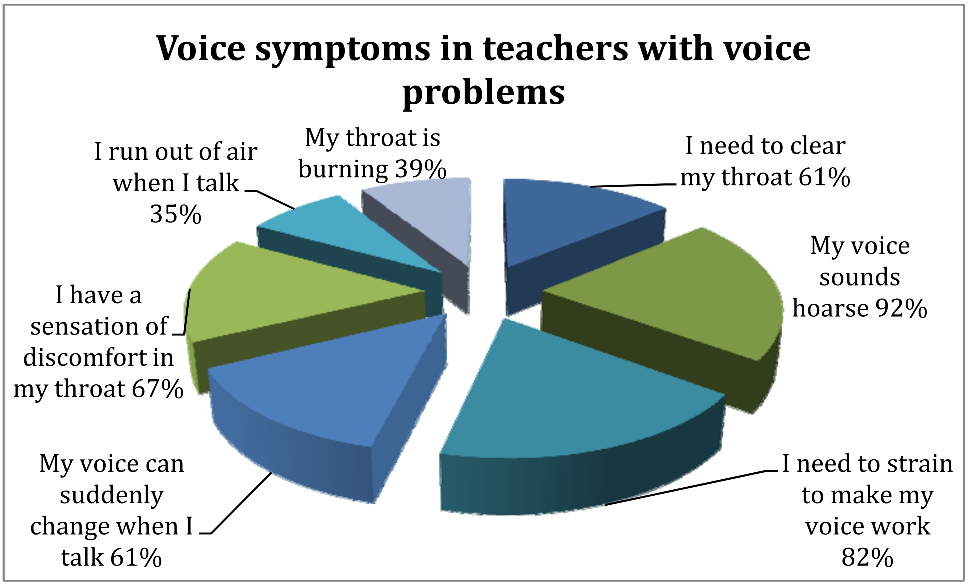

Figure 16: Perceived voice symptoms in teachers with voice problems

Figure 17 shows how the teachers the importance of the consequences of voice problems among the studied teaches. The most important one are ‘My voice upsets me’ and ‘My voice limits my work’.

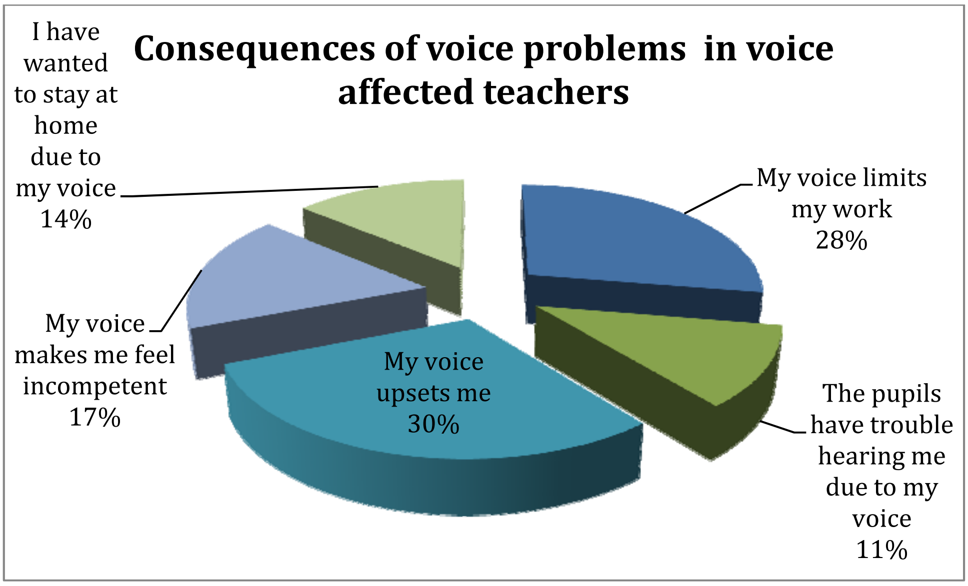

Figure 17: Perceived consequences of voice problems in voice affected teachers

4.2 Voice regulation

The components of the voice regulation has been studied in subprojects B1, B2, B3 and C, and these findings are here summarized in two pie charts, Figs.18 and 19.

Figure 18 shows the relative importance of background noise level (BNL) and voice support (ST_V) in the voice regulation and voice level. The information is extracted from the field measurements, Pelegrín-García, D., Lyberg-Åhlander, V., Rydell, R., Löfqvist, A. & Brunskog, J., (2010), 'Influence of Classroom Acoustics on the Voice Levels of Teachers With and Without Voice Problems: A Field Study'. The model of voice level (VL) versus BNL showed an average VL variation of 18 dB between the lowest and the highest measured BNL. The most important source of background noise is that of the activity noise from the pupils. The average variation of VL in the measured range of ST_V was about 9 dB.

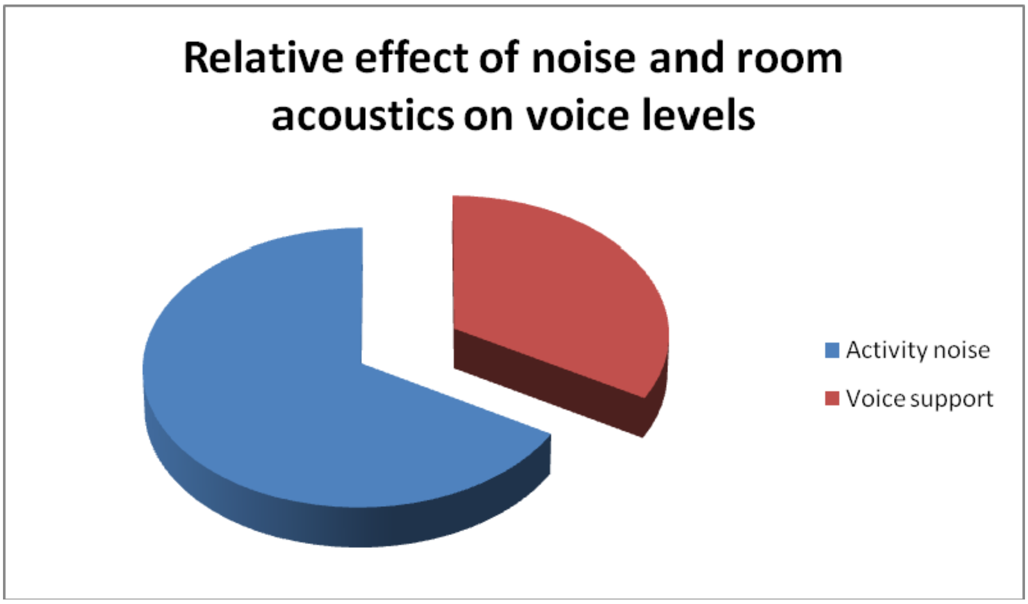

Figure 18: The relative effect of noise and room acoustics on the used voice levels

The estimated causes of the voice regulation are shown in Fig. 19, in terms of VL variation as a function of STV when background noise is not present. The estimation is taken from the slopes of the room gain (G_{RG}) – VL characteristic in the laboratory experiments in subprojects B1-3. In real classrooms, the slope is -13.5 dB/dB (Pelegrín-García, D., 2011). If the distance effect is removed, the adjustment is -3.6 dB/dB. From this amount, -1.8 dB/dB correspond to sidetone compensation (equal autophonic level) and the same amount correspond to other cognitive effects (Pelegrín-García, D., Fuentes-Mendizabal, O., Brunskog, J, and Jeong, C.H., 2011).

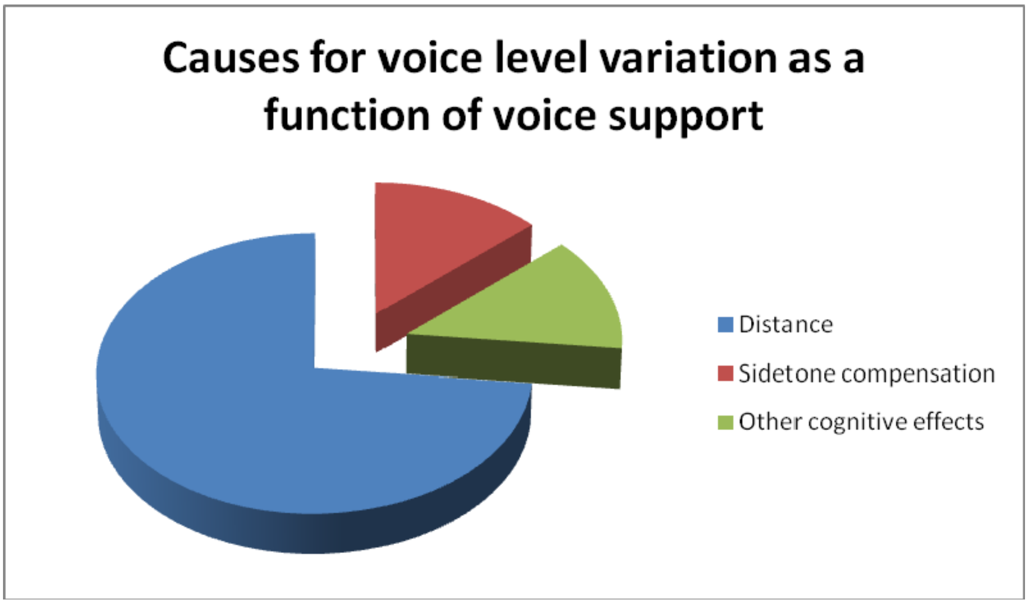

Figure 19: Estimated causes of voice regulation as function of voice support, excluding the effects due to BNL.

4.3 Recommendations regarding room acoustic design

The voice support is a measure relating the room acoustic conditions of a classroom to the teachers' use of their voice. This measure can therefore be used when designing a room. Here there are some initial recommendations for voice support:

- $ST_V < -17$ dB: **Very low**. Teaching in this room is not to be recommended, unless an amplification system is used.
- $-17 \text{ dB} < ST_V < -14$ dB: **Low**. Amplification system is highly recommended.
- $-14 \text{ dB} < ST_V < -9$ dB: **Good**. Recommended values of voice support which can deliver optimum acoustical conditions.
- $-9 \text{ dB} < ST_V < -6$ dB: **High**. Only advisable in very small classrooms
- $ST_V > -6$ dB: **Excessive**. Should be avoided. The decrease in voice level due to the sidetone compensation is remarkable and produces sensation of discomfort and decreases the voice quality. Higher values of STV than -6 dB when using electroacoustic amplification may result in a risk of feedback.

The range "Good" corresponds to the measured ST_V values in about 75% of the classrooms of our study in Sweden (which are considered acoustically satisfactory).

Using the validated prediction model in Pelegrin-Garcia, D., Brunskog, J., Lyberg-Åhlander, V., Rydell, R., & Löfqvist, A. (2011), 'Measurement and prediction of acoustic conditions for a talker in school classrooms', equation (8), is it possible to relate these recommendations to the volume and reverberation time, assuming a typical class room. This is done in Fig. 20. Assuming a typical classroom to have a volume of 150 m^3, we can say that the reverberation time should not be lower than 0.35 s in order to have ST_V > -14 dB. For larger rooms a higher reverberation time is necessary.

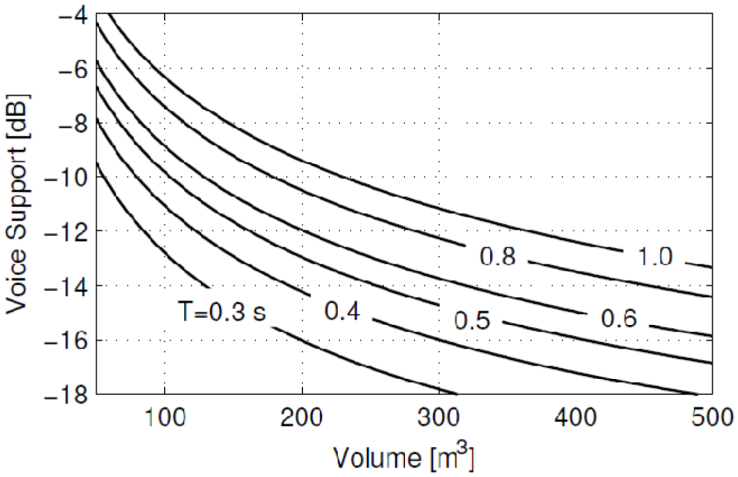

Figure 20: Voice support versus room volume for a room of proportions 28:16:10 according to the predictions of the model in Eq. (8), for different values of reverberation time.

5 Conclusions

The following conclusions can be drawn from the project:

- The room acoustic conditions themselves have an effect on voice production. This is most obvious when the talker is aware of the acoustic environment, as in the following cases: the teacher has either week voice or voice problems, or the acoustic environment is unusual (e.g. anechoic condition). In addition, high background noise levels induce an increase in vocal effort.
- New acoustic measures, namely the voice support and the room gain, are well correlated with the changes in voice level among different rooms.
- The visually perceived distance between teacher and student accounts to a great extent for changes in vocal effort.
- Voice problems in teachers arise from the interplay of the individual and the environment. Teachers with voice problems are more affected by factors in the work environment than their voice healthy colleagues. The differences between a group of teachers with self-assessed voice problems and their voice healthy colleagues were most clearly shown during field-measurements of the voice during a typical school day, while the findings from the clinical examinations of larynx and voice did not differ between the groups.
- The results from the prevalence study show that 13% of the teachers suffer from voice problems frequently or always. Most teachers however, reported occurrence of symptoms of vocal disturbances. Voice-related absence from work was common in both teachers with and without voice problems.

- Teachers with voice problems are more affected by the room acoustics and by factors adding to the background noise than their voice healthy colleagues. Any voice load is troubling for the individual who suffers from voice problems.
- The possibility for voice rest during the school day and also during teaching sessions is crucial for teachers with voice problems. The results suggest that teachers with voice problems have a higher vocal load during teaching and that this group has fewer opportunities for vocal rest and recovery during the school-day.
- When investigating or diagnosing voice dysfunction, the individual's self assessment of the problems needs to be included. No correlation was found between subjective assessment of voice problems and deviations of laryngeal morphology or voice quality.

6 Publications and diffusion of knowledge from the project

Apart from the publications discussed so far, included in the appendix of this report, several other publications and presentations of the project have been made. There has also been spread of knowledge to the public in form of interviews and articles in newspapers and popular science magazines.

Conferences. The project has been presented at several international conferences in voice, acoustics or related areas:

- The pre-study where presented at the International Conference Acoustics, 2008 in Paris (Brunskog, J., Gade, A. C., Payá-Ballester, G., Reig-Calbo, L. 2008)
- The First Nordic Conference of Voice Ergonomics and Treatment 24-25/3 2009 (Pelegrin-Garcia, D. & Brunskog, J., 2009a; Lyberg-Åhlander, V., Rydell, R.; & Löfqvist, A, 2009)
- Inter-Noise 2009, Ottawa, Canada (Pelegrín-García, D., Brunskog, J. 2009b)
- EAA Euroregio 2010, Ljubljana, Slovenia (Pelegrín-García, D., Brunskog, J., 2010)
- Inter-Noise 2010, Lisbon, Portugal (Pelegrín-García, D., Fuentes-Mendizábal, O., Brunskog, J. and Jeong, C.-H., 2010; Bottalico, P., Pelegrín-Garcia, D., Astolfi, A., Brunskog, J. 2010)
- International Occupational Hygien Association (IOHA 2010), 8th International Scientific Confernace, Rome, Italy (Brunskog, J., Pelegrín-García, D., 2010) – an invited keynote speech at a workshop
- The 2nd Pan-American/Iberian Meeting on Acoustics and 160th Meeting of the Acoustical Society of America, Cancun, Mexico, 2010 (Pelegrín-García, D., Lyberg-Åhlander, V., Rydell, R., Löfqvist, A., and Brunskog, J., 2010)

The project has been presented at Afa's noise conference in 2008, 2009 and 2010, at a meeting of the Noise Network ('Människan och bullret') in 2008 and again in 2010, and at the Voice Association's Conference April 2009. The prevalence study were presented at Voice Association's conference 24-25/4 2009, the Hearing Association's days 18/10. The Sound Environmental Center in Lund had a symposium day focusing on the project in September 2009, *Speech comfort, acoustics and learning*, with oral presentations by J Brunskog, V. Lyberg-Åhlander and D. Pelegrin-Garcia. V, together with some invited speakers. Lyberg-Åhlander presented the project at the Swedish Voice Ergonomic Network in January 2010. D. Pelegrin Garcia had an oral presentation at the seminar 'Acoustics in school', organized by the Danish Acoustical Society (DAS), Hillerød, Denmark, December 2009.

Popular science and branch magazines. An early description of the project where published in the annual acoustic edition in branch magazine for the building industry Bygg & Teknik in Mars 2008 (Brunskog 2008). The final outcome of the project where presented in the same magazine in Mars 2011 (Brunskog, J., Pelegrín-García, D., Lyberg-Åhlander, V., Rydell, R., Löfqvist, A. 2011).

Radio and newspapers. A number of interviews have resulted in articles and radio spots as follows: *Vetenskapsradion,* SR (11/12 2009); *Skolvärden,* Lärarnas riksförbund (2009); *DIKforum*[2], DIKförbundet (2009); *Röstläget,* the magazine of the Voice Association (2009); *Skolledaren,* the magazine of the principal's association (2010); *Läkartidningen* (2010); Speech Therapy Association's magazine (2010); *Forskning och Framsteg*[3] (nr.2 2010); *Skånska Dagbladet* (27/3 2010); *Lunds universitet meddelar (LUM)* (2010); *Östgöta Korrespondenten*[4] (2010); and the information magazine of Ecophon *Ecophon Acoustic Bulletin*[5] (2010).

More activities. Viveka Lyberg-Åhlander has been asked by some schools to come and talk about what can be done to solve the teachers voice problems (among other, Lars-Erik Larsson High School in Lund). The teacher education in Kristianstad have invited Viveka Lyberg-Åhlander to talk about voice care in teacher's work, and she has been in contact with Ann-Marie Körling[6], who is a teacher, lobbyist and very often referred person in teaching circles. The acoustic absorber company Ecophon (Jonas Christensson) want to involve us in training architecture students. Malmö Academy of Music would like that Viveka Lyberg-Åhlander talk to their music student teachers. J. Brunskog and D. Pelegrin-Garcia have discussed the project with influential acousticians in Denmark such as Dan Hoffmayer, DELTA (being deeply involved in regulations of, e.g., classroom), and Claus Møller Petersen, Grontmij|Carl Bro (the head of the Danish Acoustical Society), etc. The reference group of the project has also been used as a channel of informing about the project.

References

Bottalico, P., Pelegrín-Garcia, D., Astolfi, A., and Brunskog, J. (2010), Measurement of vocal doses in virtual classrooms. Proceedings of Internoise 2010, Lisbon, Portugal

Brunskog, J. (2008), Att tala i en undervisningslokal, Bygg & teknik, 99, issue: 3, p 57

Brunskog, J., Gade, A. C., Payá-Ballester, G., Reig-Calbo, L. (2008), Increase of voice level and speaker comfort in lecture rooms, Proceedings of Acoustics 08, Société Francaise d'Acoustique, Paris

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

---

[2] http://www.dik.se/www/dik/web.nsf/dx/091211-Heshet
[3] http://www.fof.se/tidning/2010/2/lararna-sliter-sonder-rosten
[4] http://www.corren.se//ostergotland/linkoping/?articleId=5205930
[5] http://www.acousticbulletin.com/EN/sound.html, David Pelegrín-García being interviewed at Internoise 2010.
[6] http://korlingsord.se/

Brunskog, J., Pelegrín-García, D. (2010). Speaking comfort and voice use of teachers in classrooms. Italian Journal of Acoustics (ISSN: 0393-1110), 34, 51-56.

Brunskog, J., Pelegrín-García, D., Lyberg-Åhlander, V., Rydell, R., Löfqvist, A. (2011), Klassrummet och lärarens röst, Bygg & teknik, 102, issue: 3

Lyberg-Åhlander, V., Rydell, R., & Löfqvist, A. (2009) Speaker's comfort in teaching environments: Prevalence of voice problems in Swedish teaching staff. Proceedings of the First Nordic Conference of Voice Ergonomics and Treatment. Publications of the Finnish Association of Speech and Language Research 41: 29-37.

Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A., (2010) Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff, J. Voice (in press). Available online 26 March 2010. DOI: 10.1016/j.jvoice.2009.12.006

Lyberg-Åhlander V, Rydell R, Eriksson J, Schalén L. Throat related symptoms and voice: development of an instrument for self assessment of throat problems. J. Voice 2010(5)

Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A. (2011), How do teachers with self-reported voice problems differ from their peers with self-reported voice health? Manuscript submitted for publication.

Lyberg-Åhlander, V., Pelegrin-Garcia, D., Rydell, R. and Löfqvist, A. (2011), Teacher's Voice Use in Teaching Environments: A Field Study Using Ambulatory Phonation Monitor (APM). Manuscript submitted for publication.

Pelegrín-García, D. (2011), Comment on "Increase in voice level and speaker comfort in lecture room"'. Journal J. Acoust. Soc. Am., 129, 1161-1164.

Pelegrín-Garcia, D, Brunskog, J., (2009a). Development of an auditory virtual environment to measure the speakers' comfort and increase of voice levels in lecture rooms. Proceedings of the First Nordic Conference of Voice Ergonomics and Treatment

Pelegrín-García, D., Brunskog, J. (2009b) Prediction of vocal effort and speakers' comfort in lecture rooms, Proceedings of Internoise 2009, Ottawa, Canada.

Pelegrín-Garcia, D., Brunskog, J. (2010a), Natural variations of vocal effort and comfort in simulated environments. Proceedings of EAA Euroregio 2010, Ljubljana, Slovenia

Pelegrín-Garcia, D, Brunskog, J. (2011), Loudspeaker-based system for real-time own-voice auralization. Manuscript.

Pelegrin-Garcia, D., Brunskog, J., Lyberg-Åhlander, V. & Löfqvist, A. (2011), Measurement and prediction of acoustic conditions for a talker in school classrooms. Manuscript.

Pelegrín-Garcia, D., Lyberg-Åhlander, V., Rydell, R., Löfqvist, A., Brunskog, J. (2010b), Influence of classroom acoustics on the voice levels of teachers with and without voice problems: a field study. Proceedings of the 2nd Pan-American/Iberian Meeting on Acoustics and 160th ASA meeting (DOI: 10.1121/1.3533839), Cancun, Mexico.

Pelegrin-Garcia, D., Smits, B., Brunskog, J, Jeong, C.-H. (2011), 'Vocal effort with changing talker-to-listener distance in different acoustic environments', J. Acoust. Soc. Am., 129, 1981-1990.

Pelegrín-García, D., Fuentes-Mendizabal, O., Brunskog, J, and Jeong, C.H. (2011), 'Equal autophonic level curves under different room acoustics conditions', submitted to J. Acoust. Soc. Am.

Pelegrín-García, D., Fuentes-Mendizábal, O., Brunskog, J. and Jeong, C.-H. (2010), Influence of the acoustic environment on the loudness of one's own voice. Proceedings of Inter-Noise 2010, Lisbon, Portugal

List of abbreviations and symbols

ANOVA	Analysis of variance
APM	Auditory phonation monitor
BNL	Background noise level
F0	Fundamental frequency
G_{RG}	Room gain
HATS	Head and torso simulator
HL	Hearing Loss
IR	Impulse response
L_{W}	Voice power level
OR	Odds ratio
p	p-value
PCA	Principal component analysis
R^2	Coefficient of determination
RIR	Room impulse response
SD	Standard deviation
SPL	Sound pressure level
ST_V	Voice support
T_{30}	Reverberation time measured from a decay of 30 dB
VHI	Voice handicap index
VHI-T	VHI-Throat
VL	Voice level
VLI	Vocal loading index
ΔL_{A}	Relative A-weighted sound pressure level
ΔL_{Z}	Relative overall sound pressure level
ΔL_{W}	Relative voice power level
σ_{F0}	Standard deviation of the fundamental frequency

Lyberg-Åhlander *et al. BMC Ear, Nose and Throat Disorders* 2010, **10**:5
http://www.biomedcentral.com/1472-6815/10/5

BMC
Ear, Nose & Throat Disorders

RESEARCH ARTICLE **Open Access**

Throat related symptoms and voice: development of an instrument for self assessment of throat-problems

Viveka Lyberg-Åhlander*[1], Roland Rydell†[2], Jacqueline Eriksson†[1] and Lucyna Schalén†[2]

Abstract

Background: Symptoms from throat (sensation of globus; frequent throat clearing; irritated throat) are common in patients referred to voice clinics and to ENT specialists. The relation to symptoms of voice discomfort is unclear and in some cases patients do not have voice problems at all. Instruments for patients' self-reporting of symptoms, and assessment of handicap, such as the Voice Handicap Index (VHI), are in common use in voice clinics. Symptoms from throat are however only marginally covered. Purpose: To develop and evaluate an instrument that could make the patients' estimation of symptoms from the throat possible. Further to facilitate the consideration of the relation between throat- and voice problems with the Throat subscale together with a Swedish translation of the Voice Handicap Index. Finally to try the VHI with the Throat subscale: the VHI-T, for test-retest reliability and validity.

Methods: A subscale with 10 throat related items was developed for appliance with the VHI. The VHI was translated to Swedish and retranslated to English. The questionnaire was tried in two phases on a total of 23+144 patients and 12+58 voice healthy controls. The reliability was calculated with Cronbach's alpha, ICC and Pearson's correlation coefficient. The validity was estimated by independent T-test.

Results: The difference in VHI-T scores between the patients and the voice-healthy controls was significant ($p = < 0{,}01$) and there was a good correlation of the test- retest occasions. The reliability testing of the entire questionnaire showed an alpha value of $r = 0{,}90$ and that for the Throat subscale separately a value of $r = 0{,}87$ which shows a high degree of reliability.

Conclusions: For the estimation of self-perceived throat and voice problems the scale on throat related problems together with the present Swedish translation of the Voice Handicap Index, (VHI) the VHI-Throat, proves to be a valid and reliable instrument. The throat subscale seems to help revealing a category of symptoms that are common in our patients. These are symptoms that have not earlier been possible to cover with the questionnaires designed for use in the voice clinic.

Background

Patient-reported symptoms together with laryngostroboscopy and perceptual analysis of the voice are essential for the evaluation of voice in logopedic and phoniatric practice [1,2]. A number of instruments for the self-rating of voice problems have been developed for use in the voice clinic. The Voice Handicap Index (VHI) [3] along with the shortened VHI: VHI-10 [4]; the Voice Activity and Participation Profile (VAPP)[5]; the Voice-Related Quality of Life (VrQoL)[6]; the Voice Outcome Survey (VOS) [7] and the Voice symptom scale (VoiSS) [8] are all designed for measuring perceived handicap and quality of life, and perceived limitations of participation and activity.

Symptoms related to the throat, such as frequent throat clearing, irritated throat, sensation of globus, or foreign body are frequently reported by patients suffering from voice disorders. These symptoms are, however, not specific and maybe due to a multitude of underlying disorders. In the area of voice, throat symptoms may be

---

* Correspondence: viveka.lyberg_ahlander@med.lu.se
[1] Department of Logopedics, Phoniatrics and Audiology, Lund University, S-221 85 Lund, Sweden
† Contributed equally
Full list of author information is available at the end of the article

interpreted either as the *cause* of functional voice disturbances and in reverse they may also be interpreted as a *consequence* of voice load or inappropriate vocal behavior [9]. Apart from vocal behavior, non-specific mucosal hyperreactivity [10], laryngo-pharyngeal reflux [11], allergy [12] and mass lesions in the throat region are often considered as causative factors. Thus, throat related problems are a rather common concern in patients referred to voice clinics. To our knowledge, among the afore-mentioned questionnaires designed for the self-evaluation of voice problems, only the VoiSS questionnaire includes a number of items addressing pharyngeal symptoms [8]. To have a complete overview of the voice-related problems, and to meet the needs of this group of patients, this type of symptoms should also be better understood. Three self-assessment scales address only the issue of throat related symptoms; however, all scales are designed to measure problems of more diagnose-specific character, the Glasgow and Edinburgh Throat Scale, designed for the evaluation of globus [13], the Reflux Symptom Index [14] and the Pharyngeal Reflux Symptom Questionnaire (PRSQ) [15], which specifically addresses reflux.

Our aim with this paper was to develop and evaluate an instrument that could simplify the patients' estimation of symptoms from the throat and to consider their relation to voice problems simultaneously. The Voice Handicap Index (VHI), a multidimensional, self administered questionnaire, developed and validated by Jacobson et al. in 1997 [3] has been translated into many languages and is widely used in clinical work and research, with at least 200 publications up till today. At our clinic, the VHI has been in use since 2000 along with a subscale designed for the measurement of throat related symptoms, VHI-T. The VHI is an instrument that is easy to distribute and to analyze. We considered it of importance for the patient to have the possibility to judge all perceived voice- and throat symptoms in the same manner and within the same "formula", by keeping to the same rating scale and number of statements as well as to the way of phrasing the statements. We therefore choose to follow the structure of the VHI, which consequently gives the possibility to use the throat-scale as a supplement to the original VHI. The present paper thus describes the construction and validation of a scale on throat symptoms in voice patients, which may be used as a supplement to the VHI.

Methods

Study design

The study was performed in two phases. During phase 1, the original VHI was translated, the Throat subscale constructed and added to the present Swedish version of the VHI. Further, the combined Voice Handicap Index-Throat (VHI-T) was tested for validity and reliability. In phase 2, the VHI-T was re-validated and retested in a large patient-control material.

Phase 1: Translation of the VHI, development of the Throat subscale. Validity and reliability testing of the VHI-T, experts and responders

An informal, diagnostic instrument with questions on throat related problems has been in use at the phoniatric department since the early nineties. Following the decision to construct an instrument that could be combined with the VHI, the ten symptoms were chosen that had been the most frequently reported during the period of use of the informal instrument. These were suggested as a subscale. The choice of the statements was made in consensus by a panel of experienced phoniatricians and speech therapists. In congruence with the VHI, the items were phrased as statements (Table 1). The statements on throat related symptoms where further commented on and changes suggested by both a panel of experienced clinicians and by patients as described below.

The VHI covers three different domains of voice problems (physical, functional, emotional) and consists of thirty statements, ten in each domain. The statements are phrased in the way the patients normally would express themselves. The occurrence of symptoms are estimated on a frequency-based scale (0 = Never, 1 = Almost Never, 2 = Sometimes, 3 = Almost Always, 4 = Always). In the original questionnaire by Jacobson et al [3], the statements are mixed. In the layout by Rosen and Murry [16], the statements are grouped into three separate domains (commonly called sub scales) with ten statements each. This layout is, in our opinion, more convenient in clinical work. For this translation and adaptation of the VHI into Swedish, the layout proposed by Rosen and Murry was used [16].

When translating an instrument, it is important not only to perform a correct translation but also to make a cross-cultural adaptation of the instrument [17,18]. The original three subscales of the VHI were translated into Swedish by a multidisciplinary experienced expert group of two speech pathologists and three phoniatricians. All items were discussed and language adjustments were made to meet the wordings normally used by our patients.

In purpose to estimate the severity of the self perceived voice problems a 10 cm Visual Analogue Scale was added to the questionnaire, where 0 = no voice problems and 10 = maximal voice problems. This parameter has been used in former studies for estimating the reliability of the VHI [3,19]. The VHI, along with the Throat subscale, were then translated and retranslated to and from English by a professional translator. After a final agreement of the expert group, the questionnaire was submitted to an external group of experts, three phoniatricians, six

Table 1: The statements of the throat subscale with corrected item-total correlation.

Statement	Corrected item-total correlation
1 Jag är torr i halsen (*My throat is dry)*	0.457
2 Jag måste harkla mig (*I need to clear my throat)*	0.625
3 Jag har mycket slem i halsen (*I have a lot of phlegm in my throat)*	0.583
4 Jag känner att det sitter något i halsen (*It feels as if something is stuck in my throat)*	0.683
5 Det svider i halsen (*My throat is burning)*	0.572
6 Jag känner ett tryck utanpå halsen (*I feel a pressure on the outside of my throat)*	0.403
7 Det känns som om jag har en klump i halsen (*It feels like a lump in my throat)*	0.675
8 Jag är irriterad i halsen (*I have an irritation in my throat)*	0.765
9 Jag har ont i halsen (*I have a sore throat)*	0.480
10 Jag har rethosta (*I have a dry cough)*	0.420

Statements are in Swedish, English within brackets

speech therapists and one singing teacher for comments on the *usefulness* and *face validity* of the questionnaire. Simultaneously, a group of ten consecutive patients were interviewed for comments on the accessibility, the degree of user-friendliness and comments on possible changes to the statements. Further, 150 consecutive patients, referred to the department for voice problems, completed the questionnaire and commented on it. Adjustments of the Throat subscale were done accordingly. For example "I have a sensation of mucus trickling down my throat" was omitted due to rarely being graded higher than 0 or 1. The patients' demographic characteristics or diagnoses were not considered during this phase.

The VHI and the Throat subscale were then submitted to the first phase of testing after minor adjustments. This test-retest procedure included another 40 consecutive patients with voice problems (20 patients with phonastenia and 20 with benign lesions of the vocal folds) and 20 voice-healthy controls from the orthopaedic out-ward department. All responders were to complete two questionnaires with at most one week in between. The first questionnaire was to be completed before the clinical examination and the second to be returned one week later. The two questionnaires were completed and returned in due time by 23 patients (16 F:7 M, median age 54 yrs, range:25-71) and 12 controls (5F:7 M, median age 39, range: 21-71). The testing revealed good reliability. However some items needed rephrasing. Examples of changes: the item "my voice causes me to lose income" (functional scale) that was changed to "my voice restricts my work-life" due to cultural differences between reimbursement systems; "my voice sounds creaky and dry" (physical scale) which was considered by patients to be difficult to answer and was changed to "my voice sounds hoarse" following the phrasing normally used by patients. Therefore the final version of the questionnaire had to be tried once more for both reliability and validity (phase 2).

Phase 2: VHI-T, revalidation and retesting

This study is based on VHI-Throat questionnaires, i.e. the three original VHI subscales (physical, functional and emotional) along with the Throat subscale. Each maximum subscale-score is 40 p and the total VHI-T is 160 p. The questionnaires were collected from 262 persons. The responders were assigned to four patient groups and one group of controls. To be included, the responders had to be older than twelve years and competent to fill out the questionnaire without help. Twelve/156 patients were excluded due to no response or late return of the second questionnaire. Of the controls, 48/106 persons were excluded due to incomplete questionnaire or late, or no, return of the second questionnaire. This paper thus reports data from 144 patients and 58 controls.

The evaluation of the patients was performed at the Department of phoniactrics, ENT clinic, Lund University Hospital, by the same three phoniatricians with long-lasting, close clinical co-operation, and consensus as to diagnostic criteria of voice disorders. The diagnoses were classified according ICD-10, Swedish version. (Svensk foniatrisk-logopedisk diagnosklassifikation, approved by the Swedish national board of health and welfare 01012000), based on clinical history, videolaryngostroboscopy or high speed filming, and perceptual voice analysis.

The patients were diagnosed with one of the following: *phonastenia* (n = 20;defined by vocal fatigue as a cardinal symptom, without any pathological laryngeal findings, with or without subjective hoarseness); *benign lesions of the vocal folds* (n = 41; 17 polyps; 6 cysts; 5 of each nodules and sulcus glottidis; 3 papillomas; two of each vascular dilatation in the mucosa or atrophy of the vocal folds; and one granuloma); *neurological laryngeal motility dis-*

order (n = 20; 18 cases with unilateral paresis of the vocal folds and two cases with spasmodic dysphonia); *benign goitre* (N = 41; all referred to the clinic for pre-surgery control), and patients referred for *throat problems* as cardinal symptoms (N = 22), not themselves complaining of voice problems. The Control group (N = 58) consisted of out-ward patients from the orthopaedic department, all reporting voice health and no former contact with voice clinicians. Table 2 presents demographic data on the included responders according diagnose.

The patients diagnosed with benign goitre and throat related problems were only included for the estimation of the validity. Retesting was not performed in these two groups. The reason for excluding the retesting of the benign goiter group was that the patients were to undergo thyreoid surgery, close after the consultation. The clinical experience is that this surgery may cause slight voice and throat complaints. The patient group with throat problems was included later in the study for the testing of validity and thus did not take part in the retesting procedure.

The reliability of the VHI-Throat was evaluated by a test-retest procedure. The distribution and collection of the questionnaires were identical to the procedure used in phase 1. The questionnaire was first administered to all patients on arriving for their primary consultation at the phoniatric department, to be completed before the clinical examination. After one week, a new questionnaire was sent to all the patients, to be completed and returned within one more week. The Controls completed the questionnaire at the orthopedic out-ward department. They were given the second questionnaire at the same occasion, and were asked to return it within two weeks. The reason for using a different way of distributing the second questionnaire to the controls, was based on earlier experience from phase one. Namely, the control persons did not return the second questionnaire when it was mailed to them. The compliance improved when the second questionnaire was handed to the controls after the completion of the first. The validity of the VHI-Throat was assessed by comparing the whole group of patients to the group of controls.

Statistics

The test-retest reliability for the VHI- Lund total scores, for the values of the subjective voice estimation, and for the Throat subscale was estimated by calculating the IntraClass Correlation coefficient (ICC). For the construct validity, independent samples t-tests were used to compare the average scores of the VHI-Throat total, subjective voice estimation values and the Throat subscale between patients and controls. The Pearson product-moment correlation coefficient was used for computing the correlations between the subscales and the VHI-Throat total score, the throat subscale and the original VHI subscales and for estimating the correlation between the subjective assessment of voice and VHI-Throat total score. The internal consistency and reliability of the total VHI-Throat subscale, as well as of the throat subscale, were calculated with inter-item correlation and Cronbach's alpha coefficient. An ANOVA was performed to further analyze the VHI-T subscales. Analyses were performed using SPSS 15.0 and 16.0 for Windows. Alpha levels were set at 0,05%. (, ICC) and 0,01% (Pearson)

Ethical aspects

The study was approved by the ethical committee at Lund University (No LU 366-01).

Results

The throat subscale, validation process

The statements of the throat subscale are presented in Table 2. The *face* and *content validity* were tested during phase 1, see Methods section above. The *test-retest reliability* of the throat subscale was estimated with ICC: r = 0,871, in 144 patients and 58 controls, proving the scale to be stable and reliable.

Construct validity and internal consistency

The average score of the throat subscale in all 144 patients (M = 13,5. Sd = 6,8) was significantly different from that in the controls (M = 6,9 Sd = 5,5), t(178) = 6,8, $p < 0.01$, proving the throat subscale to be sensitive enough to differentiate between subjects with throat problems and healthy controls (Table 3). The Cronbach's alpha coefficient for the throat subscale was r = 0,87. In Table 2, all statements of the throat sub scale are given along with the corrected item-total correlations, reflecting the degree to which each statement correlates to the total score of this scale. The criterion for inclusion of an item in a subscale is an item-total correlation of > 0.3. As shown in Table 1, the corrected item-total correlations for all statements exceeded 0.4, thus indicating satisfac-

Table 2: Demographic data for the five groups of patients and one group of voice healthy controls

	Phonastenia	Benign lesions	Neurolog. disorders	Throat rel.	Benigngoitre	Controls
N	20	41	20	22	41	58
F:M	15:5	30:11	12:8	11:11	30:11	31:27
Median Age (range)	52 (18-69)	45 (13-74)	56 (26-76)	58 (20-73)	48 (19-79)	60,5 (15-80)

Table 3: Results of T-test between patients and voice healthy controls for the VHI-Throat subscales.

		M score (Sd)	t	df	P = (2-tailed)
Throat scale	Patients	14,5 (7,3)			
	Controls	6,9 (5,5)	8,1	138	,001
Functional	Patients	9,5 (9,7)			
	Controls	1,8 (3,4)	8,3	197	,001
Physical	Patients	15,1 (9,8)			
	Controls	5,4 (5,6)	8,8	178	,001
Emotional	Patients	8,7 (9,5)			
	Controls	1,3 (3,1)	8,4	194	,001

Patients n = 144, Controls n = 58

tory correlation of the statements within this subscale. When item-total correlation was calculated for all items of the VHI-T (Appendix), the values were somewhat lower for the **t**hroat subscale, however no item scored < 0.3.

VHI-Throat: the VHI questionnaire and the throat subscale, reliability and validity

Test-retest reliability, construct validity and internal consistency

The test-retest reliability of the total VHI-T score was estimated with IntraClass coefficient (ICC): = 0,968, proving good reliability of the questionnaire. A paired samples revealed no significant differences between the first and second occasion for neither the total VHI-T scores (M = 1,6, Sd = 41,6, N = 142), $t(141) = 0{,}464$, $p = 0{,}6$ nor the individual subscale scores (Throat: (M = 0,9, Sd = 10,4, N = 142), $t(141) = 1{,}0$, $p = 0{,}2$, Functional: (M = 0,5, Sd = 12,4, N = 142), $t(141) = 0{,}526$, $p = 0{,}6$, Physical: (M = 0,3, Sd = 13,1, N = 142), $t(141) = 0{,}351$, $p = 0{,}7$, Emotional: (M = -0,3, Sd = 13,2, N = 142), $t(141) = -0{,}2$, $p = 0{,}8$) in patients and controls. The VHI-T total score in all patients (M = 47,8, Sd = 30,2, N = 144) was significantly higher than in the controls (M = 15,3SD = 15,0N = 58), $t(191) = 10{,}2$, $p < 0.05$ (2-tailed), thus indicating that the questionnaire separated persons with and without voice pathology. Independent Samples t-tests were also calculated for the subscales, showing significant differences between patients and controls for the three original subscales and the throat subscale. (Table 3) The Cronbach's alpha coefficient was $r = 0{,}90$ for the total VHI-T scale and $r = 0{,}93$ if the throat subscale would be excluded. There was a strong correlation between each of the four subscales and the total score for VHI-T, respectively, as shown by Pearson's correlation coefficient: throat subscale $r = 0{,}684$, functional scale $r = 0{,}921$, physical scale $r = 0{,}931$ and emotional scale $r = 0{,}915$. A one-way analysis of variance showed significant differences at the p < .05 level in subscale scores between the groups of patients: Throat scale: F(5,193) = 18,4, $p = .000$; Functional scale: F(5,193) = 48,1, $p = .000$; Physical scale: F(5,193) = 57,7, $p = .000$; Emotional scale: F(5,193) = 37,4, $p = .000$. Further analysis with Tukey HSD test for the Throat scale indicated statistically significant differences between the mean scores for the phonastenia group ($M = 14{,}8$, $Sd = 6{,}3$) and the control group ($M = 6{,}9$, $Sd = 5{,}7$); between the benign lesions group ($M = 15{,}8$, $Sd = 6{,}7$) and the benign goiter group ($M = 10{,}3$, $Sd = 6{,}4$) as well as the control group ($M = 6{,}9$, $Sd = 5{,}7$); between the benign goiter group ($M = 10{,}3$, $Sd = 6{,}4$) and throat related group ($M = 19{,}8$, $Sd = 5{,}6$); between the neurolog. disorder group ($M = 14{,}1$, $Sd = 8{,}1$) and the throat related group ($M = 19{,}8$, $Sd = 5{,}6$) as well as the control group ($M = 6{,}9$, $Sd = 5{,}7$).

The relation of the throat scale and the VHI

The correlation between the throat scale and the three original VHI subscales was calculated with Pearson's correlation coefficient: functional scale: $r = 0{,}356$ physical scale: $r = 0{,}544$; emotional scale: r = 0,395, thus suggesting a moderate to strong correlation with the physical scale and a moderate correlation with the functional and emotional subscales.

The relation of the throat scale to the VHI-T total score

The mean scores of the four VHI-T subscales and the VHI-T total score for each diagnose group are presented in Table 4. Table 4 also shows the relation between each subscale and the total scores of the VHI-T in percent and thus indicates the dominating subscale or subscales for each diagnose. The diagnoses follow two different patterns based on the relation between the subscale-scores. The distribution of the scores for the neurological disorders, benign lesions and phonastenia is even, with close to 25% for each subscale. The throat subscale scores for benign goiter and throat-related disorders account for more than 50% of the total VHI-T score.

Subjective estimation of the voice with Visual Analogue Scale (VAS)

Test-retest reliability and construct validity

The reliability of the subjective estimation of the voice was calculated by ICC and showed a moderate-strong correlation: r = 0,712, N = 202, p < 0,05, proving it as a satisfactory stable instrument. Calculation with independent T-test showed that the difference in the subjective estimation of the own voice between the patient-group (M = 43,8 Sd = 31,2 N = 122) and control-group (M = 14,3 Sd = 19,8 N = 58), was significant $t(163)$ = 7,7, p < 0,05. The results indicate that this instrument was sensitive enough to separate patients from controls.

Correlation between estimation of one's own voice and VHI-T total score

The correlation between the subjective estimation of the voice and the total VHI-T was a moderate when tested in all patients and controls using Pearson's product-moment correlation coefficient (r = 0.79 n = 202, p < 0.01.) For the different groups the correlation coefficient varied: Phonastenia group (0,48), Benign lesions (0,69), Neurological group (0,70), Benign Goitre (0,68), Throat related disorders (0,64), and Controls (0,32).

Discussion

The need to estimate throat problems in the voice clinic

In the Swedish healthcare system, patients with a broad spectrum of voice and voice related problems are diagnosed and treated at logopedic-phoniatric departments. In our daily practice, we have experienced that many patients report more physical aspects than those covered by the original VHI domains (functional, physical, and emotional domains). This was the impetus to create the throat subscale. Throat problems are ascribed to a multitude of etiologies, are common in voice patients and considered to be cardinal symptoms in patients with vocal fatigue. The need of a structured broader aiming instrument, for the self-assessment of the problems patients report in the voice clinic has also been emphasized by Deary et al [8] and Glas et al [20]. We share the view of these authors that the spectrum of patient-reported problems in the voice clinic is broader than the "classical" voice symptoms, and are not uncommonly symptoms that originate from throat.

VHI-Throat, a questionnaire

The VHI-Throat (VHI-T) questionnaire showed good test-retest reliability, validity and internal consistency. According to the present results, it seems that the throat subscale fends for itself as indicated by the Cronbach's alpha value as well as the corrected inter-item correlation analysis (see Table 1 and additional file 1) and by the correlation between the throat scale and the original three VHI subscales. The total score and the scores of the three original VHI subscales were comparable to those in corresponding groups of patients in other studies [4,19,21]. The VHI-T thus seems to be an appropriate tool for clinical use in Swedish speaking populations, also being patient-friendly and convenient to administer and evaluate.

Our results show that the Throat-subscale in combination with the VHI is an instrument that may make it possible to discriminate between voice and throat problems and to help the patient express both categories of concerns simultaneously. To our knowledge, until today there has been no instrument developed for the estimation of the patient's overall description of symptoms in the voice clinic, where many patients with throat-problems are referred. A deeper insight in the problems may lead to an increased understanding of the patient with throat complaints, with or without voice complaints. This knowledge may be helpful in designing the clinical intervention. However, it does not give us any indication of the origin of the problems.

Our results from the voice-healthy subjects show that it is not uncommon to report some symptoms from the throat. Moreover, our results indicate that patients who report problems mainly from the throat also have some complaints on the physical subscale. This is in accordance with the findings of Belafsky et al, who found a decrease on the physical subscale after the treatment of laryngeal reflux [14]. We believe that the VHI-T may become a use-

Table 4: Mean scores of the VHI-T subscales, percentage of the subscales of the total VHI-T scores.

	Throat		**Functional**		**Physical**		**Emotional**		**Tot VHI-T**	
	M (Sd)	**%**	**M (Sd)**	**%**	**M (Sd)**	**%**	**M (Sd)**	**%**	**M (Sd)**	**%**
Neurological N = 20	14 (8)	*20*	19 (8)	*27*	21 (6)	*30*	16 (8)	*24*	70 (22)	*100*
Ben. Lesions N = 41	16 (7)	*23*	16 (9)	*22*	29 (7)	*42*	15 (10)	*22*	70 (27)	*100*
Phonastenia N = 20	15 (6)	*30*	10 (7)	*20*	16 (6)	*34*	9 (6)	*18*	49 (19)	*100*
Ben. Goitre N = 41	10 (6)	*52*	2 (5)	*12*	6 (6)	*29*	1 (4)	*8*	20 (18)	*100*
Throat rel N = 22	20 (7)	*56*	2 (2)	*5*	10 (7)	*28*	4 (5)	*11*	36 (15)	*100*
Controls N = 58	7 (5)	*45*	2 (3)	*12*	5 (6)	*35*	1 (3)	*9*	15 (15)	*100*

ful clinical instrument that may help to discriminate the problems that might be either co-existing or occurring separately. However, sharing the opinion of Verdonck et al [21], to be able to pin-point the focus of the patient's problems it might be more rewarding to evaluate the subscale scores of the VHI, rather than the total score.

The way of collecting the second questionnaire (see methods) might of course have brought bias into the results. Based on earlier experience, the second questionnaire was given to the voice-healthy controls already at the completion of the first questionnaire, where the patients were sent the second questionnaire by mail. Even though all subjects included returned the second questionnaire within two weeks, we have no means of knowing when the second questionnaire actually was completed by the controls.

The VHI and the VHI-Throat

The Voice Handicap Index is today widely used in clinic and research. Despite some recent critical opinions that the VHI lacks statistically discrete subscales [4], it still fills the purpose of covering the self perceived voice problems and also the consequences for the quality of life that voice disorders may lead to. We have used a Swedish translation of the VHI in clinic since 2000 and it was therefore natural to choose the VHI as a base for the development of the throat subscale.

The use of VHI and other self-reporting instruments within the voice clinic has had an eye-opening effect since the patient's own estimation of the symptoms thus has come more into focus. The VHI-T is designed as an instrument for the patient to estimate the perceived problems and, in our experience the throat subscale is a good complementary tool to the VHI, allowing a better identification of actual disorders. Consequently, we can better design more appropriate therapeutic interventions. Some patients call for medical consultation specifically due to throat-related symptoms, but quite often the referring physician may interpret the symptoms as signs of a voice disorder. The use of the compiled VHI-T may thus direct the clinician to a more appropriate intervention.

Interestingly, our results indicate that it may be possible to identify two "profiles" of symptoms characterising different groups of patients. As is evident in Table 4, voice healthy controls-, benign goitre- and throat-groups report the lowest total VHI-T scores (15-36) but the percentage of their indicated throat problems is high relative to the total score. Conversely, the patients with benign laryngeal lesions report the highest VHI-T total score (70) with rather equal distribution of symptoms over the four subscales. Further studies are, however, necessary in order to estimate the usefulness of "profiles" for the clinical evaluation of individual patients. The ANOVA showed significant differences in the subscale scores between the patient groups. However, we wish to be cautious in interpretation of these findings. The VHI is a self rating instrument of symptoms and has as such not been intended as a differential diagnostic instrument. The differences between the patients' "profiles" emerging from Table 4, may however, be used for evaluating the effect of therapy within individual patients. Since the results of the validation of the original VHI-subscales within this study are in accordance with the results of other studies [3,21,22] we may suggest that the throat subscale can be used for clinical and research purposes along with any validated VHI version.

The subjective estimation of the voice with VAS

The subjective estimations of the voice with VAS showed good test-retest reliability. Correlations between the subjective estimations of the voice and the overall VHI-T score were reliable in the whole population but varied between the different diagnostic groups. Subjective estimation of voice is usually used only for proving the face validity of the VHI-questionnaire [3,5,19]. We choose to include this simple measure as a permanent item in the questionnaire. It gives a quick overview of the patient's own grading of the voice problems [1].

As in other studies [3,5,19], we also found a good correlation between the average scores from the subjective estimation of the voice and the total score of VHI-T, however with varying correlations between the diagnose groups. A discrepancy between VHI-T and VAS may be of interest since it may reflect the patient's attitude to his/her symptoms: a patient who has a combination of high VHI-T total score and a low value of self-estimation of the voice may in fact not value the symptoms as a big trouble while another individual with the reverse relationship between the self-estimation of the voice and VHI-T total values the symptoms as less tolerable. This information cannot be underestimated when taking care of the patients in voice therapy, not least since it may actually give a hint of the patient's motivation to complete the therapy.

Conclusions

The present Swedish translation of the VHI with the subscale on throat-related problems, the VHI-Throat, proves to be a valid and reliable instrument for the estimation of self-perceived voice and throat problems. The use of the throat subscale helps to reveal a category of symptoms that are common in our patients and that are only marginally covered in other available instruments. In analogy with other translations of the VHI, it can be used for both clinical purposes and for clinical research.

Authors' informations

VLÅ: Speech therapist MSc. PhD student at the dep of logopedics, phoniatrics and audiology, Lund University. LS: MD, PhD, consultant in phoniatrics, ENT dep Lund

University Hospital. RR: MD, PhD, consultant in phoniatrics and laryngology, ENT dep Lund University Hospital and dep of logopedics, phoniatrics and audiology, Lund University. JE: Speech therapist MSc, dep of logopedics, phoniatrics and audiology, Lund University.

Additional material

Additional file 1 Values of the corrected item-total correlation between the statements of the VHI-T. This file represents a table showing all statements in the VHI-T in Swedish and English, and also showing the values of the corrected item-correlation between the statements.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

This work is performed in close collaboration by the authors. VLÅ and LS initiated, designed and coordinated the study and also carried out the distribution of the questionnaires. VLÅ analysed the data and drafted the manuscript. LS helped discussing and drafting the manuscript and examined the patients. JE collected and analysed the data for the voice- healthy controls under supervision of LS and VLÅ. RR examined the patients, participated in the expert group and helped drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to express their gratitude to reg. Speech therapists Julia Lindbom, Karin Petersson and Sara Wedin for the collection and preparation of data during the first phase of this study. We are also sincerely grateful to Professor Anders Löfqvist whose valuable advice and opinions on the manuscript have been of great importance.

Author Details

[1]Department of Logopedics, Phoniatrics and Audiology, Lund University, S-221 85 Lund, Sweden and [2]ENT-Department, Lund University Hospital, S-221 85 Lund, Sweden

Received: 15 January 2010 Accepted: 27 May 2010
Published: 27 May 2010

References

1. Dejonckere PH: **Clinical implementation of a multidimensional basic protocol for assessing functional results of voice therapy. A preliminary study.** *Rev Laryngol Otol Rhinol (Bord)* 2000, **121(5)**:311-313.
2. Carding PN, Wilson JA, MacKenzie K, Deary IJ: **Measuring voice outcomes: state of the science review.** *The journal of laryngology and otology* 2009, **123(8)**:823-829.
3. Jacobson B, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger M: **The Voice Handicap Index (VHI): Development and Validation.** *American Journal of Speech-Language Pathology* 1997, **6(3)**:86-70.
4. Rosen CA, Lee AS, Osborne J, Zullo T, Murry T: **Development and Validation of the Voice Handicap Index-10.** *The Laryngoscope* 2004, **114(9)**:1549-1556.
5. Ma EP, Yiu EM: **Voice activity and participation profile: assessing the impact of voice disorders on daily activities.** *J Speech Lang Hear Res* 2001, **44(3)**:511-524.
6. Hogikyan ND, Wodchis WP, Terrell JE, Bradford CR, Esclamado RM: **Voice-Related Quality of Life (V-RQOL) following type I thyroplasty for unilateral vocal fold paralysis.** *Journal of Voice* 2000, **14(3)**:378-386.
7. Glicklich R, Glovsky R, Montgomery W: **Validation of a voice outcome survey for unilateral vocal cord paralysis.** *Otolaryngology -Head and Neck surgery* 1999, **120(2)**:153-158.
8. Deary IJ, Wilson JA, Carding PN, MacKenzie K: **VoiSS - A patient-derived Voice Symptom Scale.** *Journal of Psychosomatic Research* 2003, **54(5)**:483-489.
9. Nguyen DD, Kenny DT, Tran ND, Livesey JR: **Muscle Tension Dysphonia in Vietnamese Female Teachers.** *Journal of Voice* 2009, **23(2)**:195-208.
10. Lyberg Ahlander V, Malm L, Schalen L: **Hoarseness as a Sign of Possible Nonspecific Mucosal Hyperreactivity in Vocal Tract.** *Journal of Voice* 2009, **23(6)**:707-715.
11. Ross JA, Noordzji JP, Woo P: **Voice disorders in patients with suspected laryngo-pharyngeal reflux disease.** *Journal of Voice* 1998, **12(1)**:84-88.
12. Geneid A, Ronkko M, Airaksinen L, Voutilainen R, Toskala E, Alku P, Vilkman E: **Pilot study on acute voice and throat symptoms related to exposure to organic dust: Preliminary findings from a provocation test.** *Logopedics Phoniatrics Vocology* 2009, **34(2)**:67-72.
13. Deary IJ, Wilson JA, Harris MB, Macdougall G: **Globus pharyngis: development of a symptom assessment scale.** *Journal of Psychosomatic Research* 1995, **39(2)**:203-213.
14. Belafsky PC, Postma GN, Koufman JA: **Validity and Reliability of the Reflux Symptom Index (RSI).** *Journal of Voice* 2002, **16(2)**:274-277.
15. Andersson O, Ryden A, Ruth M, Moller RY, Finizia C, Titze K, Fau-Wollenweber S, Wollenweber S, Fau-Nell V, Nell V, Fau-Lehmkuhl U, Lehmkuhl U: **Development and validation of a laryngopharyngeal reflux questionnaire, the Pharyngeal Reflux Symptom Questionnaire.** *Scand J Gastroenterol* 2009 in press.
16. Rosen CA, Murry T, Zinn A, Zullo T, Sonbolian M: **Voice handicap index change following treatment of voice disorders.** *Journal of Voice* 2000, **14(4)**:619-623.
17. Guillemin F, Bombardier C, Beaton D: **Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines.** *Journal of Clinical Epidemiology* 1993, **46(12)**:1417-1432.
18. Sullivan M: **Konsten att läsa en artikel om livskvalitetsforskning - nu finns checklistan för att underlätta en kritisk värdering.** *Läkartidningen* 2002, **99(26-27)**:2933-2938.
19. Helidoni ME, Murry T, Moschandreas J, Lionis C, Printza A, Velegrakis GA: **Cross-Cultural Adaptation and Validation of the Voice Handicap Index Into Greek.** *Journal of Voice* 2010, **24(2)**:221-227.
20. Glas K, Hoppe U, Eysholdt U, Weber A, Rosanowski F: **Klagsamkeit von Patienten mit Dysphonien - Complaints of patients with dysphonia.** *HNO: Hals-, Nasen-, Ohrenaerzte* 2008, **56(10)**:1057-1066.
21. Verdonck-de Leeuw IM, Kuik DJ, De Bodt M, Guimaraes I, Holmberg EB, Nawka T, Rosen CA, Schindler A, Whurr R, Woisard V: **Validation of the Voice Handicap Index by Assessing Equivalence of European Translations.** *Folia Phoniatrica et Logopaedica* 2008, **60(4)**:173-178.
22. Ohlsson A-C, Dotevall H: **Voice handicap index in Swedish.** *Logopedics Phoniatrics Vocology* 2009, **34(1)**:1-7.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1472-6815/10/5/prepub

doi: 10.1186/1472-6815-10-5
Cite this article as: Lyberg-Åhlander *et al.*, Throat related symptoms and voice: development of an instrument for self assessment of throat-problems *BMC Ear, Nose and Throat Disorders* 2010, **10**:5

Speaker's Comfort in Teaching Environments: Voice Problems in Swedish Teaching Staff

Viveka Lyberg Åhlander, Roland Rydell, and Anders Löfqvist, *Lund, Sweden*

Summary: Objectives. The primary objective of this study was to examine how a group of Swedish teachers rate aspects of their working environment that can be presumed to have an impact on vocal behavior and voice problems. The secondary objective was to explore the prevalence of voice problems in Swedish teachers.

Method. Questionnaires were distributed to the teachers of 23 randomized schools. Teaching staff at all levels were included, except preschool teachers and teachers at specialized, vocational high schools. The response rate was 73%.

Results. The results showed that 13% of the whole group reported voice problems occurring sometimes, often, or always. The teachers reporting voice problems were compared with those without problems. There were significant differences among the groups for several items. The teachers with voice problems rated items on room acoustics and work environment as more noticeable. This group also reported voice symptoms, such as hoarseness, throat clearing, and voice change, to a significantly higher degree, even though teachers in both groups reported some voice symptoms. Absence from work because of voice problems was also significantly more common in the group with voice problems—35% versus 9% in the group without problems.

Conclusion. We may conclude that teachers suffering from voice problems react stronger to loading factors in the teaching environment, report more frequent symptoms of voice discomfort, and are more often absent from work because of voice problems than their voice-healthy colleagues.

Key Words: Environmental loading factors–Voice–Room acoustics–Self-evaluation–Epidemiology–Teachers–Vocal symptoms–Hoarseness.

BACKGROUND

This article presents the first phase of a project to investigate the influence of the room acoustics on teachers' voices and vocal behavior. The first aim of this study was to examine how a group of teachers rate aspects of their working environment that can be presumed to have an impact on vocal behavior and, thus, also on voice problems. The issues of classroom acoustics are of special interest. Second, the study investigates the prevalence of voice problems among Swedish teachers.

The voice is one of the most important tools that a professional educator has at hand. The voice is used for communicating with the pupils, that is, discussing, instructing, and clarifying. Thus, the teacher depends heavily on the voice, which has to be durable, flexible, and reliable. Because teachers form one of the vocational groups that have substantial vocal demands in their professional practice,[1–3] they are at risk of voice problems, also when compared with other vocations.[2–6] Fritzell[1] concluded more than 10 years ago that teachers were overrepresented in clinical populations, because 16.3% of the patients in the waiting rooms at Swedish voice clinics were teachers, but they made up only 5.9% of the Swedish working population at that time.

Vocal load is commonly seen as one of the most important causes of voice dysfunction in teachers.[7–9] Vocal loading is a combination of prolonged voice use and additional factors, such as background noise, acoustics, and air quality, which affect the fundamental frequency, mode, and intensity of phonations as well as the external frame of the larynx.[8] In a questionnaire study, McAleavy et al[10] showed that teachers rated the following three independent variables as most directly influencing the vocal health: voice-related behaviors, environmental aspects, and the teacher's anxiety. However, the concept of environmental influence may cover a variety of aspects, including room acoustics. Classroom acoustics has indeed been extensively studied, but there is a lack of studies linking the room acoustic parameters to the voice produced by the speaker.[11] According to Vilkman,[8] deficient acoustics may lead to a decrease in the intelligibility of the speech. Such a decrease in intelligibility is mainly a problem for the listener, but may, in an interactive situation, result in repetitions and probably also in increased voice intensity by the speaker to facilitate the understanding of the message. Pekkarinen et al[5] and Sala et al[12] found that the acoustic conditions in classrooms were often deficient because of noise and reverberation, and they also linked the vocal problems of preschool teachers to the room acoustics. The commonly discussed conditions in Swedish classrooms, not least in the media, are the high levels of noise caused by the students. To be heard above the background noise, the teacher unconsciously tries to increase the voice intensity, known as the Lombard effect,[13] and, thus, increases the phonatory effort. An increased phonatory effort results in an increased F_0, which seems to be one of the major contributors to vocal fatigue.[9] Brunskog et al[11] concluded that a comfortable room from the speaker's point of view is the one that provides support for the voice and helps to project the voice toward the audience. Furthermore, the changes in voice intensity were correlated with the size of the room and the gain produced by the room, where the gain is based on the relationship between the direct and early reflected sounds.

Accepted for publication December 10, 2009.

From the Department of Logopedics, Phoniatrics and Audiology, Clinical Sciences Lund, Lund University, Lund, Sweden.

Address correspondence and reprint requests to Viveka Lyberg Åhlander, Department of Logopedics, Phoniatrics and Audiology, University Hospital, SE-221 85 Lund, Sweden. E-mail: viveka.lyberg_ahlander@med.lu.se

Journal of Voice, Vol. ■, No. ■, pp. 1-11
0892-1997/$36.00

doi:10.1016/j.jvoice.2009.12.006

Several authors have studied the prevalence of voice problems in teachers.[6,14–18] There is a general agreement about the significance of the problem, although the prevalence of voice problems in professional educators varies among studies. Roy et al[15] found that 29.9% of the interviewed teachers had suffered from voice problems at some point during their professional careers. This is similar to the findings of Russell et al,[19] where 20% of the teachers were found to have voice problems. Much larger numbers are reported by McAleavy et al,[10] who found that 67% of the teachers had experienced voice or throat problems over the past year and, in a recent study by de Medeiros et al,[18] it was found that only one-third of the investigated teachers were free from voice problems. The picture is thus fragmented, and the Swedish situation is unknown. Findings from Finland suggest that voice disorders are a growing problem in teachers.[16]

METHODS AND PROCEDURES

Questionnaire

A screening questionnaire has been developed to assess teachers' ratings of their working environment and also to estimate the prevalence of voice problems in teachers. The questionnaire covered 52 items in three main domains: (1) background information; (2) room acoustics, perception of noise levels, and other issues related to the environment (items 1–13); and (3) voice problems, vocal behavior, and statements about skills in voice use (items 14–32). Most of the questions about voice were taken from a Swedish validated version of the Voice Handicap Index. Items in part 1 were answered by yes or no or description in free text. The items in part 2 were statements, for example, "The air in the classroom is dry," which were rated on a scale from 0 to 4, where 0 = completely disagrees and 4 = completely agrees. The items in part 3 were statements, for example, "I have to clear my throat," which were rated on a frequency-based scale from 0 to 4, in accordance with the scale in the Voice Handicap Index, where 0 = never, 1 = occasionally, 2 = sometimes, 3 = often, and 4 = always. The questionnaire was tested in a pilot study of 63 teachers, all permanent staff of one high school. A reference group attached to the project (experts in occupational and environmental medicine, voice, acoustics, and representatives of the teachers' unions, and building proprietors) also made comments. The validity of the questions was also discussed by a group of experienced teachers, representing the different teaching levels included in the study. Based on the pilot study and the feedback, the questionnaire was revised into its final form.

Selection of teachers and schools

Teaching staff at all levels were included, except preschool teachers and teachers at specialized, vocational high schools, because of the large variety of teaching premises; see Table 1 for the distribution of teaching levels.

TABLE 1.
Teaching Level of 467 Teachers

Teaching level	Teachers, N
Junior + intermediate school	203
Secondary school	108
High school	156

The teachers were accessed through the headmasters of a number of schools in the region. The schools were randomly selected from a restricted geographical area. The choice of the area was based on a uniform distribution of air pollution and also on an equivalent population density within the area. Of the 53 schools contacted, 22 accepted to participate in the study. A permission to distribute the questionnaire was obtained from the headmaster of each school. The teachers were informed about the study by one of the authors (V.L.Å) at regular, prescheduled, compulsory collegial meetings at each school. The questionnaire was distributed, completed, and collected during one and the same meeting. The teachers completed the questionnaire anonymously. If, however, a teacher was interested in continued participation in the project, contact information was obtained on a voluntary basis. All teachers participating in the conferences answered the questionnaire. The questionnaire was thus completed by 73% of all the teachers of all the included schools. The visits were mainly made from January to April 2009.

Subjects

The distribution of age, gender, and years in vocation for the 467 teachers is presented in Table 2. Only a small number were current smokers—36 out of 467 subjects—whereas 158 had given up smoking during a time period ranging from 0 to more than 10 years. Thirty percent of the teachers—146 out of 467—did not report any main subject in teaching, being general teachers, specialized in certain age groups. The largest group reporting a main subject was that of teachers teaching language (20%), followed by natural sciences (11%) and social sciences (9%).

Statistics and ethical considerations

A statistical power analysis based on a 20% prevalence of voice problems suggested that completed questionnaires from 398 teachers were required with a 5% margin of error. The statistical analyses were computed using *SPSS* 16.1 (SPSS, Inc., Chicago, IL). Because of the cross-sectional design of the study, no correlations are computed within the material. Factor analysis, principal component analysis (PCA), was used to uncover underlying factors and establish interactions among the answers. To assess the appropriateness of the material for PCA analysis, the correlations among the items were calculated. The eigenvalue according to Kaiser's criterion explains the amount of the total variance explained by a factor and needs to exceed 1.0. Factors were obtained with a Varimax rotation and Kaiser normalization. The chi-square test and the Mann-Whitney *U*-test were used for further statistical analyses. The alpha level for all statistical analyses was set to $P < 0.05$. The study has been approved by the Institutional Review Board at Lund University (#248/2008).

TABLE 2.
Distribution of Gender, Age, Years in Occupation, and Smoking in 467 Teachers

N	Gender, F/M	Age, Years	Smoking, N	Years in Occupation
467	336/131	M: 47 R: 23–69	36	M: 17 R: 0–43

Abbreviations: M, mean; R, range.

RESULTS

Data from 487 questionnaires were collected. Nine of the questionnaires were excluded because of incomplete data. A questionnaire was considered incomplete if lacking information on gender, age, or occupation. Eleven questionnaires were excluded, because they had been handed in by teacher students attending the collegial meeting and, thus, were not in focus of this study. A total of 467 questionnaires were evaluated in this study. Some items were left unanswered by a small number of teachers, as is common in questionnaire surveys (between zero and seven teachers/item) (Tables 3–5). These questionnaires were also included in the analysis.

Factor analysis of the questionnaire

The statements 1–13 on acoustics and environmental items and 14–32 on voice items were subjected to a PCA. Before performing the PCA, the suitability of data for factor analysis was assessed. Inspection of the correlation matrix revealed the presence of many coefficients equal to or greater than 0.3. The PCA revealed two components of eigenvalues exceeding 1 for the statements about room acoustics, explaining 29.7% and 10.7% of the variance. For the statements about the voice, four components were found explaining 39.2%, 8.1%, 7.4%, and 5.7% of the variance.

The PCA for the statements about room acoustics revealed two components, with items 1–6, 8, and 13 strongly loading on the first component and items 7, 9, and 10–12 strongly loading on the second component (Table 6). There was a moderately strong correlation between the two factors ($r = 0.542$). These findings indicate that the items listed under each component are highly loaded, specifically, onto one of these two independent underlying components.

The interpretation of the first component is the voice function and the interaction of the voice with the classroom acoustics. The second component can be interpreted as covering external sources influencing the voice use.

As shown in Table 7, the PCA for the statements about the voice revealed four components, with items 15–19, 24, and 32 strongly loading on the first component, items 22–24 loading on the second component, items 14,19, 20, 21, 23, 26–28, and 32 loading on the third component, and items 25, 29, 30 and 31 loading on the fourth component. There was a weak positive correlation between components 1 and 2 ($r = 0.338$), 1 and 4 ($r = 0.352$), and 2 and 4 ($r = 0.113$) and a weak negative correlation between components 1 and 3 ($r = -0.388$), 2 and 3 ($r = -0.306$), and 3 and 4 ($r = -0.244$). These findings indicate that the items listed under each component are highly loaded, specifically, onto one of these four independent underlying components.

The interpretation of the PCA analysis for the voice statements is that component 1 includes symptoms classically considered as early signs of voice problems and can most likely be interpreted as such also in this study, in particular, because of the inclusion of statement 32, "I have voice problems," within this component. Component 2 can be viewed as "consequences of voice problems," whereas component 3 seems to reflect functional or emotional aspects of voice problems; this component also includes statement 32. Component 4 includes symptoms from the throat.

Analysis of separate items and grouping of teachers with and without voice problems

Prevalence of voice problems. As shown in Figure 1, 37.2% of the teachers reported occasionally occurring to always-present voice problems (statement 32); 24.4% reported voice problems occurring occasionally; and 12.8% reported problems occurring sometimes–always (2–4), and of those, 3.2% reported voice problems occurring often or always (3–4).

Teachers with and without voice problems. Based on the ratings of statement 32, "I have voice problems," the participants were divided into two groups. *Group I* (N = 60) consisted of teachers having rated 2–4, that is, suffering from voice problems sometimes, often, or always. *Group II* (N = 407) included teachers having rated 0–1, that is, never or only occasionally experiencing voice problems. There were no significant differences among the groups for gender (*group I*: 80% F/20% M; *group II*: 71% F/29% M), age (*group I*: Median (Md) = 49.5, *group II*: Md = 46), smoking (*group I*: 10%, *group II*: 7%), or years of occupation (*group I*: Md = 20, *group II*: Md = 16), as shown by a chi-square test.

Items on background aspects

School level, size of class, and number of teachers in class. The teachers were distributed among the school levels investigated: 44% of the whole group of teachers taught junior/intermediate level, 23% at secondary school level, and 33% at high school level. The most common class size, reported by 81% of the teachers, was groups of 25–30 pupils. Most of the teachers taught alone (78%). However, it is increasingly common to teach in teams at the early levels. Team teaching with one or two colleagues was reported by 21% of the teachers.

The chi-square test revealed significant differences in class size between *group I*—teachers with voice problems—and *group II*—teachers without voice problems; the former had smaller classes. There were no significant differences between *groups I* and *II* in school level and number of teachers in class. The details of distributions and the results of the chi-square test are summarized in Table 3.

Possibility to rest during workday. The possibility to take a pause during the working day when needed was reported by

TABLE 3.
Comparison Between Teachers With Voice Problems (*Group I*) and Those Without Voice Problems (*Group II*) on Background Items

Question	*Group I* % (N = 60)	*Group II* % (N = 407)	χ^2	*P*-Value
School level				
Junior/intermediate	45	43	(3, N = 467) = 10.331	NS
Secondary	27	23		
High school	28	34		
Group size				
1–6	22	7	(2, N = 467) = 13.514	0.001
7–15	7	8		
15–30	71	85		
No. of teachers in classroom				
1	75	78	(3, N = 467) = 1.889	NS
2	21	16		
3	3	5		
>3	0	1		
Possibility to take a break	20	28	(1, N = 460) = 1.458	NS
Voice training	40	35	(1, N = 467) = 0.596	NS
Voice demanding spare time	22	26	(2, N = 466) = 0.595	NS
Referral for voice help	38	8	(1, N = 463) = 47.591	0.000
Sick leave	35	9	(1, N = 466) = 33.274	0.000
No. occasions of sick leave				
1	10	4		
>2	23	5	(1, N = 56) = 1.576	NS
Smoking				
Yes	10	7	(2, N = 467) = 1.424	NS
No	52	60		
Have quit smoking	38	33		
Asthma	17	8	(1, N = 466) = 5.314	0.021
Asthma medication	13	6	(1, N = 39) = 0,031	NS
Strong scents	37	21	(1, N = 464) = 8.000	0.005
Other hyperreactivity	12	7	(1, N = 40) = 1.184	NS
Hearing aid	10	2	(1, N = 464) = 11.859	0.001
Job satisfaction				
Great	52	49	(3, N = 466) = 0.897	NS
Broadly	36	41		
Not too good	12	9		
Not at all	0	1		
Voice amplification	3	1	(2, N = 464) = 4.778	NS

Differences were analyzed with the chi-square test.
Abbreviation: NS, not significant.

25%. As shown in Table 3, no significant differences were found in the possibility to take a pause between the two groups.

Previous voice training. A little more than one-third, 35%, of the teachers reported having trained their voice. The context of the training varied, but the majority, 24%, reported having had some training during their teacher education and 5% of those also in combination with singing lessons. The amount and content of the voice training is, however, unknown. Singing lessons had been attended by 11%, of whom 5% attended more than occasionally. The frequency of voice training did not differ between the two groups.

Voice load during leisure time. One-fourth, 25%, of the group reported activities during leisure time that the teachers themselves classified as vocally demanding, 12% in musical contexts (ie, choir singing/band playing), 8% within sports, and 5% in other activities (ie, study circles, acting, etc). There was no significant difference among the groups for this variable.

Earlier voice problems: sick leave and referral to professional help. Within the whole group, 12% reported having been sick-listed because of voice problems, 7% of those, more than once. Referral to professional help because of voice problems was also reported by 12%. However, in the group with

TABLE 4.
Distribution of the Ratings in Percent of Statements on Acoustics and Environment for *Group I* (N = 60)—Teachers With Voice Problems—and *Group II* (N = 407)—Teachers Without Voice Problems

Acoustical and Environmental Statements	N	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)	z	P
1. The class-room acoustics help me talk comfortably.								
Group I	60	25	30	33	7	7		
Group II	402	11	25	39	18	7	−3.319	0.001
2. There is an echo in the class-room.								
Group I	59	29	29	20	17	5		
Group II	403	36	28	23	10	3	−1.489	0.137
3. The class-room is difficult to talk in.								
Group I	60	10	19	39	25	7		
Group II	407	23	29	29	16	3	−3.521	0.000
4. I need to increase the power of my voice to make myself heard even with just a little noise in the class-room.								
Group I	60	5	14	25	37	19		
Group II	407	17	28	27	20	8	−4.595	0.001
5. The class-room air feels dry.								
Group I	60	7	17	20	26	30		
Group II	407	17	18	28	24	12	−3.377	0.001
6. My voice gets muffled by the class-room acoustics.								
Group I	58	9	14	46	26	5		
Group II	404	19	23	35	21	2	−2.584	0.010
7. There is a draught in the class-room even when the door is closed.								
Group I	60	23	22	15	27	13		
Group II	404	40	25	13	13	9	−3.114	0.002
8. The noise made by the pupils is noticeable in the class-room.								
Group I	59	5	12	19	34	30		
Group II	405	8	14	25	28	25	−1.602	0.109
9. The noise from the ventilation is noticeable.								
Group I	60	12	29	22	17	20		
Group II	404	24	24	20	20	12	−1.903	0.057
10. The noise from audio/visual resources is noticeable.								
Group I	60	35	19	21	15	10		
Group II	404	37	27	17	11	8	−1.004	0.315
11. The noise coming from outside of the class-room is noticeable.								
Group I	60	17	18	30	23	12		
Group II	405	19	24	24	22	11	−0.883	0.377
12. I have problems with my hearing.								
Group I	59	37	18	17	14	14		
Group II	406	37	21	15	13	13	−0.012	0.990
13. The class-room acoustics has influence on my way of talking (with the pupils present).								
Group I	58	21	8	14	29	28		
Group II	406	28	16	26	18	12	−3.278	0.001

Grades: 0 = completely disagree to 4 = completely agree.
The z and P values for the Mann-Whitney U-test comparing the groups are also provided.

voice problems, 38% reported having contacted professional help, and 35% had had one or more periods of sick leave. The corresponding percentages for the group without voice problems were 8% and 9%, respectively, and these were statistically significant differences.

Asthma, asthma medication, and hyperreactivity. Asthma was self-reported in 9% of the teachers and its medication in 7%. The medication consisted to 5% of corticosteroids. Hyperreactivity to strong scents was reported in 23% of the teachers, and "other hyperreactivity" was reported in 8%, most stating reaction toward pollen. The statistical analyses revealed significant differences between the group with voice problems and that without voice problems for asthma (17% vs 8%) and for hyperreactivity to strong scents (37% vs 21%). The use of medication did not differ among the groups.

Hearing. Within the whole group of teachers, 13% agreed completely that they had hearing problems (see later and Table 4), but only 3% of the entire group used a hearing aid. The statistical analysis showed a significant difference among the groups in reporting the use of hearing aid, with 10% of the group with voice problems reporting use of hearing aid versus 2% in the group without voice problems.

Job satisfaction. Most of the teachers reported great (49%) or broad (41%) satisfaction with their work. Moderate satisfaction was reported by 9%, and 0.2% reported not being satisfied at all. There were no significant differences found among the groups in terms of job satisfaction.

Voice amplification. Only eight teachers out of 464 (2%) reported that they used voice amplification—two teachers in gymnastics, four in music, one in special education, and one Montessori educator. There was no difference among the groups in terms of the use of voice amplification.

Classroom acoustics and environmental items

Within the group, 38% (ratings: 0–1) disagreed that the classroom acoustics help the teacher to talk comfortably. The distribution of the ratings is shown in Figure 2.

The results for environment and acoustics were analyzed with the Mann-Whitney *U*-test, and the results are summarized in Table 4. There were significant differences between the two groups for the following statements—1: "The class-room acoustics help me talking comfortably"; 3: "The class-room is difficult to talk in"; 4: "I need to increase the power of my voice to make myself heard even with just a little noise in the class-room"; 5: "The class-room air feels dry"; 6: "My voice gets muffled by the class-room acoustics"; 7: "There is a draught in the class-room even when the door is closed"; and 13: "The class-room acoustics has influence on my way of talking (with the pupils present)." In the news media, the noise caused by the pupils is often discussed as a problem to both staff and pupils. In the present results, 92% of the teachers agreed on the presence of noticeable noise from the pupils (Item 8). Furthermore, the perception of disturbance from other noise sources, such as ventilation noise (Item 9), noise from technical equipment (Item 10), and noise from outside the classroom (Item 11), received a moderate to strong agreement by the entire group, however, with no statistical differences between the two groups.

Within the whole group of teachers, 80% (ratings: 2–4) agreed moderately or completely on the following statements—1: the help of the classroom acoustics; 4: the need for increasing the power of the voice for making oneself heard; 5: the feeling of dryness of the classroom air; 6: the voice getting muffled by the classroom acoustics; and 11: occurrence of noise from outside the classroom.

Voice items

The results of the comparison between the groups with (*group I*) and without (*group II*) voice problems for the statements on voice are presented in Table 5. The comparisons were analyzed with the Mann-Whitney *U*-test, and they proved to be significant for all statements within the voice section (Items 14–31).

Among the statements covering early signs of voice problems, throat clearing and hoarseness were most frequently reported in the entire group (80% and 74%, respectively). Throat clearing was also the most reported symptom occurring frequently (sometimes–always) within the group (39%).

The psychosocial factors are not in focus in this phase of the project. One single item on "job satisfaction" was used to provide information about the work-related psychosocial well-being of the teachers. There were, however, no significant differences between the group with voice problems and that without voice problems, as shown in Table 3. The occurrence of stomach discomfort, occasionally discussed as stress related, differed significantly among the groups, with a higher frequency in the group with voice problems.

DISCUSSION

One aim of this study was to examine why some teachers do get voice problems, whereas others, under similar circumstances, do not, and also to explore the teachers' view on factors within their work environment that might influence the voice. The results show that the group of teachers with voice problems significantly differed from that without voice problems in their rating of most of the aspects of environment and room acoustic under investigation. This result may, thus, suggest that teachers with voice problems are influenced by these factors. Furthermore, the results indicate that the teachers with voice problems show a higher frequency of "classical" symptoms of voice attrition than the group reporting no voice problems, with symptoms, such as hoarseness, throat clearing, and a feeling that the voice limits their work. The prevalence of voice problems in this group of teachers was 13%. These findings of the prevalence of voice problems in teachers agree with those of other studies[6,19,20] and, thus, shows that the experiences of teachers in Swedish schools is similar to those of the teachers elsewhere.

Our definition of voice problems was the teacher's own rating of the statement "I have voice problems." That is, a teacher was considered to suffer from voice problems if the rating was 2 or more on the frequency-based scale, indicating that the problems occurred "sometimes," "often," or "always" (Figure 1). It is, thus, important to keep in mind that the dividing of the teacher group into two groups, one group with voice disorders and one without voice disorders, is an interpretation of the answers. Like the present study, most other studies on voice problems of teachers are based on questionnaires, which mainly indicate the subjective view of the individual teacher. It is, however, difficult to compare the results of different studies because of differences in methods and, above all, because of the definition of voice problems, as pointed out by some authors.[18,19,21]

The PCA analysis revealed two factors in the section on classroom acoustics. The first one expressed the voice function and

TABLE 5.
Distribution of the Ratings in Percent of Statements on Voice for *Group I* (N = 60)—Teachers With Voice Problems—and *Group II* (N = 407)—Teachers Without Voice Problems

Voice Statements	N	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)	*z*	*P*
14. I need voice amplification.								
Group I	58	83	3	9	5	0		
Group II	404	92	4	2	1	1	−2.410	0.016
15. I need to clear my throat.								
Group I	59	5	14	32	42	7		
Group II	406	21	45	27	7	0	−7.824	0.000
16. My voice sounds hoarse.								
Group I	60	3	15	42	38	12		
Group II	406	29	46	20	4	0	−8.771	0.000
17. My voice can suddenly change when I talk.								
Group I	59	15	24	35	24	2		
Group II	407	40	39	18	2	0	−6.263	0.000
18. I need to strain to make my voice work.								
Group I	60	10	8	37	37	8		
Group II	405	47	37	13	2	0	−9.475	0.000
19. My voice limits my work.								
Group I	59	15	25	36	20	4		
Group II	406	64	28	6	2	0	−9.139	0.000
20. I avoid certain tasks due to my voice.								
Group I	60	43	25	17	8	7		
Group II	407	83	14	1	0	0	−7.798	0.000
21. Due to my voice the pupils have trouble hearing me.								
Group I	60	35	40	20	5	0		
Group II	406	79	18	3	0	0	−7.678	0.000
22. I have wanted to stay at home due to problems with my voice.								
Group I	60	47	23	27	3	0		
Group II	407	83	14	3	0	0	−6.850	0.000
23. Others ask what is wrong with my voice.								
Group I	60	62	23	12	3	0		
Group II	404	94	5	1	0	0	−8.151	0.000
24. I have stayed at home due to problems with my voice.								
Group I	60	65	22	12	2	0		
Group II	407	85	12	2	0	0	−3.988	0.000
25. I have a sensation of discomfort in my throat.								
Group I	60	10	23	30	34	3		
Group II	405	56	30	12	2	0	−9.110	0.000
26. My voice upsets me.								
Group I	60	8	27	43	14	8		
Group II	407	83	14	3	0	0	−13.437	0.000
27. I run out of air when I talk.								
Group I	60	47	18	20	12	3		
Group II	406	79	16	4	1	0	−6.064	0.000
28. My voice makes me feel incompetent								
Group I	60	48	15	22	15	0		
Group II	401	88	9	2	0	0	−8.360	0.000
29. My throat is burning.								
Group I	59	32	29	20	19	0		

(*Continued*)

TABLE 5
Distribution of the Ratings in Percent of Statements on Voice for Group I ($N=60$)—Teachers With Voice Problems—and Group II ($N=407$)—Teachers Without Voice Problems (*Continued*)

Voice Statements	N	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)	z	P
Group II	407	71	22	6	1	0	−6.847	0.000
30. It feels like a lump in my throat.								
Group I	60	37	25	23	12	3		
Group II	407	72	20	6	2	0	−6.280	0.000
31. I have sensations of gastritis								
Group I	60	50	20	20	8	2		
Group II	407	72	14	9	4	1	−3.500	0.000
32. I have problems with my voice.								
Group I	60	0	0	75	22	3		
Group II	407	72	28	0	0	0		

Grades: 0 = never, 1 = occasionally, 2 = sometimes, 3 = often, and 4 = always.
The *z* and *P* values for the Mann-Whitney *U*-test comparing the groups are also provided.

the interaction of the voice with the acoustics of the classroom. The second component covered external sources influencing the voice. For the voice part, four factors emerged. The first factor represented early symptoms of voice problems; the second, consequences of voice problems; and the third, functional and emotional aspects of voice problems. These three correspond to the three subscales of the Voice Handicap Index.[22] The fourth factor was related to problems of the throat. Our clinical experience suggests that it is adequate to have some irregular problems from the throat without experiencing a voice problem.

Room acoustics and environmental issues

There are a small number of studies linking voice problems in teachers to the acoustic conditions of the classroom.[5,12] A few studies have also investigated the teachers' subjective opinion of factors influencing the voice.[10,23,24] Somewhat surprisingly, the room acoustics has, in these studies, been rated by the teachers as factors of minor importance. This might be the result of the concept of "acoustics" being a rather broad one, giving rise to various connotations in laymen. The statements in the present questionnaire were formulated in collaboration with both professionals in the field and teachers to produce a phrasing as close as possible to the one that would be used by a layman but still professionally adequate.

One of the more central statements of this study is statement "The classroom acoustics helps me talk comfortably." A significant difference between the ratings of the two groups was found; the group with voice problems disagreed more with this statement. This is an interesting finding underpinning the results by Brunskog et al,[11] whose findings indicate that the voice power level is correlated with the support that the room provides to the speaker, called Room Gain. It is not possible to conclude from these results that insufficient room acoustics causes voice problems, but we may speculate that a person with voice problems might be more dependent on the Room Gain and is also more sensitive to acoustic- and environment-related disturbances. The ratings of the two groups differed significantly also for items 3, 4, and 13, indicating that the classroom acoustics cause adjustments of the vocal behavior. For a person without voice training, a common strategy of adjustment is to increase the voice intensity level and also the fundamental frequency. An increase of F_0 contributes to the voice load and often leads to vocal fatigue.[9] Could it be that some of the differences between teachers with voice problems and those without problems can be partly explained by individual differences in the intuitive skills of using the inherent room acoustics for the support of the voice?

One additional factor of the environment may be the humidity of the classroom air. The dry air seems to affect teachers with voice problems more than those without them. According to Vinturri et al,[25] female speakers report more voice symptoms in dry air than in more humid air. Because the Swedish weather is cold from November to March, indoor heating often contributes to a dry indoor climate.

Voice symptoms in teachers

The differences between the teachers with and without voice problems were significant for all items in the voice section. The group with voice problems reported a frequent occurrence of symptoms of vocal fatigue, such as hoarseness, throat clearing, and voice change. The teachers in this group also stay at home more frequently or want to stay at home. This group also reports their voice as a limiting factor in their work. From the results of this study, we cannot tell anything about the number of teachers who have changed their working situation or even changed occupation. However, the surprisingly large number of teachers within *group I*, who teach in smaller groups, might indicate that there are a number of teachers who have changed their work tasks to avoid a very heavy voice load.

Interestingly, although there was a significant difference between the two groups, 87% of the entire group, who answered that they did not experience a voice problem, still reported a combination of symptoms, however, with a lower frequency, such as throat clearing, hoarseness, and voice change, clinically

TABLE 6.
Pattern Matrix From PCA Analysis for Acoustical/Environmental Statements 1–13

Statement	Component 1	Component 2
3. The class-room is difficult to talk in.	0.763	
4. I need to increase the power of my voice to make myself heard even with just a little noise in the class-room.	0.757	
13. The class-room acoustics has influence on my way of talking (with the pupils present).	0.739	
8. The noise made by the pupils is noticeable in the class-room.	0.726	
1. The class-room acoustics help me talk comfortably.	0.619	
2. There is an echo in the class-room.	0.559	
6. My voice gets muffled by the class-room acoustics.	0.532	
5. The class-room air feels dry.	0.431	
10. The noise from audio/visual resources is noticeable.		0.721
7. There is a draught in the class-room even when the door is closed.		0.625
9. The noise from the ventilation is noticeable.		0.599
11. Noise coming from outside of the class-room is noticeable.		0.586
12. I have problems with my hearing.		0.439

considered to constitute a voice disorder. Furthermore, 8% had been in contact with professional help. These findings are in accordance with the studies where comparisons have been made between teachers and other occupational groups with lower voice demands.[2,4,26,27] Teachers in general, thus, seem to report a higher frequency of voice symptoms than do the others. The definition of "problem" is, thus, crucial here. One might speculate that voice problems might be seen as an occupational hazard and not a personal matter by the teachers and, as such, nothing to really act on. But, to whom does the problem really belong? A small number of studies have shown a negative effect of the teacher's disordered voice on the pupils' learning ability.[28,29] Is then the issue of voice problems in teachers only an occupational and personal matter? This raises the issue of a thorough voice education for teacher students and for teachers already at work.

Voice-related absence from work

Our results show that teachers with voice problems are absent from work because of voice problems more than those without voice problems (35% vs 9% of sick leave). Even more striking are the 30% within the group with voice problems who have wanted to stay at home because of voice problems, compared with 3% in the other group. However, according to results from other studies, teachers who do not report voice problems also stay at home more often compared with nonteachers.[2,15,19] Additionally, *group I* reports voice problems as a limitation to its work. Both groups are broadly satisfied with their work;

TABLE 7.
Pattern Matrix From PCA Analysis for Voice Statements 14–32

Statement	Component 1	Component 2	Component 3	Component 4
16. My voice sounds hoarse.	0.803			
15. I need to clear my throat.	0.795			
17. My voice can suddenly change when I talk.	0.764			
18. I need to strain to make my voice work.	0.587			
32. I have problems with my voice.	0.378		−0.305	
24. I have stayed at home due to problems with my voice.		0.932		
22. I have wanted to stay at home due to problems with my voice.		0.859		
23. Others ask what is wrong with my voice.		0.443	−0.410	
28. My voice makes me feel incompetent.			−0.769	
21. Due to my voice the pupils have trouble hearing me.			−0.725	
20. I avoid certain tasks due to my voice.			−0.697	
19. My voice limits my work.	0.304		−0.599	
14. I need voice amplification.			−0.496	
27. I run out of air when I talk.			−0.447	
26. My voice upsets me.			−0.424	
29. My throat is burning.				0.707
31. I have sensations of gastritis.				0.668
30. It feels like a lump in my throat.				0.662
25. I have a sensation of discomfort in my throat.	0.466			0.568

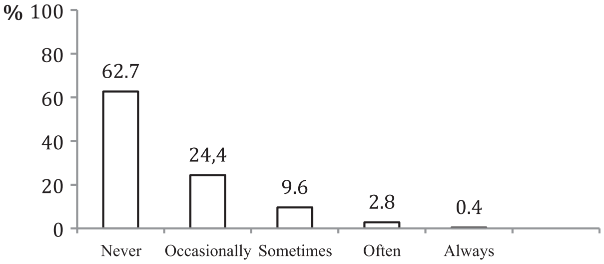

FIGURE 1. Distribution of the answers in percent for statement 32, "I have voice problems," in 467 teachers.

hence, the voice problems are probably not used as an excuse for absence. We may conclude that teachers often are bothered by the voice function. We may also conclude that absence from work because of voice problems is a reality in teachers' everyday life. In our opinion, voice-related problems in teachers should be viewed as a work-related problem and treated as such when discussing the work environment in schools.

Factors possibly influencing voice load

A number of factors are often discussed as contributing to voice load and, as such, to have an impact on voice problems. Gender is one of these, with female teachers more at risk of voice problems than male teachers.[6,19,30] Our results, however, did not show any significant differences in the aspect of gender among the groups. The number of pupils in the classroom is also considered to contribute to voice load, with an increasing numbers of pupils viewed as resulting in higher load. Our results show significant differences between the groups with and without voice problems, but somewhat surprisingly, with the largest difference in percent for the smallest size of group: 1–6 students (22% vs 7%). The explanation may be that teachers experiencing voice problems have already changed their work tasks to decrease the voice problem. Asthma and respiratory allergies are also commonly discussed factors contributing to voice problems in teachers.[27,31] The present results agree, because significantly more teachers in the group with voice problems reported asthma. However, these results have to be interpreted with caution. Asthma may be misdiagnosed and sometimes include other breathing problems that are not asthmatic. There is also a risk of asthma being auto-diagnosed. The inhalation of corticosteroids has been shown to be a possible cause of voice problems.[32,33] The voice symptoms in the investigated group might obviously be ascribed to the use of inhaled corticosteroids. However, the data in this study are self-reported; hence, the information on medicaments is sparsely detailed and has to be interpreted with caution. Moreover, there were no significant differences regarding the use of medication among the groups. Thus, it is not possible from the results of this study to tell whether the voice symptoms are caused by the medication or by the asthma and allergy themselves. Hyperreactivity to strong scents was also more prevalent in the group with voice problems. Hyperreactivity is a complex phenomenon that, in some cases, seems to lead to subjectively and, sometimes, also objectively perceived hoarseness.[34] Problems with hearing have also been shown to be a contributing factor to voice problems in teachers.[31] Even though significantly more teachers in *group I* stated the use of a hearing aid, the present results do not support the earlier findings of the influence of hearing difficulties on voice problems, because there were no significant differences among the groups for the item on problems with hearing.

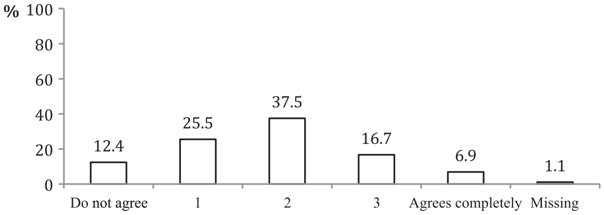

FIGURE 2. Distribution of the answers in percent for statement 1, "The class-room acoustics help me talking comfortably," in 467 teachers.

Methodological issues

A common way of investigating voice problems in teachers is to compare teachers with nonteachers.[2,4-6,19,20,26,27,35] In this study, we have used a different approach. We choose to compare the differences within the entire group of subjects, comparing teachers who reported voice problems with those who did not, a design used earlier by Gotaas and Starr[31] and Rantala and Vilkman.[36]

In analogy with Gotaas and Starr,[31] we used a frequency-based rating scale for judging the voice items. A frequency-based scale shows the absence, presence, and frequency of occurrence of a problem, but it does not tell anything about the duration of the problem. However, according to Simberg et al,[16] the memory factor may influence the results when a time-based rating is used. The reason is that the subject may better remember recent voice episodes, and this approach may, thus, result in a higher prevalence if the episodes have occurred close to answering the questionnaire. The frequency-based scale used here may reduce the influence of the temporal aspect and rather mirror the current, overall impression of the occurrence of episodes of voice problems in the individual teacher.

When comparing the self-report-based studies, it is important to consider the response rate. As discussed by Simberg et al,[16] the method for distributing the questionnaire may have a significant effect on the number of responses. In earlier studies, the response rate has varied between 29% and 98%, with higher rates in studies where interviews were made over the phone or with a questionnaire distributed "face to face." The present study had a response rate of 73% by using a face-to-face manner of distribution, with one of the authors attending prescheduled, compulsory, collegial meetings at the schools. As a consequence, the questionnaire was completed by all the teachers attending the meetings. Teachers not participating were absent from the meetings because of sickness or vacation. Although we do not have any information on the cause of the absence, it is possible that some teachers were absent because of voice problems. We may, however, presume that the present data are reasonably unbiased by individual teachers' special interest in voice or voice disorders. As to the participating schools, we are not familiar with the individual headmasters' motives of acceptance or rejection of the schools' participation in the study. However, the rejections to participate have often been accompanied by explanations of heavy workload and tight schedules and also that many investigations are currently being performed in Swedish schools.

CONCLUSIONS AND FUTURE RESEARCH

We can conclude that the teachers participating in this study agreed on several aspects of working environment as being noticeable in their work. The results also show that 13% of the teachers suffer from voice problems sometimes, often, or always. The differences within the group of teachers indicate that the group with voice problems agreed on perceiving aspects of the room acoustics and other environment-related issues to a significantly higher degree than was reported by the group without voice problems. Vocal symptoms were reported in the entire group, however, significantly more in the group considering itself to have a voice problem. Voice-related absence from work is common in both teachers with and without voice problems. The conclusion is that the teachers with voice problems are more dependent on good working conditions and in learning how to optimize the use of their voices and the room acoustics. The findings suggest that discussions about the use of the prerequisites of the classroom should be covered during voice therapy with teachers. The findings also indicate that vocal training during teacher education is a necessity. In this study, we did not make any laryngeal analyses or voice recordings. They are, however, planned for the next phase of the study to further understand the underlying factors of voice problems in teachers.

Acknowledgment

This work was supported by AFA-Insurance.

REFERENCES

1. Fritzell B. Voice disorders and occupation. *Logoped Phoniatr Vocol*. 1996;21:7-12.
2. Smith E, Lemke J, Taylor M, Kirchner HL, Hoffman H. Frequency of voice problems among teachers and other occupations. *J Voice*. 1998;12:480-488.
3. Williams NR. Occupational groups at risk of voice disorders: a review of the literature. *Occup Med*. 2003;53:456-460.
4. Ohlsson A-C, Järvholm B, Löfqvist A. Vocal symptoms and vocal behavior in teachers. *Scand J Log Phon*. 1987;12:61-69.
5. Pekkarinen E, Himberg L, Pentti J. Prevalence of vocal symptoms among teachers compared with nurses: a questionnaire study. *Scand J Log Phon*. 1992;17:113-117.
6. Roy N, Merrill RM, Thibeault S, Parsa RA, Gray SD, Smith EM. Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res*. 2004;47:281-293.
7. Vintturi J, Alku P, Sala E, Sihvo M, Vilkman E. Loading-related subjective symptoms during a vocal loading test with special reference to gender and some ergonomic factors. *Folia Phoniatr Logop*. 2003;55:55-69.
8. Vilkman E. Occupational safety and health aspects of voice and speech professions. *Folia Phoniatr Logop*. 2004;56:220-253.
9. Laukkanen AM, Ilomaki I, Leppanen K, Vilkman E. Acoustic measures and self-reports of vocal fatigue by female teachers. *J Voice*. 2008;22:283-289.
10. McAleavy GJ, Adamson G, Hazlett DE, Donegan HA, Livesey GE. Modelling determinants of the vocal health of teachers in Northern Ireland: implications for educational policy and practice. *Public Health*. 2008;122:691-699.
11. Brunskog J, Gade AC, Payá-Ballester G, Reig-Calbo L. Increase of voice level and speaker comfort in lecture rooms. *J Acoust Soc Am*. 2009;25:2072-2082.
12. Sala E, Airo E, Olkinuora P, et al. Vocal loading among day care center teachers. *Logoped Phoniatr Vocol*. 2002;27:21-28.
13. Lane H, Tranel B. The Lombard sign and the role of hearing in speech. *J Speech Hear Res*. 1971;14:677-709.
14. Jónsdóttir V, Boyle B, Martin P, Sigurdardóttir G. A comparison of the occurrence and nature of vocal symptoms in two groups of Icelandic teachers. *Logoped Phoniatr Vocol*. 2002;27:98-105.
15. Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope*. 2005;115:1988-1995.
16. Simberg S, Sala E, Vehmas K, Laine A. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *J Voice*. 2005;19:95-102.
17. de Jong FICRS, Kooijman PGC, Thomas G, Huinck WJ, Graamans K, Schutte HK. Epidemiology of voice problems in Dutch teachers. *Folia Phoniatr Logop*. 2006;58:186-198.
18. de Medeiros AM, Barreto SM, Assuncao AA. Voice disorders (dysphonia) in public school female teachers working in Belo Horizonte: prevalence and associated factors. *J Voice*. 2008;22:676-687.
19. Russell A, Oates J, Greenwood KM. Prevalence of voice problems in teachers. *J Voice*. 1998;12:467-479.
20. Smith E, Gray SD, Dove H, Kirchner L, Heras H. Frequency and effects of teachers' voice problems. *J Voice*. 1997;11:81-87.
21. Mattiske JA, Oates JM, Greenwood KM. Vocal problems among teachers: a review of prevalence, causes, prevention, and treatment. *J Voice*. 1998;12:489-499.
22. Jacobson B, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger M. The Voice Handicap Index (VHI): development and validation. *Am J Speech Lang Pathol*. 1997;6:66-70.
23. Kooijman PGC, de Jong FICRS, Thomas G, Huinck W, Donders R, Graamans K, Schutte HK. Risk factors for voice problems in teachers. *Folia Phoniatr Logop*. 2006;58:159-174.
24. Ilomäki I, Leppänen K, Kleemola L, Tyrmi J, Laukkanen A-M, Vilkman E. Relationships between self-evaluations of voice and working conditions, background factors, and phoniatric findings in female teachers. *Logoped Phoniatr Vocol*. 2009;34:20-31.
25. Vintturi J, Alku P, Lauri ER, Sala E, Sihvo M, Vilkman E. The effects of post-loading rest on acoustic parameters with special reference to gender and ergonomic factors. *Folia Phoniatr Logop*. 2001;53:338-350.
26. Sala E, Laine A, Simberg S, Pentti J, Suonpaa J. The prevalence of voice disorders among day care center teachers compared with nurses—a questionnaire and clinical study. *J Voice*. 2001;15:413-423.
27. Roy N, Merrill RM, Thibeault S, Gray SD, Smith EM. Voice disorders in teachers and the general population: effects on work performance, attendance, and future career choices. *J Speech Lang Hear Res*. 2004;47:542-551.
28. Morton V, Watson D. The impact of impaired vocal quality on children's ability to process spoken language. *Logoped Phoniatr Vocol*. 2001;26:17-25.
29. Rogerson J, Dodd B. Is there an effect of dysphonic teachers' voices on children's processing of spoken language? *J Voice*. 2005;19:47-60.
30. Sapir S, Keidar A, Mathers-Schmidt B. Vocal attrition in teachers: survey findings. *Eur J Disord Commun*. 1993;28:177-185.
31. Gotaas C, Starr C. Vocal fatigue among teachers. *Folia Phoniatr Logop*. 1993;45:120-129.
32. Ihre E, Zetterstrom O, Hammarberg B. Voice problems as side effects of inhaled corticosteroids in asthma patients—a prevalence study. *J Voice*. 2004;18:403-414.
33. Bhalla RK, Jones AS, Roland NJ. Prevalence of pharyngeal and laryngeal complications in adult asthmatics using inhaled corticosteroids. *J Laryngol Otol*. 2008;122:1078-1083.
34. Lyberg Ahlander V, Malm L, Schalén L. Hoarseness as a sign of possible nonspecific mucosal hyperreactivity in vocal tract. *J Voice*. 2009;23:707-715.
35. Simberg S, Laine A, Sala E, Ronnemaa AM. A comparison of the prevalence of vocal symptoms among teacher students and other university students. *J Voice*. 2004;18:363-368.
36. Rantala L, Vilkman E. Relationship between subjective voice complaints and acoustic parameters in female teachers' voices. *J Voice*. 1999;13:484-495.

How do teachers with self-reported voice problems differ from their peers with self-reported voice health?

Viveka Lyberg Åhlander, Roland Rydell*, and Anders Löfqvist
Dept. Logopedics, Phoniatrics and Audiology, Clinical Sciences, Lund, Lund University, Lund Sweden
*also ENT Department, Lund University Hospital, Lund, Sweden

Abstract

Objectives

This randomized case-control study compares teachers with self-reported voice problems, to age, gender and school-matched colleagues with self-reported voice health. The self-assessed voice function is related to factors known to influence the voice: laryngeal findings, voice quality, and personality, psycho social and coping aspects, searching for causative factors of voice problems in teachers.

Methods

Subjects, 31 teachers with self-assessed voice problems and 31 voice healthy colleagues recruited from a teacher-group in an earlier questionnaire study, underwent examinations of the larynx by High-Speed Imaging and Kymograms; voice recordings; Voice Range Profile; audiometry; self assessment of voice handicap and voice function; teaching and environmental aspects; personality; coping ; burnout, and work-related issues. The laryngeal and voice recordings were assessed by experienced phoniatricians and speech pathologists.

Results

The group with self-assessed voice problems had significantly longer recovery-time after voice disorders, and scored higher on all subscales of the voice handicap index.

Conclusions

The results show that the cause of voice dysfunction in this group of teachers with self-reported voice problems is not found in the vocal apparatus or in psychosocial aspect within the individual. The individual's perception of a voice problem seems to be based on a combination of the number of symptoms and of how often the symptoms occur, along with the time for recovery. The results also underline the importance of using self-assessed reports of voice dysfunction.

BACKGROUND AND AIM

This paper is a sequel to our epidemiological study of voice problems in Swedish school teachers [1] that examined the influence of working environment on teachers' voices and vocal behavior. Here, we look at etiological factors that may differentiate teachers with self-reported voice problems from teachers without any such problems. Voice problems are common in teachers and teachers are at high risk of voice disorders compared to other occupations; this has been shown in a number of studies [1-12]. There is a general agreement that vocal load is the major cause of voice problems in the teaching staff. The vocal loading that occurs in the daily-life of teachers has several causes [13]. Long teaching hours, poor room acoustics, and bad air quality are seen as the leading causes of voice problems in teachers. Psychological and emotional aspects may also contribute to voice disorders [14-18]. Teachers commonly work in a stressful environment with high vocal and psychological demands and an increasing number of students along with noticeable cut-downs of resources in Sweden. It is often argued that the physical and psychosocial environment influences voice disorders in teachers, but there are, as far as we know, only a few studies that have investigated this relationship [1, 19, 20].

Several studies [8, 11, 21, 22], have investigated the relationship between self-reported voice problems in teachers and objective findings, primarily laryngeal structures, laryngeal function, and voice quality. They suggest, however, that the relationship is not all that clear. Rantala et al. [23-25] investigated the relationship between subjective complaints and objective acoustic measures in a group of teachers and reported lack of correlation between the subjective complaints and the objective measurements. A recent study by Tavares and Martins [26], did however indicate a connection between laryngeal findings and reports of hoarseness in a teaching population and Gotaas and Starr [27] found voice quality to correlate to reports of voice fatigue at certain time-points.

Teachers have high occupational voice demands. They need a flexible voice to instruct, discipline, clarify, and for attracting interest and attention. The increased voice load and the voice load's impact on the voice are evident when teachers are compared to occupational groups with lesser occupational voice demands [8, 11, 28]. However, commonly, not all the staff at a work place is affected by voice problems. Thus, the aim of the present study is to compare teachers with self-reported voice problems to age, gender and school-matched colleagues with self-reported voice health. We relate the self-assessed voice function to factors known to influence the voice function: laryngeal findings, voice quality, and personality and psycho social circumstances. In addition, we investigate the teachers' estimation of their voice function and test their hearing, aiming at investigating possible causative factors of voice problems in teachers. The study has a case-control design with the source population being the group of teachers investigated in the earlier study [1].

MATERIALS AND METHODS

Subjects

All participants were recruited from the population of teachers who participated in our earlier cross-sectional questionnaire survey. All the teachers at 22 schools (n=467) present at pre-scheduled collegial meetings, completed a questionnaire on environment and voice [1]. Planned continuation of the project was explained and the teachers were asked if they were interested in participating, and 220 of them were.

Matching of subjects and controls

Among the subjects willing to participate in the present study, one group of the subjects was recruited among the 41 teachers who, in the questionnaire study, rated themselves as suffering from voice problems. Since it was important to find pairs of cases and controls working at the same school, we searched for possible case-control pairs at the included schools. One subject was excluded due to lack of any control at his school. In addition, two smokers were excluded since it was not possible to find a gender- and age matched smoking control at the school. The remaining 38 subjects were contacted by phone and were informed both orally and in writing about the examination procedures. One was not possible to reach and six subjects declined to participate. Two subjects had changed occupation and no longer worked as teachers. Four declined further participation due to lack of time or interest. A total of 31 teachers with voice problems ended up in the study.

For each subject a control-subject (n=31) was selected from the same school, among those teachers who had estimated no voice problems in the questionnaire study (n=179). The pairs were matched for gender and, as closely as possible, for age.

Two paired groups of teachers were thus formed: Group I (N=31, 26F/5M) included teachers with self-assessed voice problems, with a mean age of 48,7 years (Sd=10,7) and a median time in occupation of 15 years (range 1-40); Group II (N=31, 26F/5M) had teachers without voice problems with a mean age of 44,6 years (Sd=9,9) and median time in occupation of 14 years (range 2-39). All the participants had given their written consent to participate in the study. The pairs came from 12 of the 23 schools from the earlier epidemiological study.

Examination procedure

The teachers were examined at the Department of Logopedics, Phoniatrics and Audiology at Lund University Hospital between May 2009 and February 2010. Written information about the examination procedures was e-mailed to all teachers before the examination and was repeated orally by one of the authors (VLÅ) at the time of the examination. All teachers were subjectively free from upper airway infections and allergies at the time of examination. In most teachers, the examinations were performed during ordinary work weeks and after school-hours; however, three came for the examination at a day off work. The order of examinations followed the same routine of voice recordings, laryngeal examination and last the phonetogram/voice range profile (VRP). There was no fixed order between answering the questionnaires and the audiometry.

Recordings and analyzes

Larynx and the vocal folds

The teachers underwent examination of the larynx and vocal folds with a 70 degree rigid laryngoscope. A digital documentation system was used, HRES Endocam (Wolf, Germany). First, high resolution mode was used for evaluation of organic lesions, adduction and abduction. In high-speed mode 2000 frames/s were recorded for male subjects and 4000 frames/s for female subjects. These recordings were used to evaluate mode and symmetry of vibration at the glottic level. Kymograms were calculated at the mid portion of the membraneous vocal fold. The examinations were performed without local anesthetic in 56/62 subjects, but in six cases: three subjects and three controls Xylocain spray was used (1-3 doses of 10 mg each). All examinations were performed by one of the authors (RR), who was unaware of which group each participant belonged to.

Analyzes of Larynx and vocal folds

The recordings were coded and randomized. The final evaluation of the recordings was made in consensus by two experienced phoniatricians unaware of the grouping of the subjects. Following clinical practice, the guidelines by the Committee on Phoniatrics of the European Laryngological Society (ELS)[29], and suggestions by Kendall [30] for high-speed imaging, a protocol was constructed to assess the following (brackets refer to the presentation of results):

- The morphological structure of the vocal folds (Table 1)
- Asymmetry of posterior larynx: The position of the corniculate tubercles during phonation and rest [31]. (Tables 1 and 2)

- The symmetry and periodicity of vocal fold vibration of the right and left vocal fold separately. (Table 1)
- The activity of the false vocal folds (Table 1)
- The degree and type of glottal opening at maximal closure (Tables 3 and4)
- The propagation and amplitude of the mucosal wave of the right and left vocal fold separately. (Table 1)
- The symmetry and periodicity of vocal fold vibration of the right and left vocal fold separately. (Table 1)
- The phase difference/periodicity: variations in the vibratory cycle, possibly causing asymmetrical closure. (Table 1)
- The Open Quotient in percent of the glottal cycle (time of open phase/time of vibratory cycle). (Fig.1)

The glottic open phase and phase difference were assessed from kymograms. All parameters were judged on a four-point scale (0, no deviance; 3, severe deviance) except for the degree of glottal closure which was judged on a six point rating scale according to Södersten and Lindestad [32] and the pattern of glottal closure which was also categorized according Södersten and Lindestad [32]:

- A: spindle shaped incomplete closure, closure at the vocal processes.
- B: spindle-shaped incomplete closure at the posterior third of the folds, closure at the vocal processes.
- C: Spindle-shaped incomplete closure at the anterior third of the folds, closure at the vocal processes.
- D: Spindle shaped incomplete closure at the posterior and the anterior thirds of the folds, closure at the vocal processes and at the middle of the membranous portion (“hourglass”).

To assess inter rater reliability, eight randomly selected recordings were analyzed twice.

Voice

The voice was signal was digitized at 16 kHz with 16 bit resolution in a sound-proof booth during the reading of a standard text (the Northwind and the Sun) using Soundswell Core 4.0 + Soundswell Voice 4.0, (Hitech Development AB, Täby, Sweden) and a head-worn microphone (MkE2 Sennheiser, www.sennheiser.com), placed 30 cm from the mouth. Due to a change of computer equipment, five of the voices were recorded on MiniDisc (Sony MDS-101), with the same microphone. All recordings were made by one of the authors (VLÅ).

Perceptual rating of voice quality

The voice recordings with a total duration of app. 45 s each were organized in three differently randomized “lists” so that all 62 voices were presented in different order on each list. A panel of three experienced voice-pathologists rated all voices in consensus on a Visual Analogue Scale (VAS) which was presented through the Spruce listening test: Judge 2.0 (Hitech Medical, Täby Sweden). The voices were judged for five parameters, defined according to Hammarberg: hyperfunction, breathiness, vocal fry, hard glottal onsets, and instability [33]. In addition, Grade of Voice Disorder was estimated in analogy with the GRBAS scale [34]. The choice of parameters was limited by the number of parameters possible to present in the Judge application. The judges were given written information with instructions to listen to each voice at a maximum of three times. They were also instructed not to return to a voice that already had been rated. The judges were further instructed to comment on other aspects than those presented through the Judge application, and in such cases add the comments to a protocol. The results were then calculated for overall differences and intra class correlations.

Voice Range Profile

A maximum phonetogram (Voice Range Profile, VRP) was performed with the teacher standing in front of a laptop computer and recorded on a real-time phonetograph Phog 2.5 (Hitech Medical, Täby Sweden) with a head-worn microphone (AKG C420) at a distance of 7 cm from the lips. The phonetogram (VRP) was always recorded last during the examination process to avoid possible laryngeal fatigue.

According the guidelines by the European Union of Phoniatricians [29] the signal was corrected to equal 30 cm distance from the mouth. The teachers phonated with glissandos on the vowel /a/ trying to cover as large an area as possible in frequency and SPL with connected contours. The teachers started at a habitual fundamental frequency gliding downwards to the softest phonation and thereafter, keeping as soft phonation as possible, working upwards through the frequency range towards the highest possible frequency. The procedure was then repeated in loud voice. When this was completed, the teacher was asked to fill out blank spots and try to "connect" the contours. The teachers were free to take the time they needed to complete the VRP. The glissando was practiced a few times before the recording started. All instructions and prompting was carried out by the same author (VLÅ).

The analysis of the VRP followed the procedure described in Ma et al. [35]. All VRPs were measured by the same author (VLÅ). Four boundary points were analyzed for each recording: the highest frequency, the lowest frequency, the maximum and minimum intensity. The maximum area, in semitones x dB, and the frequency ranges were automatically calculated by the Phog 2.5 software.

Analyzes of F0 and LTAS

The sound-files were explored with the help of Soundswell Voice™ and the fundamental frequency was calculated for each voice. A long-time average spectrum was made to obtain information on the voice source, in particular the tilt of the source spectrum [36]. For the analysis, silence and periods of unvoiced sounds were eliminated. For the latter, a comparison was made of the spectral levels below and above 1 kHz. If the lower frequency band dominated a frame, this frame was retained as voiced; otherwise, it was discarded. The ratio of energy in the frequency bands 0-1kHz and 1-5 kHz was calculated. This measure provides information on the tilt of the source spectrum, i.e., how rapidly the amplitude of the higher partials decreases. The second one was the energy in the frequency band 5-8 kHz. A large amount of energy in this band can be a sign of noise due to an incomplete glottal closure [37].

Audiometry

Audiograms were obtained by the same audiologist. The equipment used was a GSI16 (Grason-Stadler Inc.) audiometer together with one pair of Telephonics TDH-39P supra-aural earphones with MX-41/AR cushions. The equipment was calibrated in accordance with IEC 60318-3 and ISO 389-1(IEC, 1998c;ISO 2003). Test stimuli were pure tones of 1-2 seconds duration (35 ms rise and fall times). The following test order was used: 1000, 1500, 2000, 3000, 4000, 6000, 8000, 500, 250 and 125 Hz. Audiometry was conducted in accordance with ISO 8252-1[38] using the manual descending technique (-10/+5 dB). The threshold was defined as the lowest level where three responses had been recorded. The test was performed in a double-walled soundproof booth (complying with the maximum permissible ambient sound pressure level as specified in ISO 8252-1) during one session [38]. The mean value of 500, 1000, 2000, 4000 Hz was calculated for each ear. The sound pressure levels for 3000, 4000 and 6000Hz were also analyzed separately.

Subjective assessments

Questionnaires

Voice Handicap, Self assessment of voice, voice- and teaching related aspects and environment

The teachers were asked to complete the Voice Handicap Index-Throat (VHI-T) [39], which consists of the original three VHI subscales (physical, functional and emotional aspects on voice problems [40], along with a subscale on throat related problems. Each subscale consists ten statements and the occurrence of symptoms are estimated on a frequency-based scale (0=Never, 1=Almost Never, 2=Sometimes, 3=Almost Always, 4= Always). The total sum of this scale might thus be 160 p. A self-assessment of current voice problems and was included, assessed on a 100 mm VA-Scale. In addition, the subjects were asked about demographics and teaching circumstances (posture, native tongue and the language(s) of the students); voice problems during teaching (frequency of voice problems, time of voice recovery, if problems occur with or without a simultaneous cold), and teaching environment (changes made in teaching style or teaching environment due to voice problems, smell in classroom). These questions were answered on a separate questionnaire.

Demand-control and support

Aspects related to work were measured with *the Job Content Questionnaire (JCQ)*. The JCQ is a self-administrated instrument designed to measure social and psychological characteristics of work according to the high demand/low control model of job strain development and covering issues relevant to work demands such as decision making, social interaction etc.[41, 42]

The 26 questions, rated on a four-graded rating scale (1=disagrees completely, 4= agrees completely), comprise the dimensions of job control, job demands, and job support. The job demands, control, and support variables are further dichotomized into high and low categories based on current means from a large population study [43].

JCQ has been widely used for research, at least 70 publications are presented up to date, however only two in teachers [44, 45] and none in relation to voice problems. The JCQ has been translated and assessed for stability in 23 languages until today [42].

Burnout or exhaustion disorder

A frequently discussed problem the society today is burnout or exhaustion disorders [46]. Melamed et al.[47] cite the definition by Shirom [48] of burnout “as the chronic depletion of an individual’s coping resources”(47, pp 1). He characterizes burnout by the constellation of emotional exhaustion, physical fatigue, and cognitive weariness. This syndrome does not overlap with any other clinical syndromes such as depression or anxiety [48] and it is conceptually distinct from a temporary state of fatigue, which passes after a resting period. To investigate the possible symptoms of burnout *the Shirom-Melamed Burnout Questionnaire (SMBQ)* was used [48]. This self-administered instrument consists of 22 questions rated on a frequency based eight graded rating-scale (0-7). The overall burnout index is computed as the mean value of four subscales comprising cognitive weariness, emotional and physical exhaustion, tension, and listlessness.

Coping

The way the individual copes with stressful situations has also been discussed to be a cause of voice problems [49] and an effect on emotions caused by the vocal disabilities [50]. *The Utrechtse Coping List (UCL,* 51*)* in its short form with 22 questions was used to investigate this aspect. Muelenbroek et al [52] have used the longer version for investigations of voice problems in teacher students. The subscales used in the present paper were passive avoidance, depressive reactions, and active reactions.

Personality.

Baker [53] notes that the role of personality in the origin of voice problems has long been of great interest and various measuring methods have been used to investigate this issue. To investigate the possible role of personality in this population of teachers, the two subscales “Psychic Trait Anxiety” and “Adventure seeking” from the *Swedish Universities Scale of Personality (SSP,* 54) were used, providing a rough estimate of the commonly used dimensions of neuroticism and extraversion, respectively. The SSP items were rated on a four-grade scale, ranging from ‘does not apply at all’, to ’applies completely’.

The questionnaires were registered and analyzed in SPSS and the results compared within the pairs with paired samples t-tests; chi2 tests and in SAS for Exact Odds Ratios (OR).

Statistics and ethical considerations

The statistical analyses were computed using SPSS 18.1. For most continuous variables, paired samples t-tests were calculated, for the comparison of the assessment of voice quality the Wilcoxon signed rank test was used due to skewed distributions. For the discrete outcomes variables, 2-sided χ^2 tests were used, with exception for the aspect “Thoughts about change of work”, which was analyzed by Fisher’s exact test due to the expected frequency in one cell being below the recommended frequency of five. The OR calculations for paired samples were performed by SAS® 9.2 for Windows with the lowest level as reference. The inter rater reliability was calculated for each parameter separately, with Intra Class Correlation (ICC). Spearman’s rho was used for the computing of correlations. The alpha level for all statistical analyses was set to $P < 0.05$. The study has been approved by the Institutional Review Board at Lund University (#248/2008).

RESULTS

Demographics

A paired samples t-test revealed significant differences in age between the groups: Group I (M=48,7 Sd=10,7) and Group II (M=44,6 Sd 9,9) t(30)=2,503, *p*=0,018. There were no significant differences found between the groups for time in occupation as concluded by a paired samples t-test.

Larynx and vocal folds

Most aspects could be rated in all subjects. However, and as shown in Tables 1-6, the number varies somewhat between parameters. The inter-rater reliability of the doubled recordings was r=0,851, calculated with Intra Class Correlation. There were no statistically significant differences between the pairs for any aspect. Morphological changes (Table 1) were found in eight subjects (13%), five in Group I (scaring of mucosa, left vf; paresis of left vf; hypoplasia of hemilarynx; contact granuloma, left; vf thickening of the lower border and hypertrophy false vf, left side) and three in Group II (Dry and hyperemic mucosa; minimal thickening of right vf; false left vf hypertrophy/cyst) shown not significant. Tables 1-4 and Fig. 1 present the results of the assessment of the high-speed recordings.

Figure 1. *Distribution of the open quotient in 31 teachers with voice problems (Group I) and 31 teachers with healthy voices (Group II*

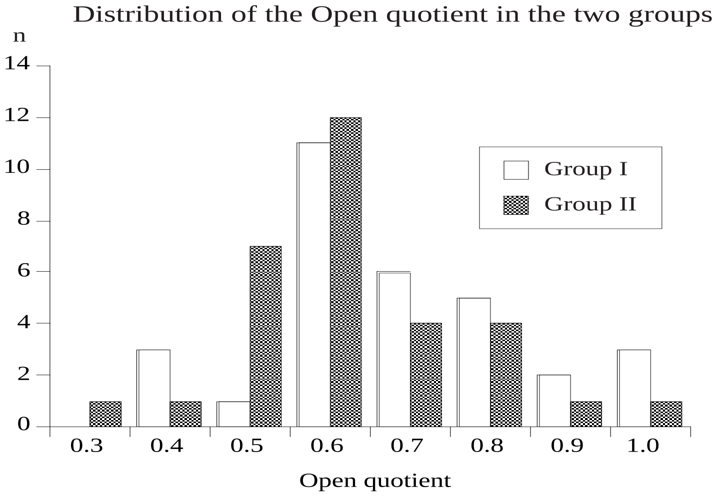

Table 1. *Number of subjects' assessed laryngeal status according high-speed recordings in 31 teachers with voice problems (Group I) and 31 teachers with healthy voices (Group II). 0= no deviance, 3= severe deviance.*

Parameter	**0**	**1**	**2**	**3**	**Total**
Morphological changes					
GI	26	3	2	0	31
GII	25	6	0	0	31
Ab-adduction of VFR					
GI	27	1	2	0	30
GII	27	3	0	0	30
Ab-adduction of VF L					
GI	24	3	2	0	29
GII	26	4	0	0	30
Corniculate tub. Rest					
GI	19	6	2	2	29
GII	23	7	0	0	30
Corniculate tub. Phon					
GI	16	13	1	1	31
GII	10	20	1	0	31
Mucosal wave ampl. R					
GI	18	8	5	0	31
GII	13	14	4	0	31
Mucosal wave ampl. L					
GI	20	6	2	3	31
GII	18	12	1	0	31
Mucosal wave, propagation R					
GI	15	9	7	0	31
GII	12	11	8	0	31
Mucosal wave, propagation L					
GI	18	4	6	3	31
GII	17	9	5	0	31
Phase difference					
GI	23	6	0	1	30
GII	23	8	0	0	31
False vocal cords act R					
GI	19	9	2	1	31
GII	21	6	3	0	30
False vocal cords act L					
GI	13	15	2	1	31
GII	16	11	2	1	30

Table 2. *Position of the most anterior corniculate tubercle in 31 teachers with voice problems (Group I) and 31 teachers with healthy voices, (Group II)*

	Right	Left	No difference	Total
Group I	8	7	16	31
Group II	12	7	11	30

Table 3. *Distribution of assessed degree of closure in two groups of teachers: N=31 teachers with voice problems (Group I) and N=31 teachers with healthy voices (Group II). 1-6 denotes increasing degree of incomplete closure.*

Degree of closure	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Group I	6	17	5	1	1	1	31
Group II	8	10	11	1	1	-	31
Total	14	27	16	2	2	1	62

Table 4. *Number of subjects with deviating pattern of glottal closure in two groups of teachers. Group I: teachers with voice problems N=31 and Group II: teachers with healthy voices N=31 A: spindle shaped incomplete closure, closure at the vocal processes. B: spindle-shaped incomplete closure at the posterior third of the folds, closure at the vocal processes. C: Spindle-shaped incomplete closure at the anterior third of the folds, closure at the vocal processes. D: Spindle shaped incomplete closure at the posterior and the anterior thirds of the folds, closure at the vocal processes and at the middle of the membranous portion ("hourglass").*

Type of closure	**A**	**B**	**C**	**D**	**Total**
Group I	2	4	0	3	9
Group II	1	2	0	3	6
Total	3	6	0	6	15

Voice

The results are based on 31 teachers in Group I and 30 teachers in Group II. Unfortunately, the voice recording of one of the controls could not be analyzed due to technical problems. The inter rating reliability of the three voice-lists was calculated for each parameter and varied between r=0,728-r=0,886 according to the ICC. The ICCs for all parameters are shown in Table 5.

Table 5*. Intra Class Correlations (ICC) of the Inter-rating reliability of the auditory perceptual voice ratings.*

Parameter	**ICC**
Hyperfunction	0,886
Breathiness	0,861
Vocal fry	0,879
Hard Glottal Attacs	0,728
Instability	0,801
Grade of voice disorder	0,853

The assessment of voice quality is summarized in Table 6, presenting the average values for the groups. The assessments were made on a 100 mm Visual Analogue Scale, however the software returns the ratings as of 1000 mm. As is evident from Table 6, there were no significant differences found between the groups for any of the voice quality aspect, as shown by the Wilcoxon signed Rank test.

Table 6 *Mean values of voice parameter judgments for N=30 teachers with voice problems (Group I) and 31 teachers with healthy voices (Group II), assessed on a 1000 mm VA-Scale (see text for clarification).*

Parameter	**Group I, Mean (Sd)**	**Group II Mean (Sd)**
Hyperfunction	46 (98)	61 (106)
Breathiness	95 (128)	45 (67)
Vocal Fry	67 (69)	103 (101)
Hard glottal attacs	23 (62)	13 (18)
Instability	11 (29)	8 (37)
Grade of voice disorder	78 (124)	65 (65)

Voice Range Profile, F0 and Long Time Average Spectrum analyzes

Table 7 presents the measurements of the VRP and F0. No significant differences were found between the pairs, neither for women nor for men for any of the measures.

Table 7. *Values of Voice Range Profiles (VRP) and for F0 in running speech: Area dB (semitones*dB), minimum and maximum dB, minimum and maximum F0 (Hz) F0 in running speech for women and men in two groups of teachers. Teachers with voice problems (Group I) and teachers without voice problems (Group II). All average values.*

			Group I		Group II	
			F N=26	M N=5	F N=26*	M N=5
VRP	Area	Area	828 (254)	822 (246)	868 (198)	906, (131)
	F0 statistics	F0	362 (67)	230 (44)	370 (54)	230 (31)
		Min F0	118 (18)	73, (13)	115 (19)	67 (11)
		Max F0	1004 (277)	750 (221)	1006 (204)	666 (146)
	SPL statistics	SPL	69 (7)	(70) (8)	69 (4)	72 (4)
		Min dB	50 (4)	56 (18)	48 (3)	50 (4)
		Max dB	94 (9)	94 (11)	93 (7)	98 (7)
Running speech	F0		203 (21)	131 (12)	199 (13)	127 (12)

*running speech: n=25, see text.

Audiometry

The results of the audiograms are based on audiograms from 22 teachers from Group I and 29 controls from Group II. Thus, a group-wise comparison was made. There were no significant differences between the groups at any other level. Tinnitus was reported by seven persons, three in Group I, and four in Group II. The use of hearing aid was reported by three participants, two in Group I, and one in Group II.

Questionnaires

Voice Handicap Index-Throat, Self-assessment of voice problems and VAS judgment

The paired samples t-test revealed statistically significant differences for all four subscales of the VHI-T as well as for VHI-T total. The results are summarized in Table 8.

Table 8*. Mean and t and p values for paired samples t-test along with Odds Ratios for VHI-T in two groups of teachers: Teachers with voice problems (Group I, N=31) and teachers without voice problems (Group II, N=31).*

Subscale	**Group I M(Sd)**	**Group II M(Sd)**	***t(df)***	**p**	**OR**
Throat	15,3 (5,9)	8,7 (5,0)	5,451 (29)	0,0001	1,43
Physical	13,8 (8,6)	6,7 (6,6)	4,394 (29)	0,0001	1,27
Functional	8,5 (7,0)	2,5 (3,6)	4,199 (29)	0,0001	1,26
Emotional	9,0 (9,5)	1,7 (3,2)	4,248 (29)	0,0002	2,03
VHI-T Total	46,7 (22,2)	19,3 (15,0)	6,406 (29)	0,0005	1,93

The teachers rated their over-all voice problems on a 100 mm Visual Analogue Scale. A paired samples t-test revealed significant differences between the groups, Group I (M=34, Sd=23,0) and Group II (M=13 Sd=16,3), t(25)=4,890 p=<.001, OR= 1,12.

For the frequency of occurrence of voice problems, a chi-square test showed significant differences between the two groups: $\chi 2$, (5 n=60)=20.138, *p*=0,01, OR= 3.99, the OR indicating that teachers with voice problems were close to four times as likely to rate a high frequency of voice problems. The occurrence of voice problems is shown in Table 9 . There were also significant differences between the groups for voice problems occurring without a concurrent upper-airway infection, $\chi 2$, (2 n=60)=18,670 *p*=0.0008. OR=3.60, as shown in Table 10.

Table 9 *Occurrence of voice problems in two groups of teachers. Teachers with voice problems (Group I) and teachers without voice problems (Group II).*

%	% No voice probl	Every year	<once a month	>once a month	Every week	Every day	%
Group I (N=31)	0	32	6	26	19	16	100
Group II (N=29)	34	41	7	14	3	0	100

Table 10 *Occurrence of voice problems in teachers who have voice problems without a simultaneous upper-airway infection. Teachers with voice problems (Group I) and teachers without voice problems (Group II), in percent.*

	Every year	<once a month	>once a month	Every week	Every day	%
Group I (N=26)	27 (7)	4 (1)	27 (7)	23 (6)	19 (5)	100 (26)
Group II (N=10)	40 (4)	20 (2)	30 (3)	0	10 (1)	100 (10)

A Chi-square test also revealed significant differences between the groups for the time-span for voice recovery $\chi 2$, (7 n=60)=17.608, *p*=0,014, cf. Table 11 with OR= 2.03.

Table 11 *Time for recovery from voice problems in two groups of teachers, teachers with voice problems (Group I) and teachers without voice problems (Group II), in percent.*

	No voice probl	One hr or less	A couple of hrs	Over night	Weekend	Holiday	Never	%
Group I (N=31)	0	13	10	27	23	17	10	100
Group II (N=29)	34	17	7	24	7	10	0	100

Teaching and environmentally related issues

Fisher's exact test showed differences (*p*=0,029) between the case-control pairs for considerations about changing occupation due to voice problems, where 18% in group I had considered a change of occupation but none in group II, OR=2,03. No further differences were found within the pairs for either social status, number of children, age, or time in occupation. Nor were there any differences in most teaching related aspects. Most teachers taught in their native language and stood up during teaching. Most students were speaking the same language as the teacher. Similarly, there were no differences in changes in teaching methods or teaching environment due to possible voice problems.

Control-demand-support, burnout, coping and personality

No differences were found within the pairs for symptoms of burnout syndrome (SMBQ), personality traits (SSP), or for coping strategies (UCL) using paired-samples t-tests. The mean values for SMBQ-global were: Group I=2,7 (Sd1,0) and Group II=2,5 (Sd1,1) which can be compared to reference scores of 3.2 for females and 2.9 for males in a Swedish, healthy population [55]. However, among the three main dimensions Job Demand, Job Control and Job Support of the Job Content Questionnaire (JCQ), significant differences were found for the sub-scale "Job-Control": Group I (M= 3,48 Sd=,20) and Group II (M=3,27 Sd=,29), t(28)=3,047 *p*=0,005. The ratings of Job Demand and Job Support were found non significant. Job Demand: Group I (M= 2,84 Sd=,51) and Group II (M=2,72 Sd=,45), t(28)= 0,946 *p*=0,352 Job Support: Group I (M= 3,79 Sd=,35) and Group II (M=3,78 Sd=,06), t(28)=3,047 *p*=0,888.

As shown by the t-test, the ratings of "Job Demands" are moderately and equally high in both groups while "Job Control" is significantly higher in Group I. The JCQ results were summarized through combinations of the dichotomized ratings of the three main dimensions Demand, Control and Support, in order to define a specific work situation: High demands and low control is defined as "Job Strain", high demands and high control form the category "Active", low demands and low control is defined as "Passive", and low demands and high control form the category "Relaxed". In addition, low support (support from colleagues and management) in combination with "Job Strain" is defined as "Iso-strain", a particularly unfavorable work situation. Table 12 shows the distribution of the subjects according this classification. A larger number of teachers from Group I are found in the "Active" category, where a combination of high demands and high control is represented, while more of the teachers in Group II are found in "Job Strain" category due to ratings of high demands and low control. However, the chi-square and Fisher's exact test showed no significant differences in Job strain (*p*=0,056) between the groups. Iso-strain was not found for any teacher.

Table 12. *Number of teachers for each category of the JCQ. Group I: teachers with voice problems and Group II: voice-healthy teachers. Percentages in brackets. For further explanation, see text.*

	Job strain	Relaxed	Active	Passive	Total
Group I	1 (3,2)	11 (35)	18 (58)	1 (3,2)	31 (100)
Group II	6 (20)	10 (33)	11 (36)	3 (10)	30* (100)

*The result of Group II is based on questionnaires from 30 teachers, due to one questionnaire not completed.

Correlations

Correlations were computed with Spearman's rho for aspects that could be expected to correlate: frequency of symptoms, voice quality ratings, age, morphological findings and recovery time. Almost all correlations were below .5 for most aspects. These are weak to moderate correlations and are thus not presented. However, in individuals from Group I who had deviant morphological laryngeal structure (top row in Table I), the correlation between the rating of morphological structure and Grade, VHI-T, and Recovery time were: structure and grade: 0.577; structure and recovery-time: 0.866 and structure and VHI-T: 0.881. Grade, VHI-T and recovery for the controls with remarks on laryngeal structure did not correlate.

DISCUSSION

Voice function is a complex phenomenon and has an undisputable relation to the voice load and occupational demands. As far as we know, this study differs from earlier studies with respect to the matching of the participants. To isolate the possible influences from environment and the persons' behavior in the classroom, we selected gender- and age-match pairs from the same schools and examined differences in their laryngeal, vocal, hearing and psycho-social aspects. By selecting subjects from the same schools, we wanted to control the influence from the work-environment. Overall, the present results show very small differences within the pairs. The most noteworthy differences are the findings of VHI-T and the time it takes to recover from voice problems. Apart from these differences, there were no statistically significant differences in structure or function that may explain why the teachers that do have voice problems actually have them in contrast to their peers.

The selection of the case-control pairs within this study was based on the teachers' own assessment of the statement "I have voice problems" in the earlier questionnaire survey [1]. The definition of "voice problem" is thus based on the individuals' conception of their own voice. Despite the large number of studies of teachers' voices today, there is still no consensus about the criteria for defining a voice disorder [56]. Commonly, the definition has been based on the number and frequency of symptoms of voice disorders [5, 8, 21] or on the clinician's observations of laryngeal findings or on remarks on the voice quality [21, 22]. The question of the individual's perception of the symptoms has seldom been raised. In analogy with others, our results show that even the teachers who assess themselves as being voice-healthy reports a number of symptoms. There were as many morphological laryngeal findings in the controls as in the group of teachers with voice problems. However, the control subjects obviously don't view their voice symptoms – or the effect of them - severe enough to call them problematic. We thus consider it to be very important to include the subjects' own conception of the voice function, not least in clinic. According Deary et al: "People's ratings of their symptoms are an important guide in gauging the severity of medical disorders, and are specially useful in assessing the response to treatment" (15, p . 374).

Laryngeal findings

High speed digital imaging was used for the laryngeal examinations. One reason is that this is the current standard technique at our department and another one is that it is a new tool in the voice clinic, and there is thus a need of compiling normative data from high-speed examinations [57]. Kendall [30] conclude that the use of High-speed filming offers benefits over standard videostroboscopy for studying aperiodic vocal fold motion which is often thought to be a contributing factor in voice disorders. All subjects could be examined which is probably due to the short time of the examination. Due to the high frame rate (2000 for males and 4000 in females) only a very short sequence is needed. However, there is not yet any golden standard for the assessment of high-speed digital image recordings.

Most subjects in our study were found to be normal in all laryngeal aspects. Findings of asymmetry and structural deviations were made, but without any significant differences within the matched pairs. The importance of asymmetrical vocal fold adduction movements as an explanatory factor in voice disorders has been long discussed [31]. Lindestad et al. found that laryngeal adduction asymmetries were frequent in normal voices (ibidem), but no findings of morphological deviations are mentioned.

It is, however, no surprise that there are no differences between the groups. Most studies that have included laryngeal examinations in investigations of teachers' voices have been unable to establish a connection between the laryngeal status and the subjective symptoms: Urrutikoetxea et al. [4] examined 1 046 teachers and found structural deviations in 20,8%. Ilonmäki et al. [21] found severe organic changes in 14% of the 78 pre-school teachers investigated. Sala et al. [8] made organic findings in 29% of 262 teachers. None of these studies found a correlation between laryngeal findings and subjective symptoms of voice disorders. So, does a laryngeal deviation have no impact on vocal behavior? There is firm clinical evidence about such a relationship, but little is known about an individual's capacity to cope with the effects. This calls for further comparative studies with non-teachers. The findings of Sala et al.[8] indicate that there may be differences in the occurrence of laryngeal findings between teachers and voice healthy non-teachers. They found 29% of the teachers at day-care centers to have laryngeal deviations but only 7% in a group of nurses. In a recent study on 882 patients referred to ENT clinics, van Houtte et al. [58] found 50% of voice professionals, including teachers, to have some kind of structural deviations, compared to 60% in the entire group. However, this was found in a treatment–seeking group in contrast to other studies and little is known about the prevalence of laryngeal deviations in a voice healthy population without a heavy voice load.

Some clues might be found in our results. In the five teachers with voice problems where morphological findings were made, correlations were found for the voice quality parameter Grade of voice disorder and for both VHI-T and Recovery time. None of these aspects correlated in the controls. However, the methods of exploring laryngeal aspects vary between studies and the results are thus hard to compare.

Voice

Similarly, there were no differences within the pairs with respect to voice quality assessments and the acoustic measurements, F0, VRP and LTAS. This is in line with the findings by Ohlson et al. [11] who compared a group of teachers with a group of nurses and found no differences between the groups in LTAS, voice quality, or VRP. In contrast, voice quality differences between teacher-groups were found in a recent study by Tavares and Martins [26], but this might be explained by the large amount of laryngeal pathology in their material.

Gotaas and Starr [27] compared teachers experiencing vocal fatigue to teachers, who did not experience vocal fatigue, and concluded that there were no voice-quality differences between the groups on non-vocal fatigue days. With three exceptions, all teachers in our study were examined after their workday. There were significant differences within the pairs in their own assessment of current voice problems and voice quality, but we did not ask about their views on vocal effort during their past workdays, and a lack of voice load can thus be a confounding factor in the results. It is important to emphasize that the present perceptual ratings of voice quality were all on low grades on the VA-scale and thus have to be interpreted with caution. A finding underlining the lack of correlation between symptoms and findings was some of the ratings of Grade (>200) that was assessed in subjects who subjectively rated their voice problems to 0. Obviously, there are difficulties in assessing quality aspects of normal or nearly normal voices.

The results of the VRP and the LTAS showed no significant differences between the groups. However, Subsinskiene [59] did find differences in VRP results between healthy trained and non-trained professional speakers: pitch range and area of high frequencies differed significantly. The VRP shows the physiological and acoustical constraints [60]. Thus, the difference in findings between studies may have its' explanation in the compared groups. In contrast to the present study Siupsinskiene compared well- and non-trained professionals. It may thus be assumed that the voice training had influenced on the vocal possibilities. The effect of voice training is also supported by the conclusion of Holmberg et al. [61] in their study of changes across voice therapy for patients with vocal fatigue.

It is important to note that this is not a field study but rather a snapshot of the status of the teachers. In other studies, the voice has been measured during a workday. In these studies [22-25, 62], differences have been individuals who report only few symptoms of voice problems. Field measurements with a voice accumulator have been made in the current subject pool.

Audiometry

Generally, the pure-tone hearing thresholds showed no differences between the groups. However, the present finding is inconclusive due to the variation of number of performed measurements in the two groups (Group I: n=22, Group II: n=29) and further research is required to elucidate any relationship between hearing thresholds and voice problems. Further, little is known about the relationship between individuals' hearing and the perception of the own voice in relation to the sound environment. Hearing is most likely important for the relation between voice and the perception of the acoustical properties of the room. Further research is warranted within this area.

Subjects' assessment of voice handicap and voice function

The main differences between the pairs in this study were the subjects' own assessment of their voice, voice handicap, and in the recovery time. The VHI, and the VHI-T (the VHI with a subscale on throat problems [39], have been shown to separate subjects with and without voice disorders [39, 40, 63]. It is noteworthy that the highest OR of the VHI-T subscales was found on the emotional subscale, which indicates that teachers with voice problems are twice as likely as their voice healthy colleagues to score high on this subscale. This higher scoring on the emotional subscale may be indicating that if the individual considers the symptoms as communicatively hindering and even embarrassing, (s)he is more apt to consider the symptoms problematic.

Furthermore, the discrepancy within the pairs in terms of the recovery-time from symptoms of voice problems is very interesting. Similar findings were made by Sala et al. [8] where the day-care centre teachers reported a longer time for the symptoms to disappear than the group of nurses. This might indicate micro-structural changes in the larynx that we are not able to detect with today's technology and thus warrants further studies.

Control-demand-support, burnout, coping and personality

There is an increasing number of studies linking psychological factors to functional dysphonia [15, 64]. These factors include higher levels of anxiety, lower levels of sense of control, quality of life and coping [15, 16, 18]. Roy et a. [65] found that the majority of people with functional dysphonia were introverts. Andersson and Schalén [17] noted that interpersonal conflicts related to family and work were one of the important contributing factors in psychogenic voice disorders, and Gassull et al. [49] in a recent study that teachers with voice problems were highly reactive to stress.

We used a battery of questionnaires to investigate the various aspects that have been found to contribute to the etiology of dysphonia and also the Job Content Questionnaire (JCQ) to cover aspects of demand-control-support. The JCQ was the only scale that showed some differences between the groups. The underlying theory of the JCQ is that a combination of high demands and low control/low support causes *job strain* which is defined as harmful. That is, when there is a combination of high psychological demands and a low worker's decision latitude there is an increased risk of harmful job strain. If the social support at the work-place is low, this further increases the risk. However, the active or passive behavior of the employee needs to be taken into account. An active behavior gives rise to "good stress", predicting motivation, new learning behavior, and new coping strategies [41]. The differences within the pairs did not support the hypothesis of a higher degree of job-strain (high demands and low control) in the teachers with voice problems. Instead the results showed that both groups rated high degrees of job demands and job support but differed in the aspect of job control, where the group with voice problems rated significantly higher values.

The results may be due to a selection bias. In the questionnaire study, we asked the respondents who wanted to further take part of the project to mark this on the questionnaire. This may have caused the more active teachers with feelings of control of their social life and work situation to step forward. The non-difference within the pairs may also depend on the normality of the data, there were no big differences in any scale as compared to a normal population. Buck et al, [66] found differences between groups of dysphonics, functional v s organic, but only a minority (17 %) of patients in the functional group showed clinically significant levels of psychological distress. The difference between the present study and others might also be due to the use of different instruments. We used a battery of tests that have been developed for a Swedish population (Swedish Universities Scale of Personality [54], or had been tried and on a Swedish population (Job Content Questionnaire and Shirom-Melamed Burnout Questionnaire, [41, 48]. The Utrechtse Coping Lijst, measuring coping, has been used in teachers with voice problems [52]. It was however a time-consuming battery of tests, and took the most part of the examinations to complete. There is no consensus about which questionnaire/questionnaires to use for investigating psychological factors in dysphonic patients or in research-groups and further studies are thus warranted in this area. However, for the investigation of work-related issues we found the Job Content-model very useful, and thus recommend it for further investigations of work-related dimensions in connection to voice problems.

Conclusion

For the two groups in this study the main differences were found for the VHI-T and time for recovery after voice problems. Thus, the combination of the number of symptoms and of how often the symptoms occur, along with the time it takes to recover, seems to underlie the individual’s perception of the voice problem. The results also underline the importance of investigating the individual’s view of the severity of the voice dysfunction. The main conclusion of this study is that the cause of voice dysfunction in the group of teachers with self-reported voice problems is not found in the vocal apparatus or within the individual. It may instead be found in the interplay of the individual’s behavior and the work-environment which we plan to study in a future project.

Acknowledgment

The kind and persistent cooperation of the teachers is gratefully acknowledged. We are greatly indebted to Associate Professor Kai Österberg, division of Occupational and Environmental Medicine, Lund University for kind and generous guiding and assistance with choice and interpretation of the instruments measuring psychological and psychosocial aspects. We also like to express our thanks to Audiologist, MSc Ingrid Lennart for performing the audiometry and to Audiologist, PhD Jonas Brännström for valuable comments on the manuscript. This work was founded by grants from AFA insurance, Stockholm, Sweden.

REFERENCES

1. Lyberg Åhlander V, Rydell R, Löfqvist A. Speaker’s comfort in Teaching Environments: Voice problems in Swedish Teaching Staff. *J Voice.* 2010. [Epub ahead of print].
2. Fritzell B. Voice disorders and occupations. *Logoped Phoniatr Vocol.* 1996;21:7–12.
3. Titze IR, Lemke J, Montequin D. Populations in the U.S. workforce who rely on voice as a primary tool of trade: A preliminary report. *J Voice.* 1997;11(3):254-259.

4. Urrutikoetxea A, Ispizua A, Matellanes F. Pathologie vocale chez les professeurs: Une étude vidéo-laryngo-stroboscopique de 1.046 professeurs [Vocal pathology in teachers: A videolaryngostroboscopic study in 1046 teachers]. *Rev Laryngol Otol Rhinol (Bord).* 1995;116:255 – 262.
5. Smith E, Gray SD, Dove H, Kirchner L, Heras H. Frequency and effects of teachers' voice problems. *J Voice.* 1997;11(1):81-87.
6. Simberg S, Sala E, Vehmas K, Laine A. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *J Voice.* 2005;19(1):95-102.
7. Sapir S, Keidar A , Mathers-Schmidt B. Vocal attrition in teachers: Survey findings. *Eur J Disord Commun.* 1993;28:177 – 185
8. Sala E, Laine A, Simberg S, Pentti J, Suonpaa J. The Prevalence of voice disorders among day care center teachers compared with nurses - A questionnaire and clinical study. *J Voice.* 2001;15(3):413-423.
9. Russell A, Oates J, Greenwood KM. Prevalence of voice problems in teachers. *J Voice.* 1998;12(4):467-479.
10. Roy N, Merrill RM, Thibeault S, Gray SD, Smith EM. Voice disorders in teachers and the general population: Effects on work performance, attendance, and future career choices. *J Speech Lang Hear Res.* 2004;47:542 – 551.
11. Ohlsson AC, Järvholm B, Löfqvist A. Vocal symptoms and vocal behaviour in teachers. *Scand J Logoped Phoniatr.* 1987;12:61-69.
12. Morton V, Watson D. The teaching voice: Problems and perceptions. *Logoped Phoniatr Vocol.* 1998;23(3):133-139.
13. Vilkman E. Occupational safety and health aspects of voice and speech professions. *Folia Phoniatr Logop.* 2004;56(4):220-253.
14. Kooijman PGC, Thomas G, Graamans K, de Jong FICRS. Psychosocial Impact of the teacher's voice throughout the career. *J Voice.* 2007;21(3):316-324.
15. Deary IJ, Wilson JA, Carding PN, Mackenzie K. The dysphonic voice heard by me, you and it: Differential associations with personality and psychological distress. *Clin Otolaryngol Allied Sci.* Aug 2003;28(4):374-378.
16. Roy N, Bless DM. Personality Traits and psychological factors in voice pathology: A foundation for future research. *J Speech Lang Hear Res.* 2000;43(3):737-749.
17. Andersson K, Schalén L. Etiology and treatment of psychogenic voice disorder: Results of a follow-up study of thirty patients. *J Voice.* 1998;12(1):96-106.
18. Wilson JA, Deary IJ, Millar A, Mackenzie K. The quality of life impact of dysphonia. *Clin. Otol.* 2002;27(3):179-182.
19. McAleavy GJ, Adamson G, Hazlett DE, Donegan HA, Livesey GE. Modelling determinants of the vocal health of teachers in Northern Ireland: Implications for educational policy and practice. *Public Health.* 2008;122(7):691-699.
20. Sala E, Airo E, Olkinuora P, Simberg S, Strom U, Laine A, Pentti J, Suonpaa J. Vocal loading among day care center teachers. *Logoped Phoniatr Vocol.* 2002;27(1):21-28.
21. Ilomäki I, Leppänen K, Kleemola L, Tyrmi J, Laukkanen A-M, Vilkman E. Relationships between self-evaluations of voice and working conditions, background factors, and phoniatric findings in female teachers. *Logoped Phoniatr Vocol.* 2009;34(1):20 – 31.
22. Laukkanen AM, Ilomaki I, Leppanen K, Vilkman E. Acoustic measures and self-reports of vocal fatigue by female teachers. *J Voice.* 2008;22(3):283-289.
23. Rantala L, Paavola L, Korkko P , Vilkman E. Working-day effects on the spectral characteristics of teaching voice. *Folia Phoniatr Logop.* 1998;50(4):205-211.
24. Rantala L, Vilkman E. Relationship between subjective voice complaints and acoustic parameters in female teachers' voices. *J Voice.* 1999;13(4):48-495.
25. Rantala L, Vilkman E, Bloigu R. Voice changes during work - subjective complaints and objective measurements for female primary and secondary schoolteachers. *J Voice.* 2002;16(3):344-355.
26. Tavares ELM, Martins RHG. Vocal evaluation in teachers with or without symptoms. *J Voice.* 2007;21(4):407-414.
27. Gotaas C, Starr CD. Vocal fatigue among teachers. *Folia Phoniatr Logop.* 1993;45:120-129.

28. Smith E, Lemke J, Taylor M, Kirchner HL, Hoffman H. Frequency of voice problems among teachers and other occupations. *J Voice.* 1998;12(4):480-488.
29. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van De Heyning P, Remacle M, Woisard V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques -Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch of Oto-Rhino-Laryngology.* 2001;258(2):77-82.
30. Kendall KA. High-speed laryngeal imaging compared with videostroboscopy in healthy subjects. *Arch Otolaryngol Head Neck Surg*. 2009,135;(3):274-281.
31. Lindestad P-Å, Hertegård S, Björck G. Laryngeal adduction asymmetries in normal speaking subjects. *Logoped Phoniatr Vocol.* 2004;29(3):128-134
32. Södersten M, Lindestad PA. Glottal closure and perceived breathiness during phonation in normally speaking subjects. *J Speech Lang Hear Res.* 1990;33(3):601-611.
33. Hammarberg B. Voice research and clinical needs. *Folia Phoniatr Logop.* 2000;52(1-3):93-102.
34. Hirano M. *Clinical examination of voice.* New York: Springer; 1981.
35. Ma E, Robertson J, Radford C, Vagne S, El-Halabi R, Yiu E. Reliability of speaking and maximum voice range measures in screening for dysphonia. *J Voice.* 2007;21(4):397-406.
36. Löfqvist A, Mandersson B. Long-time average spectrum of speech and voice analysis. *Folia Phoniatr Logop.* 1987;39(5):221-229.
37. Hanson HM, Stevens KN, Kuo HKJ, Chen MY, Slifka J. Towards models of phonation. *Journal of Phonetics.* 2001;29(4):451-480.
38. ISO 8253-1 Acoustics: Audiometric test methods part 1. Basic pure tone air and bone conduction threshold audiometry. *ISO 8253-1.* Vol 389-1: International Organization for Standardization; 1998.
39. Lyberg-Ahlander V, Rydell R, Eriksson J, Schalén L. Throat related symptoms and voice: development of an instrument for self assessment of throat-problems. *BMC Ear Nose Throat Disord.* 2010;10:5. http://www.biomedcentral.com/1472-6815/10/5
40. Jacobson B, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger M. The Voice Handicap Index (VHI): Development and validation. *Am J of Speech-Lang. Path.* 1997;6(3):86-70.
41. Karasek R, Brisson C, Kawakami N, Houtman I, Bongers P, Amick B. The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. *J Occup Health Psychol.* 1998;3(4):322-355.
42. Job Content Questionnaire center DoWE. http://www.jcqcenter.org/.2010. Massachusetts, Lowell, USA.
43. Lindeberg SI, Rosvall M, Choi B, Canivet C, Isacsson SO, Karasek R, Ostergren PO. Psychosocial working conditions and exhaustion in a working population sample of Swedish middle-aged men and women. *Eur J Public Health.* [Epub ahead of print].
44. Azlihanis HA, Naing L, Aziah DB. Sociodemographic, occupational and psychosocial factors associated with job strain among secondary school teachers in Kota Bharu, Kelatan. *Malays. J. med. sci.* 2006;13:18-19.
45. Brown DE, James GD, Mills PS. Occupational differences in job strain and physiological stress: female nurses and school teachers in Hawaii. *Psychosom Med.* 2006;68(4):524-530.
46. Socialstyrelsen, ed. *Utmattningssyndrom (Burn-out syndrome)*. Stockholm: Swedish National Board of Health and Welfare; ed. Bjurner & Bruno AB; 2003. Welfare Stockholm.
47. Melamed S, Ugarten U, Shirom A, Kahana L, Lerman Y, Froom P. Chronic burnout, somatic arousal and elevated salivary cortisol levels. *J Psychosom Res..* 1999;46(6):591-598.
48. Shirom A, ed. *Burnout in work organizations*. New York: John Wiley & Sons; 1989. Cooper CL RI, ed. International review of industrial and organizational psychology; pp.25-48.
49. Gassull C, Casanova C, Botey Q, Amador M. The Impact of the Reactivity to Stress in Teachers with Voice Problems. *Folia Phoniatr Logop.* 2010;62(1-2):35-39.
50. Epstein R, Hirani SP, Stygall J, Newman SP. How do individuals cope with voice disorders? Introducing the voice disability coping questionnaire. *J Voice.* 2009;23(2):209-217.

51. Scheurs PJG, van de Willege G, Brosschot JF, Tellegen B , Graus GMH. De Utrechtse coping Lijst: UCL Omgaan met problemen engebeurtenissen. [The Utrecht Coping List: to cope with problems and incidents] Lisse, The Netherlands: Swets & Zeitlinger; 1988.
52. Meulenbroek LFP, Thomas G, Kooijman PGC , de Jong FICRS. Biopsychosocial impact of the voice in relation to the psychological features in female student teachers. *J of Psychosom Res.* 2010;68(4):379-384.
53. Baker J. The role of psychogenic and psychosocial factors in the development of functional voice disorders. *Int J Speech Lang Pathol.* 2008;10(4):210 – 230.
54. Gustavsson JP, Bergman H, Edman G, Ekselius L, von Knorring L, Linder J. Swedish universities Scales of Personality (SSP): Construction, internal consistency and normative data. *Acta Psychiatr Scand.* 2000;102(3):217-225.
55. Stenlund T, Ahlgren C, Lindahl B, Burell G, Knutsson A, Stegmayr B, Birgander LS. Patients with burnout in relation to gender and a general population. *Scand J Public Health.* 2007(35):516-523.
56. Mattiske JA, Oates JM, Greenwood KM. Vocal problems among teachers: A review of prevalence, causes, prevention, and treatment. *J Voice.* 1998;12(4):489-499.
57. Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE. Clinical Implementation of Laryngeal High-Speed Videoendoscopy: Challenges and Evolution. *Folia Phoniatr Logop.* 2008;60(1):33-44
58. Van Houtte E, Van Lierde K, D'Haeseleer E , Claeys S. The prevalence of laryngeal pathology in a treatment-seeking population with dysphonia. *Laryngoscope.*120(2):306-312.
59. Siupsinskiene N. Quantitative analysis of professionally trained versus untrained voices. *Medicina (Kaunas).* 2003;39(1):36-46.
60. Titze IR. Acoustic interpretation of the voice range profile (phonetogram). *J Speech Lang Hear Res.* 1992;Feb;35(1):21-34
61. Holmberg E, Ihre E, Södersten M. Phonetograms as a tool in the voice clinic: Changes across voice therapy for patients with vocal fatigue. *Logoped Phoniatr Vocol.* 2007;32(3):113-127.
62. Jonsdottir V, Laukkanen AM, Vilkman E. Changes in teachers' speech during a working day with and without electric sound amplification. *Folia Phoniatr Logop.* 2002;54:282-287.
63. Verdonck-de Leeuw IM, Kuik DJ, De Bodt M, Guimaraes I, Holmberg EB, Nawka T, Rosen CA, Schindler A, Whurr R , Woisard V. Validation of the Voice Handicap Index by assessing equivalence of european translations. *Folia Phoniatr Logop.* 2008;60(4):173-178.
64. Yiu EML. Impact and prevention of voice problems in the teaching profession - embracing the consumers' view. *J Voice* 2002;16(2):215-228.
65. Roy N, Bless DM, Heisey D. Personality and voice disorders: A multitrait-multidisorder analysis. *J Voice.* 2000;14(4):521-548.
66. Buck F, Drinnan M, Wilson J , Barnard IS. What are the illness perceptions of people with dysphonia: a pilot study. *J Laryngol Otol.* 2007;121(1):31-39.

Increase in voice level and speaker comfort in lecture rooms

Jonas Brunskog[a] and Anders Christian Gade
Acoustic Technology, DTU Electro, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark

Gaspar Payá Bellester
C/Sènia 1, ler. Monòver, E-03640 Alicante, Spain

Lilian Reig Calbo
C/Aitana, 48, BJ-A Muro de Alcoy, E-03830 Alicante, Spain

(Received 11 March 2008; revised 20 January 2009; accepted 20 January 2009)

Teachers often suffer from health problems related to their voice. These problems are related to their working environment, including the acoustics of the lecture rooms. However, there is a lack of studies linking the room acoustic parameters to the voice produced by the speaker. In this pilot study, the main goals are to investigate whether objectively measurable parameters of the rooms can be related to an increase in the voice sound power produced by speakers and to the speakers' subjective judgments about the rooms. In six different rooms with different sizes, reverberation times, and other physical attributes, the sound power level produced by six speakers was measured. Objective room acoustic parameters were measured in the same rooms, including reverberation time and room gain, and questionnaires were handed out to people who had experience talking in the rooms. It is found that in different rooms significant changes in the sound power produced by the speaker can be found. It is also found that these changes mainly have to do with the size of the room and to the gain produced by the room. To describe this quality, a new room acoustic quantity called "room gain" is proposed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081396]

PACS number(s): 43.55.Hy, 43.70.Aj [NX] Pages: 2072–2082

I. INTRODUCTION

The primary means of communication in most educational settings are speech and listening. The acoustics of the lecture room can restrict or support the speaker and improve the sound of the voice and the intelligibility of speech. The room acoustics in lecture rooms is therefore an important issue when considering the productivity and working environment in schools and other teaching situations. Thus, a large amount of work has been carried out within this field. However, the large body of published articles focuses on the point of view of the listener. It is therefore easy to find works on speech intelligibility in the room and advisable reverberation times (RTs) and background noise levels (BNLs) in order to achieve good learning conditions, etc., (see, e.g., Bistafa and Bradley[1]). There are also standards and recomendations,[2–4] indicating how well established this field is.

However, it is known that teachers often suffer from health problems or tension related to their voice. Recent works made it evident that the teacher's labor is one of the professions with high vocal demands.[5] Examples of other professions with high vocal demands are actors, singers, journalists, telephone operators, and military personnel. Studies show that a majority of teachers have experienced vocal problems, about one-tenth have severe problems, and 5% have experienced such severe, numerous, and frequent voice problems that their working ability is challenged.[5] For the teacher, in the long run, this voice load due to speaking in the classroom can result in voice disorders such as hoarseness and voice fatigue and can even force teachers to retire early from their profession. Lubman and Sutherland[6] disclosed that this is an important economic problem for governments and private schools.

Most teachers have probably experienced that different rooms vary in comfort when one speaks in them. However, even though the vocal problem is so important, just a few studies about the speaker and his behavior in and impression of the lecture room have been accomplished. One example is Kleiner and Berntson,[7] where the early reflections of the sound produced by the speaker were studied in a synthetic experimental setup. A system of loudspeakers in an anechoic chamber was used to simulate different rooms. All settings simulated rooms with different shapes but the same volume. The interest was in the effect of lateral and vertical early reflections on the speakers' comfort. Different combinations of delayed simulated reflections were tested. A paired-comparison test was used in order to find the setting preferred by the speakers. It was concluded that symmetrical settings were preferred over asymmetrical ones. There was however no significant difference between the different symmetrical settings, and perfectly symmetrical settings are not realistic in real rooms with a movable speaker. It can be noted that this was an entirely subjective study—no objective values were calculated from the simulated impulse responses. Kob *et al.*[8] presented results from a study where the voice status of 25 teachers were investigated using standard methods as applied by audimetrisists, phoniatricians, and speech therapists, in addition to an acoustic analysis of speech and voice samples. The acoustics of some rooms was

---

[a] Electronic mail: jbr@elektro.dtu.dk

also investigated, and the result of speaking in different rooms was analyzed dependent on the voice status. The results indicate an influence of both the room acoustics and the voice status on the voice quality of the teachers. But the study used RT and speech transmission index as the parameters describing the room acoustic environment. Thus, no clear distinction was made between the problem perceived by the listener and the speaker.

Several studies in which different voice parameters were measured in real classrooms have been reported, e.g., Rantala *et al.*[9,10] or Jonsdottir *et al.*[11] However, in these studies the influence of the room was not included. Instead, the focus here was to study different subgroups of speakers, e.g., with and without voice problems. The voice parameters were primarily the voice level [defined as the sound pressure level a distance of 1 m from the speaker] and pitch (more specifically the fundamental frequency F_0 of the voice signal) and fluctuations in these parameters.

Thus, the literature relating the room to the speaker and the voice signal produced is rather thin; not much information is available on how to design or improve the room in order to make a better environment for the speaker. However, such information is available in the field of acoustics of rooms for music performance. Also here, the majority of works deal with the conditions for the audience, but there have also been studies concerning how musicians experience and react to the room acoustics. An important example is Gade,[12] who, in a laboratory experiment in an anechoic chamber equipped with a loudspeaker system similar to that of Kleiner and Berntson,[7] let musicians play in and react to simulated sound fields. Gade[13] also carried out corresponding subjective and objective studies in real concert halls. In both cases the subjective response answered by the musicians were correlated with different objective measures. Gade found that the "support" provided by the room—the sensation that the room responds to his instrumental effort—is important for the musicians. Gade defines an objective measure, called *ST*, that correlates well with the sensation of support. *ST* is determined as

$$ST = 10 \log \frac{E_{20-x}}{E_{\text{dir}}}, \qquad (1)$$

where E_{20-x} is the energy in the impulse response from 20 ms to x ms (x being either 100 or 200 ms, or even infinity) [see Eq. (2)] and E_{dir} is the energy in the direct path, defined as $E_{\text{dir}}=E_{0-10}$, which is the energy within the first 10 ms. The impulse response is to be measured with a source-receiver distance of 1 m. Obviously, 1 m distance is larger than the typical distance between the musician's ear and his instrument, but this distance was still chosen to obtain a measure with sensible variation and dynamic range. *ST* is thus the fraction of energy coming later than 20 ms relative to the direct sound. In the absence of reflected sound *ST* equals $-\infty$ dB, and a zero support, $ST=0$ dB, means that the total contribution from the reflections equals the direct sound. This definition works well in large rooms where the direct part of the impulse response is clearly separated from the reflections, but measurements of ST is problematic for smaller rooms. Another problem with the definition in Eq. (1) is that it does not clearly reflect what happens close to the source, which at the same time is the position to be studied. In the real situation, e.g., in case of singing or speaking, the source is the mouth and the receiver position is the ear, just a few centimeters away. The direct path is thus described by the transfer function (or impulse response) from the mouth, around the head, to the ear in absence of reflections. How to deal with this is not obvious in case of the definition in Eq. (1). A third problem is that an anechoic chamber is included in the present study, and *ST* is undefined in such a room. Thus, in the present study we have made use of another definition using the measured impulse response of a setup with an artificial dummy head torso and taking as reference the measured value in an anechoic room. The new quantity is called room gain, with abbreviation RG and variable G_{RG} (see Sec. II C).

It seems likely that the vocal problems of teachers are due to the voice level being increased in different situations when teachers feel uncomfortable with the environment. The environment here not only includes the physical environment of the lecture room, but also the students and the overall working conditions. There are two hypotheses here, one being that vocal health problems are related to an environment where the speaker feels that he must increase his voice, the other being that the physical environment itself can cause the speaker to increase his voice. Only the latter will be tested in the present paper. The aim of this project was thus to find some of the parameters that cause the speaker to force their voice and situations when it is uncomfortable to speak.

Aspects not taken into account in this study are the influence of the audience, including the background noise (BN) produced by them, the change in voice during the day, the influence of voice problems of the subjects and other aspects related to the subjects (e.g., mood or attitude toward teaching), and the speech intelligibility in the rooms, subjectively or objectively. Moreover, the study only deals with nonamplified voices.

One question is then which objectively measurable parameters to include in the study. Real rooms are to be used, and the focus is on the speaker, not the listener. Thus, the parameters should be related to what the speaker experiences at the position where he speaks. Parameters related to speech recognition and intelligibility are therefore left out. The impulse response contains all information of the transfer path from source to receiver, and most measures can be calculated from it. It is however important that the source and receiver positions are as correct as possible. Parameters that are extracted from the impulse response are the RT and the RG. Parameters not included in the impulse response are those not directly related to the acoustic transfer path—that is, BN and the size of the room. Thus, four basic parameters are chosen to characterize each room—RT, RG, BNL and *volume*. However, different variants of these parameters were tested as well.

In the subjective study, most of the questions were related to the objective parameters. Thus, the subjects were asked about the impression of reverberation and support, as well as background level in the rooms studied. They were also asked about the general impression of speaking in the

TABLE I. The rooms used in the experiments and their objective values. All rooms are located at DTU, Lyngby, Denmark. ◊, number of questionnaire answers for each room.

Name	Abbrev.	V (m^3)	T_{30} (s)	T_{EDT} (s)	G_{RG} (dB)	$L_{BN,A}$ (dB)	nr.◊
Auditorium 81	A81	1900	1.06	1.12	0.28	41.8	14
Auditorium 21	A21	1220	1.53	1.72	0.29	53.5	19
Lecture r. 019	LR	190	0.46	0.40	0.42	47.5	21
Meeting r. 112	MR	94	0.42	0.33	0.58	47.5	17
Large anechoic ch.	ACH	1000	0.06	0.01	0	45.9	17
IEC listening room chamber	IEC	100	0.34	0.32	1.12	46.7	16

room and if they raised their voice when speaking. A question about echo phenomena was also included in order to be able to say if this parameter influences the general impression of the room.

The main findings in this paper is that the different rooms significantly change the sound power produced by the speaker. It is found that these changes mainly have to do with the size and the RG of the room.

II. METHOD

A. Method overview

Both subjective responses and objective measures of the room and of the voice level are collected. A selection of different natural acoustic environments are used—opposite of using a synthetic sound field. In simulated sound fields the variables can be changed rapidly and with precision in wide ranges. However, the sound quality is still limited due to the need for real time processing of the signals produced by the speaker. Moreover, the visual impression of the room cannot easily be included—this might be a positive aspect in many cases, but here it is important to get the visual size of the room and the distance to the audience right. Therefore, real rooms were chosen to be used—six in total. The range in the physical parameters of the rooms used was wide, including small meeting and listening rooms, a medium size lecture room, two larger auditoria, one with high RT and one with low RT, and a large anechoic room.

In the six rooms the sound power level produced by six speakers was measured. Each of the speakers held a short lecture (about 5 min). Objective room acoustic parameters were measured in the rooms as well, and a subjective questionnaire was handed out to about 20 persons who had experience in speaking in the rooms. A statistical analysis was then used to find relationships between the subjective responses and the objective measures.

B. The subjects

In the objective study six speakers were used. Three of these were teachers at Acoustic Technology, Ørsted*DTU; the other three were students in acoustics. In one of the rooms (meeting room 112, building 352), only five speakers were present. The speakers had no known voice pathology. Each speaker was instructed to give the same lecture in all rooms. However, as the speakers did not have a written text to read, the lectures were not identical. Most speakers used a laptop computer with a Powerpoint presentation as the basis of the speech. In order to get the background level identical, a laptop and a video projector (if available in the room) were present also for those not using it. All speakers were male, age about 20–55. There is a possibility that the speakers do not fully represent all relevant speakers, as it consisted of those finding it interesting to participate. Actually, the teachers participating were known to have weak voices (low voice power). However, most of the analysis are made on a relative voice power level (VPL) (see Sec. II C), which decreases the variance in the data. Another subset problem might be that all subjects were acousticians, a fact that might influence the result—we choose to believe that this has a minor influence only.

In the subjective study 21 subjects participated (between 14 and 21 responses were collected for each room, see Table I). The subjects were teachers and students in acoustics—the participants in the objective part were also present in the subjective part. Both male and female subjects aged between about 20 and 60 participated.

C. Objective measurements and equipment

1. Impulse response measurements

The impulse response $h(t)$ of the rooms is measured to calculate RT and RG. The equipment used for the measurements were power amplifier LAB 300 from LAB Gruppen, microphone unit type 4192-L-001 Brüel & Kjær (B&K), conditioning preamplifier Nexus type 2690 B&K, and sound level calibrator type 4231 B&K. In case of the reverberation measurements, an omnidirectional dodecahedron loudspeaker was used, and in case of the RG measurements a dummy head torso was used, as described below. The DIRAC software[14] was used with e-sweep excitation signal. The sweep length was 21.8 s.

2. Reverberation time

Generally, the most important room acoustic parameter is the RT (variable T_{30}) (see ISO 3382).[15] The *early decay time* (EDT) (variable T_{EDT}), is the RT determined from the first 10 dB range of the decay curve. The EDT is known to be more closely related to the subjective impression of reverberation than RT. In the analysis EDT was mainly used. (A reference of these basic room acoustic parameters is Kuttruff.[16])

The RT is calculated from the impulse response using the Schroeder method.[16] The RTs were calculated in octave bands. In order to describe the RT as a single number, the arithmetic mean of the RT in the octave bands centered in 500 and 1000 Hz is used.

3. Room gain

The transmission path from the mouth to the ear has three parts: bone conduction, a direct airborne part, and a room reflection part; it is the latter path that is of interest here. The perceived beneficial increase in the loudness caused by the room is assumed to be due to the early reflections as compared to the direct response without reflections, perceived as one's ability to hear oneself properly in the room. This is here denoted as a gain, or support, caused by the room. The parameter used in the present study is called RG (variable G_{RG}). It is defined as the energy (in decibels) of the impulse response measured between the mouth and the ear of a dummy head torso, taking as reference the corresponding measurement in the anechoic chamber where only the direct sound is present. As explained earlier, the reason for not using the support measure ST is that small rooms are also to be included in the present study, and then the definition of the ST is not appropriate, as the direct part of the impulse cannot be separated from the rest of the impulse response. Moreover, an anechoic chamber is included in the study, and here $ST=-\infty$.

The energy of an impulse response in a time interval t_1 to t_2 can be calculated as

$$E_{t_1-t_2} = \int_{t_1}^{t_2} h^2(t)dt, \tag{2}$$

where $h(t)$ is the impulse response. The energy in the entire impulse response is in the same way,

$$E = \int_{0}^{\infty} h^2(t)dt. \tag{3}$$

The corresponding impulse energy level is $L_E = 10 \log E/E_{\mathrm{ref}}$, where E_{ref} is the reference value. The RG is then defined as the energy in decibel in the signal relative to the direct energy as measured in the anechoic chamber,

$$G_{\mathrm{RG}} = L_E - L_{E,\mathrm{ach}} = 10 \log E/E_{\mathrm{ach}}, \tag{4}$$

where $L_{E,\mathrm{ach}}$ and E_{ach} are the impulse energy level and energy in the anechoic chamber, respectively.

The RG is related to the support ST, as defined in Eq. (1). If it is assumed that $E_{\mathrm{dir}} \approx E_{0-20} \approx E_{\mathrm{ach}}$ and $E_{20-x} \approx E_{20-\infty}$, then

$$ST \approx 10 \log \frac{E - E_{0-20}}{E_{\mathrm{ach}}} \approx 10 \log(10^{G_{\mathrm{RG}}/10} - 1). \tag{5}$$

A support value of $ST=0$ thus corresponds to $G_{\mathrm{RG}}=3$ dB, meaning that the reflections contribute with the same energy as the direct sound. It should, however, be noticed that the source/receiver distance is different in the definition of ST as compared to G_{RG}.

The equipment used was the same as described under the impulse response above, with the following changes: dummy head, head and torso simulator type 4128 with right ear simulator type 4158 and left ear simulator type 4159 B&K, and power amplifier for the sound source (the dummy mouth).

The dummy head was placed in the area where the speaker normally stands during the lecture (next to the blackboard or similar). The average of six different positions of the dummy head was used. Moreover, the average RG of the left and right channels was calculated and used in the data analysis.

The RG was calculated from the impulse response by means of postprocessing in MATLAB. All signals have been normalized with a maximum amplitude of the signal to 1 (amplitude of the first peak of the impulse response). Some problems with the signals were found during the analysis. Noise was found in all the signals. In order to reduce the effect of this problem, all the impulse response signals were truncated (cutted) so as to avoid the last part of the signal, which mainly contained noise. Thus, the noise effect was minimized, and it is judged that its influence can be disregarded.

The RG was calculated per octave band. In order to define the RG of the room with one characteristic value, the arithmetic mean of the RG in the octave bands between 125 Hz and 4 kHz is used.

4. Background noise level

In a speech situation the BNL (variable $L_{BN,A}$) is important. BNL can be defined as the sound pressure level of the noise measured in the absence of the sound under investigation—in this case the speech. The BN can originate from the ventilation systems, the outdoor environment and traffic, equipment such as computers and projectors, and the students/audience. As the BNL increases, the speaker may increase his voice to compensate and overcome the noise in order to be heard. The voice will be affected by the mental pressure due to the failure of being heard. The frequency content in the voice signal will then be changed—there will be more high frequency content due to an increased fundamental frequency. These changes are known as the Lombard effect; an early reference is Lane and Tranel.[17] The effect is included in ANSI-S3.5.[2] (Sometimes, the term "Lombard effect" is restricted to just the increase.) This is also closely related to the fact that in a situation with several people talking to each other, they increase their voice to overcome the BNL that is produced by all the persons speaking, producing a nonlinear feedback loop, see, e.g., Hodgson *et al.*[18] Naturally, the number of students and their behavior during the lecture also may play an important role here—the students will contribute to the background level and probably react in relation to the Lombard effect. However, this aspect is not part of the present work (due to schedule reasons and time limits); the present project is focused on the characteristics of the room only, leaving this important aspect to further research. The number of listeners present in the room was just a few (3–5) and adult, so there contribution to the BNL is assumed to be low. The BNL naturally present in the rooms (from the ventilation system, video projector, computers, etc.) was, however, registered.

The equipment used to measure L_{BN} is the same as for the impulse response measurements for the reverberation. The measurement duration is 21.8 s. The mean value of six microphone positions have been used in all rooms. The positions were in the area the teacher was using. To get a single value, the A-weighted level $L_{BN,A}$ is used. The equipment used by the speakers (laptop computer and projector) was present in the room during the measurement.

5. Room volume

Of the objective parameters describing the rooms, finally the size or *volume* (variable V) has also been used. The hypothesis here is that the speakers unconsciously adjusts the level of the voice depending on the room size and the distance to the audience, so that everyone is likely to hear. However, it is not clear if it is the volume by itself or a typical length scale in the room that is the primary variable here. Thus, V, $\log V$, and $\sqrt[3]{V}$ were all tested.

6. Voice power level

With the rooms defined, the last step is to define the behavior of the speaker in the room. In this project, this is described by the strength of the speaker's voice. The quantity used here is the *voice power level* (VPL) (variable L_W) that is the source power in decibel. Thus, the sound power level produced during speech by the different test speakers was measured in the different rooms.

The measurement of the VPL is a central issue of this paper. The measurements are made with a computer phone conversation headset, placed on the speaking subjects. The experimenter made sure that the position of the headset was fixed to the same position in all measurements, about 3 cm from the mouth. The equipment consisted of Headset Creative HS-390 and sound analyzer DIRAC. The signals were measured while the speaker was lecturing for about 5 min. An average of 15 signal segments of 21.8 s were used for each subject.

A calibration procedure was needed to transfer the measured signals to sound power level L_W. The dummy head torso equipped with a loudspeaker in the mouth was placed in a reverberation chamber with the headset attached in the same position as described above. A broad band noise signal was fed to the loudspeaker and measured simultaneously by the headset and with microphones in the reverberant field of the room according to sound power level standard measurements (ISO 3743-2). The measurements and calibrations were performed in octave bands. The relation between the sound power of a source and the sound pressure level in one position determined by a microphone can generally be expressed as $L_W = L_p + G$, where G is a gain constant for the setup (depending on the source-receiver distance and source directivity) and L_p is the sound pressure level as measured by the headset. It is now assumed that the microphone is so close to the source that only the direct field is present (i.e., the signal to noise ratio is assumed to be so good that the room response can be neglected). Moreover, it is also assumed that all speakers had the same directivity, equal to that of the dummy head. It is thus assumed that G is constant during all measurements in all rooms. (Note that this quantity obviously is different from G_{RG}.) Finally, having determined both L_W and L_p at the same time in the reverberation chamber, the gain constant G is determined.

The VPL is determined in octave bands from 125 Hz to 4 kHz. In order to have a single value, three different methods are tested: linear ($L_{W,l}$) and A-weighted ($L_{W,A}$) absolute VPL and linear VPL relative to the VPL in the anechoic chamber (ACH), ΔL_W. Note that the subtraction is made for each speaker, so that ΔL_W is made relative to the VPL for that speaker in the ACH. In this way the variance is reduced. The ACH room was chosen as it was the room with the highest average VPL. (The room with the lowest VPL, the meeting room (MR), was also considered to be used as a reference, but this idea was dropped as not all speakers spoke in this room.)

D. The rooms

To get good statistic results, it is important to apply a wide range and even distribution of the different physical variables defining the room. The rooms and the values of the objective measures are given in Table I. The rooms were a small MR and an IEC listening room (IEC), a medium size lecture room (LR), two larger auditoria, one with high RT (A21) and one with low (A81), and a large anechoic room (ACH). Including the anechoic room means that the subjects have a very clear reference for RT and RG—which both are zero in this room. Besides, ACH is relevant as it represents out door surroundings. The range covered by the volume, the RT and the RG can be considered large in comparison to what can be found in real life situations. For the BN, only the naturally present BN was included. Thus, this variation is small as compared to what can be found in real life situations.

E. Questionnaire and subjective response

In an attempt to relate the objective parameters of the room and the VPL to the subjective experience of the rooms, a questionnaire was designed. The questions were formulated after a first interview with a few teachers. The parameters considered are described below.

The questions were answered for each of the rooms in which the subject had experience talking. Thus, the subjects were not necessarily in the room when the questions were answered—in an attempt to increase the number of answered questionnaires. 21 subjects answered the questions; the number of answers for each room varied between 14 and 21 (see Table I). The questions were answered on a scale from 1 to 7. Only the natural numbers were used. Taking the arithmetic average of these answers, a subjective response variable S_i was formed, where the index i is the abbreviation of the question (see below).

The questions are the following (the questions are given in italic fonts)—it should, however, be noted that the these are not exactly the questions used (due to poor English).

Do you consider this room to be good to speak in? This question is referred to the degree of comfort and how easy it is to speak in the room. The rank is between low if the room

TABLE II. Significance test of the subjective response parameters and VPL parameters (different versions) using ANOVA. The following symbols are used: * means significant at the 5% level, *** means significant at the 0.1% level, and — means no significance at the standard levels.

Question	GSI	TR	ECHO	BN	IV	ES	$L_{W,l}$	$L_{W,A}$	ΔL_W
p-value	$<10^{-6}$	$<10^{-6}$	0.046	0.16	$<10^{-6}$	$<10^{-6}$	0.13	0.11	0.036
Significance	***	***	*	—	***	***	—	—	*

is not good to speak in and high if it is good to speak in. This parameter is labeled GSI, variable S_{GSI}.

Do you think the RT is too long in the room? This question clearly refers to the objective parameter of RT. The rank in this case goes from "no" if the reverberation is not too long or "yes" if it is too long. This parameter is labeled TR, variable S_{TR}.

Have you noticed echo phenomena in the room? The sensation of echo might influence the general impression of the room, so this response is introduced even though it is not represented in the objective parameters. The answers should be covered between low if no echo is noticed and high if there is too much echo. This parameter is labeled ECHO, with variable S_{ECHO}. A low score is considered good.

Is the BN too high in the room? The subjects' response might be from "yes" if they think there is a lot of BN in the studied room to "no" if they think that there is no noise in the room. This parameter is labeled BN, variable S_{BN}. A low score is considered good.

Do you have to increase your voice in this room to be heard? This question is interrelated to the sound power level. The answer is between "no" if the subjects think they did not increase the voice, to "yes" if they did have to increase the voice. This parameter is labeled IV, variable S_{IV}. A low score is considered good.

Is there enough support in this room? This has to do with whether the room helps the speaker to hear himself. The rank is between bad support if they believe that the room does not yield support at all and good support if the support is sufficient. This parameter is labeled ES, variable S_{ES}. A high score is considered good.

F. Data analysis

The statistical analysis of the data was carried out in MATLAB. This analysis incorporates analysis of variance (ANOVA), correlation coefficients, and linear regressions.

In order to find relationships between the subjective responses and the objective parameters—a psychometric function—some postprocessing has been done. The psychometric function, relating a subjective parameter S with upper limit $S_{\max}$ and lower limit $S_{\min}$, and an objective parameter d (or a linear combination between such parameters) should be an S-shaped function. The reason for this is that the objective parameter is not bounded, $d \in [-\infty\infty]$, but the subjective parameter is bounded, $S \in [S_{\min} S_{\max}]$. One choice of such a function is

$$S = \frac{S_{\max} - S_{\min}}{1 + e^{-d}} + S_{\min} \tag{6}$$

(this choice of psychometric function is taken from paired-comparison theory[19,20]). The point of using such a relation is that S has a finite domain $S \in [S_{\min} \cdots S_{\max}]$, whereas d might have an infinite domain $d \in [-\infty \cdots \infty]$. In the present case $S_{\max}=7$ and $S_{\min}=1$. Solving for d in Eq. (6), a suitable transformation from the finite S-domain to the infinite d-domain of the objective measures is found,

$$d_S \equiv -\ln \frac{S_{\max} - S}{S - S_{\min}}. \tag{7}$$

The parameter d_S can be used as the dependent variable in regressions connecting objective measures to subjective response.

However, in some cases the objective parameter is non-negative, $d > 0$. That is the case for the RT and the RG. Moreover, in the present study the extreme situation of zero RT and RG is included in the study due to the use of the anechoic chamber. In these cases Eqs. (6) and (7) have to be modified. The following equations then applies:

$$S = \frac{2(S_{\max} - S_{\min})}{1 + e^{-d}} + 2S_{\min} - S_{\max} \tag{8}$$

and

$$d_S \equiv -\ln \frac{S_{\max} - S}{S - 2S_{\min} + S_{\max}}. \tag{9}$$

However, in many cases the range of the objective parameter is so small that the error of using a linear regression directly between d and S is small. That is actually the case in the present study, and in the result section below, the regressions are often performed both using the psychometric function and directly between S and d.

III. RESULTS

A. Validity and quality of the data

An ANOVA is used to examine if the variations in the data are sigificant. The left part of Table II presents these results concerning the subjective parameters. The variations are significant except for BN, where no significant variations are found at the 5% level or better (p-value of 0.16), and for detection of echo ECHO, where the variations are significant at the lower level of 5% (p-value of 0.046), but not higher. It should here be noted that the variation in the background level of the rooms was small and that there are no known problems with echo or flutter echo in the rooms used. In the

TABLE III. The rooms used in the experiments and their objective values.

Abbrev.	$L_{W,l}$ (dB)	$L_{W,A}$ (dB)	ΔL_W (dB)
A81	62.9	60.0	−1.30
A21	63.9	60.9	−0.08
LR	62.9	60.1	−1.93
MR	58.7	55.2	−4.33
ACH	65.0	62.1	0
IEC	59.8	57.0	−4.32

same way, the right part of Table II presents the significance test of different versions of the VPL. Here the significance of the variations in the data is less, probably due to the lower number of subjects participating. However, taking VPL relative to the result in the anechoic chamber, ΔL_W, yields significant variations at the 5% level (p-value of 0.036).

In the further analysis, only $L_{W,l}$ and ΔL_W will be used describing the VPL. $L_{W,A}$ is disregarded as it does not increase the significance much and is not as straightforward as $L_{W,l}$. Moreover, results depending on the subjective responses BN and absolute VPL, $L_{W,l}$, should be considered only as trends.

B. Relationships among objective parameters

The objective parameters used to describe the rooms were presented in Table I. The objective changes in the VPL are presented in Table III. The correlation matrix between these parameters is given in Table IV. It should be noted that the VPL measures correlate well with the volume, especially log V, and the RG G_{RG}. There is no significant correlation between the VPL measures and RT and BN. It should also be noted that the RT measures and the BN measure do not correlate significantly with any other measure.

Note that the correlation between support ST as calculated in Eq. (1) and the other parameters is not included here as the support is undefined in the anechoic chamber due to the lack of reflections (the value would be $-\infty$).

The results of single variable linear regression are found in Table V. Only results with $p < 0.1$ are shown. It is shown once again that log V and G_{RG} correlate well with VPL. A multiple linear regression model using these two variables is

TABLE V. Single variable linear regression. Only regressions with $p < 0.1$ are shown. Left: between VPL ΔL_W and the objective parameters. Right: between S_{GSI} and the subjective parameters. The variables b_0 and b_1 are the regression constants, the constant term and the linear term, respectively.

Dependent variable	ΔL_W		S_{GSI}	
Independent variables	log V	G_{RG}	S_{IV}	S_{ES}
R^2	0.78	0.74	0.73	0.61
p	0.02	0.03	0.03	0.07
b_1	2.94	−4.40	−0.90	0.72
b_0	−9.64	−0.021	8.30	1.91

$$\Delta L_W = -5.68 + 1.81 \log V - 2.28 G_{\mathrm{RG}}, \qquad (10)$$

with $R^2 = 0.86$ and $p = 0.05$. The improvement of using two parameters is described by the fact that R^2 increases from 0.78 to 0.86 and at the same time the model is at the limit of significance. The model is shown in Fig. 1.

C. Relationships among subjective parameters

The subjective response parameters are presented in Table VI. The correlation matrix for these parameters is given in Table VII. Using the objective domain transformation according to Eqs. (7) and (9) yielded similar results.

The results of single variable linear regressions are found in the right part of Table V. Only results with $p < 0.1$ are shown. It can be seen that S_{IV} and S_{ES} correlate well with S_{GSI}; these regressions are also shown in Figs. 2 and 3. A multiple linear regression model using these two variables is

$$S_{GSI} = 6.82 - 0.715 S_{IV} - 0.189 S_{ES}, \qquad (11)$$

with $R^2 = 0.74$ and $p = 0.13$. Thus, the improvement of the two parameter model was not large, and the model is not significant. This is probably due to a high linear dependency between S_{IV} and S_{ES}.

D. Relationships between subjective and objective parameters

Table VIII shows the correlation between the objective parameters and the subjective responses (the number of objective parameters has been reduced as T_{30} and $\sqrt[3]{V}$ have been

TABLE IV. Correlation matrix for the objective measures, including the VPL. Only correlations with p-values lower than 0.2 are shown. In parentheses: $0.2 > p > 0.1$; roman upright: $0.1 > p > 0.05$; italic: $0.05 > p > 0.01$; boldface: $p < 0.01$.

Objec.	$L_{W,l}$	ΔL_W	T_{30}	T_{EDT}	V	log V	$\sqrt[3]{V}$	L_{BN}	G_{RG}
$L_{W,l}$	**1**	**0.97**	—	—	(0.63)	*0.82*	0.76	—	−0.81
ΔL_W	**0.97**	**1**	—	—	(0.72)	*0.88*	*0.84*	—	*−0.86*
T_{30}	—	—	**1**	**1.00**	—	—	—	—	—
T_{EDT}	—	—	**1.00**	**1**	—	—	—	—	—
V	(0.63)	(0.72)	—	—	**1**	**0.96**	**0.98**	—	(−0.63)
log V	*0.82*	*0.88*	—	—	**0.96**	**1**	**1.00**	—	−0.76
$\sqrt[3]{V}$	0.76	*0.84*	—	—	**0.98**	**1.00**	**1**	—	(−0.72)
L_{BN}	—	—	—	—	—	—	—	**1**	—
G_{RG}	−0.81	*−0.86*	—	—	(−0.63)	−0.76	(−0.72)	—	**1**

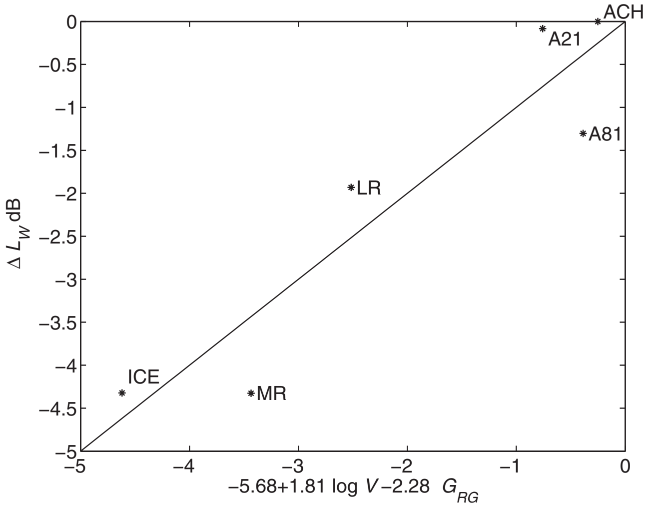

FIG. 1. Regression model (10) versus the real data of increase in VPL ΔL_W. Room abbreviation according to Table I.

ignored). Using the objective domain transformation according to Eq. (7) and (9) again yields similar results (a slightly better correlation on average).

The results from single variable linear regressions are found in Table IX. Only the regressions with $p<0.1$ are shown. The regression between IV and ΔL_W is shown in Fig. 4, and that between TR and T_{EDT} is shown in Fig. 5. A multiple linear regression model for IV using two variables is

$$S_{IV} = -0.198 + 1.73 \log V - 1.11 G_{\mathrm{RG}}, \qquad (12)$$

with $R^2=0.90$ and $p=0.03$. The improvement of using two parameters is described by the fact that R^2 increases from 0.86 to 0.90 while the model is still significant.

IV. ANALYSIS AND DISCUSSION

The ANOVA test in Table II indicates that in general the statistical quality of the subjective data is better than in the VPL data. One reason for this is probably the higher number of participants in the subjective questionnaire (about 20) as compared to the VPL measurements (about 6). However, it is known that it is difficult to get statistically consistent data for the voice strength (see, e.g., Rantala *et al.*[9]). Anyway, in the present study significant variations in the VPL data are found in case of the relative VPL, ΔL_W, using just six subjects. One reason for this is the normalization procedure of the data by taking the value relative to the anechoic chamber. In this way the natural variation in VPL among the subjects is reduced, and only the increments for different rooms are studied. Moreover, using a wide range of different rooms—including the anechoic chamber, large auditoriums, and small meeting rooms—is likely to increase the variation in VPL.

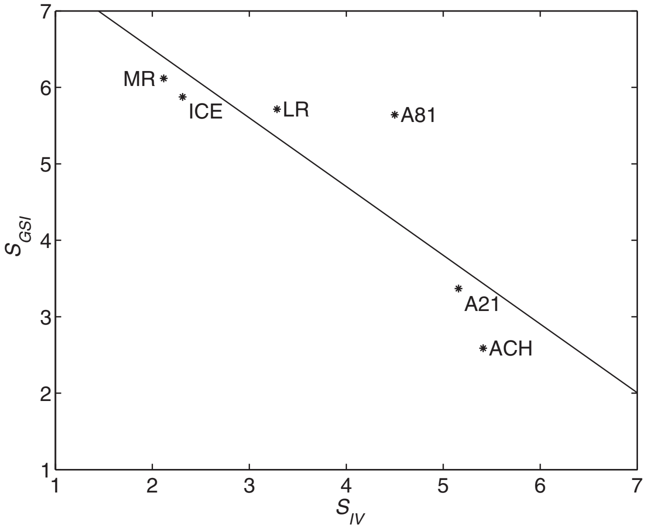

FIG. 2. Regression model between subjective variables S_{GSI} (good to speak in) against S_{IV} (increase voice), according to right part of Table V. Room abbreviation according to Table I.

Considering Table IV, room volume and RG show high correlation with the VPL. An increase in volume increases the VPL, and an increase in RG decreases the VPL. These results are significant if considering ΔL_W related to log V and G_{RG}. Of the size measures, the logarithm of the volume, log V, has the highest correlation. One can regard $V^{1/3}$ to be a typical length scale of the room and log V to be related to the average sound pressure level in the room for a given source power level. Thus, the fact that the increase in VPL is better correlated to log V than $V^{1/3}$ suggests that the aural cues might be more important than the visual cues. The VPL relative to the value in the anechoic chamber, ΔL_W, correlates in general better than the absolute linear VPL, $L_{W,l}$. This is probably linked to the fact that ΔL_W has higher significance than $L_{W,l}$ in the ANOVA test in Table II. Equation (10) expresses the relationship between ΔL_W, log V, and G_{RG}, also shown in Fig. 1. In Table VIII there is a trend that ΔL_W is correlated with ES, the question related to support in the room. Moreover, log V and G_{RG} are correlated to IV, the

TABLE VI. The rooms used in the experiments and their subjective response values. The scale is between 1 and 7. The notation is $\bar{S}/s$, where $\bar{S}$ is the average value and s is the standard deviation. In the further analysis the average value is used then denoted as S.

Abbrev.	S_{GSI}	S_{TR}	S_{ECHO}	S_{BN}	S_{IV}	S_{ES}
A81	5.64/0.74	2.64/1.34	1.93/1.64	4.00/1.52	4.50/1.34	3.29/0.83
A21	3.37/1.54	5.16/1.50	3.42/2.11	3.74/1.59	5.16/1.26	4.16/0.96
LR	5.71/0.78	1.76/0.54	2.95/2.01	4.33/1.43	3.29/1.27	5.05/0.86
MR	6.12/1.27	2.00/1.00	2.53/2.03	4.59/1.80	2.12/1.05	5.53/0.94
ACH	2.59/2.03	1.00/0	1.41/1.46	5.29/2.73	5.41/2.12	1.29/0.99
IEC	5.88/1.54	1.63/1.08	2.38/2.31	5.06/2.38	2.31/1.01	5.50/0.97

TABLE VII. Correlation matrix for the subjective measures using the subjective scale S. Only correlations with p-values lower then 0.2 are shown. In parentheses: $0.2 > p > 0.1$; roman upright: $0.1 > p > 0.05$; italic: $0.05 > p > 0.01$; boldface: $p < 0.01$.

Subj.	GSI	TR	ECHO	BN	IV	ES
GSI	**1**	—	—	—	*−0.85*	0.78
TR	—	**1**	(0.71)	*−0.84*	—	—
ECHO	—	(0.71)	**1**	(−0.66)	—	0.66
BN	—	*−0.84*	(−0.66)	**1**	—	—
IV	*−0.85*	—	—	—	**1**	*−0.85*
ES	0.78	—	0.66	—	*−0.85*	**1**

question if the subject had to increase the voice to be heard. There is also a trend that log V and G_{RG} are correlated to ES. These results confirm the results above.

Considering again Table IV, RT and BNL did not show any correlation with the VPL. Both of these results can seem surprising; RT is the generally most frequently used room acoustic measure, and BN is known to increase the speech level in other circumstances, e.g., in connection with the Lombard effect.[18] However, there is an important difference between these parameters in the present study. The variation in the RT data is rather large, T_{EDT} from 0.01 s in the anechoic room to 1.53 s in auditorium 21, but the variation in background level is small, from 41.8 dB (A) in auditorium 21 to 53.5 dB (A) in auditorium 21 (see Table I). "Large" and "small" should be understood as relative to what is normally found in lecture rooms. Moreover, the BNL in the room used was too low to influence speech. It is thus quite likely that a dependency in BN could be found if more extreme values had been included. The same conclusion does not apply for the RT. Moreover, in Table VIII it can be noted that ΔL_W is not correlated with the corresponding subjective responses TR or BN, which confirms the discussion above.

Considering the correlation among the subjective responses (Table VII), it can be noted that the question of whether the room is good to speak in, GSI, is correlated with the question about increase in voice level to be heard, IV. Thus, the ability to make oneself heard is judged to be important in the general judgment of the room. This is confirmed in Table VIII where GSI is correlated with ΔL_W. There is also a trend that GSI is correlated to ES, the question of whether there is enough support in the room. The other questions (TR, ECHO, and BN) do not show any correlation. It can thus be concluded that a room is good to speak in if it has support, and it is not necessary to increase the voice too much.

In Table VII it can also be seen that the question of whether the RT is too long, TR, is correlated to the question of whether there is too much BN (with negative sign due to the orientations of the subjective scales). Moreover, in Table VIII it is found that also T_{EDT} is correlated to BN but L_{BN} is not. This might seem strange. However, it should be remembered here that the questionnaire was not answered at the same time as the measurements, and that the subjects had the option to answer it while being elsewhere. Thus, BN is rather the experience of the BN as they could remember it. The most severe source of BN is probably the students present during the lecture. In the light of the Lombard effect, it is likely that this noise increases with increasing RT. It is thus not so surprising that T_{EDT} turns out to correlate well with BN. Thus, the subjective response BN does not refer to and is not related to the measured BN.

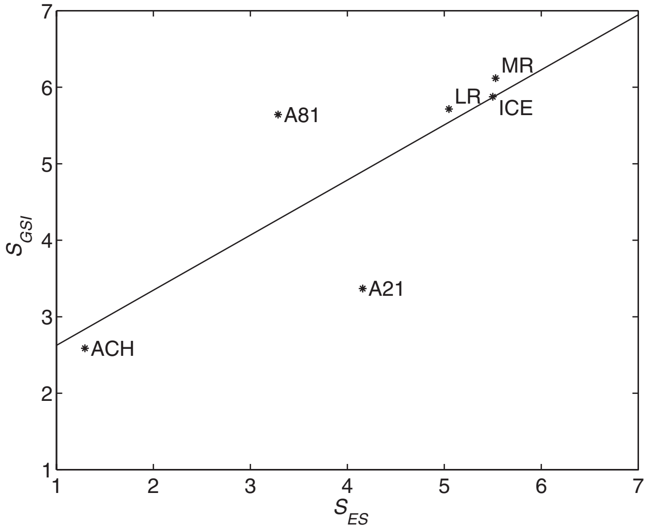

FIG. 3. Regression model between subjective variables S_{GSI} (good to speak in) against S_{ES} (enough support), according to right part of Table V. Room abbreviation according to Table I.

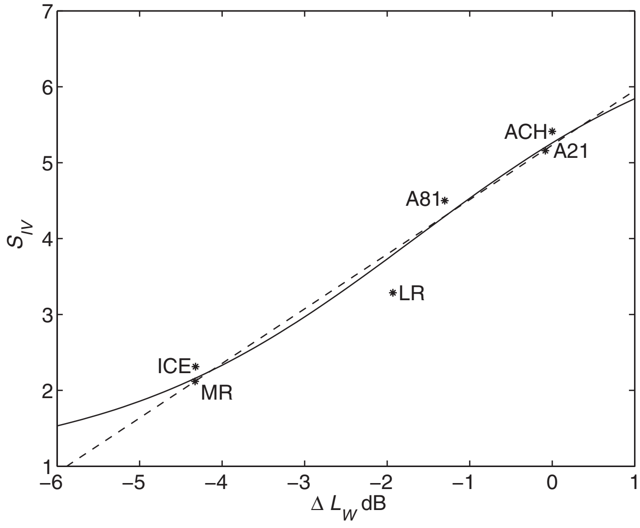

FIG. 4. Regression model between subjective variable S_{IV} (increase voice) against increase in VPL ΔL_W according to Table IX. Room abbreviation according to Table I. Solid line: Using objective domain transformation equation (7). Dashed line: Linear regression.

TABLE VIII. Correlation matrix for the objective and the subjective measures using the subjective scale S. Only correlations with p-values lower than 0.2 are shown. In parentheses: $0.2 > p > 0.1$; roman upright: $0.1 > p > 0.05$; italic: $0.05 > p > 0.01$; boldface: $p < 0.01$.

Obj. and subj.	S_{GSI}	S_{TR}	S_{ECHO}	S_{BN}	S_{IV}	S_{ES}
$L_{W,l}$	−0.80	—	—	—	**0.94**	−0.80
ΔL_W	*−0.82*	—	—	—	**0.98**	−0.79
T_{EDT}	—	**0.96**	—	*−0.90*	—	—
V	—	—	—	—	0.79	(−0.65)
$\log V$	(−0.63)	—	—	—	**0.93**	−0.77
L_{BN}	—	(0.65)	0.78	—	—	—
G_{RG}	0.68	—	—	—	*−0.83*	0.80

In Table VII it is also found that there is a trend that the question of whether echo is noticed, ECHO, is correlated to the question of whether there is enough support in the room, ES. This can be interpreted as follows: the reflections that contribute to the RG and support also might be imagined to cause echo phenomena, e.g., flutter echo. However, ECHO does not show big influence on any other parameter and is not correlated with GSI or IV, so it is judged that echo phenomena have not influenced the results. None of the rooms are known to have problems with flutter echo.

In Table VII the question of whether there is enough support in the room, ES, is correlated to the question of whether the subject had to increase the voice to be heard, IV. This seems natural, and it is also reflected in the correlation between ΔL_W and G_{RG} among the objective measurements (Table IV).

The strong correlation between the subjective response of increasing the voice, S_{IV}, and the objectively measured VPL should be noticed in Table VIII. This can be interpreted as the subjects being aware that they have to increase the voice in the room.

In Table VIII T_{EDT} is strongly correlated to TR. Thus, the subjects are aware of the RT. It should then be remembered that all subjects were teachers or students in acoustics and therefore familiar with the concept of RT.

Concerning the frequency rang of RT and RG: the frequency rang used (the octave bands from 125 Hz to 4 kHz for the RG and 500 Hz and 1 kHz octave bands for the RT) has in this study been assumed to be most responsible for the impression of the two measures. Different versions of the parameters have been tested, but not reported, and the chosen definitions and frequency range give good correlation. However, there probably is a need for more research in order to finetune the measures.

Using the regression between ΔL_W and IV (Table IX and Fig. 4), some preliminary design guidelines can be proposed. If a subjective response of $S_{IV} \leqslant 3$ is regarded as a good room, the model yields that this corresponds to $\Delta L_W \leqslant -3.1$ dB. Now, using the model in Eq. (10) (see Fig. 1), this corresponds to $G_{RG} \geqslant 0.80 \log V - 1.1$ dB. Thus, for a room with volume 100 m^3 the RG should be $G_{RG} \geqslant 0.5$ dB, and for a room with volume 1000 m^3 the RG should be $G_{RG} \geqslant 1.3$ dB. It should however be noted that such guidelines are preliminary, and should not be used before further evidence has been obtaind. Also note that the recommended values might be difficult to realize in reality for large auditoriums. Thus, these guidelines are limited to smaller rooms and rooms without voice amplification systems.

V. CONCLUSIONS

The voice power relative to the value in the anechoic chamber varies significantly between room.

The increase in the voice power produced by a speaker lecturing in a room is correlated with the size of the room (especially $\log V$) and the gain produced by the reflections in the room, G_{RG}. These relations are significant.

No significant correlation is found between the increase in the voice power and the RT or background level of the

TABLE IX. Single variable linear regression between subjective and objective variables. Only regressions with $p < 0.1$ are shown. The upper part uses the subjective domain S, and the lower part uses the objective domain d_S according to Eqs. (7) and (9).

Dependent variable	S_{GSI}	S_{TR}	S_{IV}		
Independent variables	ΔL_W	T_{EDT}	ΔL_W	$\log V$	G_{RG}
R^2	0.68	0.92	0.96	0.86	0.69
p	0.04	0.003	0.0006	0.007	0.04
b_1	−0.64	2.20	0.72	2.27	−3.13
b_0	3.61	0.94	5.23	−2.12	5.20
R^2	0.71	0.89	0.97	0.86	0.69
p	0.03	0.005	0.0004	0.008	0.04
b_1	−0.50	0.903	0.538	1.68	−2.32
b_0	−0.27	−0.075	0.895	−4.55	0.863

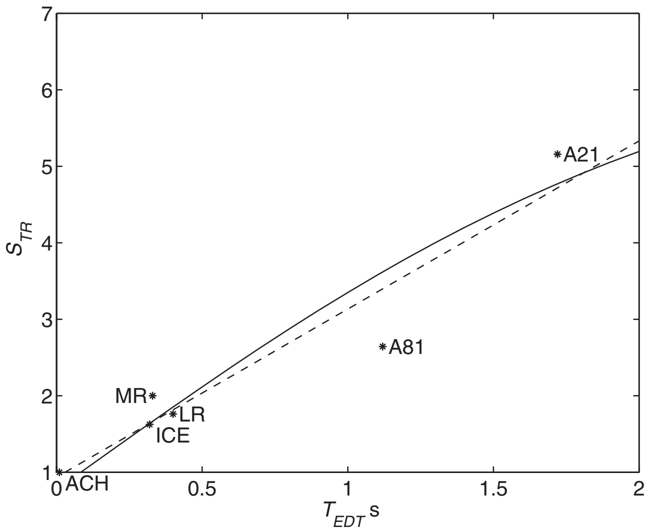

FIG. 5. Regression model between subjective variable S_{TR} (reverberation) against early RT T_{EDT} according to Table IX. Room abbreviation according to Table I. Solid line: Using objective domain transformation equation (9). Dashed line: Linear regression.

room in this study. The latter is probably due to the too small variations in the background levels in the rooms studied.

The general impression of whether a room is good to speak in is linked to the impression of whether it is necessary to increase the voice in the room and if the room provides support to the speaker. The former relation is significant, and the latter is only a trend.

There is a significant correlation between the question of whether the subject had to increase the voice and the actual increase in voice power. There is also a significant correlation between the question about the reverberation in the room and the measured RT. This means that the subjects participating were aware of these parameters.

ACKNOWLEDGMENTS

The authors are grateful to the colleagues and students at DTU participating in the study. Most of the practical work presented in this paper were carried out by two students at DTU during their MSc projects, Gaspar Payá Bellester and Lilian Reig Calbo.

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

[2] ANSI-S3.5, Methods for the calculation of speech intelligibility index, ANSI, American National Standard Institute, New York, 1997.

[3] *Building Bulletin 93, Acoustic Design of Schools: A Design Guide*, edited by C. Hopkins (Dep. Education and Skills, London, UK, 2004).

[4] IEC-60268-16, Sound system equipment—Part 16: Objective rating of speech intelligibility by speech transmission index, IEC, International Electrotechnical Commission, Geneva, 1998.

[5] E. Vilkman, "Voice problems at work: A challenge for occupational safety and health arrangement," Folia Phoniatr Logop **52**, 120–125 (2000).

[6] D. Lubman and L. C. Sutherland, "Good classroom acoustics is a good investment," in *17th ICA Proceedings* (Universita di Roma "La Sapienza," Rome, Italy, 2001), Vol. **V**, pp. 138–139.

[7] M. Kleiner and A. C. O. Berntson, "Early frontal plane reflections preferred for talkers," in *12th ICA Proceedings* (Canadian Acoustical Association, Toronto, Canada, 1986).

[8] 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," Acoust. Sci. & Tech. **29**, 86–94 (2008).

[9] L. Rantala, L. Paavola, P. Körkkö, and E. Vilkman, "Working-day effects on the spectral characteristics of teaching voice," Folia Phoniatr Logop **50**, 205–211 (1998).

[10] L. Rantala, E. Vilkman, and R. Bloigu, "Voice changes during work: Subjective complaints and objective measurements for female primary and secondary schoolteachers," J. Voice **16**, 344–355 (2002).

[11] V. Jonsdottir, A.-M. Laukkanen, and E. Vilkman, "Changes in teachers speech during a working day with and without electric sound amplification," Folia Phoniatr Logop **54**, 282–287 (2002).

[12] A. C. Gade, "Investigations of musicians' room acoustic condition in concert halls. Part I: Method and laboratory experiments," Acustica **69**, 193–203 (1989).

[13] A. C. Gade, "Investigations of musicians' room acoustic condition in concert halls. Part II: Field experiments and synthesis of results," Acustica **69**, 249–262 (1989).

[14] Brüel &Kjær, Product data, Dirac Room Acoustics Software, type 7841, www.bksv.com (Last viewed 2/21/09).

[15] ISO-3382, Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters, ISO, International Organization for Standardization, Geneva, 1997.

[16] H. Kuttruff, *Room Acoustics*, 4th ed. (Spoon, London, 2000).

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

[18] M. Hodgson, G. Steininger, and Z. Razavi, "Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments," J. Acoust. Soc. Am. **121**, 2023–2033 (2007).

[19] P. V. Rao and L. L. Kupper, "Ties in paired-comparison experiments: A generalization of the Bradley-Terry model," J. Am. Stat. Assoc. **62**, 194–204 1967.

[20] A.-C. Johansson, P. Hammer, and E. Nilsson, "Prediction of subjective response from objective measurements applied to walking sound," Acta. Acust. Acust. **90**, 161–170 (2004).

Comment on “Increase in voice level and speaker comfort in lecture rooms” [J.Acoust.Soc.Am. 125, 2072-2082 (2009)]

David Pelegrín-García[a]
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark

(Dated: January 5, 2011)

Recently, a paper written by Brunskog *et al.* “Increase in voice level and speaker comfort in lecture rooms” [J.Acoust.Soc.Am. 125, 2072-2082 (2009)] related teachers’ variation in vocal intensity during lecturing to the room acoustic conditions, introducing an objective parameter called “room gain” to describe these variations. In a failed attempt to replicate the objective measurements by Brunskog *et al.*, a simplified and improved method for the calculation of room gain is proposed, in addition with a new magnitude called “voice support”. The new measurements are consistent with those of other studies and are used here to build two empirical models relating the voice power levels measured by Brunskog *et al.* to the room gain and the voice support.

PACS numbers: 43.55.Hy, 43.70.Mn

Brunskog *et al.* published “Increase in voice level and speaker comfort in lecture rooms” previously in this journal.[1] Their work showed a possible influence of room acoustics (through a new parameter named “room gain”) on the vocal intensity used by teachers for talking in rooms. In addition, different subjective aspects regarding the perceived acoustic conditions while talking were studied by means of questionnaires. The work extended its relevance to the areas of ergonomics and occupational health, as it described an interaction between man and environment with possible consequences for voice health originated from working conditions. A recent epidemiological study has shown that teachers with voice problems rate classroom acoustics as an element affecting their voice much more often than those without voice problems.[2] In this context, the work of Brunskog *et al.* could offer a reference dataset to compare the vocal performance of teachers’ with and without voice problems under different acoustic conditions. However, it has been impossible to replicate the room gain measurements of Brunskog *et al.* in the original rooms of their study. The aim of this paper is to provide a more accurate and replicable dataset relating the voice power levels measured by Brunskog *et al.* to the objective parameters “room gain” and “voice support” derived with an alternative method. The first section presents the definition of room gain according to Brunskog *et al.*’s method, pointing out some potential limitations, and it is followed by the definition of room gain and voice support according to an alternative method. The second section compares the objective measurements in the rooms of Brunskog *et al.* as they appear in the original study and with the alternative method. The last section describes two empirical models relating the voice power level to the room gain and the voice support.

NOTE: The terms vocal intensity, voice level, and voice power level L_W are used in this paper to express the total radiated speech power from a talker. While the first term is used as a qualitative description, the other two terms are used indistinctly to express a quantitative magnitude.

I. ROOM ACOUSTIC PARAMETERS FOR A TALKER

Two equivalent metrics that characterize the effect of room acoustics as perceived by a talker are used: “room gain” (G_{RG}) and “voice support” (ST_V).

Brunskog *et al.* defined the room gain as the degree of amplification produced by the room on the talker’s voice, as perceived by the talker himself. The calculation of room gain proposed in Brunskog *et al.* requires the measurement of two impulse responses (IR) corresponding to the sound transmission path between the mouth and the ears of a dummy head: one at the room of interest $h(t)$, and another one at an anechoic chamber $h_{ach}(t)$. From these two measurements, the energy levels of the IR at the position of interest, L_E, and at the anechoic room, $L_{E,ach}$ are calculated:

$$L_E = 10 \log \frac{\int_0^\infty h^2(t)\ \mathrm{d}t}{E_0}, \tag{1}$$

$$L_{E,ach} = 10 \log \frac{\int_0^\infty h_{ach}^2(t)\ \mathrm{d}t}{E_0}, \tag{2}$$

where E_0 is an arbitrary energy reference. The room gain is calculated as the difference between these two energy levels,

$$G_{RG} = L_E - L_{E,ach}. \tag{3}$$

The room gain is conceptually related to Gade’s objective support[3], which is widely used in stage acoustics to compare the energy of early sound reflection patterns from a music instrument to the player’s ears among different rooms for music performance. Gade’s objective support is used to characterize many different kinds of instruments, with different distances from the source to

---

[a] Electronic address: `dpg@elektro.dtu.dk`

the ears of the musicians and different directivity patterns. In the case of voice, the path between mouth and ears is rather well defined.

Brunskog *et al.*'s method for calculating room gain is conceptually and theoretically correct, and can be used to calculate the room gain at positions with very close reflecting surfaces. However, an important limitation of the method is the required measure of an IR in anechoic conditions, which can be an obstacle for many professionals. Additionally, in practice, the IR in anechoic conditions might differ from the direct sound in the measuring conditions due to changes in temperature, humidity, background noise, and distortion artifacts when measuring. The practical limitations lead to measurement error, which is illustrated in the following example.

Nine IRs in a small room, corresponding to the acoustic path between the mouth and the left ear of a Head and Torso Simulator (HATS) B&K type 4128 with left ear simulator B&K type 4159, were measured with the 01dB Symphonie system. The measurements corresponded to three repetitions at three different reproduction gains, keeping the HATS position fixed. The signal-to-noise ratio (SNR), calculated from the peak level to the noise floor level, was at least 60 dB in all impulse responses. The impulse responses were trimmed to the intersection of the exponential decay curve with the noise floor of the measurement with the lowest SNR (the intersection time was noted as t_{min}). The impulse responses were normalized to a peak amplitude of 1, and the energy levels L_E in the interval (0–t_{min}) were calculated. The estimated standard deviation of L_E was 0.02 dB, whereas the maximum difference between two measurements of L_E was 0.06 dB. This error is not usually regarded as important, but as defined in Eq. (3), the room gain can be significantly biased by such an amount, since typical values lie between 0 dB and 0.6 dB.

It would be beneficial to derive the room gain from a single impulse response measurement and increase the sensitivity of the method. For this, the author proposes the measurement of the impulse response using a HATS with a mouth simulator according to recommendation ITU-T P.58[4] and ear simulator with ear canal, according to recommendation ITU-T P.57[5] Type 3. The source should be at least 1 m away from all boundaries, including the floor, using a stand to appropriately place the HATS at the height of the head of an average standing person. The distance gap of 1 m allows for a time gap free of reflections of approximately 5.8 ms. The direct sound $h_d(t)$ is obtained by applying a window $w(t)$ to the measured impulse response $h(t)$ (see Figure 1),

$$h_d(t) = h(t) \times w(t), \tag{4}$$

where $w(t)$ is

$$w(t) = \begin{cases} 1 & t < 4.5 \text{ ms} \\ 0.5 + 0.5\cos\left(2\pi(t-t_0)/T\right) & 4.5 \text{ ms} < t < 5.5 \text{ ms} \\ 0 & t > 5.5 \text{ ms} \end{cases} \tag{5}$$

with $t_0 = 4.5$ ms and $T = 2$ ms. The reflected sound $h_r(t)$ is the complementary signal

$$h_r(t) = h(t) \times (1 - w(t)) = h(t) - h_d(t) \tag{6}$$

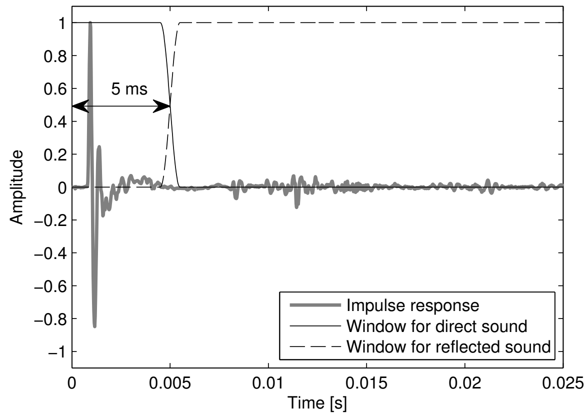

FIG. 1. Example of an impulse response $h(t)$ and the windowing applied to extract the direct and reflected sound.

The energy levels corresponding to the direct sound ($L_{E,d}$) and the reflected sound ($L_{E,r}$) are calculated as

$$L_{E,d} = 10\log\frac{\int_0^\infty h_d^2(t)\,\mathrm{d}t}{E_0}, \tag{7}$$

$$L_{E,r} = 10\log\frac{\int_0^\infty h_r^2(t)\,\mathrm{d}t}{E_0}. \tag{8}$$

The voice support ST_V, in analogy to Gade's objective support, is defined as the difference between the reflected sound and the direct sound from the HATS' mouth to ears IR,

$$ST_V = L_{E,r} - L_{E,d}, \tag{9}$$

which is related to the room gain through the formula

$$G_{RG} \approx 10\log\left(10^{ST_V/10} + 1\right). \tag{10}$$

This formula is obtained under the assumption that the total energy is approximately the sum of the energies corresponding to the direct and the reflected sound after windowing,

$$L_E \approx 10\log\left(10^{L_{E,d}/10} + 10^{L_{E,r}/10}\right). \tag{11}$$

Gade's objective support is intended for big rooms, so the early reflections are counted from 20 ms, and the first 10 ms in the impulse response are regarded as direct sound. This parameter cannot be used in small rooms (e.g. rooms for speech), as the early reflections are much closer to the direct sound than in large halls, and may fall in the direct sound interval or in the interval from 10 ms to 20 ms, which is ignored by the definition. With the present definition of direct and reflected paths, it is possible to calculate room gain and voice support in many rooms. The only limitation is that all boundaries of the room should be 1 m away from the measurement equipment.

The indirect calculation of room gain after measuring the voice support with Eq. (10) reduces the deviation in the results. Using the same impulse responses of the previous example, the standard deviation in the measured room gain was reduced from 0.02 dB to 0.004 dB, and the maximum differences between two measurements did not exceed 0.01 dB.

II. ABOUT THE MEASURED PARAMETERS

Table I shows the six rooms used in the study of Brunskog *et al.* with their volume and the original measurements of reverberation time T_{30} and room gain, noted as G'_{RG}. Inspecting the original G'_{RG} data, the value of 1.12 dB measured in the IEC listening room appears too high in comparison to that measured in the meeting room (0.58 dB), which is smaller and more reverberant than the IEC room. These values imply that the reflected sound pressure level in the IEC room would be about 3 dB higher than in the meeting room, using Eq. (10).

The room impulse responses in the six rooms of the study were measured again, following the procedure described in the previous section. No filtering, other than the intrinsic response of the loudspeaker, was applied to the signals for deriving the objective parameters. The values of voice support ST_V and room gain G_{RG}, measured for each room as the average of six repetitions, are shown in Table I. The differences between old and new room gain values are indicated as ΔG_{RG}.

The new measurements confirm the initial suspicions. The room gain in the IEC listening room is indeed lower than in the meeting room. The room gain in the anechoic chamber was 0 dB in the original study by definition, and it is 0.01 dB by the present method described here. In general, the room gain values are lower than in the original study ($\Delta G_{RG} > 0$ in all cases), a fact that has been already reported.[6] None of the room gain values was higher than 0.5 dB. The voice support has a greater dynamic range and might be more suitable for use in architectural acoustics. However, in anechoic rooms, $ST_V \to -\infty$, and the finite values measured under these conditions must be treated carefully.

III. REVISED EMPIRICAL MODELS

The new room gain values differ considerably from the original values. In order to enable reliable comparison with future studies, the empirical model relating voice power level from the study of Brunskog *et al.* to the room gain has to be recomputed. The relative voice power level (ΔL_W) is defined as the difference between the overall L_W in a certain room and the overall L_W measured in the anechoic room. A simplified linear model of only one explanatory variable is preferred,

$$\Delta L_W[\text{dB}] = 0.5 - 13.5 \times G_{RG}, \tag{12}$$

The model predicts a decrease in the expected voice power level with increasing room gain ($R^2 = 0.83,\ p = 0.01$). Alternatively, rooms with low room gain demand higher vocal intensity from talkers. The measured values, and the regression model (12), are shown in Figure 2. A two-variable model, similar to the one proposed in Brunskog *et al.*, which describes the relative voice power level as a function of the room gain and the logarithm of the volume, is not significant at the 5% level ($R^2 = 0.83,\ p = 0.07$) and shows marginal or no influence of the logarithm of the volume on the voice levels.

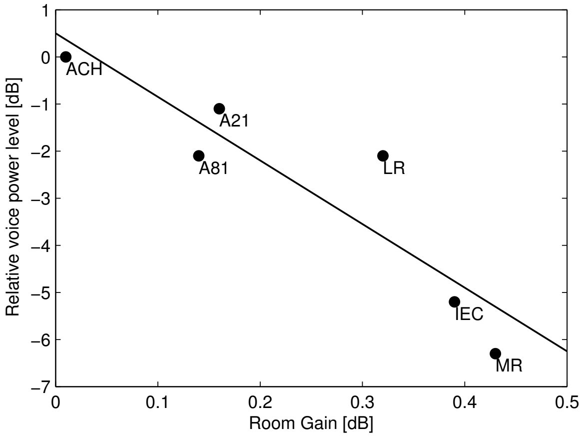

FIG. 2. Relative L_W produced by talkers in the study by Brunskog *et al.* as a function of the room gain. The reference L_W is the average overall L_W measured by Brunskog *et al.* in the anechoic chamber.

Figure 3 shows the relative values of voice power level measured by Brunskog *et al.* versus the voice support. The critical dependence of ST_V value on the measurement SNR in the anechoic chamber suggests that voice level does not change much for very negative values of ST_V, also shown with the transformed regression model using the room gain (dotted curve in Figure 3). A linear dependence of ΔL_W and ST_V for all the conditions studied is not a good approximation. This approximation does not exclude the possibility of modeling a linear dependence between L_W and ST_V in a limited range of ST_V, as has been done in recent studies,[7,8] while approaching an asymptotic L_W value for very negative ST_V (dashed line in Figure 3). Excluding the measurement in the anechoic chamber, the best linear model (solid line in Figure 3) is

$$\Delta L_W[\text{dB}] = -13 - 0.78 \times ST_V. \tag{13}$$

The accuracy of the predictions decreases with this parameter ($R^2 = 0.66,\ p = 0.09$). It would not be wise to conclude that the voice support is less valid than the room gain to describe the changes in voice level due to the acoustic conditions perceived by the talker. More conditions are needed to assess the robustness of room gain and voice support as explanatory variables of voice level variations due to changes in the auditory perception of one's own voice elicited by the room.

TABLE I. Rooms in the study by Brunskog *et al.*, and measured objective parameters. The volume V, reverberation time T_{30}, and room gain G'_{RG} are taken from Brunskog *et al.*'s paper. The room gain G_{RG} and voice support ST_V correspond to new measurements. The differences between old and new room gain values are indicated as ΔG_{RG}.

Name	Abbrev.	V [m^3]	T_{30} [s]	G'_{RG} [dB]	G_{RG} [dB]	ΔG_{RG} [dB]	ST_V [dB]
Auditorium 81	A81	1900	1.06	0.28	0.14	0.14	-14.9
Auditorium 21	A21	1220	1.53	0.29	0.16	0.13	-14.2
Lecture r. 019	LR	190	0.46	0.42	0.32	0.10	-11.1
Meeting r. 112	MR	94	0.42	0.58	0.43	0.15	-9.8
Large anechoic ch.	ACH	1000	0.06	0	0.01	0.01	-27.3
IEC listening room ch.	IEC	100	0.34	1.12	0.39	0.73	-10.3

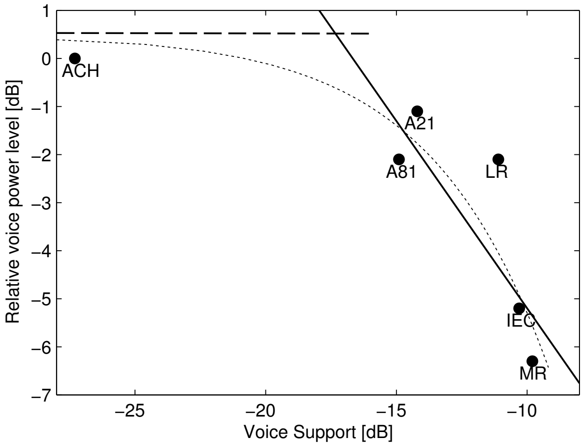

FIG. 3. Relative L_W produced by talkers in the study by Brunskog *et al.* as a function of the voice support. Solid line: regression model excluding the measurements in the anechoic chamber. Dashed line: expected asymptotic relative L_W value. Dotted line: regression model for room gain. The reference L_W is the average overall L_W measured by Brunskog *et al.* in the anechoic chamber.

REFERENCES

[1] J. Brunskog, A. Gade, G. Payá-Ballester, and L. Reig-Calbo, "Increase in voice level and speaker comfort in lecture rooms.", J. Acoust. Soc. Am. **125**, 2072–82 (2009).

[2] V. Lyberg-Åhlander, R. Rydell, and A. Löfqvist, "Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff", J. Voice (2010), available online 25 March 2010.

[3] A. Gade, "Investigations of musicians room acoustic conditions in concert halls. Part I: Methods and laboratory experiments", Acustica **69**, 193–203 (1989).

[4] International Telecommunication Union, "ITU-T P.58. Head and torso simulator for telephonometry", Recommendation, Geneva (1996).

[5] International Telecommunication Union, "ITU-T P.57. Artificial ears", Recommendation, Geneva (2009).

[6] K. Ueno, K. Kato, and K. Kawai, "Effect of Room Acoustics on Musicians' Performance. Part I: Experimental Investigation with a Conceptual Model", Acta Acustica united with Acustica **96**, 505–515 (2010).

[7] D. Pelegrin-Garcia and J. Brunskog, "Prediction of vocal effort and speakers' comfort in lecture rooms", in *Proceedings of Inter-Noise 2009* (Ottawa, Canada) (2010).

[8] D. Pelegrin-Garcia and J. Brunskog, "Natural variations of vocal effort and comfort in simulated environments", in *Proceedings of EAA Euroregio congress on Sound and Vibration 2010* (Ljubljana, Slovenia) (2010).

Vocal effort with changing talker-to-listener distance in different acoustic environments

David Pelegrín-García,[a] Bertrand Smits, Jonas Brunskog, and Cheol-Ho Jeong
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark

(Dated: January 17, 2011)

Talkers adjust their vocal effort to communicate at different distances, aiming to compensate for the sound propagation losses. The present paper studies the influence of four acoustically different rooms on the speech produced by thirteen male talkers addressing a listener at four distances. Talkers raised their vocal intensity by between 1.3 and 2.2 dB per double distance to the listener and lowered it as a linear function of the quantity "room gain" at a rate of -3.6 dB/dB. There were also significant variations in the mean fundamental frequency, both across distance (3.8 Hz per double distance) and among environments (4.3 Hz), and in the long-term standard deviation of the fundamental frequency among rooms (4 Hz). In the most uncomfortable rooms to speak in, talkers prolonged the voiced segments of the speech they produced, either as a side-effect of increased vocal intensity or in order to compensate for a decrease in speech intelligibility.

PACS numbers: 43.55.Hy, 43.70.Mn

I. INTRODUCTION

In face-to-face communication, a talker makes a decision about the desired vocal output based on the given communication scenario. Some factors affecting this decision are the intention of the talker (dialog, discipline, rebuke…), the distance between talker and listener, and special requirements of the listener, due to hearing impairment or language disorders. Once the decision is made, the talker starts to speak and uses a series of feedback mechanisms (auditory, tactile, proprioceptive, and internal) to grant that the actual vocal output matches the desired vocal output.[1]

Speaking in various rooms leads to different experiences or sensations for a talker, due to changes in auditory feedback. The vocal effort required for communicating with a listener at different distances changes with room acoustic conditions, as does also the feeling of vocal comfort. One should differentiate between the concepts of vocal effort and vocal comfort. Vocal effort, according to Traunmüller and Eriksson,[2] is a physiological magnitude different from vocal intensity, that accounts for the changes in voice production required for the communication at different distances. This definition of vocal effort can be extended to also include the changes in voice production induced by noise or the physical environment. These changes include vocal intensity, fundamental frequency (F0), vowel duration, and the spectral distribution of speech. Vocal comfort, according to Titze,[3] is a psychological magnitude determined by those aspects that reduce the vocal effort. Vocal comfort reflects the self-perception of the vocal effort by the feedback mechanisms listed above.

The maximization of vocal comfort should be a priority in situations of very high vocal demands, which are hazardous for the vocal health, such as teaching environments. A recent study revealed that around 13% of teachers suffer from voice problems.[4] Indeed, the prevalence of voice problems among teachers is much higher than it should, compared to their representation in overall population.[5–7] Vilkman[8] points out "bad classroom acoustics" as one of the hazards for voice health from the testimonies of teachers who had suffered from voice disorders. These disorders are related, in many cases, to the intensive use of the voice as an occupational tool.

To characterize the amount of voice use, and estimate the risk of suffering from voice problems, Titze *et al.*[9] introduced a set of measures of the accumulated exposure of vocal fold vibration, called vocal doses. The vocal doses are calculated from the phonation time, F0, and the vocal fold vibration amplitude. In the present work, the variations of vocal intensity (as a rough estimate of the vocal fold vibration amplitude), F0, and phonation time are reported without going further into a detailed risk analysis, leaving this task to future studies and more advanced analytical models. As in Rantala *et al.*[10] both the mean and the standard deviation of F0 are measured as indicators of vocal effort.

Although bad classroom acoustics might be hazardous for voice health, only a few works have attempted to relate classroom acoustics to voice production. Hodgson[11] suggested a simple empirical prediction model to calculate average voice levels used by teachers in university lecture rooms, depending on individual factors, acoustical characteristics of the room and student activity noise. Brunskog *et al.*[12] found that the average vocal intensity used by teachers in different classrooms is closely related to the amplification of the room on the talker's perceived own voice (defined as "room gain"). From this study, it appears that teachers speak louder in rooms with a low room gain and softer in rooms with a high room gain, at a rate of -13.5 dB/dB (dB of voice level per dB of room

---

[a] Electronic address: `dpg@elektro.dtu.dk`

gain).[13] However, none of the two previous studies took into account the distance between teachers and students, which could explain by itself some of the changes in voice level. From a different perspective, Kob *et al.*[14] found that teachers with voice disorders were more affected by unfavorable classroom acoustics than their healthy colleagues.

In a more general communication context, several investigations have analyzed the vocal intensity used by a talker to address a listener located at different distances. One general finding is that the vocal intensity is approximately proportional to the logarithm of the distance. The slope of this relationship is in this paper referred to as the compensation rate (in dB/dd), meaning the variation in voice level (in dB) each time that the distance to the listener is doubled (dd). Warren[15] found compensation rates of 6 dB/dd when talkers produced a sustained vocalization (/a/) addressing listeners at different distances, suggesting that talkers had a tacit knowledge of the attenuation of sound with distance. However, a sound attenuation of 6 dB/dd is only found in free-field or very close to the source. Warren did not provide information on the experimental acoustic surroundings. Michael *et al.*[16] showed that the speech material (natural speech or bare vocalizations) influenced the compensation rates and found lower values than Warren: 2.4 dB/dd for vocalizations and 1.3 dB/dd for natural speech. Healey *et al.*[17] obtained compensation rates in a range between 4.5 dB/dd and 5 dB/dd when the task was to read a text aloud to a listener at different distances. Liénard and Di Benedetto[18] found an average compensation rate of 2.6 dB/dd in a distance range from 0.4 m to 6 m using vocalizations. Traunmüller and Eriksson[2] carried out their experiments with distances ranging from 0.3 m to 187.5 m to elicit larger changes in vocal effort, finding a compensation rate of 3.7 dB/dd with spoken sentences. In general, there is a substantial disagreement among the results of different studies.

Each of the previous experiments analyzing voice production with different communication distances was carried out in only one acoustic environment. Michael *et al.*[16] pointed out that unexplained differences among experimental results might be ascribed to the effect of different acoustic environments, because the attenuation of sound pressure level (SPL) with distance depends on the room acoustic conditions. Zahorik and Kelly[19] investigated how talkers varied their vocal intensity to compensate for the attenuation of sound with distance in two acoustically different environments (one indoor and one outdoor), when they were instructed to provide a constant SPL at the listener position. When uttering a sustained /a/, the talkers provided an almost uniform SPL at each of the listener positions, which indicated that talkers had a sophisticated knowledge of physical sound propagation properties. The measured compensation rates laid between 1.8 dB/dd for an indoor environment, and 6.4 dB/dd for an outdoor environment.

In addition, some of the studies investigated further indicators of vocal effort at different communication distances. Liénard and Di Benedetto[18] also found a positive correlation between vocal intensity and F0, and significant spectral changes in vowels. Traunmüller and Eriksson[2] observed that the duration of vocalic segments increased with communication distance, and thus, with vocal effort.

In summary, there have been many studies reporting vocal intensity at different communication distances, as well as other descriptors of vocal effort: F0 and vowel duration. Only one study[19] analyzed the additional effect of the acoustic environment on the vocal intensity, although the instruction—*provide a constant sound pressure level at the listener position*—and the speech material—vocalizations—were not representative of a normal communication scenario. The aim of the present study is to analyze the effect of the acoustical environment on the natural speech produced by talkers at different communication distances in the absence of background noise, reporting the parameters which might be relevant for the vocal comfort and for assessing the risks for vocal health.

II. EXPERIMENTAL METHOD

The speech from thirteen talkers speaking to one listener at four different distances in four different rooms was recorded. The speech signals were processed to calculate measures of vocal intensity, F0, and the relative duration of the phonated segments.

A. Subjects

Thirteen male talkers participated in the experiment as talkers. Two of the talkers were acting as listeners and experimenters at different times. All thirteen subjects had ages between 23 and 40, and had neither hearing impairment, visual impairment, nor vocal disorder. None of the subjects was a native English speaker, but nevertheless all of them used English as the spoken language during the tests.

B. Instruction

Before the start of the tests, the listener/experimenter explained the instructions verbally to each talker at a close distance. The talkers were given a map which contained roughly a dozen of labeled items (e.g. “diamond mine”, “fast flowing river”, and “desert”), starting and ending point marks, and a path connecting these two points. They were instructed to describe the route between the starting point and the finish point, indicating the items along the path (e.g. “go to the west until you find the harbor”), while trying to enable eye-contact with the talker. There were sixteen maps in total, and a different map was used at each condition. The order of the maps was randomized differently for each subject. These maps have been used extensively in previous research to obtain a dialog-based speech corpus.[20] The object of using maps was evoking natural speech from the talkers in a

TABLE I. Physical volume, reverberation time, room gain, speech transmission index (mouth-to-ears), and A-weighted background noise level measured in the 4 environments: anechoic chamber, lecture hall, corridor, and reverberation room.

	V [m^3]	T_{30} [s]	G_{RG} [dB]	STI	$L_{N,Aeq}$ [dB]
Anechoic Room	1000	0.04	0.01	1.00	< 20
Lecture Hall	1174	1.88	0.16	0.93	28.2
Corridor	410	2.34	0.65	0.83	37.7
Rev. Room	500	5.38	0.77	0.67	20.6

very specific context and mode of communication. An alternative method for obtaining natural speech could have been instructing talkers to speak freely. However, there would have been different modes of communication and contexts among subjects, which would have introduced higher variability in the data.

After explaining the task to the talker, the listener stood at different positions and indicated the talker nonverbally when to start talking. The listener gave no feedback to the talker, either verbally or non-verbally, about the voice level perceived at his position.

At the end of the the experiment, the subjects were asked about the experience of talking in the different rooms and they could answer openly.

C. Conditions

For each subject, the experiment was performed in a total of 16 different conditions, resulting from the combination of four distances (1.5, 3, 6, and 12 m) and four different environments: an anechoic chamber, a lecture hall, a long, narrow corridor, and a reverberation room. The environments were chosen so as to represent a wide range of room acoustic conditions, while being large enough to allow distances between talker and listener of up to 12 m. However, not all of these rooms were representative of everyday environments. The order of the rooms was randomized for each subjects, but the distances from talker-to-listener were always chosen from closest to furthest. Talker and listener stood further than 1 m from the walls and faced each other.

The volume V, reverberation time T_{30}, room gain G_{RG}, speech transmission index (STI) between talker's mouth and ears, and A-weighted background noise levels $L_{N,Aeq}$, measured in the rooms are shown in Table I.

1. Reverberation time

The reverberation time T_{30} was measured according to ISO-3382,[21] using a dodecahedron loudspeaker as an omnidirectional sound source and a 1/2" microphone, Brüel & Kjær (B&K) type 4192. The measurements were carried out with DIRAC[22], using an exponential sweep as the excitation signal. The T_{30}, obtained from the impulse response using Schroeder's method[23] and averaging the measurements in the 500 Hz and 1 kHz one-octave bands, is shown in Table I.

2. Room gain

The room gain G_{RG} was measured with the method proposed by Pelegrin-Garcia[13] in the empty rooms, using a Head and Torso Simulator (HATS) B&K type 4128 with left ear simulator B&K type 4159 and right ear simulator B&K type 4158. The software measurement DIRAC was used to generate an exponential sweep as an excitation signal and extract the impulse responses from the received signals on the microphones at the ears of the HATS. The HATS was placed at the talker position, with the mouth at a height of 1.6 m, and more than 1 m away from reflecting surfaces. The G_{RG} values reported for each room correspond to the average of the values at the two ears and three different repetitions and are shown on Table I. No filtering was applied to the impulse response to calculate G_{RG}.

3. Speech transmission index

The STI was derived with the Aurora software suite[24] from the same mouth-to-ears impulse responses used for the G_{RG} measurements, and ignoring the effect of background noise. The values resulting from averaging three repetitions and the two channels (left and right) at each environment are shown on Table I. One should note that the STI parameter was not originally intended to explain the transmission of speech between the mouth and the ears of a talker, as in this case, but to characterize the transmission channel between talker and listener. The STI values presented here are used only as rough indicators of the perceived degradation in one's own voice due to reverberation and ignoring completely the bone-conducted component of one's own voice.

4. Background noise level

The A-weighted, 20-second equivalent background noise levels ($L_{N,Aeq}$) were measured in the empty rooms using a sound level meter, B&K type 2250. The results from averaging the measurements across four positions in each room are shown in Table I. Possible noise sources contributing to the reported levels are ventilation systems, traffic, and the activity in neighboring areas. All the measured background noise levels were below 45 dB(A) so, according to Lazarus,[25] the produced voice levels were not affected by the noise.

5. Speech sound level

The *speech sound level*[26] S is defined as the difference between the sound pressure level L_p produced by a source with human voice radiation characteristics at a certain

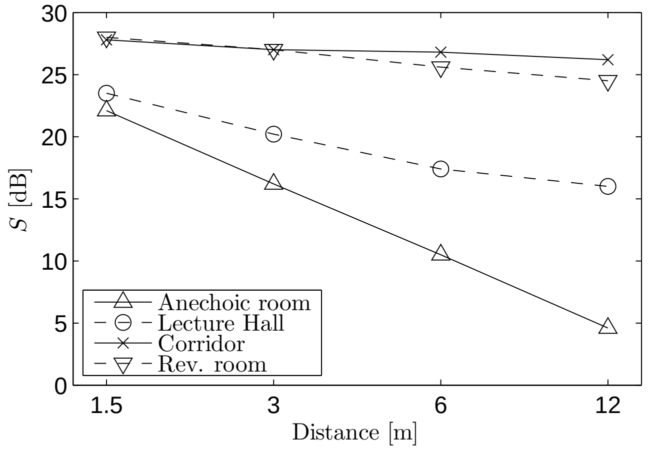

FIG. 1. Speech sound level S as a function of distance.

position and the level L_{ref} produced by the same source at 10 m in free-field, averaged over all directions in space,

$$S = L_p - L_{\mathrm{ref}} \quad . \tag{1}$$

A directive loudspeaker JBL Control One was used as the sound source, and was placed at the talker position, with the edge of the low frequency driver at a height of 165 cm above the floor and pointing toward the listener. The sound pressure level L_p produced by the loudspeaker reproducing pink noise was analyzed in one-octave bands with a sound level meter, B&K type 2250, at the listener position for each of the four distances in each room.

The reference sound pressure level L_{ref} was calculated as the average of 13 measurements in an anechoic chamber with a distance of 10 m between the sound level meter and the loudspeaker. For each measurement, the loudspeaker was turned at steps of 15^o from 0^o to 180^o and reproduced the same pink noise signal with the same gain settings as used for the measurement of L_p.

The resulting S, as a function of distance, averaged across the one-octave mid-frequency bands of 500 Hz and 1 kHz, is presented in Fig. 1.

D. Processing of the voice recordings

The acoustic speech signal was picked up with a DPA 4066 headworn microphone, placed on the talker's cheek at a distance of 6 cm from the lips' edge. The signal was recorded with a Sound Devices 722 digital recorder in 24 bits/44.1 kHz PCM format, and later processed with MATLAB. The length of the recordings varied between one and two minutes, depending on the map and the talker.

1. Voice power level

Vocal intensity is related to the strength of the speech sounds. There are many ways to represent this magnitude, e.g., on-axis SPL at different distances in free-field, sound power level (L_W), or vibration amplitude of the vocal folds. Among these parameters, the sound power level appears to be the most appropriate one to characterize the total sound radiation from a source. Indeed, it is possible to determine the sound power level if the on-axis SPL in free-field conditions and the directivity of the speaker are known. Following the works of Hodgson[11] and Brunskog *et al.*,[12] the sound power level was chosen as the main index of vocal intensity and is also referred to as voice power level.

TABLE II. Increase of SPL (in dB) at the headworn microphone due to sound reflections (used as correction factor), measured with a dummy head. The reference situation is the measurement of SPL in anechoic conditions. Abbreviations are used instead of the complete name of the rooms: LH for the lecture hall, COR for the corridor, and REV for the reverberation room.

	Frequency (Hz)					
Room	125	250	500	1000	2000	4000
LH	0.27	0.05	0.12	0.22	0.07	0.15
COR	0.58	0.32	0.46	0.54	0.59	0.69
REV	0.30	0.18	0.38	0.49	0.43	0.51

To determine the voice power level of the recordings, the equivalent SPL in the one-octave bands between 125 Hz and 4 kHz was first calculated. A correction factor due to the increase of SPL at the headworn microphone in the different rooms was applied (see values in Table II). The correction factor was measured by analyzing the SPL produced by the HATS, reproducing pink noise with a constant sound power level in the different rooms, at the headworn microphone, which was placed on the HATS. The SPL readings from the anechoic chamber were subtracted to the readings in each room. The difference between the corrected SPL at the headworn microphone and the voice power level was determined by performing sound power measurements in a reverberation room in a similar way as described in Brunskog *et al*[12]. However, instead of using a dummy head (as in Brunskog *et al.*), the speech of six different talkers, one by one, was recorded simultaneously using a headworn microphone DPA 4066 and a 1/2" microphone, B&K type 4192, positioned in the far field, where the sound field is assumed to be diffuse. The difference between the mean corrected SPL measured at the headworn microphone and the voice power level as a function of frequency is shown in Fig. 2.

2. Fundamental frequency

F0 was extracted from the recordings with the application Wavesurfer[27] using the Entropic Signal Processing System method at intervals of 10 ms. Taking a sequence with the F0 values of the voiced segments (the only segments for which the algorithm gave an estimation of F0), the mean (noted as $\bar{F_0}$) and the standard deviation (noted as σ_{F_0}) were calculated.

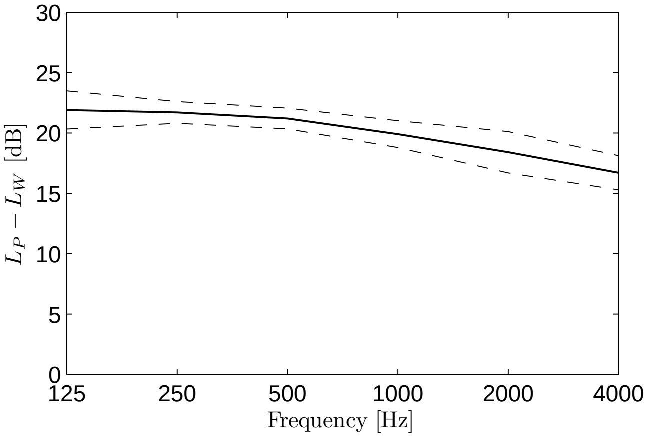

FIG. 2. Difference between the SPL measured at the head-worn microphone, corrected for the increase in SPL due to sound reflections, and L_W. Bold line: mean value. Dashed lines: one standard deviation above and below the mean value.

3. Phonation time ratio

Due to the large variations in the length of speech material among subjects and conditions, the absolute phonation time is not reported, but the ratio of the phonation time t_P to the total duration of running speech t_S in each recording, referred to as *phonation time ratio* (PTR). The calculation procedure is shown in Fig. 3. First, the original speech signal (Fig. 3a) is processed to obtain the running speech signal (Fig. 3b). Then, this signal is split into N non-overlapping frames or segments of a duration $t_F = 10$ ms (Fig. 3c). In the i-th frame, the logical variable k_i ($k_i = 0$ if the segment is unvoiced; $k_i = 1$ if it is voiced) is determined with Wavesurfer. The total duration t_P of phonated segments is $t_F \times \sum_{i=1}^{N} k_i$. Thus,

$$\mathrm{PTR} = \frac{\text{Phonation time}}{\text{Running speech time}} = \frac{t_F \sum\limits_{i=1}^{N} k_i}{t_S}, \qquad N = \left\lfloor \frac{t_S}{t_F} \right\rfloor \tag{2}$$

The floor operator $\lfloor \cdot \rfloor$ results in the closest integer not larger than the operand.

E. Statistical method

For each parameter (L_W, $\bar{F_0}$, σ_{F_0}, and PTR), a linear mixed model[28] was built from a total of 208 observations (13 subjects × 4 distances × 4 rooms), using the `lmer` method in the library `lme4`[29] of the statistical software `R`.[30] The "full model" included the logarithm of the distance as a covariate and the acoustic environment (or room) as a factor, and the interaction between the distance and the room. In the present paper, the mixed model for a response variable y which depends on the i-th subject, the j-th distance d_j, and the k-th room, is

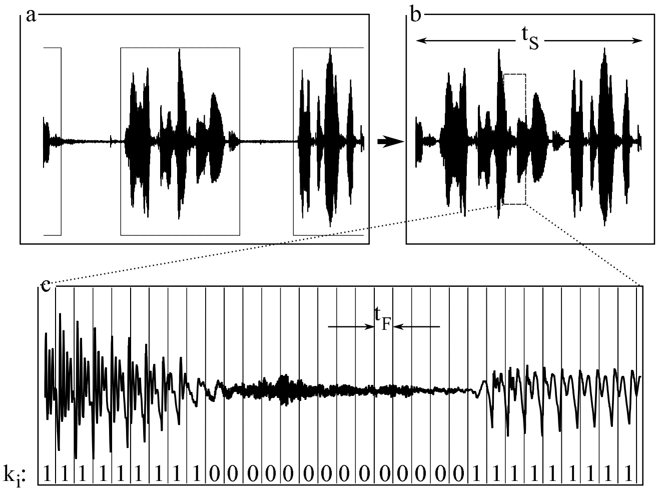

FIG. 3. Post-processing of the recordings and computation of the phonation time ratio. a) Original speech signal. b) Running speech signal of duration t_S, obtained from the original signal by removing 200 ms-long frames with very low energy. c) Calculation of the phonation time by splitting the running speech signal in frames of length $t_F = 10$ ms, determining whether each segment i is phonated ($k_i = 1$) or not ($k_i = 0$) and summing up the time of all phonated segments.

presented in the form

$$y_{ijk} = a_k + \alpha_i + (b_k + \beta_i) \times \log_2(d_j/1.5) + \epsilon_{ijk}. \tag{3}$$

The fixed effects are written on Roman characters (a_k and b_k) and the random effects are written on Greek characters (α_i, β_i, and ϵ_{ijk}). The random effects are stochastic variables normally distributed with zero mean. The distance dependence is contained in the parameters b_k and β_i (fixed slope and random slope, respectively). On the fixed part, the subscript k indicates an interaction between room and distance. If there is no interaction, b_k becomes a constant b. The presence of β_i indicates that the dependence of the response variable y on the distance d is different for each subject. The intercept ($a_k + \alpha_i$) adjusts the overall value of y, and it has a fixed part a_k and a random part α_i. The fixed intercept contains the effect of the room k on the response variable. The random part is also referred to as intersubject variability. The residual or unexplained variation ϵ_{ijk} is also regarded as a random effect. The standard deviations of the random effects α_i, β_i, and ϵ_{ijk} are notated as σ_α, σ_β, and σ_ϵ, respectively.

The actual models were built as simplifications of the "full model". First, the significance of the interaction (room-dependent slope b_k) was tested by means of likelihood ratio tests (using the function `anova` in `R`), comparing the outcomes of the full model and a reduced model without the interaction (constant slope b). If the full model was significantly better than the reduced model, the first one was kept. Otherwise, the reduced model was used. Another test for the suitability of random slopes was made by comparing the full model to another one with fixed slopes by means of a likelihood ratio test. In the same way, if the model with random slopes was significantly better than the one with fixed slopes, the first one was chosen. The suitability of including the basic variables (room and distance) was assessed by comparing the chosen model from the previous tests to a reduced version that only contained one variable (room or distance) with likelihood ratio tests. However, all the parameters showed dependence on the room and the distance. The models did not include a random effect for the room due to the subject.

The p-values for the overall models were calculated by means of likelihood ratio tests comparing the fit of the chosen model to the fit of a reduced model which only contained the random intercept due to the effect of the subject (and no dependence on room or distance). The p-values associated to each predictor and the standard deviations of the random effects were obtained with the function `pvals.fnc(...,withMCMC=T)` of the library `languageR`[31] in `R`, which makes use of the Markov Chain Monte Carlo (MCMC) sampling method.

The choice of mixed models has the following basis: a considerable amount of the variance in the observations is due to the intersubject differences (which could be revealed with an analysis of variance table), so the subject is regarded as a random effect. Conceptually, it is similar to applying a normalization for each subject, or regarding the subject as a factor in traditional statistical modeling.

III. RESULTS AND ANALYSIS

The measurements of L_W, $\bar{F}_0$, σ_{F_0}, and PTR were used to build four different linear mixed models according to (3). The coefficients for the intercepts and slopes corresponding to the fixed-effects of the models, together with the standard deviations of the random effects, are presented in Table III. The statistical significance (p-value) of the fixed effects and interactions included in each model, along with the overall significance levels, is shown in Table IV.

A. Voice power level

The measured L_W, as a function of the distance and for each of the rooms, averaged across all subjects, is shown in Fig. 4. In the same figure, the lines show the fixed-effects part of the empirical model described in (3) and Table III. L_W depends almost linearly on the logarithm of the distance (with slopes between 1.3 dB and 2.2 dB per doubling distance) and changed significantly among rooms (intercepts between 54.8 dB and 56.8 dB). At each distance, the highest L_W was always measured in the anechoic room. A significant interaction was found between the room and the logarithm of the distance, because the variation of L_W with distance in the reverberation room (1.3 dB per doubling distance) was lower than the variation in the other rooms (1.9 to 2.2 dB per doubling distance). The standard deviation of the intersubject variation was estimated to be 2.7 dB, whereas the individual differences in the variation of L_W with distance had a standard deviation of 0.76 dB per doubling distance.

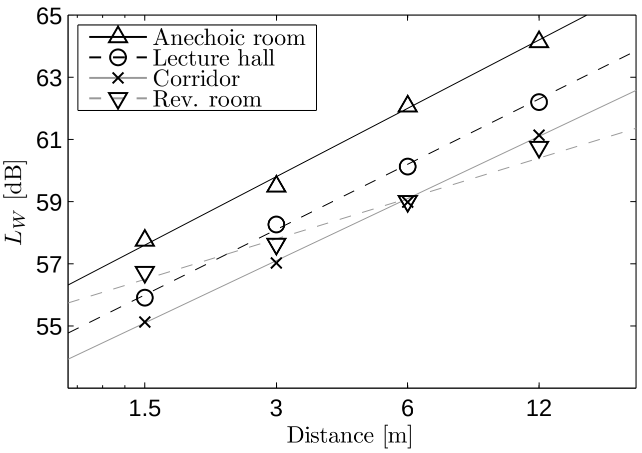

FIG. 4. Average voice power level used by the talkers at different distances to the listener. The lines show the predictions of the empirical model. The different slopes of the lines show an interaction between the room and the distance.

B. Fundamental frequency

Figure 5 shows the subject-averaged measured $\bar{F}_0$ (data points) and the corresponding empirical model (lines) described in (3) and Table III, for the different distances and rooms. $\bar{F}_0$ changed significantly among rooms (intercepts between 119.3 Hz and 123.6 Hz) and had an almost linear dependence on the logarithm of the distance, with a slope of 3.8 Hz per doubling distance, identical for all the rooms. However, by visual inspection of Fig. 5, in the anechoic and reverberant rooms, there was less variation between the distances of 1.5 m and 3 m than at further distances. $\bar{F}_0$ in the anechoic room was about 4 Hz higher than in the other rooms for all distances. The standard deviation of the intersubject variation was estimated in 16.3 Hz, whereas the individual differences in the variation of $\bar{F}_0$ with distance had a standard deviation of 2.95 Hz per doubling distance.

The measured σ_{F_0}, as a function of the distance and for each of the rooms, averaged across all subjects, is shown in Fig. 6. The lines in the figure show the fixed-effects part of the empirical model described in (3) and Table III. σ_{F_0} changed significantly among rooms (intercepts between 19.2 Hz and 23.2 Hz) and had a weak linear dependence on the logarithm of the distance, with a slope of 0.63 Hz per doubling distance, equal among the rooms. The standard deviation of the intersubject variation was estimated in 5.22 Hz, whereas the individual differences in the variation of σ_{F_0} with distance had a standard deviation of 1.29 Hz per doubling distance. The latter value is larger than the fixed-effect slope (0.63 Hz) which means that, for a number of subjects, σ_{F_0} decreased with distance. This is the reason for the low statistical significance of the σ_{F_0} dependence with the logarithm of the distance shown on Table IV. Therefore,

TABLE III. Fixed and random effects included in the mixed models. The fixed effects are characterized for the intercepts a and slopes b, whereas the random effects have zero mean and only their standard deviation is shown. Abbreviations are used instead of the complete name of the rooms:ACH for the anechoic room, LH for the lecture hall, COR for the corridor, and REV for the reverberation room. Note that the b values for $\bar{F}_0$, σ_{F_0}, and PTR are independent of the room.

	Fixed effects								Random effects		
	a_k (Intercept)				b_k (Slope)				Intercept	Slope	Residual
Parameter	ACH	LH	COR	REV	ACH	LH	COR	REV	σ_α	σ_β	σ_ϵ
L_W [dB]	56.8	56.0	54.8	56.2	2.2	2.0	1.9	1.3	2.74	0.76	1.33
$\bar{F}_0$ [Hz]	123.6	120.1	119.8	119.3	3.8				16.3	2.95	3.6
σ_{F_0} [Hz]	23.2	22.0	20.6	19.2	0.63				5.22	1.29	2.77
PTR	0.65	0.55	0.56	0.67	0.026				0.059	-	0.062

TABLE IV. Statistical significance and p-values of the fixed effects and interactions considered in the empirical models and overall significance of the models. N.S.: Non-significant.

	Main effects		Interaction	
	log(Distance)	Room	Room × log(Distance)	Overall
L_W	< 0.001	< 0.001	0.009	< 0.001
$\bar{F}_0$	< 0.001	< 0.001	N.S.	< 0.001
σ_{F_0}	0.10	< 0.001	N.S.	< 0.001
PTR	< 0.001	< 0.001	N.S.	< 0.001

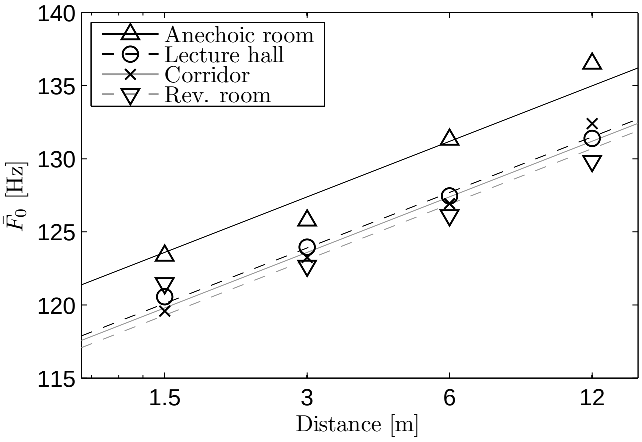

FIG. 5. Average mean fundamental frequency used by talkers at different distances to the listener. The lines show the predictions of the empirical model.

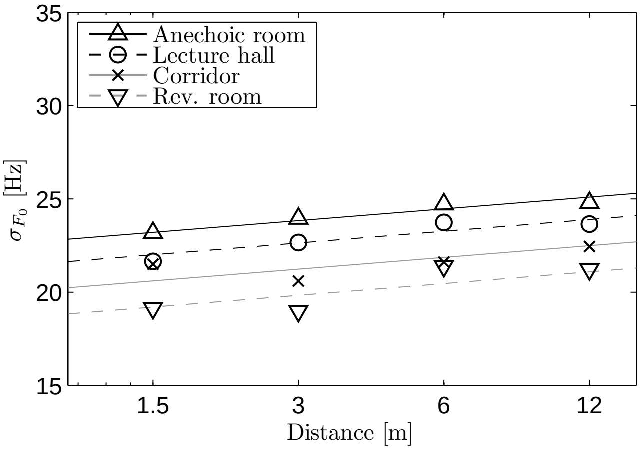

FIG. 6. Average long-term standard deviation of the fundamental frequency used by talkers at different distances to the listener. The lines show the predictions of the empirical model.

the amount of σ_{F_0} change as a function of distance was mainly an individual factor.

C. Phonation time ratio

The measured PTR, as a function of the distance and for each of the rooms, averaged across all subjects, is shown in Fig. 7. In the same figure, the lines show the fixed-effects part of the empirical model described in (3) and Table III. PTR had a weak linear dependence on the logarithm of the distance (with a slope of 0.026 per doubling distance, equal for all rooms) and changed significantly among rooms, especially between two groups: one formed by the anechoic room and the reverberation room (intercepts 0.65 and 0.67) and a second group formed by the lecture hall and the corridor (intercepts 0.55 and 0.56). The standard deviation of the intersubject variation was estimated in 0.059. The change in PTR with distance was not significantly different among subjects, so the model does not include a random slope.

D. Subjective impressions

The talkers expressed their opinions verbally about the experience of talking in the different rooms. One general comment was that the anechoic chamber was an unnatural place to speak in, due to the lack of sound reflections, and that they felt moved to raise their vocal intensity to make themselves heard at the listener location, and for this reason, it was not a comfortable environment for talking. The reverberation room was very unpleasant for speaking, due to the excessive reverberation. Talkers admitted that they had to modify their speech strat-

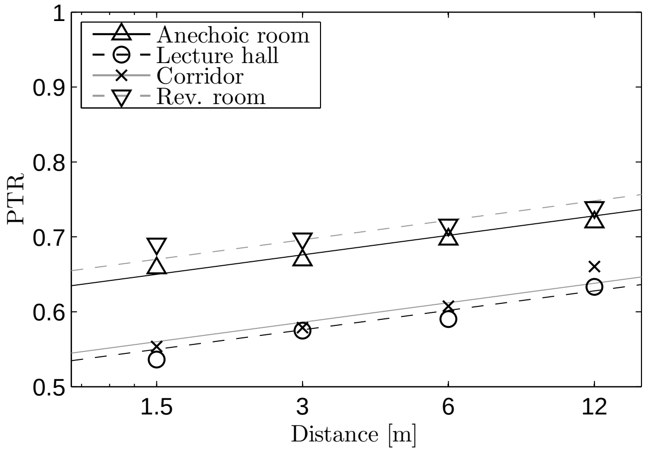

FIG. 7. Average phonation time ratio (relative appearance of voiced segments in running speech) used by talkers at different distances to the listener. The lines show the predictions of the empirical model.

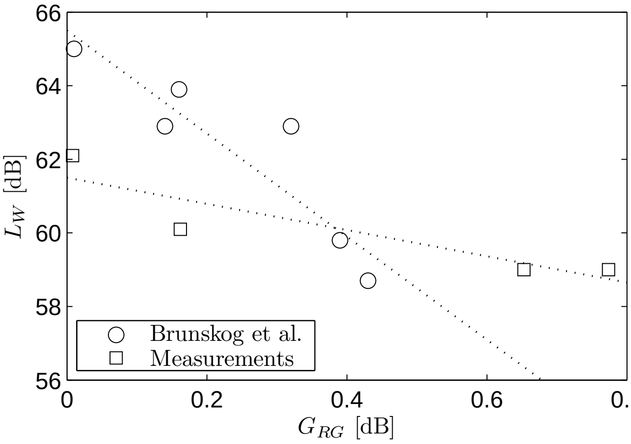

FIG. 8. Average L_W at 6 m versus room gain G_{RG}, as compared to the results of Brunskog *et al.*

egy to compensate for the poor acoustic conditions. A few of the subjects preferred overall the corridor, due to the sensation of support or being helped by the room to reach longer distances without having to increase their voice level too much, although they pointed out some acoustical deficiencies like a noticeable echo. Most of the subjects preferred the lecture hall for speaking. However, they admitted that it was demanding to talk at the longest distance (12 m). Many subjects commented that the acoustic conditions of the experimental rooms were not the desirable ones in rooms for speech.

IV. DISCUSSION

Figures 4 to 7 show the variation of the measured parameters (L_W, $\bar{F}_0$, σ_{F_0}, and PTR) with distance and across environments. As all of the measured parameters indeed have variation with distance and acoustic environment, they are potential indicators of vocal effort.

The measurements shown in Fig. 4 reveal that the average variations of L_W when the distance increases from 1.5 m to 12 m are in the range between 3.9 dB in the reverberation room and 6.6 dB in the anechoic room. These variations are mainly the consequence of a conscious decision of the talker to raise the voice level as a response to a change in communication distance. However, the fact that the compensation rates differ among rooms shows the influence of auditory feedback in voice level adjustment. Furthermore, the effect of room on L_W varies between 2 dB at 1.5 m and 3.3 dB at 12 m. These values are smaller but comparable to the effect of distance on L_W. Thus, the perception of one's own voice via reflections in the room boundaries is important for voice level regulation, together with the direct air transmission and the bone conducted components, as Siegel and Pick[32] stated.

Brunskog *et al.* used G_{RG} as a metric to quantify the importance of the reflected sound from one's own voice. This measure is indeed a measure of sidetone (one's own voice reaching the ears) amplification. Taking the subject-averaged L_W values measured at 6 m, a distance which is representative of a lecturing scenario, the least squares regression model using G_{RG} as a predictor is

$$L_{W,6} = 61.5 - 3.56 \times G_{RG}. \tag{4}$$

The R^2 for this regression model is 0.82, whereas the p-value is 0.09. The L_W values, with the regression line (4), are compared to the results of Brunskog *et al.*[12,13] in Fig. 8. The slope of the regression line in the current measurements is much lower than the slope obtained by Brunskog *et al.* (-3.6 dB/dB vs. -13.5 dB/dB). The difference between slopes might be explained by the fact that the distance was not taken into account by Brunskog et al. In their study, the rooms with high G_{RG} values were small rooms where the listeners stood close to the talker whereas the rooms with low room gain were larger and the listeners stood far from the talker. Thus, there is an unwanted correlation between the room gain and the distance, due to the experimental design, but which is found in typical real rooms. The model from Brunskog *et al.* predicts L_W in a general situation with varying distance to the listeners, but the model (4) accounts for the variation due exclusively to changes in auditory feedback.

As in some studies of sidetone amplification,[33] L_W decreases with increasing sidetone amplification (estimated by G_{RG}). However, there are two differences between these studies and the present study. One is the range of L_W variation and the second is the magnitude of the effect. In the present study, talkers raised L_W by 3.2 dB on average while speaking in the anechoic room at a distance of 12 m, compared to the reverberant room. In other studies of voice production with altered sidetone, variations in voice level of up to 20 dB were reported. In these studies, the sidetone was altered by inducing temporary hearing loss on the subjects, thus decreasing all components of sidetone (direct, reflected and bone conducted sound), or attenuating the airborne sound while bone conduction is preserved. The significantly different ranges of voice level variation obtained in the previous studies (up to 20 dB) and in this study (approximately 3.2 dB by the effect of room) might be due to the fact that only the reflected component was changed in this study, while the direct and bone conducted components of the talker's own voice were kept unchanged. Therefore, the overall sidetone variations were much smaller than in the other studies. The magnitude of the effect on traditional sidetone compensation was in the range between -0.25 dB/dB and -0.57 dB/dB, whereas in the present study the magnitude of the effect was -3.6 dB/dB, as can be seen in (4). These differences could be explained by two alternative hypotheses. The first is that the changes in L_W are purely due to the Lombard effect, and that the room reflections alter the loudness of one's own voice to a greater extent than indicated by the single figure G_{RG}. The second is that there are additional psychological attributes related to room perception affecting the voice regulation at a cognitive level, through internal feedback mechanisms.

The measured compensation rates for L_W due to changes in distance between talker and listener were between 1.3 dB/dd in the reverberation room and 2.2 dB/dd in the anechoic chamber. These compensation rates are much lower than the ones obtained by Warren,[15] Healey *et al.*,[17] and Traunmüller and Eriksson[2]. However, they are closer to other studies[16,18], and especially close to the 1.8 dB/dd measured indoor by Zahorik and Kelly.[19] Differences from the previous studies might arise from the selection of subjects or different instruction. In the present study, there were significant differences in vocal behavior among subjects, indicated by the random slope effect in Table III, which predicts a standard deviation of 0.76 dB/dd over the fixed slopes 1.3 to 2.2 dB/dd. In any case, the individual compensation rates were not as large as 6 dB/dd.[15,19] In addition, natural speech was evoked in the present experiment by means of the map task, which resulted in lower compensation rates than would be obtained by using short vocalizations, as Michael *et al.*[16] stated.

Figure 9 shows the relationship between the L_W produced by the talkers and the sound speech level S at the listener position, which is an alternative representation of the data in Fig. 4. The dashed line in Fig. 9 represents the theoretical L_W values that would keep the SPL constant at the listener position. According to Zahorik and Kelly,[19] if talkers accurately compensated for the sound propagation losses—providing an almost constant average SPL at the listener position—the expected L_W would lay exactly on top of a line with the same slope as the dashed line, meaning that a talker would lower L_W by 1 dB whenever S increases by 1 dB. The L_W data points in Fig. 9 follow approximately straight lines with different slopes for each room: -0.4 dB/dB in the anechoic chamber, -0.8 dB/dB in the lecture hall, -1.1 dB in the reverberation room, and -3.8 dB/dB in the corridor. In the lecture hall and the reverberation room, talkers approximately compensated for sound propagation losses.

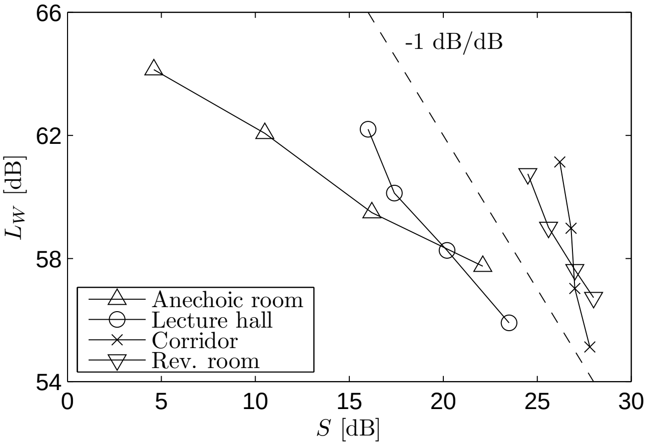

FIG. 9. Voice power level versus speech sound level S at the listener's position. The dashed line has a slope of -1 dB/dB. If the L_W values laid in a line with the same slope, talkers would be providing a constant SPL at the listener position.

However, there was an undercompensation in the anechoic chamber, meaning that the SPL produced at the listener position decreased with distance, and an overcompensation in the corridor, where the SPL increased with the distance. Undercompensation appears to take place in rooms with big differences of S between short and long distances, i.e. rooms with dominating direct sound. Overcompensation takes place in rooms where differences in S at short and long distances were small, i.e. rooms with strong reverberant field. Undercompensation and overcompensation were present because the talkers were not explicitly asked to compensate for sound propagation losses, and many of the talkers were not used to talk in the environments of the study. It is presumed that talkers would be able to compensate for sound propagation losses with an explicit instruction and training to get acquainted with the acoustical properties of each room.

Compensation rates have a meaning when the distance between talker and listener is well defined, such as in a face-to-face conversation. In the case of a distributed audience, as in the usual teaching context, the situation is more complex and it is not clear what is the distance estimation of the talker. In that case, according to Brunskog *et al.*[12,13], talkers apparently adjust their voice levels guided by the room gain or degree of amplification provided by the room at their ears (Fig. 8).

The changes in $\bar{F}_0$ were similar to those in L_W, as both parameters increased linearly with the logarithm of the distance, and it was in the anechoic room where the highest $\bar{F}_0$ were obtained at each distance. Table III shows that $\bar{F}_0$ changed 3.8 Hz by doubling the distance and was 4 Hz higher in the anechoic room than in the other rooms. In simplified terms, the extra vocal effort demanded to speak in the anechoic room is comparable to the effect of doubling the distance to the listener in other rooms. However, the changes among other rooms (maximum of 0.8 Hz) were not as important so as to ascribe a significant effect to the room. It seems more likely that the unfamiliarity of talkers with the anechoic room accentuated some changes in speech production too much, which are not observed in everyday rooms. Nevertheless, $\bar{F}_0$ is an important measure of vocal effort to show that, at long communication distances, the number of vocal fold vibrations (or collisions) increases, which leads to higher vocal doses that might eventually result in vocal fold trauma.

The talkers had the general remark that the anechoic room and the reverberation room were the most uncomfortable environments to speak in. Both environment were the two most extreme rooms in terms of T_{30}, STI, and G_{RG}, as shown in Table I. The anechoic chamber demanded an increased vocal effort due to lacking support, with a G_{RG} value of 0.01 dB. On the other hand, it was very unpleasant and stressing to speak in the reverberation room, which could be explained by the remarkably lower STI value (only 0.67) corresponding to the transmission between mouth and ears. Talkers' comments suggest that there is a compromise between STI and G_{RG}, in order for rooms to be comfortable. The poor vocal comfort rating for the reverberation room cannot be explained by the measured L_W or $\bar{F}_0$, as the L_W and $\bar{F}_0$ in this room were not higher than the values measured in the lecture hall and the corridor, the most preferred rooms. This observation supports the idea that the concepts of vocal effort and comfort are not exactly opposite.

As shown in Fig.6 and Table III, the model predicted significant differences in σ_{F_0} among the environments for all distances. The highest σ_{F_0} was found in the anechoic room, followed by those in the lecture hall, the corridor, and the reverberation room, in reverse order to the reverberation times: the reverberation room, the corridor, the lecture hall, and the anechoic chamber (in decreasing order), or in the same order as the STI. According to this observation, speech produced in acoustically live rooms is more monotonous (meaning low variability in F0) than in acoustically dry rooms. The extreme values of σ_{F_0} were obtained in the least preferred rooms. The highest σ_{F_0} in the anechoic room might be an indication of increased vocal demands (increased L_W and $\bar{F}_0$), whereas the low σ_{F_0} in the reverberant room might be an observable feature of the speech produced under low STI conditions. However, this assertion needs to be proved in a broader range of acoustic conditions.

In Fig. 7, the average PTR was remarkably different between two groups of environments, and correlated well with the subjective impressions of talkers regarding vocal comfort. The highest PTR values were measured in the most uncomfortable rooms (0.67 in the reverberation room and 0.65 in the anechoic room), whereas the PTR in the other two rooms was significantly lower (0.55 in the lecture hall and 0.56 in the corridor). The increased voice levels or vocal efforts explain the high values obtained for the anechoic chamber, as Liénard and Di Benedetto[18] also reported. However, the high PTR obtained in the reverberation room might be due to the adaptation of the talker to the environment. It seems that talkers tried to improve the speech intelligibility in such a reverberant environment by separating the consonant segments of their speech, resulting in longer vocalic segments.

V. CONCLUSIONS

The present paper studies the changes in different speech parameters (voice power level, fundamental frequency, phonation time ratio) describing vocal effort when talkers addressed a single listener at different distances under various room acoustic conditions in the absence of background noise. The main conclusions are:

- The decision of using a certain voice level depends on the visually perceived distance to the listener and varies between 1.3 and 2.2 dB per double distance to the listener.
- The room acoustic conditions modify the auditory feedback of the talker's own voice, inducing significant changes in voice level with an approximately linear dependence on the amplification of the room to one's own voice, given by the magnitude "room gain", at a rate of -3.6 dB/dB.
- The mean fundamental frequency increases with distance at a rate of 3.8 Hz per double distance to the listener and is 4 Hz higher in anechoic conditions.
- A room that provides vocal comfort requires a compromise between room gain and STI, supporting the voice from a talker but not degrading the perceived speech quality.
- The standard deviation of the fundamental frequency and the relative duration of voiced segments in a running speech signal might be symptomatic indicators of vocal comfort in a room.

Acknowledgments

The authors wish to express their gratitude to all the participants in the measurements, to one anonymous reviewer that enabled significant improvements in this work, and to Finn Jacobsen and Torben Poulsen for their positive feedback toward improving the quality of the paper. This research has been partially funded by the Swedish company AFA Försäkring as a part of the project "Speaker's comfort and voice health in classrooms".

[1] L. Raphael, G. Borden, and K. Harris, *Speech science primer: Physiology, acoustics, and perception of speech*, 167–176, fifth edition (Lippincott Williams & Wilkins, Baltimore and Philadelphia) (2007).

[2] H. Traunmüller and A. Eriksson, "Acoustic effects of variation in vocal effort by men, women, and children", J. Acoust. Soc. Am. **107**, 3438–51 (2000).

[3] I. Titze and D. Martin, *Principles of voice production*, 49–51 (National Center for Voice and Speech, Iowa City) (1998).

[4] V. Lyberg-Åhlander, R. Rydell, and A. Löfqvist, "Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff", J. Voice (2010), available online 25 March 2010.

[5] J. Preciado-López, C. Pérez-Fernández, M. Calzada-Uriondo, and P. Preciado-Ruiz, "Epidemiological study of voice disorders among teaching professionals of La Rioja, Spain.", J. Voice **22**, 489–508 (2008).

[6] N. Roy, R. Merrill, S. Thibeault, R. Parsa, S. Gray, and E. Smith, "Prevalence of voice disorders in teachers and the general population", J. Speech Lang. Hear. Res. **47**, 281–93 (2004).

[7] A. Russell, J. Oates, and K. M. Greenwood, "Prevalence of voice problems in teachers.", J. Voice **12**, 467–79 (1998).

[8] E. Vilkman, "Voice problems at work: A challenge for occupational safety and health arrangement.", Folia Phoniatr. Logo. **52**, 120–25 (2000).

[9] I. R. Titze, J. G. Svec, and P. S. Popolo, "Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues.", J. Speech Lang. Hear. Res. **46**, 919–32 (2003).

[10] L. Rantala, E. Vilkman, and R. Bloigu, "Voice changes during work: Subjective complaints and objective measurements for female primary and secondary school teachers", J. Voice **16**, 344–355 (2002).

[11] M. Hodgson, "Experimental investigation of the acoustical characteristics of university classrooms", J. Acoust. Soc. Am. **106**, 1810–19 (1999).

[12] J. Brunskog, A. Gade, G. Payá-Ballester, and L. Reig-Calbo, "Increase in voice level and speaker comfort in lecture rooms.", J. Acoust. Soc. Am. **125**, 2072–82 (2009).

[13] D. Pelegrin-Garcia, "Comment on "Increase in voice level and speaker comfort in lecture rooms" [J.Acoust.Soc.Am. 125, 2072-2082 (2009)]", J. Acoust. Soc. Am. (2010), submitted for publication.

[14] M. Kob, G. Behler, A. Kamprolf, O. Goldschmidt, and C. Neuschaefer-Rube, "Experimental investigations of the influence of room acoustics on the teachers voice", Acoust. Sci. Tech. **29**, 86–94 (2008).

[15] R. Warren, "Vocal compensation for change in distance", in *Proceedings of the 6th International Congress of Acoustics*, 61–64 (International Commission for Acoustics, Tokyo) (1968).

[16] D. Michael, G. Siegel, and H. Pick Jr, "Effects of distance on vocal intensity", J. Speech Hear. Res. **38**, 1176–83 (1995).

[17] E. C. Healey, R. Jones, and R. Berky, "Effects of perceived listeners on speakers' vocal intensity.", J. Voice **11**, 67–73 (1997).

[18] J. S. Liénard and M. G. Di Benedetto, "Effect of vocal effort on spectral properties of vowels.", J. Acoust. Soc. Am. **106**, 411–22 (1999).

[19] P. Zahorik and J. W. Kelly, "Accurate vocal compensation for sound intensity loss with increasing distance in natural environments.", J. Acoust. Soc. Am. **122**, EL144–50 (2007).

[20] A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert, "The HCRC map task corpus", Lang. Speech **34**, 351–366 (1991).

[21] International Organization for Standardization, "ISO-3382:2009, Acoustics - Measurement of room acoustic parameters – Part 1: Performance spaces", (2009).

[22] Acoustics Engineering, "Measuring impulse responses using Dirac", Technical Report, Acoustics Engineering (2007), Technical Note 001. Available for download at `http://www.acoustics-engineering.com/support/technotes.htm` (date last viewed 10/16/10).

[23] M. Schroeder, "New Method of Measuring Reverberation Time", The Journal of the Acoustical Society of America **37**, 409–12 (1965).

[24] A. Farina, "Aurora plug-ins", appearing on the World Wide Web at URL site `http://www.aurora-plugins.com` (date last viewed 10/12/10).

[25] H. Lazarus, "Prediction of Verbal Communication in Noise-A Review: Part 1", Appl. Acoust. **19**, 439–63 (1986).

[26] M. Barron, *Auditorium acoustics and architectural design*, chapter Acoustics for speech, 223–240 (Taylor & Francis, London) (1993).

[27] K. Sjölander and J. Beskow, "Wavesurfer", Stockholm: Centre for Speech Technology (CTT) at KTH. Available for download at `http://sourceforge.net/projects/wavesurfer/` (date last viewed 10/16/10) (2000).

[28] D. Hedeker, "Generalized linear mixed models", in *Encyclopedia of Statistics in Behavioral Science*, edited by B. Everitt and D. Howell, second edition (John Wiley & Sons, New York) (2005).

[29] D. Bates and M. Maechler, *lme4: Linear mixed-effects models using S4 classes* (2010), URL `http://CRAN.R-project.org/package=lme4`, (date last viewed 10/16/10), R package version 0.999375-33.

[30] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2010), URL `http://www.R-project.org`, (date last viewed 10/16/10). ISBN 3-900051-07-0.

[31] R. H. Baayen, *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics".* (2009), URL `http://CRAN.R-project.org/package=languageR`, (date last viewed 10/16/10), R package version 0.955.

[32] G. Siegel and H. Pick Jr, "Auditory feedback in the regulation of voice", J. Acoust. Soc. Am. **56**, 1618–24 (1974).

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

Abstract

In order to investigate the influence of room acoustic conditions on voice production, a system for the real-time auralization of one's own voice has been designed. This system combines computerized room acoustic models, psychoacoustic processing, short-delay convolution techniques, mixed-order Ambisonics encoding/decoding, and loudspeaker reproduction. Equalization filters are used on an individual basis to adjust the performance of the system to each particular talker, including the ratio between direct and reflected sound. The auditory cues of delay, amplitude, frequency response, and directionality corresponding to each sound reflection are preserved. Thus, this system is suitable for psychoacoustics and cross-modality research, integration in multimodal virtual reality systems or room acoustics enhancement.

Loudspeaker-based system for real-time own-voice auralization

David Pelegrín-García Jonas Brunskog
dpg@elektro.dtu.dk jbr@elektro.dtu.dk

February 28, 2011

1 Introduction

Recently, the field of voice ergonomics, which is defined as *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* [1], has received some attention. One of the topics of research in this field is the study of the interaction between room acoustics, noise, and voice production. This is of special relevance in the case of teachers, for whom the prevalence of voice problems is significantly higher than in the rest of the population [2, 3].

In a recent investigation, Brunskog *et al.* studied the effect of the classroom acoustic conditions on the voice levels of a number of teachers [4]. The same teacher had to move to a number of different rooms, which were sometimes located far away from each other. This methodology may have introduced some bias in the results by making comparisons among rooms or judgments about them slightly difficult. In addition, the acoustic conditions of the rooms were given beforehand, with very limited possibilities of adjusting them.

The present paper describes a tool to accurately recreate the acoustics of different rooms in a controllable way inside a loudspeaker array, located in a highly damped room. This will enable a more careful and flexible design of experimental conditions in research. The recreation of different room acoustic conditions is based on the reconstruction of the sound field of the simulated room using the method proposed by Favrot and Buchholz [5], although introducing slight modifications for the requirements of real-time performance. The reconstruction of the sound field is focused mainly on the voice of a talker at his own ears, so he/she has the experience of being in an acoustic environment different from the actual room. According to Kleiner *et al.* [6], this system aims to "auralize" the talker's voice in real time.

Previous auralization systems with the same aim have been reported in the literature. Kleiner and Bertson [7] used a system with nine loudspeakers that could provide up to 50 early reflections obtained from delay lines. Shearer and Torres [8] used a two-channel, headphone-based auralization system able to convolve in real time the voice of a talker with an impulse response calculated with a room acoustic simulation software [9]. In a more recent work, Cabrera *et al.* [10] used a pair of earspeakers to render a binaurally recorded sound-field, with the possibility of accounting for head rotations by means of head-tracking in the horizontal plane.

Similar systems have been built to investigate the importance of room acoustic conditions for singers. Marshall and Meyer [11] used a system with 7 loudspeakers and 4 microphones that simulated 4 early reflections and late reverberation, with the particularity of allowing the presence of several performers at the same time. Noson *et al.* [12] studied the preference of singers after introducing an additional reflection in realistic environments, with the aid of a microphone, a delay line and a loudspeaker. In more recent works, Yuen *et al.* [13] and Stetson and Braasch [14] used a two-channel convolution system able to recreate binaural sound fields through binaural impulse response measurements in real halls.

Other investigations, not focused on the talker's voice, but on the effect of room acoustics on musical performance and subjective preference of musicians in stage, have used similar setups. Gade [15] used a system with five loudspeakers and a microphone to generate sound fields consisting of a single reflection and a reverberation tail. Ueno and Tachibana [16] designed a 6-loudspeaker system to simulate sound fields obtained through the measurement of the corresponding impulse responses in real rooms.

During the past few years, many technological advances have made it possible to implement techniques which were previously known but not technically possible. As an example, state-of-the-art PCs have sufficient processing capability to perform a number of simultaneous convolutions efficiently, without expensive and dedicated DSP, as required one decade ago [17]. There are several free software open source solutions available to perform efficient multiple channel convolutions with very low delay [18, 19]. The release of new multi-channel digital audio standards such as MADI [20], in combination with multi-channel sound cards, has simplified the connections from the system, expanded the possibilities of centralized convolution systems, and made the technology affordable for a larger number of people. In addition, state-of-the-art room acoustics simulation software provides fairly accurate predictions of the sound-fields in rooms [21, 22]. The system presented in this paper takes advantage of all these innovations to perform the real-time convolution of the own voice with a 29-channel simulated impulse response that, reproduced through 29 loudspeakers, generates the reflected 3D sound field of one's own voice. These components is added to the sound of one's own voice propagated directly through the air or through the body. The reconstruction of the reflected sound field is made according to a realistic approach. It combines the output of a room acoustics simulation program [23] with the spatial and psychoacoustic decoding scheme proposed by Favrot and Buchholz [5], thus preserving delay, amplitude, spectrum, and directional cues of the simulated reflections. Very long impulse responses can be used, so the system does not put a restriction of the maximum length for practical use in room acoustics.

2 System description

2.1 Overview

A block diagram illustrating the overall real-time auralization system is shown in Fig. 1. As can be seen, there are two main parts, namely the pre-processing stage and the real-time processing, acquisition, and reproduction stage. The first part includes all the necessary steps to obtain the impulse responses of the environment that will be used in the auralization. This includes the design of a computerized room acoustic model, the calculation of an impulse response with room acoustics software and its encoding and decoding with mixed-order Ambisonics techniques for the given layout of the loudspeaker reproduction system. The second part contains all the elements of the system that apply the desired room impulse response to a talker's speech signal in real time. These are: an acquisition part with a microphone and a sound card, a real-time section with a software convolver and an equalizer filter, and a reproduction system based on 29 loudspeakers.

2.2 Pre-processing

A very important part of the auralization system is the offline calculation, decoding, and storage of an accurate set of impulse responses ready to be used in the second block, which applies the room effect to a talker's voice in real-time. This part of the system is an adaptation of the LoRA toolbox designed by Favrot and Buchholz [5]. The LoRA toolbox is a software application that uses the output (impulse response with directional information) of an acoustics simulation program to encode it in Ambisonics and decode it to a particular reproduction layout, producing an IR for each loudspeaker. However, some modifications in the procedure and calculation are needed in order to

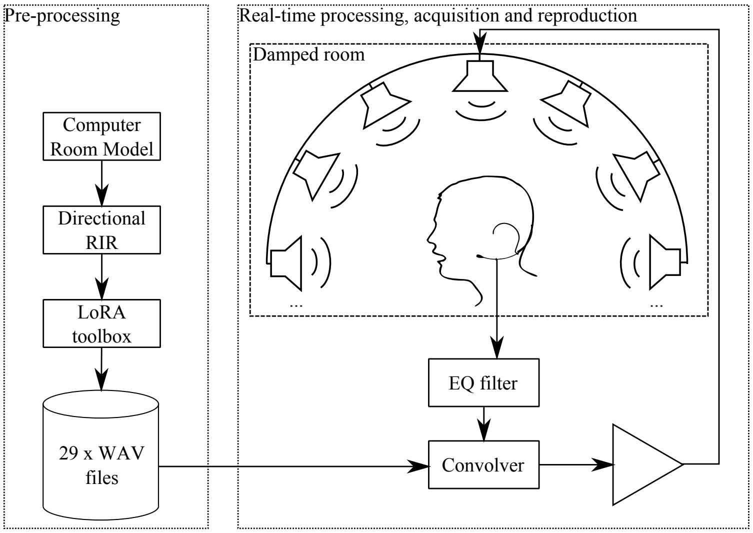

Figure 1: Block diagram of the real-time convolution system.

match the requirements for self-voice auralization.

First, a computer-based room acoustic model is needed, which is then loaded into an acoustic simulation program. In the proposed system, Odeon is used [23], although other alternative solutions may also be used, as long as the interface with the LoRA toolbox is implemented satisfactorily. In the acoustic simulation, the source is located at the talker's position, avoiding positions too close to the boundaries that could not be satisfactorily reproduced by the system due to the inherent latency (analyzed in section 3.2). The receiver point is located 1 m in front of the source. Note that this position does not correspond to the position of the ears relative to the mouth (sound source). However, the reflection pattern is reasonably similar to the reflection pattern experienced at the position of the ears. In addition, the proposed calibration method takes advantage of this approximation, as will be discussed in section 3.3. The source is oriented toward the audience and has a directivity pattern similar to the average human speech [24, 25].

For rooms in a volume range of approximately 100 m^3 $< V <$ 1000 m^3, the used simulation parameters are 5000 rays, a maximum reflection order of 2000, a transition order of 3 reflections between early reflections and late reverberation, and a histogram resolution of 10 ms for the late reverberation. The length of the response is adjusted to correspond at least to the largest reverberation time among all frequency bands for the simulated room. The early part of the response is calculated through the image source method and the late part by ray-tracing. Although 5000 rays are usually a low number in this kind of simulations, it is not of critical importance here, since the fine structure of the late reverberation is not of interest, but only the envelope of the energy-time curve.

The acoustic simulation program exports the discrete early reflections separately, each one with its delay, direction of incidence, and attenuation per frequency band. The late reverberation is exported as vectorial intensity (i.e., in first order Ambisonics format WXYZ) in each of the standard octave frequency bands from 63 Hz to 8 kHz at the defined time intervals. The combination of these two components is referred to as the *Directional IR* in Fig. 1. The LoRA toolbox is adapted to omit the direct sound from these files, because it will be produced by the talker himself during the real-time auralization. The early reflections are then encoded in fourth order Ambisonics and decoded into the corresponding loudspeaker layout for reproduction (see Fig. 6 in [5]). The envelope of the reverberation tail is decoded with a lower directional accuracy (first order Ambisonics) than the early reflections, which leads to a higher degree of diffuseness in the resulting multichannel IR. The decoded envelopes are filled with noise sequences uncorrelated among the different channels, in order to avoid coherent interference effects and coloration of the sound. The late reverberation is added to the early reflections and the resulting impulse responses for each loudspeaker are stored as separate WAV files with a 32 bit precision and a sampling frequency of 44.1 kHz.

2.3 Real-time acquisition, processing, and reproduction

The real-time operations in the system can be separated into signal acquisition, processing (convolution), and reproduction.

2.3.1 Signal acquisition

The talker's speech signal is picked with a headworn microphone DPA 4066-F, placed on the talker's cheek, then digitalized at 44.1 kHz/24 bit with a Behringer ADA8000 and sent into a PC with a RME HDSP MADI audio interface connected to a RME ADI648 (MADI/ADAT converter). Although other placements of the microphone could be more suitable for research, as e.g. in Cabrera *et al.* [10], the built-in fitting accessory was quite ergonomic and well adapted to the placement on the cheek. The microphone capsule is close enough to the mouth to avoid any severe influence of feedback (see analysis later in the paper). As in Pörschmann [26], the spectral distortion introduced by this placement of the microphone is corrected with an equalizer filter $h_{\mathrm{EQ}}(t)$, which adjusts the spectrum of the speech signal to match the spectrum of the on-axis speech signal at 1 m in front of the mouth. The calculation of the equalizer filter is done on an individual basis, as the placement of the microphone in relation to the mouth of the speaker differs among users. The measurement of the equalizer filter is used also to calibrate the system, as detailed in the next section. The justification for applying the equalizer filter is that the calculation of the impulse response in the simulation program assumes an on-axis source signal to provide a spectrally correct output. For practical reasons, the equalizer filter was pre-convolved with the stored multichannel room impulse responses, reducing the overall delay in the system during run-time operation. Nevertheless, the conceptual representation of Fig. 1 is still valid.

2.3.2 Convolution

The convolver is the most technically demanding element of the system. It should provide high quality audio, both regarding bit depth and sampling frequency, introduce the lowest possible delay between input and output, and convolve a number of long impulse responses. Lengths of hundreds of thousands of taps are typical for room impulse responses. In the present system, 29 simultaneous convolutions are required (one for each loudspeaker).

To perform the convolutions, a free software convolver—*jconvolver*—is used [19]. *Jconvolver* is a multichannel software implementation of the variable block-size convolution scheme proposed by Gardner [27]. It runs in a four-core PC under Fedora 8 Linux, patched with the real-time kernel module from Planet CCRMA and uses JACK audio server with ALSA sound driver architecture. The convolver is configured with a simple script that defines the input (the speech signal from the microphone), the 29 impulse responses, and adjustments of gain and delay to account for the position of the loudspeakers in the actual arrangement, which are at different distances from the center of the layout. With JACK, each of the outputs of the convolver are assigned to physical outputs of the audio interface.

In order to investigate the demands of the DSP software in relation to the process capability of the hardware (Quad core Intel PC with 8 GB of RAM), a small benchmark study was carried out. In Table 1, the CPU load is measured as a function of the minimum block size (identical for JACK and *jconvolver*) and length of the impulse response, while calculating 29 impulse responses. In Table 2, the CPU load is indicated for each combination of number of channels and minimum block size, for an impulse response of 65536 samples. The CPU load increases with the number of channels and the length of the impulse response, whereas it decreases with the block size. The drawback of the decrease in CPU load is an increase in latency, which is not desirable for real-time convolution. The measured low values of CPU load show that it is possible

Table 1: Benchmarking. CPU load versus different combinations of minimum convolution block size and impulse response length, for 29 parallel convolutions and a 44.1 kHz sampling frequency. The latency introduced by *jconvolver* is indicated in parentheses.

IR length	Block size (latency)		
	64 (2.9 ms)	128 (5.8 ms)	256 (11.6 ms)
22050	8.7 %	7.3 %	6.6 %
44100	9.2 %	7.9 %	7.2 %
88200	10.2 %	9.0 %	8.2 %
176400	13.4 %	11.8 %	11.0 %

to run in parallel alternative processes to record or monitor an input or output signal, or also to run multiple instances of *jconvolver* in the same computer, so as to simulate more complex auditory environments, for example, adding a second sound source at a different position in the simulated room.

2.3.3 Reproduction

The output signals are converted into the analog domain with a MADI/ADAT converter RME ADI-648 and four Behringer ADA8000 devices, amplified, and sent to 29 DYNAUDIO BM6 loudspeakers. The loudspeakers are arranged on the surface of a quasi-sphere with distances in the range 1.5 m–2.0 m from the center of the arrangement (see Fig. 2 for specific details of this layout). As the frequency response of the loudspeakers is fairly flat in the frequency range of interest for voice (100 Hz–10 kHz), no equalizers are introduced, as these could be detrimental for the audio quality with small displacements from the equalized position [28].

Table 2: Benchmarking. CPU load versus different combinations of minimum convolution block size and number of channels, for an impulse response of 65536 samples and a 44.1 kHz sampling frequency. The latency introduced by *jconvolver* is indicated in parentheses.

Number of channels	Block size (latency)		
	64 (2.9 ms)	128 (5.8 ms)	256 (11.6 ms)
4	2.1 %	1.6 %	1.5 %
8	3.1 %	2.7 %	2.4 %
16	6.0 %	5.0 %	4.4 %
32	10.8 %	9.3 %	8.6 %

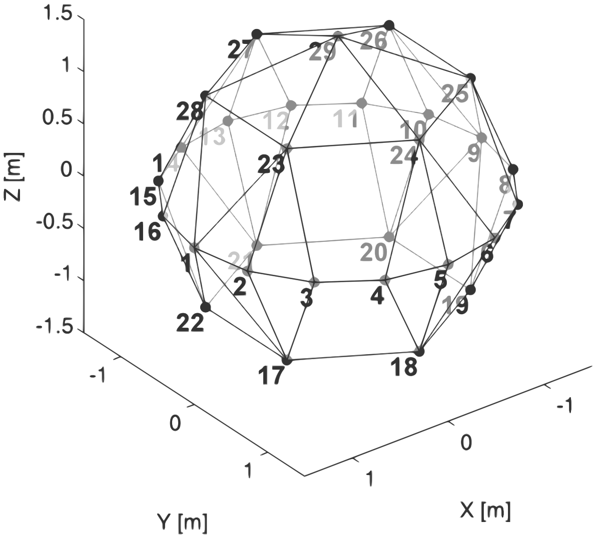

Figure 2: Position of the 29 loudspeakers in the array used for reproduction (from Favrot and Buchholz [5]).

Table 3: Reverberation time T_{30} of the test room.

Frequency [Hz]	125	250	500	1000	2000	4000
T_{30} [s]	0.16	0.09	0.08	0.07	0.07	0.07

3 Practical considerations

There are some practical issues that should be addressed so that this auralization system works as intended.

3.1 Acoustic conditions of the reproduction room

In the first place, the real-time auralization system requires an acoustically dry environment, ideally anechoic, so that the loudspeakers reproduce what they are meant to and not a combination of the simulated room and the test room itself (due to the sound reflections). The physical reproduction room, with dimensions 4.7m × 4.6m × 3.4m, is covered in its whole majority with sound absorbing materials, and its reverberation time, measured according to the standard ISO-3382 [29], is shown in Table 3. The value of 0.16 s at 125 Hz could seem a bit high for this application, but due to the fact that the reflected component of one's own voice in this frequency band is much lower than the sound transmitted directly through the air or through the body, the influence is negligible. As with the loudspeakers, no inverse filtering of room acoustics is applied.

3.2 Delay / Latency

The term *real-time* applied to this system can lead to some confusion or misunderstanding, as there is actually a certain latency or delay in the system. In a room impulse response, there is usually a time gap between the arrival of the direct sound and the first reflection. If the latency of the system is shorter than this gap, then it is possible to remove a number of samples corresponding to the latency, compensating for this delay without missing any reflection. In our system, the measured latency was 11.5 ms. This delay included the block size used in JACK (64 samples) at the input and the output, the block processing in *jconvolver* (64 samples), the time of sound propagation from the loudspeakers to the ears, and smaller delays in other processes (A/D, D/A, etc).

Considering the sound propagation between the mouth and the ears, a time gap of 11.5 ms between the arrival of the direct sound and the first reflection corresponds to a reflection coming from a boundary at a distance of 2 m. Thus, reflections coming from walls closer than this distance cannot be simulated properly, with precise timing, level, and direction. As a consequence, the smallest volume of a box-shaped room with the source at its center that can be accurately simulated is $(2 \times 2)^3 = 64\text{m}^3$. However, smaller rooms are highly dominated by modal effects in a broad frequency range and acoustic simulations with ray-tracing and image-source methods do not perform very accurately for these situations. As a rule of thumb, one limitation of the system is that it cannot simulate rooms smaller than the laboratory room.

Shorter latencies would be desired in this system, although a very obvious limit in our system is imposed by the distance to the loudspeakers. Reducing the distance to the loudspeakers might not be a good solution, because the number of loudspeaker would need to be reduced, reducing the accuracy on directional reproduction, or the loudspeaker would produce much more noticeable physical sound reflections, which should be avoided.

3.3 Calibration

A correct calibration is crucial for providing a convincing experience while using the system. The calibration of this system has two goals: on the one hand, adjusting the frequency response in order to compensate for the location of the acquisition microphone in relation to the talker's mouth, and on the other hand, adjusting the direct-to-reflected energy level difference to match that of a realistic situation.

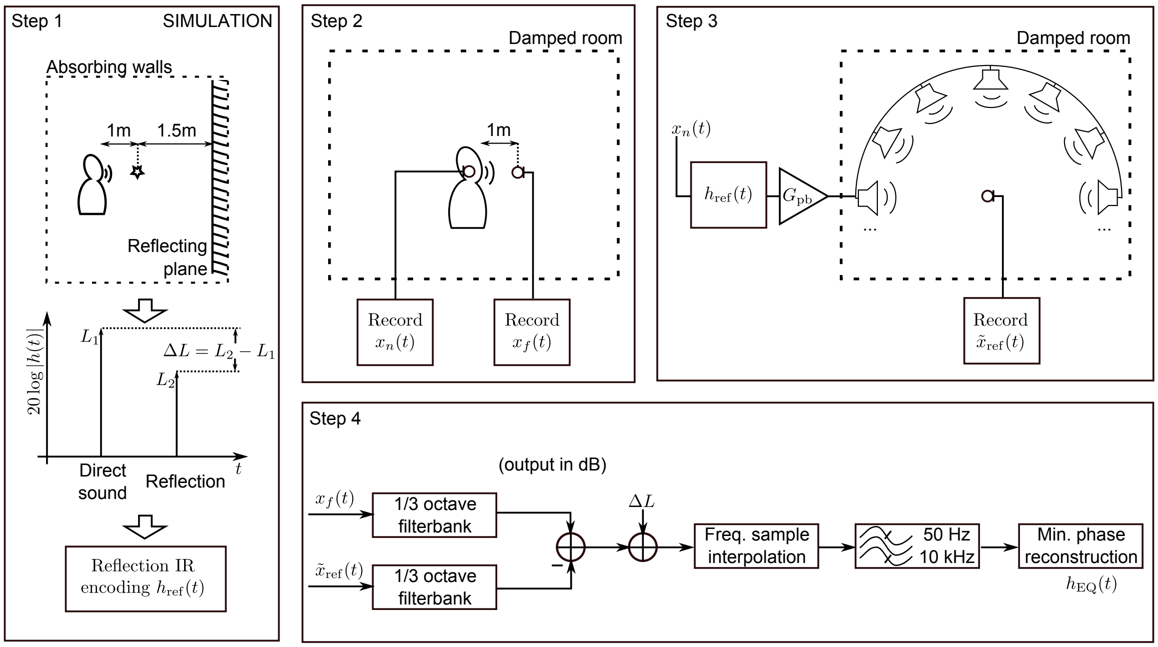

Figure 3: Steps involved in the calibration process. Step 1: calculation, by means of simulation, of a reference impulse response consisting of the direct sound and a single reflection. Step 2: Measurement of the on-axis speech signal at 1 m. Step 3: Playback, processing, and recording of the reflection. Step 4: Comparison of the direct sound and the reflection to obtain the personalized equalizer filter.

Given a human speaker, the level difference between the direct sound at a distance of 1 m in front of him and the reflected sound (direct-to-reflected ratio) is noted as ΔL. This difference must be the same, regardless of the fact that it is obtained by simulation or in a real scenario. The proposed calibration method, summarized in Fig. 3 aims at replicating the level difference obtained by simulation in the real-time auralization system.

The first step, only performed once, is the calculation of an impulse response $h(t)$ produced by a single reflecting plane in front of a human speaker by means of acoustic simulation,

$$h(t) \approx \delta(t - t_d) + 10^{\Delta L/20}\delta(t - t_r), \tag{1}$$

where $h(t)$ is calculated with the same calculation parameters as for an arbitrary room. That is, the source has the directivity features of a human speaker, and the receiver is located 1 m in front of it. The reflecting plane is in this case located 1.5 m in front of the receiver point. The plane is orientated normally to the line that connects source and receiver. In this way, the direct sound (with delay t_d and level L_1, regarded as the reference) and the reflection (with delay t_r and level L_2) originate from the same direction and have the same spectral distribution, ignoring the effect of air absorption and the finite size of the plane. The level difference between the two components is $\Delta L = L_2 - L_1$ dB. The IR corresponding to the single reflection, excluding the direct sound, is noted as $h_{\text{ref}}(t)$, processed with the LoRA toolbox, and stored,

$$h_{\text{ref}}(t) \approx 10^{\Delta L/20}\delta(t - t_r), \tag{2}$$

and the corresponding Fourier transformed version:

$$20\log_{10}|H_{\text{ref}}(f)| = \Delta L \text{ dB}. \tag{3}$$

The second step requires the presence of a human talker in the loudspeaker room. The talker is equipped with the headworn microphone, which requires a careful fixing to the talker's head in order to preserve the relative position to the mouth throughout the operation. A measurement microphone B&K type 4192 is placed 1 m in front of the mouth. Next, the talker is asked to speak continuously during 30 s, staring at a reference sign so that the mouth is aligned with the measurement microphone. Both signals from the measurement microphone $x_f(t)$ and the headworn microphone $x_n(t)$ are recorded simultaneously. The goal of the calibration procedure is to obtain an ideal equalizer filter $\tilde{h}_{\text{EQ}}(t)$ that applied to $x_n(t)$ and reproduced through the system (with a gain symbolized G_{pb}, where "pb" stands for "playback") produces $x_f(t)$ at the center position,

$$G_{\text{pb}}x_n(t) * \tilde{h}_{\text{EQ}}(t) = x_f(t), \tag{4}$$

or in the frequency domain,

$$G_{\mathrm{pb}} X_n(f) \tilde{H}_{\mathrm{EQ}}(f) = X_f(f), \tag{5}$$

from which the ideal filter results,

$$\tilde{H}_{\mathrm{EQ}}(f) = \frac{X_f(f)}{G_{\mathrm{pb}} X_n(f)}. \tag{6}$$

The gain of the system G_{pb} is still unknown and requires another measurement. The room should be empty and the measurement microphone has to be moved to the center of the laboratory room, so that its position corresponds to the point between the two ears when a talker would be present. The previously recorded signal from the headworn microphone $x_n(t)$ is routed to the input of the convolver, which is loaded with the single reflection, $h_{\mathrm{ref}}(t)$. The output of the convolver is sent to the amplifiers and reproduced through the loudspeakers. At the same time, the measurement microphone records the resulting signal, $\tilde{x}_{\mathrm{ref}}(t)$,

$$\tilde{x}_{\mathrm{ref}}(t) = G_{\mathrm{pb}} x_n(t) * h_{\mathrm{ref}}(t), \tag{7}$$

and the corresponding Fourier transform:

$$\tilde{X}_{\mathrm{ref}}(f) = G_{\mathrm{pb}} X_n(f) H_{\mathrm{ref}}(f). \tag{8}$$

From the previous signals, it is possible to calculate the filter $\tilde{H}_{\mathrm{EQ}}$:

$$\tilde{H}_{\mathrm{EQ}}(f) = \frac{X_f(f) H_{\mathrm{ref}}(f)}{\tilde{X}_{\mathrm{ref}}(f)}. \tag{9}$$

Making use of eq. (3) and using logarithms:

$$20 \log_{10} |\tilde{H}_{\mathrm{EQ}}(f)| = 20 \log_{10} |X_f(f)| - 20 \log_{10} |\tilde{X}_{\mathrm{ref}}(f)| + \Delta L. \tag{10}$$

The fourth step is a practical implementation of eq. (10). It uses the signals corresponding to the on-axis direct sound, $x_f(t)$, and the reflection, $\tilde{x}_{\mathrm{ref}}(t)$, as inputs. The signals

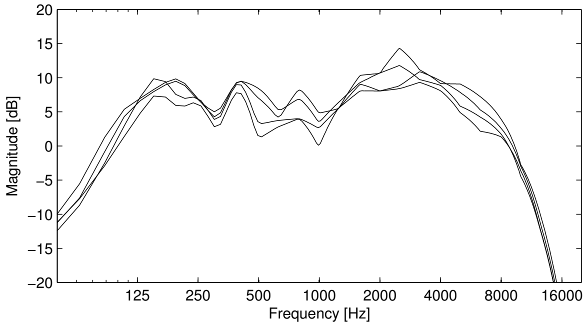

Figure 4: Magnitude response of the equalizer filters $h_{\mathrm{EQ}}(t)$ calculated in repeated measurements with different placement of the headworn microphone on the same talker.

$x_f(t)$ and $\tilde{x}_{\mathrm{ref}}(t)$ are processed with a spectral analyzer that calculates the energy level of the signals in one-third octave frequency bands between 31.5 Hz and 16 kHz. The level difference between the two components is calculated and the target level difference ΔL is added. The result is the magnitude frequency response (in one-third octave bands) of the equalizer filter. The magnitude frequency response at frequencies other than the standardized one-third octave center frequencies are obtained by interpolation. The response is band-pass filtered to eliminate frequencies lower than 50 Hz, which are not likely to have been produced by the human voice, and frequencies higher than 10 kHz, to prevent unstable feedback in the system. The resulting filter $h_{\mathrm{EQ}}(t)$ (slightly different from the ideal $\tilde{h}_{\mathrm{EQ}}(t)$) is a 2048-tap FIR filter obtained by minimum phase reconstruction of the magnitude frequency response described in the previous steps.

As an example, Fig. 4 shows the magnitude frequency response of the equalizer filter $h_{\mathrm{EQ}}(t)$ calculated for the same talker with slightly different microphone positions. As can be seen, these filters are fairly consistant, with a standard deviation of about 1.8 dB (averaged across frequency).

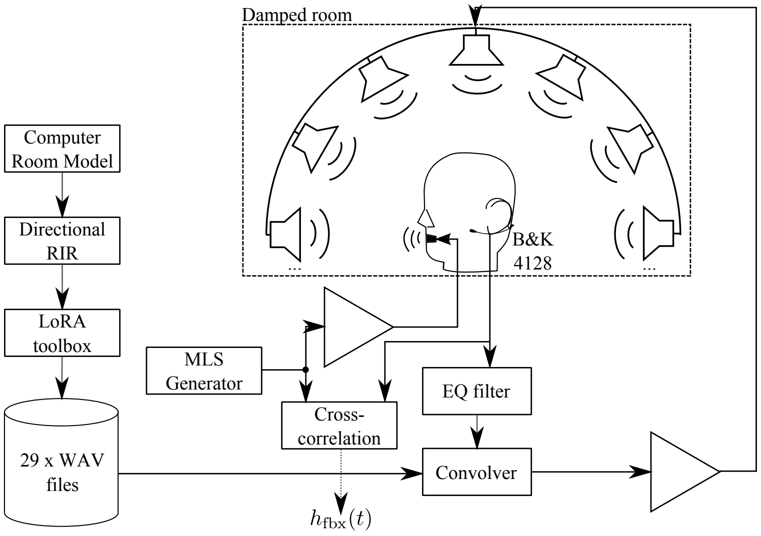

Figure 5: Setup used to measure the impulse response $h_{\mathrm{fbx}}(t)$ between the mouth and the headworn microphone, from which the feedback in the system is evaluated.

3.4 Feedback

The presence of the acquisition microphone and the loudspeakers in the same room generates a closed loop which introduces some feedback (unstable or not) in the system. Inspired by the method of Rokutanda *et al.* [30], the feedback is derived from an IR measurement $h_{\mathrm{fbx}}(t)$ at the headworn microphone using a Head And Torso Simulator B&K type 4128 (HATS) while the auralization system is running an arbitrary room simulation (see the complete system in Fig.5). The mouth-loudspeaker of the HATS is driven with an amplified pseudo-random noise signal (MLS). By calculating the cross-correlation of this signal and the signal at the headworn microphone, $h_{\mathrm{fbx}}(t)$ is obtained. It also contains the effect of the mouth radiator. The early part of $h_{\mathrm{fbx}}(t)$, in this case, contains the direct sound plus some reflections from the loudspeaker room and the torso, and the rest of $h_{\mathrm{fbx}}(t)$ is the feedback component. Figure 6 provides an example measured response. Comparing the early (direct) part with the feedback, the feedback-to-direct ratio (FDR) is calculated.

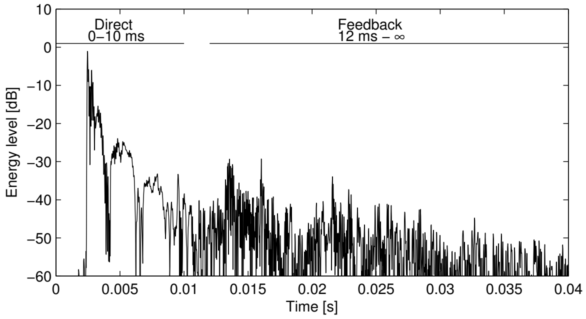

Figure 6: Feedback impulse response $h_{\text{fbx}}(t)$

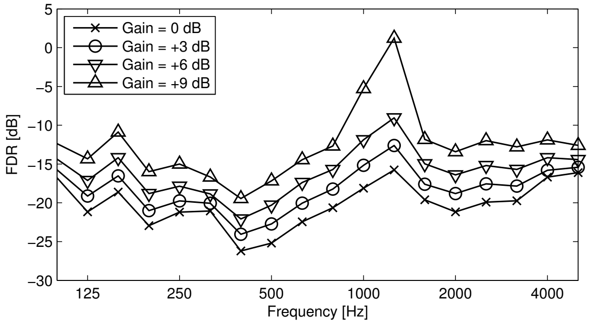

Figure 7: Feedback-to-direct as a function of the frequency for different gains of the simulated room impulse responses

The FDR, calculated for the same simulated room impulse response at different gains, and for different frequency bands, is shown in Fig. 7. As can be seen, when the gain of the system is increased more than 6 dB, the feedback component increases non-linearly at frequencies around 1.25 kHz. This non-linear increase of the feedback component results in instability and oscillation. In normal operation of the system (a gain of 0 dB), the feedback is 15 to 25 dB lower than the direct sound, depending on the frequency band. These curves have been calculated for a central position of the talker, facing to the front, and a particular simulated room IR. For other simulated room IR, orientations or positions of the talker, the curves of Fig. 7 would appear different.

3.5 Misalignment error

The system for the real-time auralization of one's own voice assumes a speaker at the exact center of the loudspeaker array facing to the front. In this case, the directivity of the human talker matches the simulated directivity pattern, which has been chosen with this orientation. When the human talker turns around, the directivity pattern which is simulated by the system is still facing to the front. There is a mismatch between the actual and the simulated directivity pattern which emphasizes reflections from some directions and attenuates reflections from other directions. As a result, the simulated sound field is wrong. However, slight movements of the head do not give rise to a serious error. A measure of the error ϵ produced by head rotations in azimuth ϕ_0 and elevation θ_0, could be quantified by the following formula:

$$\epsilon(\phi_0, \theta_0) = 1 - \frac{\iint_{4\pi} D(\phi, \theta) D(\phi - \phi_0, \theta - \theta_0)\, \mathrm{d}\Omega}{\iint_{4\pi} D^2(\phi, \theta)\, \mathrm{d}\Omega}, \tag{11}$$

where D is the linear directivity pattern of the simulated human head (assumed to be equal to an average talker long term speech directivity), ϕ and θ are the spherical coordinates (see Fig. 8), $\phi = 0, \theta = 0$ is the design orientation of the talker, and Ω indicates the solid angle. The head rotations in the radial direction are ignored in this analysis. Figure 9 shows the error graphically on a logarithmic scale: $10 \log(1 - \epsilon)$. As expected, the accuracy decreases with frequency, as the voice becomes more directive. It is worthwhile to point out that the error produced by azimuthal rotations is lower than the error that would be produced by the same rotations in elevation. Azimuthal head rotations are more likely to occur than elevational ones. The contour lines at -3 dB show that azimuthal head rotations in the range $-30^o \leq \phi_0 \leq 30^o$ do not introduce severe inaccuracy of the simulated sound field. However, this error could be minimized, and the accuracy drastically improved by introducing a head tracking system that used the information about the head orientation to dynamically update the multichannel

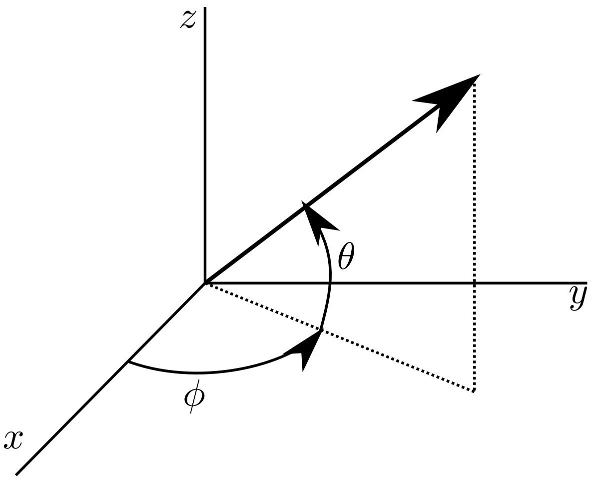

Figure 8: Definition of angles in spherical coordinates.

impulse response.

3.6 Summary of assumptions

The real-time auralization system for the own voice has been built-up on a number of assumptions which are worthwhile summarizing:

- Only the reflected corresponding to the own voice is simulated, as the direct airborne sound and the body-conducted sound are present in our voice when the ears are not blocked.
- The directivity of the human voice is kept fixed, independently of the phoneme, the phonation mode and the subject. In reality, the directivity pattern changes noticeably with these variables.
- The reflected sound field at 1 m in front of the mouth is fairly similar (in the statistical sense) to the sound field at the ears. This assumption, although questionable, is necessary in order to apply the proposed calibration method.

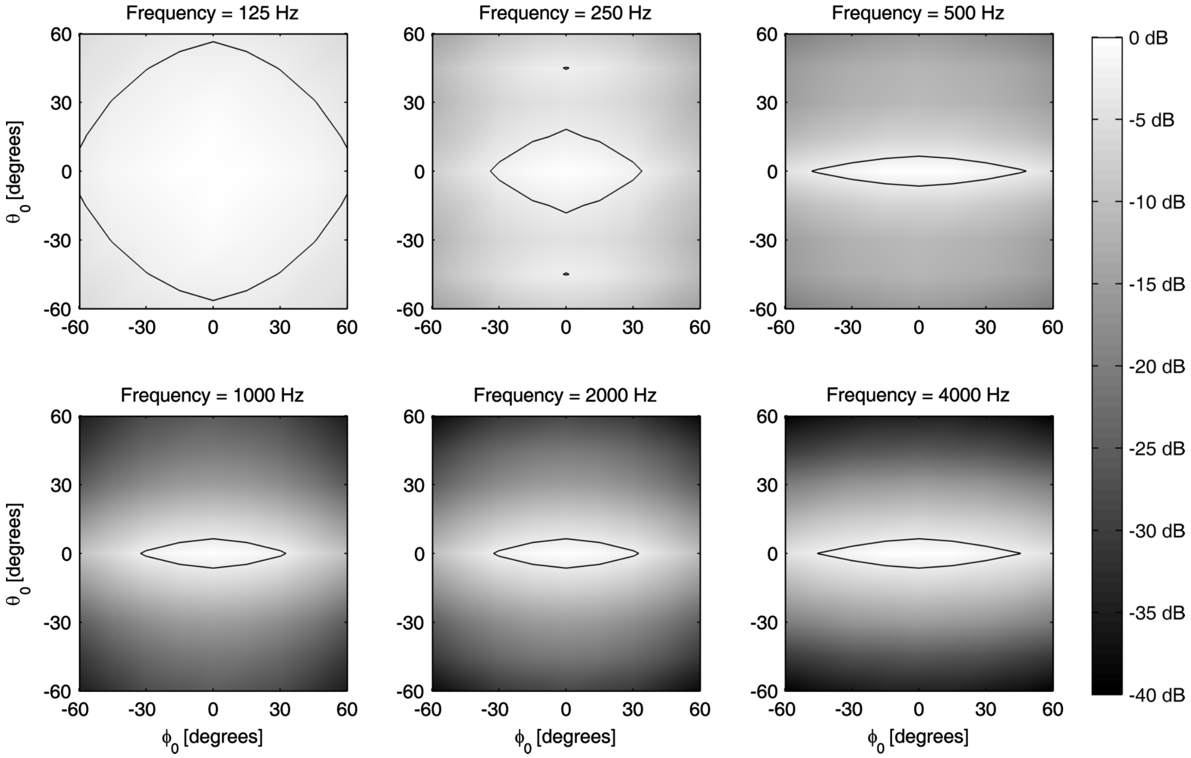

Figure 9: Accuracy of the simulation $10\log(1-\epsilon)$ as a function of the azimuth (ϕ_0) and elevation (θ_0) head rotations for the different frequency bands. The 0 dB value corresponds to the perfect alignment of the talker to the simulated orientation. The -3 dB contour lines are indicated in solid.

- The effect of room acoustics in the physical room is ignored. It would be better, however, to install this system in a completely anechoic chamber.
- Feedback during operation (closed loop) has not been taken into account during calibration (open loop).
- The misalignment error is acceptable for use in $\pm 30^o$ around the front direction.

4 Applications

The first experiences trying the system have been very positive, in the sense that it generates a convincing impression of being in environments different from the reproduction room, matching the expectations that talkers have about the simulated environments.

The system for real-time auralization of one's own voice finds one of its main applications in psychophysics or cross-modality research. It is possible to investigate how people perceive environments by using exclusively aural cues produced with their own voices, study the subjective effects of the acoustic environment on voice production, or study the preference of theater actors in different acoustical settings. The system described in this article is being used at the time of publication in a research project where the relation between classroom acoustical conditions and the vocal behavior of a teacher is investigated.

The system could easily be adapted for use with music instruments. In this case, it would be necessary to make the computer acoustic simulation with the directivity pattern of the desired musical instrument, and perform the calibration exactly described in this paper, but replacing the headworn microphone with a microphone to pick the sound from the instrument. However, this microphone needs to be mounted on the instrument to reject feedback and avoid the variation of the acoustic path between source and acquisition microphone during operation.

Furthermore, this system could be used as a part of larger virtual reality systems, in order to achieve a more immersive experience [26]. Applying some of the techniques here described and simplifying the reproduction method, this kind of system might also find place in digital entertainment.

Acknowledgments

Thanks to Sylvain Favrot and Jörg Buchholz for the development of the hardware arrangement and the LoRA toolbox, which allowed us to obtain the present system. Thanks to Anders Christian Gade for his valuable ideas regarding calibration and quality of the system. This research is financed by the Swedish organization AFA Försäkring.

References

[1] V. Jónsdottir and L. Rantala, "Nordic cooperation in the field of voice ergonomics," in *Proceedings of the first Nordic conference of voice ergonomics and treatment*, pp. 8–9, 24-25th April 2009 2009.

[2] N. Roy, R. Merrill, S. Thibeault, R. Parsa, S. Gray, and E. Smith, "Prevalence of voice disorders in teachers and the general population," *J. Speech Lang. Hear. R.*, vol. 47, pp. 281–93, 2004.

[3] V. Lyberg-Åhlander, R. Rydell, and A. Löfqvist, "Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff," *J. Voice*, 2010. In press.

[4] J. Brunskog, A. Gade, G. Payá-Ballester, and L. Reig-Calbo, "Increase in voice level and speaker comfort in lecture rooms.," *J. Acoust. Soc. Am.*, vol. 125, pp. 2072–82, april 2009.

[5] S. Favrot and J. M. Buchholz, “LoRA - A loudspeaker-based room auralisation system,” *Acta Acust. united Ac.*, vol. 96, pp. 364–375, 2010.

[6] M. Kleiner, B. Dalenbäck, and P. Svensson, “Auralization - an overview,” *J. Audio Eng. Soc.*, vol. 41, pp. 861–861, 1993.

[7] M. Kleiner and A. Berntson, “Early frontal plane reflections preferred for talkers,” *12th ICA Proceedings (Canadian Acoustical Association, Toronto, Canada, 1986).*

[8] J. Shearer and R. Torres, “Auralization studies on talker comfort,” *J. Acoust. Soc. Am.*, vol. 114, p. 2305, 2003.

[9] B. Dalenbäck, “CATT-Acoustic, Room Acoustic Prediction Software,” 2006.

[10] D. Cabrera, H. Sato, W. L. Martens, and D. Lee, “Binaural measurement and simulation of the room acoustical response from a person’s mouth to their ears,” *Acoustics Australia*, vol. 37, no. 3, pp. 98–103, 2009.

[11] A. Marshall and J. Meyer, “The directivity and auditory impressions of singers,” *Acustica*, vol. 58, no. 130-140, p. 178, 1985.

[12] D. Noson, S. Sato, H. Sakai, and Y. Ando, “Singer responses to sound fields with a simulated reflection,” *J. Sound Vib.*, vol. 232, no. 1, pp. 39–51, 2000.

[13] C. Yuen, P. Calamia, and N. Xiang, “Investigation of voice stage support: A subjective preference test using an auralization system of self-voice,” *J. Acoust. Soc. Am.*, vol. 121, no. 5, p. 3174, 2007.

[14] K. Stetson and J. Braasch, “Singers preferences for acoustical characteristics of performing spaces.,” *J. Acoust. Soc. Am.*, vol. 125, p. 2586, 2009.

[15] A. Gade, “Investigations of musicians room acoustic conditions in concert halls. Part I: Methods and laboratory experiments,” *Acustica*, vol. 69, pp. 193–203, 1989.

[16] K. Ueno and H. Tachibana, "Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system," *Acoust. Sci. & Tech.*, vol. 24, no. 3, pp. 130–138, 2003.

[17] D. McGrath, "Huron-a digital audio convolution workstation," *Presented at the 5th Australian AES regional convention*, p. preprint 4023, 1995.

[18] A. Torger and A. Farina, "Real-time partitioned convolution for Ambiophonics surround sound," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 21–24.

[19] F. Adriaensen, "Design of a convolution engine optimised for reverb," in *LAC2006. 4th Linux Audio Conference*, 2006.

[20] AES10-1991, "AES recommended practice for digital audio engineering - serial Multichannel Audio Digital Interface (MADI)," *J. Audio Eng. Soc.*, vol. 39, no. 5, pp. 368–377, 1991.

[21] I. Bork, "Report on the 3rd Round Robin on Room Acoustical Computer Simulation - Part 1: Measurements," *Acta Acust. united Ac.*, vol. 91, no. 4, pp. 740–752, 2005.

[22] I. Bork, "Report on the 3rd Round Robin on Room Acoustical Computer Simulation - Part II: Calculations," *Acta Acust. united Ac.*, vol. 91, no. 4, pp. 753–763, 2005.

[23] C. L. Christensen, *Odeon room acoustics program, version 9.1, user manual, industrial, auditorium and combined editions*, January 2008.

[24] W. Chu and A. Warnock, *Detailed directivity of sound fields around human talkers.* Institute for Research in Construction, National Research Council Canada, Tech. Rep, Canada, 2002.

[25] C. Hopkins, *Building Bulletin 93. Acoustic design of schools, a design guide.* Department of Education and Skills, London, UK, 2004.

[26] C. Porschmann, "One's own voice in auditory virtual environments," *Acta Acust. united Ac.*, vol. 87, no. 3, pp. 378–388, 2001.

[27] W. Gardner, "Efficient convolution without input-output delay," *J. Audio Eng. Soc.*, vol. 43, no. 3, pp. 127–136, 1995.

[28] S. Norcross, G. Soulodre, and M. Lavoie, "Subjective investigations of inverse filtering," *J. Audio Eng. Soc.*, vol. 52, no. 10, pp. 1003–1028, 2004.

[29] International Organization for Standardization, "ISO-3382, Acoustics - Measurement of the reverberation time of rooms with reference to other acoustical parameters," 1997.

[30] M. Rokutanda, T. Kanamori, K. Ueno, and H. Tachibana, "A sound field simulation system for the study of ensemble performance on a concert hall stage," *Acoust. Sci. & Tech.*, vol. 25, no. 5, pp. 373–378, 2004.

1st EAA – EuroRegio 2010
Congress on Sound and Vibration
15 - 18 September 2010, Ljubljana, Slovenia
With Summer School for Young Researchers from 13 - 15 September 2010

Natural variations of vocal effort and comfort in simulated acoustic environments

David PELEGRIN GARCIA and Jonas BRUNSKOG

[a] Acoustic Technology, Dept. of Electrical Engineering, Technical University of Denmark
Oersteds Plads 352, 2800 Kgs. Lyngby, DENMARK, e-mail: dpg@elektro.dtu.dk

ABSTRACT

Many teachers suffer from voice problems related to the use of their voices in the working environment. The noise generated by students and external sound sources (like traffic noise or neighboring classrooms) is a major problem, as it leads to an increased vocal effort. In the absence of high levels of background noise, the room has also an effect on the talker’s voice. In order to quantify the relative importance of the acoustic environment on the vocal demands for teachers, a laboratory investigation was carried out. Thirteen teachers had to read a text aloud under ten different room acoustic conditions, artificially generated by electroacoustic means. The vocal intensity decreased with the objective parameter *support*, which quantifies the amount of sound reflections provided by the room at the talker’s ears,relative to the direct sound, at a rate of -0.21 dB/dB. The reading pace decreased with the reverberation time at a rate of 5 words/minute per s. The sensation of comfort and suitability of the rooms for a talker was investigated using a questionnaire. A non-linear relationship between this magnitude and the reverberation time was observed, defining an optimum range around 0.85 s.

.

1. INTRODUCTION

Teachers suffer from voice problems in a greater proportion than in the rest of the population [1]. These problems are in many cases originated from the intensive use of their voices as an occupational tool [2]. Background or activity noise in classrooms makes teachers increase their vocal effort (and thus, the vocal intensity), as a consequence of the Lombard reflex and the need of keeping themselves heard on top of the noise [3].

In the absence of high levels of background noise, the classroom acoustic conditions can condition the vocal intensity produced by teachers [4]. Kob et al.[5] also showed an effect of classroom acoustics on teachers with and without voice problems.

In laboratory experiments, Pelegrin-Garcia and Brunskog [6] observed a decrease of the vocal intensity (measured as sound power level, or anechoic sound pressure level) with the ratio of reflected sound to direct sound (from the own voice) measured at the ears, at a rate of -0.65 dB/dB. Ten times the logarithm of the ratio between reflected and direct sound energy was defined as *support*, and it was extracted from the impulse response measured with a dummy head between a loudspeaker located at its mouth and a microphone at the eardrum position. However, that experiment lacked of significant results and used very few conditions.

The present paper reports the result of a more extensive laboratory experiment with more subjects (thirteen) and simulated sound field conditions (ten), analyzing other speech properties other than the vocal intensity, and studying the subjective impressions of talking in the rooms. The goals are two. First, determining more accurately what the relationship between room acoustics and voice production is, and second, observing which properties of a sound field make a room good to speak in there.

2. METHOD

Thirteen teachers (4 females, 9 males) of secondary school, high school, and university, aging 30 to 67 years, participated in the experiment. The teachers did not have known voice problems (according to their statements) or hearing loss greater than 25 dB HL below 4 kHz. Once they were in the laboratory room, and for each condition, they were instructed to read a text (Goldilock's passage [7]) during 2.5 minutes, addressing a listener located at a distance of 2 m. A dummy head was located at that position to provide the visual distance cue. After reading the text, the teacher had to rate a set of questions regarding the experience of talking in that condition, by making a vertical tick in a continuous horizontal line.

The different sound fields or experimental conditions were generated in a laboratory facility with a loudspeaker-based real-time auralization system [6]. It consisted of 29 loudspeakers placed in a quasi-sphere around a subject in a highly damped room. The speech signal from the talker in the center was picked with a headworn microphone, convolved in real time with the impulse response (IR) of the environment, and recorded for analysis. The different IR were obtained by computer acoustic simulation (Odeon) and mixed-order Ambisonics encoding/decoding using the LoRA toolbox [8].

There were ten experimental conditions, consisting of nine different simulated IR and the condition '0' of no IR simulated (and thus corresponding to the actual acoustic conditions of the laboratory room). The nine experimental conditions were the combination of three different classroom geometries (A,B, and C) and three different placement of absorptive materials in those rooms (1,2, and 3). Table 1 summarizes the geometrical volume, the reverberation time (T_{30}) and the support (*ST*) derived from objective IR measurements in the laboratory facility using a dummy head B&K HATS type 4128, with a loudspeaker at its mouth as a source and two microphones at the eardrums as receivers. The T_{30} was calculated as the average of the 500 Hz and 1 kHz octave bands, after removing the first 5 ms of the IR, in order to avoid the strong influence of the direct sound. The *ST* was calculated without frequency-weighting. Both T_{30} and *ST* were calculated for the left and the right ear and the results were averaged. The conditions were presented in random order for each subject.

Table 1. Experimental conditions and objective parameters

	A1	*A2*	*A3*	*B1*	*B2*	*B3*	*C1*	*C2*	*C3*	*0*
V (m^3)	1174	1174	1174	344	344	344	130	130	130	72
T_{30} *(s)*	1.62	1.40	0.64	1.06	0.66	0.62	1.02	0.66	0.40	0.05
ST (dB)	-15.6	-15.8	-17.4	-15.0	-15.6	-16.5	-12.1	-14.7	-16.8	-18.2

There were eight questions (Q1 to Q8) that the subjects had to answer. Q1 to Q6 were answered with the degree of agreement (left: totally disagree, right: totally agree) with the statement. Q7 was rated from very low (left) to very high (right) and Q8 with “no voice problems” (left) to “very severe voice problems” (right).

Q1. I would feel exhausted if I were talking in this classroom for a whole lesson
Q2. The classroom is good to speak in
Q3. The classroom enhances and supports my speech
Q4. I must raise my voice in order to be heard in the classroom
Q5. The sound system makes my voice sound unnatural
Q6. I noticed echo phenomena in the classroom
Q7. Rate the degree of reverberance that you perceived in the classroom
Q8. Rate how you perceive your voice now

As in [6], the sound power level L_W was extracted from the recordings. The fundamental frequency was extracted with the ESPS method, in intervals of 50 ms. The mean and standard deviation of the F0 sequences for each recording were calculated. However, there were no significant differences among different conditions. The number of words (nwords) completed during each recording period of 2.5 minutes was also counted. The statistical analysis of the data was performed with the statistical software package R.

3. RESULTS AND DISCUSSION

Figure 1 shows the measured L_W against *ST* values (left), and the number of words versus the T_{30} (right).

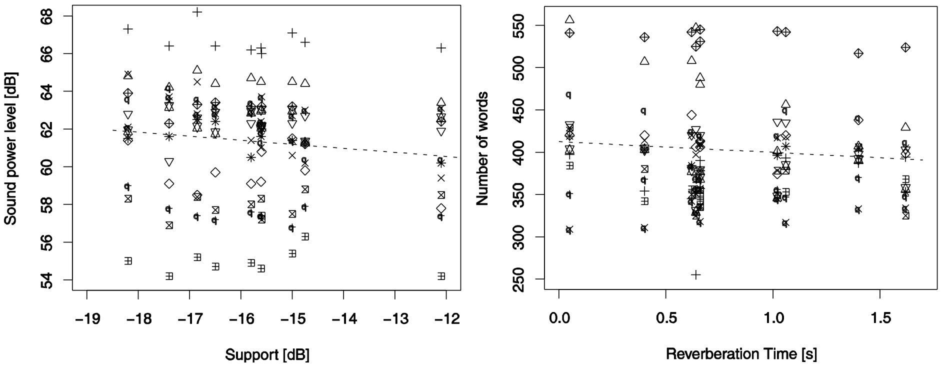

Figure 1. (Left) L_W versus *ST*. (Right) *nwords* versus T_{30}. Different symbols correspond to different subjects. The dashed lines correspond to regression lines calculated with linear mixed models

Figure 1 shows a big spread among observations. Most of them are related to individual factors which only shift the absolute values, while keeping similar variations among conditions. The factor "subject" was considered a random effect, and a linear mixed model [9] was used to evaluate the dependence of L_W with ST, finding a significant relationship (p=0.004). An identical procedure was followed to analyze n_{words} (p=0.045). The regression lines shown in Figure 1 correspond to the output of the linear mixed models, which are expressed in Eq. (1) and (2) for L_W and n_{words}, respectively

$$L_W = 58.0 - 0.21 \cdot ST \tag{1}$$

$$n_{\text{words}} = 412 - 12.6 \cdot T_{30} \tag{2}$$

As can be seen, the sound power level of the voice decreases with the ST, at a rate of -0.21 dB/dB. This rate is smaller (in absolute value) than reported in [6]. This deviation can be due to the different instructions given to the subjects. One reason for this might be that asking the talker to read a text aloud for a listener located at 2 m does not lead to the same voice adjustment as it would be required for addressing a group of people at further distances with spontaneous speech.

From equation (2), the number of words decreases consistently with T_{30}. The average reading rate in the extreme conditions, predicted by the model, are 164.5 words/minute (T_{30}=0.05s) and 156.6 words/minute (T_{30}=1.62s).

The answers to the questions Q1 to Q8 (in cm from the beginning of the line) are shown in Figure 2. The horizontal axes show the objective parameter (T_{30} or ST) that correlates best with the answers. The best fitting (and significant at the 5% level) models are also indicated on the figures. No model is shown for Q8 because an ANOVA test on the answers to this question suggests just random variation on the data.

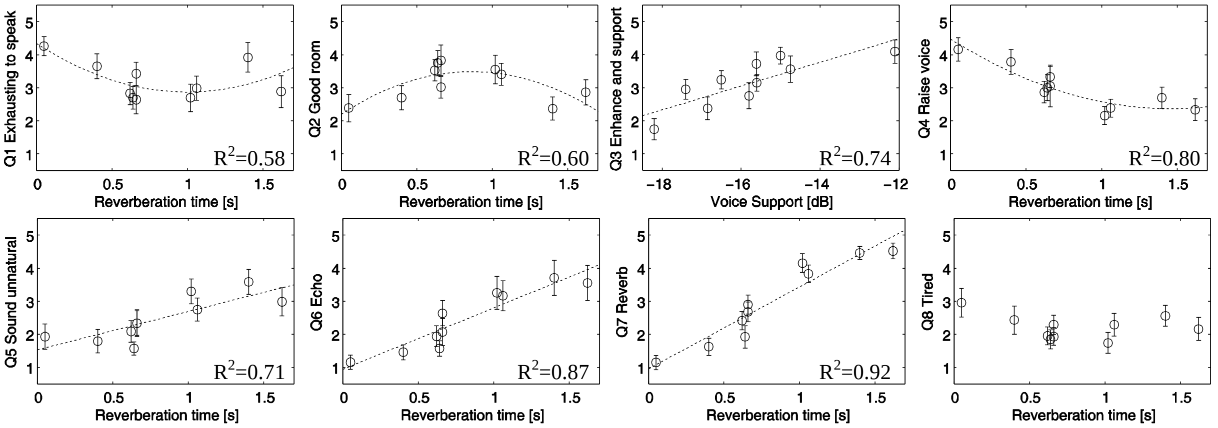

Figure 2. Ratings to the different questions Q1 to Q8 as a function of the objective room acoustic parameters T_{30} or ST

The questions Q1 and Q2, which ask questions related to the comfort show a non-linear dependance with the T_{30}. The regression model for Q2 is:

$$Q_2 = 2.2 + 2.98 \cdot T_{30} - 1.72 \cdot T_{30}^2 \tag{3}$$

This model reveals the presence of an optimum T_{30} range (around 0.85 s) which maximizes the vocal comfort in a classroom. However, this statement should be read carefully, as the effect of background noise is not included. In Q1, the minimum (least discomfort) is located at T_{30}=0.99s, slightly higher than with Q2.

It is worthwile to remark the high correlation between the perceived reverberance (Q7) and T_{30}. In a similar way, the answer to the question Q3, regarding the support from the room, is highly correlated with the objective parameter *ST*.

4. CONCLUSIONS

The conclusions of the laboratory experiment analyzing teachers' speech and subjective impressions under different room acoustic conditions are the following:

- The sound power level used by talkers decreases with the *support* of the room at a rate of -0.21 dB/dB.
- The reading pace (numbers of words per minute) decreases with the *reverberation time*, at a rate of 5 words/minute per each s.
- In the absence of high levels of background noise, a *reverberation time* around 0.85 s defines the talker's preferred acoustic conditions of a classroom in terms of vocal comfort.
- The sensation of support from the room is related to the objective parameter *support*.

ACKNOWLEDGMENTS

This work has been supported by AFA Försäkring. Thanks to Sylvain Favrot and Jörg Buchholz, who made the reproduction system described in this paper. The authors also want to express their gratitude to all the teachers participating in the tests.

5. REFERENCES

[1] V. Lyberg Åhlander, R. Rydell, and A. Löfqvist, "Speaker's comfort in Teaching Environments: Voice problems in Swedish Teaching Staff". Journal of Voice, 2010. In press.

[2] 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.

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

[4] J. Brunskog *et al.*, "Increase in voice level and speaker comfort in lecture rooms," J. Acoust. Soc. Am. Vol. 125(4), 2009, pp. 2072-2082.

[5] M. Kob *et al.*, "Experimental investigations of the influence of room acoustics on the teacher's voice," Acoustical Science and Technology, Vol. 29 (1), 2008, pp. 86-94.

[6] D. Pelegrín-García and J. Brunskog, "Prediction of vocal effort and speakers' comfort in lecture rooms". Proceedings of Internoise 2009, Ottawa, 2009.

[7] J.G. Švec, P.S. Popolo, I.R. Titze, "The Goldilocks passage and scripts for frequency extraction, voicing detection, SPL calculation and vocal dose determination in speech". The National Center for Voice and Speech Online Technical Memo, No. 1, April 2003, ver. 1.4.

[8] S. Favrot and J. M. Buchholz, "LoRA - A loudspeaker-based room auralisation system", Acta Acust. united Ac. Vol. 96, 2010, pp. 364–375.

[9] D. Hedeker, "Generalized linear mixed models", in Everitt, B. and Howell, D.C. (Editors) "Encyclopedia of Statistics in Behavioral Science, 2nd ed.", John Wiley & Sons, 2005.

Invited Paper

inter.noise 2010
noise and sustainability
13-16 June 2010 . Lisbon . Portugal
www.internoise2010.org

Measurement of vocal doses in virtual classrooms

Pasquale Bottalico[A]; David Pelegrin Garcia[B]; Arianna Astolfi[A]; Jonas Brunskog[B]

Affiliation: {[A] Department of Energetics, TEBE group Politecnico di Torino; Corso Duca degli Abruzzi, 24 10129 Turin, Italy; [B] DTU Electrical Engineering; Technical University of Denmark; Ørsteds Plads; Bygning 352, 2800 Kgs. Lyngby}
e-mail: {pasquale.bottalico@polito.it; dpg@elektro.dtu.dk; arianna.astolfi@polito.it; jbr@elektro.dtu.dk}

Abstract

This work shows the results of a preliminary study about the determination of the optimal acoustical conditions for speakers in small classrooms. An experiment was carried out in a laboratory facility with 22 untrained talkers, who read a text passage from “Goldilocks” during two minutes under 13 different acoustical conditions, that combined different kind of background noise and virtual classroom acoustics. Readings from the vocal fold vibrations were registered with an Ambulatory Phonation Monitor device. The speech signal from the talker in the center of the facility was picked up with a head-worn microphone, convolved in real time with the impulse response of the chosen classroom, and reproduced through 29 loudspeakers placed around the subject.

In particular, two different primary school classrooms were selected, with very low and very high reverberation time and, for each of them, two speaker positions. The acoustic of the classrooms was simulated with Odeon 9.0. Environmental noise recorded in quiet classrooms, traffic, and babble noise were considered as sources of disturbing noise during experimentation.

Several acoustical parameters were calculated from the impulse response measured with an artificial head (corresponding to the mouth-ears path) placed at the talker position while simulating the classrooms. Time histories of the vocal fold vibration readings, with the trend of the fundamental frequency and an estimation of the sound pressure level, sampled every 50 ms, were obtained. From these data the vocal doses Time dose, Vocal Loading Index, Distance Dose, Energy Dissipation Dose, and Radiated Energy Dose were calculated and correlated with the acoustical features of the classrooms.

Keywords: vocal dose, voice ergonomics, classroom acoustics.

1 Voice parameters

1.1 Fundamental frequency and sound pressure level

The two vocal parameters analyzed in this paper were the sound pressure level (SPL), and the fundamental frequency (F_0). These parameters were collected with an Ambulatory Phonation Monitor (APM) model 3200 from KayPentax, which consisted of an accelerometer to be attached to the talker's neck and an acquisition device that processed the accelerometer signal to provide F_0 and an estimation of the SPL (anechoic, at 0.15 m on-axis) under phonation.

1.2 Vocal doses

Five different vocal dose measures are used as indicators of vocal effort and of the exposure of the vocal fold tissue to vibrations.

- **Time dose** (Dt). Measures the accumulated time of vocal fold vibration (phonation).
- **Vocal Load Index** (VLI). Measures the accumulated number of vibrations of the vocal folds, taking into account the fundamental frequency and the time they are vibrating.
- **Distance Dose** (Dd). This dose is a measure of the displacement accumulated by the vocal folds due to vibrations.
- **Energy dissipation dose** (De). Measures the total amount of energy dissipated into heat over a unit volume of vocal fold tissues
- **Radiated energy dose** (Dr). Measures the total energy radiated from the mouth during phonation.

For a complete formulation of these doses, see a thorough presentation in [1].

2 Experimental method

2.1 Setup Overview

The experiment was carried out in one of the facilities of the Acoustic Technology Group, at the Technical University of Denmark. This facility was a highly damped room with dimensions 4.7 m x 4.6 m x 3.4 m, where a system for the real-time auralization of a talker's voice was installed. In the absence of such an auralization system, the talker would listen himself only via direct airborne sound and by bone conduction, without listening the own voice reflected at the boundaries. This system generates artificially the reflections of any room (called a *virtual room*, as it does not correspond to the actual laboratory) by picking the talker's voice with a head-worn microphone, convolving it with a multi-channel room impulse response, and reproducing the resulting signals via 29 loudspeakers distributed around the talker on the surface of a quasi-sphere. As a result, the talker has the impression of being speaking in another room. This system has also been used in previous research [2].

Additionally, a noise signal was played back from a single loudspeaker (so as to minimize interference effects) for some of the experimental conditions. A control system decided which impulse responses and noise signals to use. The overall experimental setup is shown in Figure 1. The speech signal picked with the headworn microphone is equalized to obtain an on-axis signal.

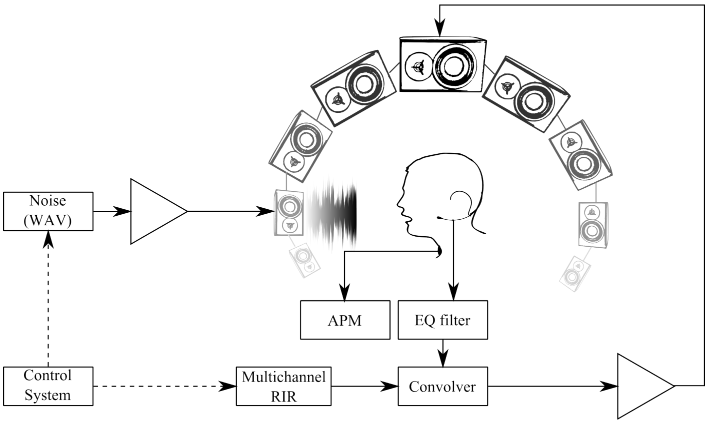

Figure 1 – Experimental setup

2.2 Subjects

A total of 22 subjects, 11 males and 11 females, with ages between 20 and 28 years, and without particular voice training, participated in the experiment. None of the subjects had known problems with their hearing or their voice, according to self-reports. Many different nationalities – and thus different cultures of using voice – were represented.

2.3 Experimental conditions

There were 13 experimental conditions in each test, and none of them was repeated. The conditions were obtained by a combination of an acoustically simulated room, a talking position, and a specific kind of noise. Two rooms, Fontana (a) and (b), were simulated with Odeon [3]. These rooms were school classrooms with a floor area of 46 m^2 and different reverberation times; Fontana (a) had 1.45 s, and Fontana (b) had 0.35 s (average of the octave frequency bands 500 Hz – 1 kHz). In each of the classrooms, two different positions v1 and v2 were chosen, so as to obtain different early sound reflection patterns, thus resulting in four different acoustic environments.

The early reflection patterns and the late reverberation were processed with Favrot and Buchholz's method [4] so as to extract a multichannel room impulse response that could be convolved with the talker's voice to generate the "room effect" by playing it back through the 29 loudspeakers. In each acoustic environment, three different kinds of noise were alternatively added: multitalker or babble noise (recording of children activities in a classroom) with L_{eq}=57 dB, ambient noise with L_{eq}=37-40 dB, and traffic noise with L_{eq}=54 dB. The last condition was designed to be the absence of any simulated acoustic element. Therefore, no reflections were added to the talker's voice and the background noise corresponded to the background noise in the test room.

Table 1 shows a summary of the acoustic parameters of the different experimental conditions. They were obtained by placing a Head and Torso Simulator (HATS) B&K type 4128, with a mouth simulator and microphones at the ears, in the middle of the room, as if it was a human talker, and measuring the impulse response corresponding to the path between the mouth and the ears (including the loudspeaker response). The support is calculated as ten times the logarithm of the ratio between the reflected sound energy (E_r) and the sound energy arriving directly to the ears (E_d) [2].

$$\text{Support (dB)} = 10\log\frac{E_r}{E_d} \qquad (1)$$

Table 1 – Experimental conditions

#	Room	Position	Noise type	Noise level [dB]	Rev. time [s]	Support [dB]
1	Fontana(a)	v1	Babble	57	1.43	-13.6
2	Fontana(a)	v1	Ambient	40	1.43	-13.6
3	Fontana(a)	v1	Traffic	54	1.43	-13.6
4	Fontana(a)	v2	Babble	57	1.47	-15.3
5	Fontana(a)	v2	Ambient	40	1.47	-15.3
6	Fontana(a)	v2	Traffic	54	1.47	-15.3
7	Fontana(b)	v1	Babble	57	0.36	-17.5
8	Fontana(b)	v1	Ambient	37	0.36	-17.5
9	Fontana(b)	v1	Traffic	55	0.36	-17.5
10	Fontana(b)	v2	Babble	57	0.33	-17.8
11	Fontana(b)	v2	Ambient	37	0.33	-17.8
12	Fontana(b)	v2	Traffic	55	0.33	-17.8
13	Void	-	Background	28	0.08	-18.1

2.4 Procedure

Before starting the tests, an accelerometer was attached to the subject's neck, below the glottis. A calibration measurement was performed with the help of a reference microphone, so that it would be possible to estimate the SPL from the accelerometer signal. The head-worn microphone was also attached to the subject, so that the capsule was on the cheek, between the ear and the mouth, at 5 cm from the lips' edge.

The subjects sat at a chair in the middle of the room. They were instructed to read the *Goldilock's passage* [5], so a listener at 1.6 m could hear it (a dummy head was placed at that specific position), despite the room acoustics and the noise that was to be presented. The subjects were asked to start from the beginning of the passage for each condition.

There were a total of 13 conditions, presented in random order for each subject, so that the bias introduced by vocal loading during testing was distributed evenly among the different conditions. For each condition, the reading time was 2 minutes. The start and ending of each presentation was indicated by means of an audio message reproduced through the loudspeakers. Between presentations, subjects had a short break of 30 to 60 seconds approximately. The total experiment time was about 35 minutes on average.

3 Results and analysis

In order to test whether there were significant variations in the different measured vocal doses (Dt, VLI, Dd, De, Dr) across environments and noises, two sets of ANOVAs (ANalysis Of VAriance) were performed, each one for male and female. The first one tested the hypothesis of equal means between the conditions I) Babble and traffic noise and II) Ambient noise, for each of the environments and vocal doses. The *p*-values resulting from the ANOVA tests are shown in Table 2. There were significant variations at 5% level in VLI across noises in all the environments. There were also significant variations in Dd across noises in all cases, except in Fontana (b) for female subjects, where nevertheless there were significant variations at 10% level. In the case of De, there were significant variations at 5% level for male and at 10% level for female. Furthermore there were also significant variations in Dt across noises only for Fontana (a). No significant variations with noises were found for the Dr.

The second set of ANOVAs tested the hypothesis of equal means among the two simulated classrooms, for the different doses and noise typologies. The *p*-values resulting from these ANOVA tests are shown in Table 3, where “no-noise” includes the conditions of background and ambient noise, and “noise” refers to conditions with babble or traffic noise. For male speakers, there were significant variations at the 5% level on all doses, except on Dr, among the different environments, with “no-noise”. Only VLI shows a significant variations at the 10% level in “no-noise” conditions. No significant variations among environments were found in the case of “noise” conditions.

Table 2 – Statistical significance of the variation of measured vocal doses across noise. Numerical values correspond to *p*-values from ANOVAs testing the hypothesis of no variation in means across noise, for each combination of vocal dose and environment, and for both genders.

Gender	**Male**		**Female**	
Environment	Fontana (a)	Fontana (b)	Fontana (a)	Fontana (b)
Dt	0,01	0,39	0,01	0,23
VLI	< 0,01	< 0,01	< 0,01	< 0,01
Dd	< 0,01	< 0,01	0,04	0,08
De	< 0,01	0,02	0,08	0,09
Dr	0,29	0,24	0,15	0,14

Table 3 – Statistical significance of the variation of measured vocal doses across environments. Numerical values correspond to *p*-values from ANOVAs testing the hypothesis of no variation in means across environments, for each combination of vocal dose and noise, and for both genders. “No-noise” contains the conditions of background and ambient noise.

Gender	**Male**		**Female**	
Noise	No-Noise	Noise	No-Noise	Noise
Dt	0,05	0,41	0,13	0,11
VLI	0,02	0,32	0,09	0,97
Dd	0,01	0,50	0,84	0,21
De	0,04	0,33	0,77	0,10
Dr	0,18	0,18	0,15	0,13

The Dt for the different configurations, averaged for 22 subjects, and its standard deviation, are plotted in Figure 2. The Dt accumulated by talkers in “no-noise” conditions was significantly lower than in the other conditions with babble and traffic noise, in good agreement with the low *p*-values obtained in the ANOVA tests. No significant variations were found between traffic and babble noise.

Figure 3 shows the trend of VLI versus the different configurations, averaged for 11 male and 11 female subjects, and its standard deviation. The trend for both genders is the same. The VLI accumulated by talkers in no-noise conditions was significantly lower than in the other conditions with babble and traffic noise, in good agreement with the low p-values obtained in the ANOVA tests. This effect is related with an increase of the fundamental frequency in conditions with a high level of noise.

The Dr, averaged for 22 subjects, and its standard deviation, are shown in Figure 4. High values of this dose present higher standard deviations. This dose represents the radiated energy from the mouth, and the results show a maximum for both environments in the case of babble noise.

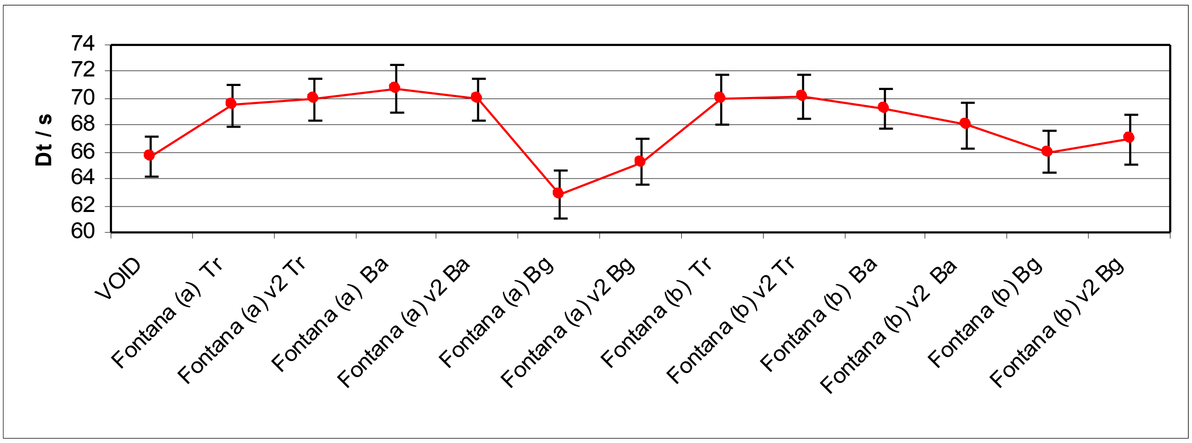

Figure 2 – Dt as a function of the different configurations, averaged for 22 subjects, and its standard deviation.

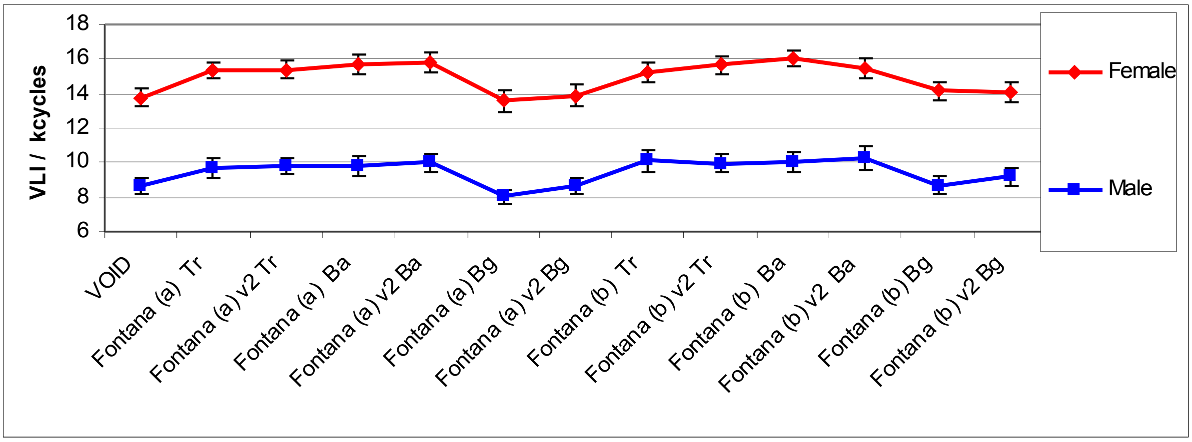

Figure 3 – VLI versus the different configurations, averaged for 11 male and 11 female subjects, and its standard deviation.

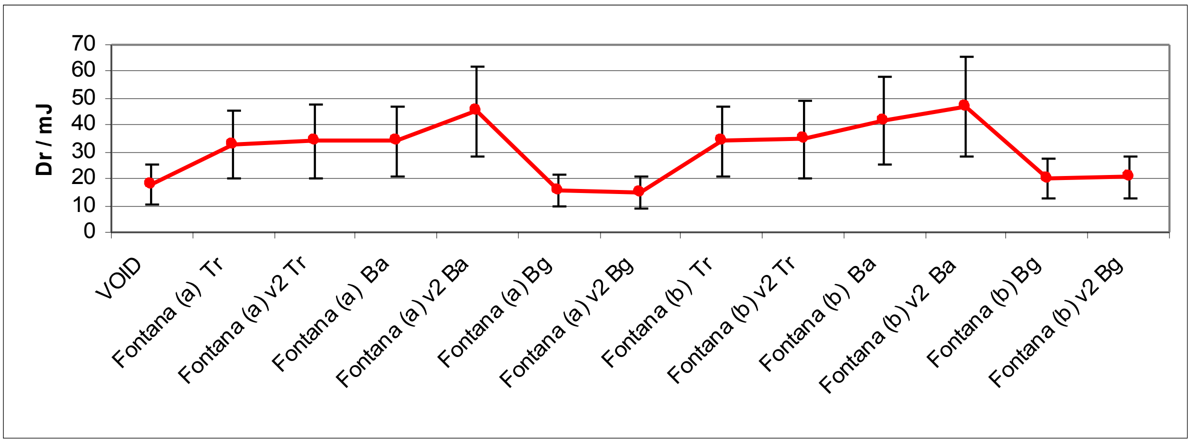

Figure 4 – Dr as a function of the different configurations, averaged for 22 subjects, and its standard deviation.

The values of VLI accumulated by speakers were also considered as a function of the two classrooms, Fontana (a) and (b), in order to investigate the effect of reverberation on this dose. The VLI values were averaged for female and male subjects, across the conditions containing ambient noise. TheVLI scores for both classrooms, along with the standard deviations, are plotted in Figure 5, separately for the results of male and female speakers. In agreement with the *p*-values obtained in the ANOVA tests, there was a significant main effect of the environment on the VLI accumulated by subjects. In particular, a lower reverberation time resulted in a higher VLI, more for male than for female speakers.

The De values were averaged for female and male subjects, across the conditions with ambient noise. The De scores for both classrooms, along with the standard deviations, are plotted in Figure 6, separately for male and female speakers. There was a significant main effect of the environment on the De accumulated by male subjects, but not for females, in good agreement with the *p*-values obtained in the ANOVA tests. This dose represent the power dissipated by the vocal fold during the phonation, and it is inversely related to the reverberation time.

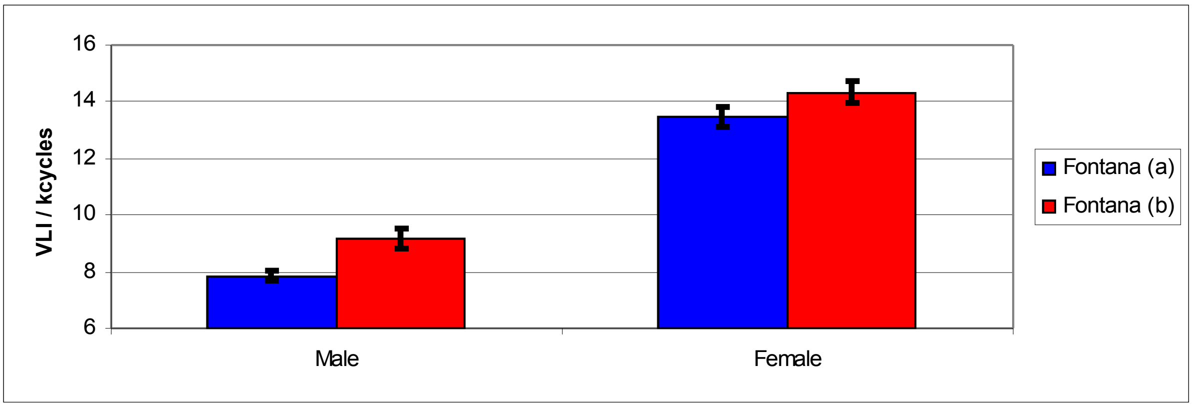

Figure 5 – VLI accumulated by speakers for the two classrooms.

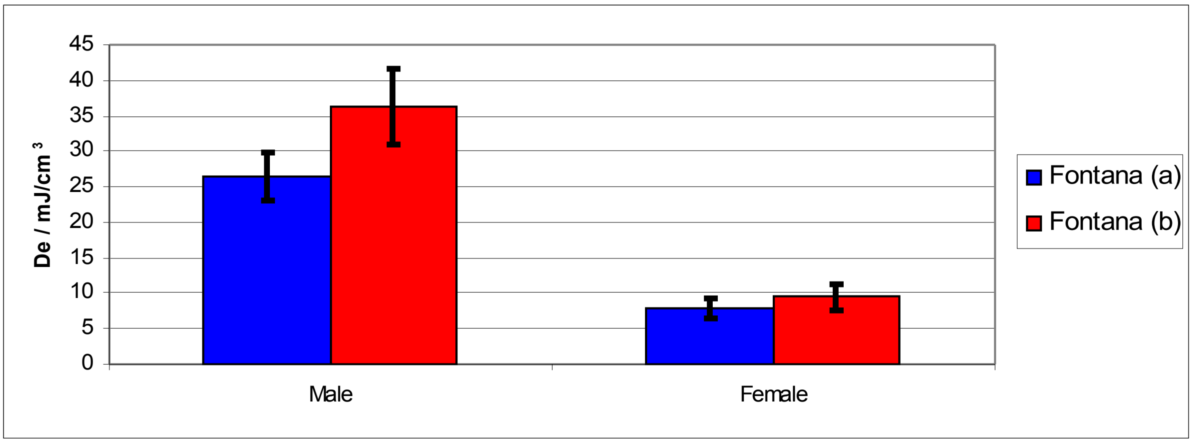

Figure 6 – De accumulated by speakers for the two classrooms.

To investigate possible additional effects of room acoustics on vocal doses, multiple regression analyses of the vocal doses accumulated by the speakers were performed, regarding the values of noise level and the room acoustics parameters (reverberation time and support) as explanatory variables. Table 4 summarizes the results in terms of the resulting R^2 (coefficient of determination) values for the multiple regressions. The R^2 values for the noise level are first given for each dose, where the subdivision between male and female is considered if the dose is a function of the fundamental frequency. If there were significant additional effects of room acoustics, the R^2 value would be expected to increase by adding values of one of the room acoustics parameters to the regression analysis. Adding the support to the prediction resulted in modest but significant increases in the prediction accuracy of the doses Dt, VLI, Dr and De. Only for Dd, a higher R^2 value was achieved by adding the RT to the prediction.

Table 4 – Resulting R^2 values of multiple regression analyses of the vocal doses accumulated by the speakers, regarding the values of noise level L_n and the room acoustics parameters (reverberation time RT and support) as explanatory variables.

R^2	**L_n $(L_n)^2$**	**L_n $(L_n)^2$ RT**	**L_n $(L_n)^2$ Support**
D_t	0.755	0.758	***0.764***
VLI_Male	0.855	0.903	***0.924***
VLI_Female	0.934	0.936	***0.940***
D_d_Male	0.904	***0.920***	0.914
D_d_Female	0.934	***0.979***	0.971
D_e _Male	0.894	0.903	***0.912***
D_e _ Female	0.866	0.904	***0.945***
D_r _Male	0.657	0.667	***0.687***
D_r _Female	0.560	0.633	***0.686***

The equations (2) show that, when the reflected sound energy arriving to the ears becomes fainter, in comparison with the direct sound energy, the VLI accumulated (in kcycles) by speakers increases. Figure 7 illustrates the resulting multiple regression equations (2) for combinations of noise level and three levels of Support. For example, a decrease of 3 dB in support leads to an average increase of 350 oscillatory periods of the vocal folds, for male speakers.

$$\begin{aligned} VLI\ Female &= 0{,}003 \cdot L_n^2 - 0{,}18 \cdot L_n - 0{,}04 \cdot Support + 15{,}89 \\ VLI\ Male &= 0{,}003 \cdot L_n^2 - 0{,}12 \cdot L_n - 0{,}11 \cdot Support + 8{,}32 \end{aligned} \tag{2}$$

In the case of Dd, RT values tended to be slightly more effective than support in increasing the R^2 values. The equations (3) show that, an increase of 0.5 s in RT generates a decrease of 1 m in the Dd accumulated by a female and of 0.7 m by a male speaker. Figure 8 illustrates the resulting multiple regression equations (3) for combinations of noise level and four values of RT.

$$\begin{aligned} D_d\ Female &= 0{,}009 \cdot L_n^2 - 0{,}25 \cdot L_n - 2{,}35 \cdot RT + 24{,}80 \\ D_d\ Male &= 0{,}027 \cdot L_n^2 - 1{,}85 \cdot L_n - 1{,}32 \cdot RT + 58{,}59 \end{aligned} \tag{3}$$

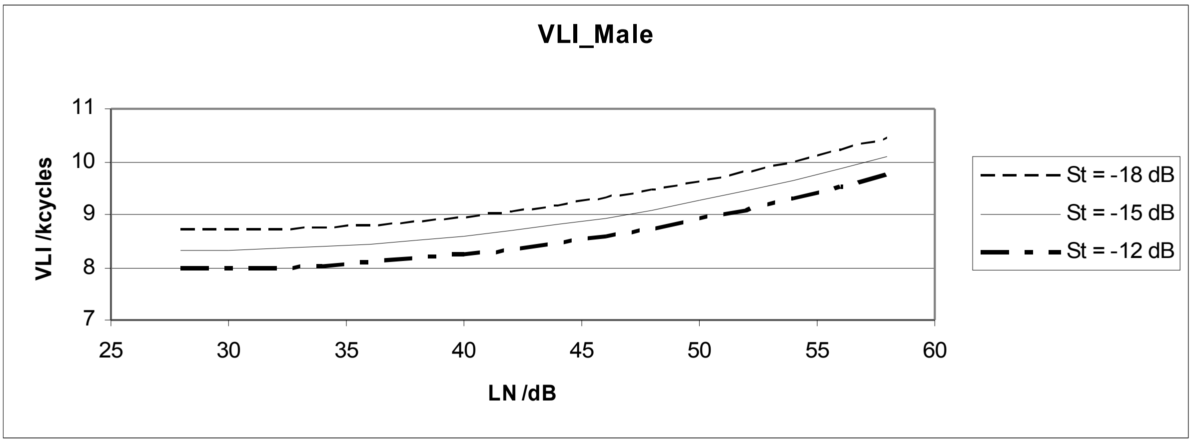

Figure 7 – Multiple regression equations for combinations of noise level and Support, for VLI.

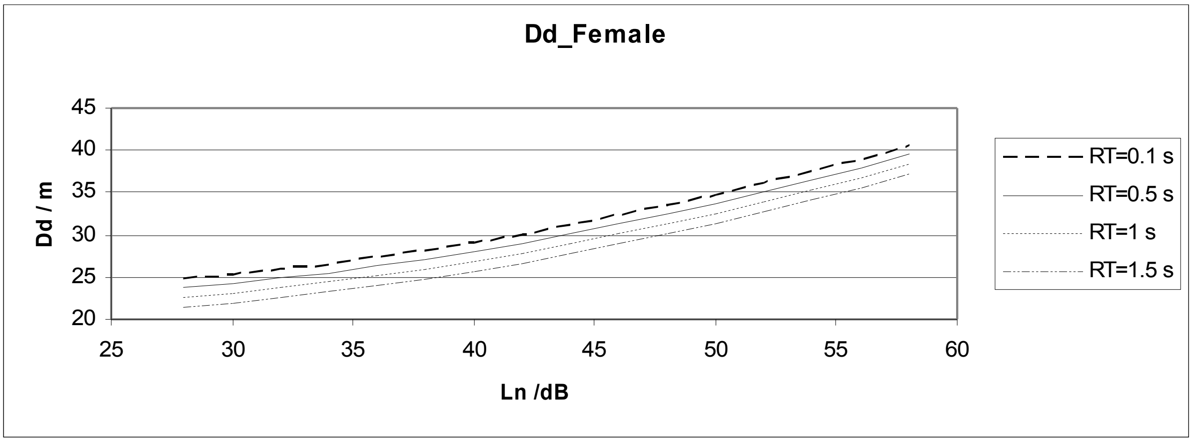

Figure 8 – Multiple regression equations for combinations of noise level and RT, for Dd.

4 Conclusions

From the present paper the following conclusions can be drawn:

- All the analyzed vocal doses are mostly influenced by the noise level, and only in conditions with a low noise level, it is possible to observe the effects of different room acoustic conditions on vocal doses.
- The VLI accumulated by speakers in each environment is significantly higher in noise conditions, this effect means that each subject increase his fundamental frequency with the noise level, as a side-effect of a voice level increase, in good agreement with the Lombard reflex.
- In order to quantify the increase in F_0 due to noise level, a Dt equal to 55% and a noise level of 50 dB are considered. A Dt equal to 55% represents, on average, the phonation time in percentage for a sample of continuous speech, and 50 dB is the typical noise level in primary classrooms [6]. A variation of +5 dB on noise level leads to an increase of 16 Hz in the fundamental frequency, averaged over female and male speakers. In the same conditions, a decrease of 6 dB on support leads to an

increase of 8 Hz in the fundamental frequency, averaged over female and male speakers.

- A higher RT results in a lower De, that represents the power dissipated by the vocal folds during phonation.
- Dr values show a maximum for both classrooms in the case of babble noise.
- An increase of 0.5 s in RT generates a decrease of 1 m in the Dd accumulated by a female and of 0.7 m by a male speaker.

References

[1] Titze, I.R.; Švec, J.G.; Popolo, P.S. Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues, *Journal of Speech, Language and Hearing Research*, Vol. 46, 2003, pp. 919–932

[2] Pelegrín-García, D.; Brunskog, J. Prediction of vocal effort and speakers' comfort in lecture rooms. *The 38th International Congress and Exposition on Noise Control Engineering (Internoise 2009)*, Ottawa, Canada, August 23-26, 2009 (only electronical version available)

[3] Christensen, C.L. *Odeon Room Acoustics Program, Version 9.1, User Manual, Industrial, Auditorium and Combined Editions*, Denmark, 2008

[4] Favrot, S.; Buchholz, J.M. LoRA - A loudspeaker-based room auralisation system, Acta Acustica united with Acustica, 2010, in press

[5] Švec, J.G.; Popolo, P.S.; Titze, I.R. The Goldilocks passage and scripts for frequency extraction, voicing detection, SPL calculation and vocal dose determination in speech, The National Center for Voice and Speech Online Technical Memo, No. 1, April 2003, version 1.4

[6] Picard, M.; Bradley, J. S. Revisiting Speech Interference in Classrooms, *Audiology* 40, 2001, pp.221-244.

Speaking comfort and voice use of teachers in classrooms

Jonas Brunskog and David Pelegrin Garcia

DTU Electrical Engineering; Technical University of Denmark; Ørsteds Plads; Bygning 352, 2800 Kgs.
Lyngby. e-mail: jbr@elektro.dtu.dk, dpg@elektro.dtu.dk;

Abstract

Teachers suffer from voice problems more often than the rest of the population, as a consequence of the intensive use of their voices during teaching. Noise and classroom acoustics have been defined as hazards eventually leading to voice problems. In order to make a good classroom acoustic design to preserve the teachers' voices and maximize their comfort, it is necessary to understand the underlaying relationship between classroom acoustics and teachers' voice production. This paper presents a brief summary of investigations looking into this relationship. A pilot study, carried out in different rooms, measured significant changes in the voice power levels as a function of the objective quantities volume and support or room gain, which measure the degree of amplification of the own voice in a room. A virtual acoustics system was designed in order to recreate different acoustic conditions in laboratory. Different experiments using this system showed that the talkers' voice levels decreased linearly with the support of the simulated room. Another experiment aimed to report the relative importance of visual and auditory cues by measuring the voice levels used by talkers to address a listener located at various distances, in rooms with very different acoustics. A field study in schools of southern Sweden found out that teachers with and without voice problems, during actual teaching, are affected differently by the support of the classroom. A last laboratory experiment was carried out to introduce the use of vocal doses for the investigation of voice production under different classroom acoustics and noise conditions, finding that the relative effect of noise on voice production is more important than the effect of room acoustics.

Keywords: voice ergonomics, classroom acoustics, support, auralization.

1 Introduction

The most common means of communication in a classroom is speaking and listening. The teacher's voice is thus the tool for communicating with the students. The room acoustics in the classroom is the communication channel from the speaker to the listener. It affects the quality of the speech signal and thus the ability to understand what the teacher says. There have been many studies trying to optimize the acoustical conditions for the students, based on measures of intelligibility of speech, signal-to-noise ratios or optimum reverberation time (e. g. Bistafa and Bradley, 2000, Sato and Bradley, 2008 and Yang & Bradley 2009). Most of the studies have focused on the listener's point of view, but Sato and Bradley, (2008) points out that a low reverberation may affect teachers' voice. Pekkarinen, & Viljanen, 1991, found that the sound environment in the classrooms often were poor for optimum audibility, and also notes that the teacher increases the voice effort in high background noise (the Lombard effect). Hodgson (1999) provides a simple empirical prediction model to calculate average voice levels used by teachers in university lecture rooms, depending on individual factors, acoustical characteristics of the room and student activity noise. Kob et al (2008) found that teachers with voice disorders are more affected by unfavorable classroom acoustics than their healthy colleagues.

The pre-study of the present project, Brunskog et al. (2009), studies the classroom acoustics from the point of view of the speaker, and therefore tries to relate the voice production process to different measurable parameters of the classroom, including the size of the room, acoustical parameters and background noise. It is shown that the voice power used is related to the volume of the room and to the support or room gain provided in the position of the speaker. The results also indicate that the auditory cues may be more important than visual cues in this voice regulation.

In the field of voice therapy and phoniatrics the teachers’ voice health problems is of major concern, due to the required clinical assistance, but also due to the financial impact that the teachers’ absence produces in the global budget of the country, Verdolini and Ramig (2001). There seems to be a consensus that voice load is an important factor for voice problems (see e.g. Vintturi et al. 2003). Laukkanen et al. (2008) identifies higher fundamental frequency (F0) and higher sound pressure level (related to the voice power) as signs of increased muscle tone, a consequence of adaptation to the load after a day of work. They also note that the higher F0 correlates well to the increased feeling of tiredness in the throat.

This paper present some experimental results from an ongoing project on teacher’s use of there voice in different rooms. An overall aim of this project is to investigate the teacher's voice use and voice effects in relation to the room acoustic conditions, and to study whether the speakers take into account auditory cues in order to regulate the voice levels they use to speak, even in the absence of background noise.

The project, financed by *AFA försäkringar* in Sweden, is done in close cooperation between the Acoustic Technology group at the Department of Electrical Engineering, at the Technical University of Denmark (DTU), which is responsible for the technical-acoustic experiments and analyzing them, and the Voice Research Group at the Department of Logopedics, Phoniatrics and Audiology, Lund University (LU), Sweden. The study has been assessed and approved by the Regional Ethics Committee in Lund. This paper mainly present results form the room acoustic part of the project. For recent findings in the voice therapist part, we refer to Lyberg Åhlander et al (2010).

2 Method overview

The experiments reported here are mainly laboratory experiments, using real or virtual environments. A virtual acoustic system has been developed for this project. Moreover, new or developed metrics describing the room acoustic conditions for a talker are introduced.

2.1 Development of the virtual acoustic system

A real-time self-voice auralization system has been developed. The room, called SpaceLab, consists of 29 loudspeakers placed in a quasi-sphere around a subject in a highly damped room, Favrot et al (2010a), Favort (2010b)), and Pelegrín-Garcia et al (2009a). 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. As a result, the talker has the impression of being speaking in another room.

A block diagram of the system is shown in Figure 1. Here, the IR (stored in 29 WAV files, one for each loudspeaker) is loaded into the convolution software *jconvolver*. This requires the computer modeling of the desired room and the calculation of the different transmission paths with a room acoustics simulation software (Odeon). The output of Odeon is decoded and encoded in Ambisonics, adjusted to the requirements of the system. An equalizer filter is used to correct the biased spectral distribution of the picked signal.

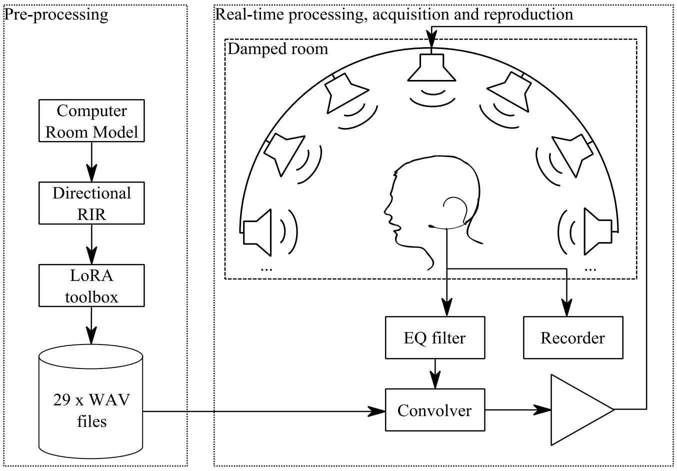

Figure 1: Block diagram of the virtual acoustic system.

2.2 Room acoustic parameters for a talker

The room acoustic parameters for a talker are related to the possible ways in which his own voice reaches his ears. They require the measurement of the airborne acoustic path between the mouth and the ears, which is characterized by an impulse response (IR) $h(t)$. This airborne path has two components: the direct sound, transmitted directly from the mouth to the ears, and the indirect sound, coming from reflections at the boundaries. For this reason, the last component is also referred to as reflected sound. Two parameters are derived from the IR measurement and the relation between the direct and the reflected sound: *room gain* and *support*. The background of the support measure comes from musical room acoustics, where the concept is used in connection to the stage, and is related to the possibility for the musicians to hear themselves when playing, Gade (1989a, 1989b).

The IR has to be measured with a dummy head that contains a loudspeaker at its mouth, used as source, and microphones at its ears, used as receivers. So as to ensure a correct separation of the direct and the reflected sound components, it is necessary to place the dummy head at more than 1 m from reflecting or scattering surfaces, measured from the center of the mouth. In this way, there is a time gap free of reflections after the arrival of the direct sound. Then, the direct sound component is extracted by applying a window $w_d(t)$, of 5 ms duration, to $h(t)$. The complementary window, $w_r(t) = 1 - w_d(t)$ is applied to $h(t)$ in order to extract the reflected component arriving to the ears. An illustration of the signal and the windows is shown in Figure 2.

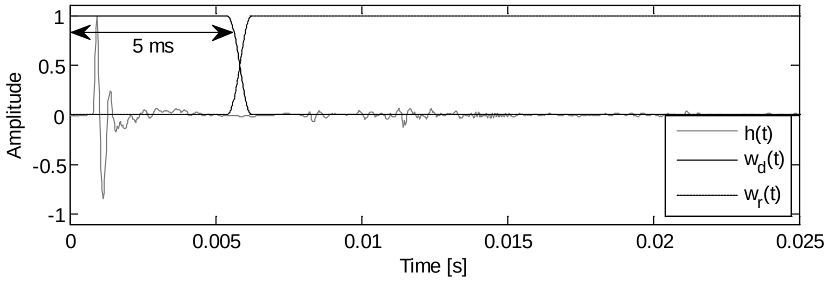

Figure 2: Example of an IR and windowing applied to extract direct and reflected components.

The windowed signals $h(t)w_d(t)$ and $h(t)w_r(t)$ can be filtered using one-octave bandpass filters with center frequencies between 125 Hz and 4 kHz so as to study the importance of directed and reflected sound in the octave bands of interest in room acoustics. These bandpass filters are here generically called $h_f(t)$. 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.)

The room gain G_{RG} was defined by Brunskog et al. (2009) as the degree of amplification provided by the room to one's own voice, disregarding the contribution of the own voice which is transmitted directly through the body. This is the difference between the total energy level in a room and the direct energy level. However, Brunskog et al. calculate the direct energy level with an IR measurement in an anechoic environment.

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

As in Pelegrin-Garcia (2009b), the *support* is chosen as an alternative measure for the degree of amplification of a room to one's own voice. In this case, the *support* compares the energy level of the reflections $L_{E,ref}$ with the energy level of the direct sound, extracted from the impulse response corresponding to the path between the mouth and the ears.

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

The *support* ranges from -18 dB to -5 dB in normal rooms, whereas the room gain is limited to a range between 0 dB and less than 2 dB.

3 Results

3.1 The pilot study

The pilot study for this project, Brunskog et al. (2009) showed correlation between the physical characteristics of the rooms and voice power, and with perceived quality, such that the room is perceived good or bad to talk in. It was show that the parameters in the room that primarily affect the voice power are the size of the room and the support for the speaker (or room gain) provided by the room. The results also indicate that the auditory cues may be more important than visual cues in this voice regulation: the measured changes in the voice power is correlated with the logarithm of the volume (which means a compensation for changes in average noise level outside the room) and not equally well with the cube root of volume (which estimates the mean distance to the audience, that is a visual reference).

3.2 Pre-experiment in the virtual environment

This experiment aimed to investigate the voice used by a teacher to address a group of imaginary students under different simulated acoustics, Pelegrín-Garcia and Brunskog (2009b). Five subjects, aging 23-35 with normal hearing and voice status, talked freely in 5 different simulated acoustic environments during 3 minutes in each of them. The goal was to give a lecture of a familiar topic to a group of 30 students located in front of them. In addition, they had to answer a small questionnaire after speaking in each simulated room. (A more thorough study can be found in Pelegrin-Garcia et al. (2010a))

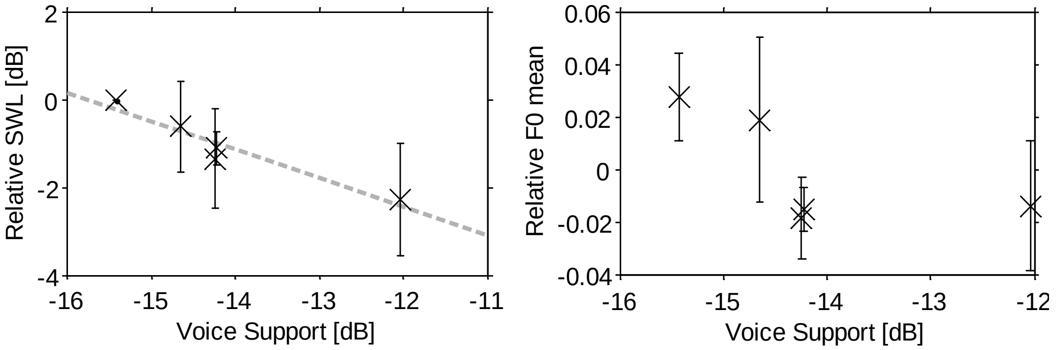

Figure 3: Left: Voice power level versus support and regression line. Right: Fundamental frequency variations versus voice support.

The results in Figure 3 show a significant linear dependence (R^2=0.92) between the changes in voice power level used by the speaker and the support provided by the room to the talker's voice, with a slope of -0.65 dB/dB, although the absolute mean variations were between 2 and 3 dB. The fundamental frequency used by the talkers changed significantly between environments, although it did not follow a linear trend.

3.3 Modified gain

The goal of this experiment was to measure the vocal output when the support was changed, keeping the reverberation time fixed. Thus, the different stimuli did not correspond to actual simulated rooms, but to a single impulse response with different gains. Five teachers talked freely in 10 different simulated acoustic conditions during 3 minutes in each of them. The goal was to give a lecture of a familiar topic to a group of 30 students located in front of them. In addition, they had to answer a small questionnaire after speaking in each simulated room.

The measured variations in voice power level used by subjects are shown in Figure 4. The trend of the voice power level, indicated by the dashed red line, lays very close to the voice power level measured in the pre-experiment (green line). The slope of the line is in this case -0.58 dB/dB. This indicates that the experiment is fairly repeatable, and that the acoustic environment can systematically change the vocal behavior.

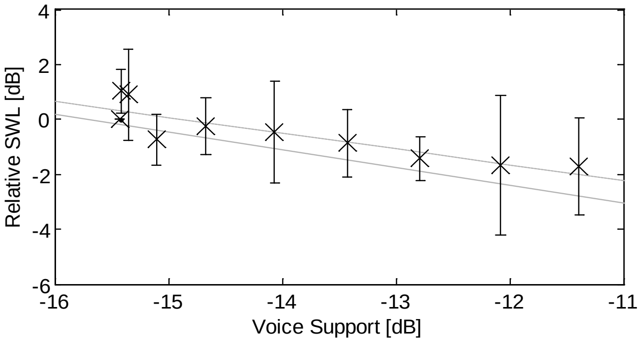

Figure 4: Measured relative voice power level versus support. Dashed red: regression line. Green: regression line of the pre-experiment.

3.4 Distance to talker

In this subproject the voice used by talkers where analyzed when they had to address a listener at different distances under different acoustic conditions in real rooms. Six talkers, aging 21 to 30, had to talk freely to a listener about a familiar topic during 2 minutes. The listener moved alternately at positions located at 1.5 m, 3 m, 6 m and 12 m away from the talker. The voice of the speaker was recorded from a headworn microphone. This experiment was repeated in four rooms: an anechoic chamber, a reverberation room, a long narrow corridor and a big lecture room.

The measurements show that speakers raise their vocal power when the distance to the listener increases, at a rate of 1.5~2.0 dB per double distance (see Figure 5, left). The voice power level produced in the anechoic room differed significantly from the other rooms.

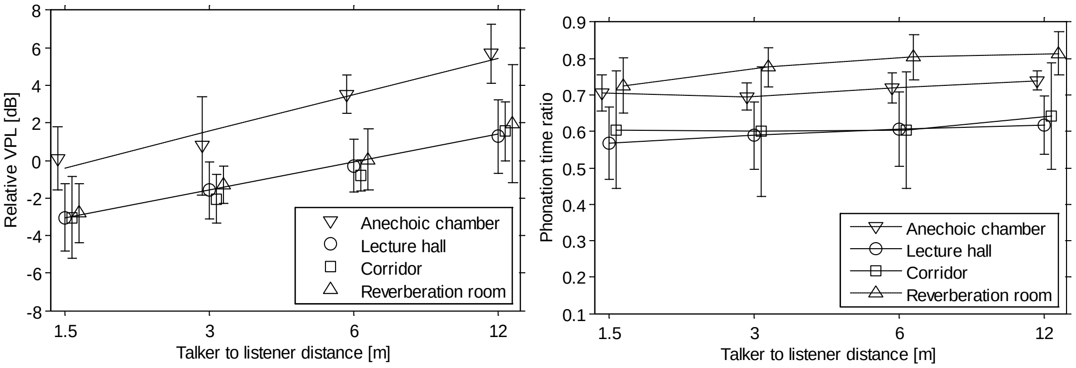

Figure 5: Left: Variations in Voice Power Level versus distance. Right: Phonation time ratio versus distance.

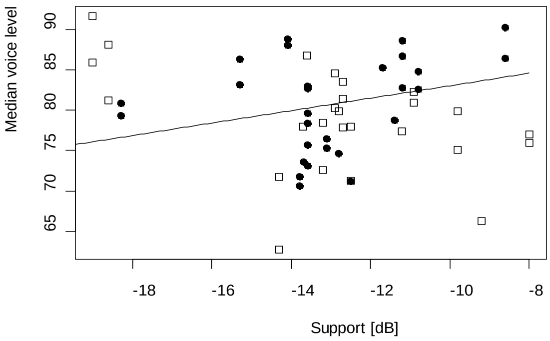

Figure 6: Comparison of the model and the measured values. Left: Median voice level vs Support. Right: Median voice level vs. Median noise level.

Furthermore, the subjects expressed their preference about vocal comfort, stating that the least comfortable environments were the anechoic room and the reverberation room. While the analysis of the voice levels cannot account for this preference, other parameters might be better suited. The phonation time ratio (ratio between duration of voiced segments and total duration of running speech) might be appropriated for this purpose, as it can be seen from Figure 5, right. The subjects produce longer vowels in the anechoic room and the reverberation room, compared to the two other rooms, either to overcome the poorer speech intelligibility at the listener location (in the reverberation room) or due to the raised voice levels (in the anechoic room).

3.5 Field study

The field study examines how classroom acoustics interacts with the voices of 14 teachers without voice problems and 13 teachers with voice problems, Pelegrin-Garcia et al. (2010b). The assessment of the voice problems was made with a questionnaire and a laryngological examination. During teaching, the sound pressure level at the teacher's position was monitored. The teacher's voice level and the activity noise level were separated using mixed Gaussians. In addition, objective acoustic parameters of *Reverberation Time* and *Room Support* were measured in the 30 empty classrooms of the study. An empirical model shows that the measured voice levels (see Figure 6) depend on the activity noise levels and the *Support*. Teachers with and without voice problems were differently affected by the *Support* of the classroom. The results thus suggest that teachers with voice problems are more aware of classroom acoustic conditions than their healthy colleagues and make use of the more supportive rooms to lower their voice levels. This behavior may result from an adaptation process of the teachers with voice problems to preserve their voices.

3.6 Experiments including noise

Another experiment was carried out at DTU in collaboration with the Politecnico di Torino. The goal was to measure vocal doses of speakers under different conditions of room acoustics and noise, Bottalico et al (2010). Vocal doses are a set of measures derived from an estimation of the SPL and the fundamental frequency used by a talker during phonation, Titze et al. (2003). They are measured with an accelerometer attached to the talker's neck and an Ambulatory Phonation Monitor (APM).

Table 1: Experimental conditions, including noise

#	Noise type	L_N [dB]	T_{30} [s]	ST [dB]	#	Noise type	L_N [dB]	T_{30} [s]	ST [dB]
C1	Background	28	0.08	-18.1	C8	Traffic	55	0.36	-17.5
C2	Ambient	40	1.43	-13.6	C9	Traffic	55	0.33	-17.8
C3	Ambient	40	1.47	-15.3	C10	Babble	57	1.43	-13.6
C4	Ambient	37	0.36	-17.5	C11	Babble	57	1.47	-15.3
C5	Ambient	37	0.33	-17.8	C12	Babble	57	0.36	-17.5
C6	Traffic	54	1.43	-13.6	C13	Babble	57	0.33	-17.8
C7	Traffic	54	1.47	-15.3					

In the SpaceLab, 22 untrained talkers (11 males, 11 females), without self-reported known problems with their hearing or their voice, had to read aloud a text passage from "Goldilocks", Svec et al. (2003), during two minutes under 13 different acoustical conditions. These conditions combined different kinds of background noise (traffic, ventilation, or babble noise), at levels ranging from 37 dB to 57 dB, and different room impulse responses, obtained by simulation of medium-sized classrooms with T_{30} in the range between 0.33 s to 1.47 s and ST in the range from -17.8 dB to -13.6 dB. The conditions C1 to C13 are summarized in Table 1. C1 is the condition in which no noise is played back and no impulse response is simulated.

The most remarkable differences among conditions were observed in the vocal dose VLI (Vocal Loading Index). The results are shown in Figure 7. There were significant differences in VLI between the conditions with low L_N (C1 to C5) and the conditions with higher L_N (C6 to C13). Only when the background noise is sufficiently low ($L_N < 40$ dB), there is an effect of different values of ST on the VLI. In this situation, conditions with high ST values (C2 and C3) result in lower Vocal Loading than in conditions with low ST (C4 and C5).

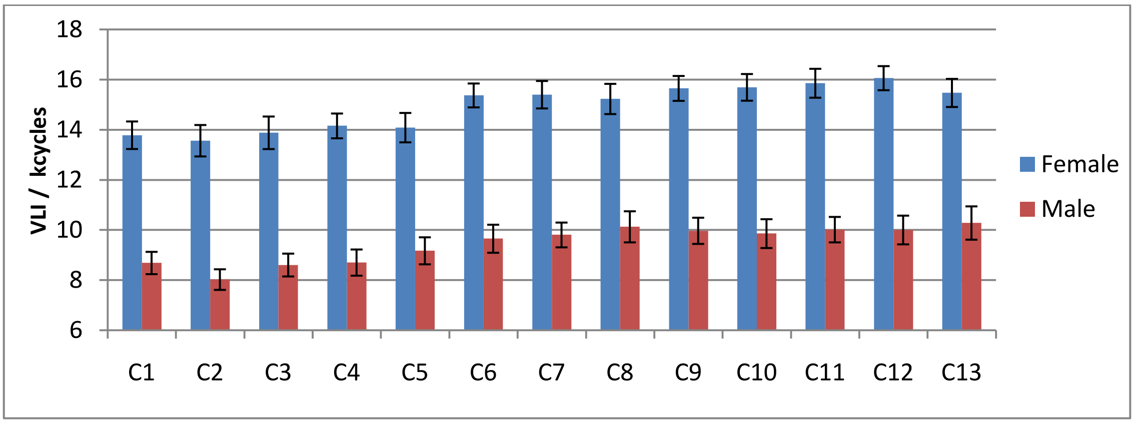

Figure 7: VLI versus the different configurations, averaged for 11 male and 11 female subjects, and its standard deviation.

4 Conclusions

From the present paper the following conclusions can be drawn:

- High background noise levels induce an increase in vocal effort.
- The visually perceived distance between teacher and student accounts to a great extent for changes in vocal effort.
- The room acoustic conditions themselves have an effect on voice production. This is obvious when the talker is aware of the acoustic environment, in the following cases: there are no other cues (like visual), the teacher has either week voice or voice problems, or the acoustic environment is unusual (e.g. anechoic condition).
- New acoustic measures, namely the support and the room gain, are well correlated with the changes in voice level among different rooms.

References

Bistafa S. and Bradley, J. (2000), Reverbation time and maximum backgroundnoise level for classrooms from a comparative study of speech intelligibilty metrics, J. Acoust. Soc. Am. , 107, , 861-875.

Bottalico, P., Pelegrín-Garcia, D., Astolfi, A., and Brunskog, J. (2010), Measurement of vocal doses in virtual classrooms. Proceedings of Internoise 2010, Lisbon, Portugal

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

Favrot, S.; Buchholz, J.M. (2010a), LoRA - A loudspeaker-based room auralisation system, Acta Acustica united with Acustica, 96, 364-375.

Favrot, S, (2010b), A loudspeaker-based room auralization system for auditory research, PhD thesis, Technical University of Denmark.

Gade, A. C., (1989a), Investigations of musicians' room acoustic condition in concert halls. part I: Method and laboratory experiments, Acustica 69, 193–203.

Gade, A. C., (1989b), Investigations of musicians' room acoustic condition in concert halls. part II: Field experiments and synthesis of results, Acustica 69, 249–262.

Hodgson, M., (1999), Experimental investigation of the acoustical characteristics of university classrooms, J. Acoust. Soc. Am., 106, 1810-1819.

Kob, M., Behler, G., Kamprolf, A., Goldschmidt, O., and Neuschaefer-Rube, C., (2008), Experimental investigations of the influence of room acoustics on the teachers voice, Acoustical Science and Technology 29, 86–94.

Laukkanen, AM, Ilomaki, I., Leppanen, K., & Vilkman, E. (2008), Acoustic measures and self-reports of vocal fatigue by female teachers, J. Voice , 22, 283-289.

Lyberg-Åhlander, V., Rydell, R. and Löfqvist, A., (2010) Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff, J. Voice (in press).

Pekkarinen, E. and Viljanen, V. (1991), Acoustic conditions for speech-communication in classrooms, Scandinavian Audiology, 20 (4), 257-263

Pelegrín-Garcia, D, Brunskog, J., (2009a). Development of an auditory virtual environment to measure the speakers' comfort and increase of voice levels in lecture rooms. Proceedings of the First Nordic Conference of Voice Ergonomics and Treatment

Pelegrín-García, D.; Brunskog, J. (2009b) Prediction of vocal effort and speakers' comfort in lecture rooms, Proceedings of Internoise 2009, Ottawa, Canada.

Pelegrín-Garcia, D. and Brunskog, J. (2010a), Natural variations of vocal effort and comfort in simulated environments. Proceedings of EAA Euroregio 2010, Ljubljana, Slovenia

Pelegrín-Garcia, D., Lyberg-Åhlander, V., Rydell, R., Löfqvist, A., and Brunskog, J. (2010b), Influence of classroom acoustics on the voice levels of teachers with and without voice problems: a field study. Proceedings of the 2nd Pan-American/Iberian Meeting on Acoustics and 160th ASA meeting, Cancun, Mexico.

Verdolini, K. and O'Ramig, L. (2001), Review: Occupational risks for voice problems, Logopedics Phoniatrics Vocology, 26(1), 37.

Vintturi, J., Alku, P., Sala, E., Sihvo, M., & Vilkman, E. (2003), Loading-related subjective symptoms during a vocal loading test with special reference to gender and some ergonomic factors, Folia Phoniatrica et Logopaedica , 55, 55-69.

Sato, H. and Bradley, J. (2008), Evaluation of acoustical conditions for speech communication in working elementary school classrooms. J. Acoust. Soc. Am., 123(4), 2064–2077.

Švec, J.G., Popolo, P.S., Titze, I.R. (2003), The Goldilocks passage and scripts for frequency extraction, voicing detection, SPL calculation and vocal dose determination in speech, The National Center for Voice and Speech Online Technical Memo, No. 1, April 2003, version 1.4.

Titze, I.R., Švec, J.G., Popolo, P.S., (2003), Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues, J. Speech, Language and Hearing Research, 46, 919–932.

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

Equal autophonic level curves under different room acoustics conditions

David Pelegrín-García,[a] Oier Fuentes-Mendizábal, Jonas Brunskog, and Cheol-Ho Jeong
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark

(Dated: February 11, 2011)

The indirect auditory feedback from one's own voice arises from sound reflections at the room boundaries or from sound reinforcement systems. The relative variations of indirect auditory feedback are quantified through the room acoustic parameters room gain and voice support, rather than with the reverberation time. Fourteen subjects matched the loudness level of their own voice (the autophonic level) to that of a constant and external reference sound, under different synthesized room acoustics conditions. The matching voice levels are used to build a set of equal autophonic level curves. These curves give an indication of the amount of variation in voice level induced by the acoustic environment as a consequence of the sidetone compensation or Lombard effect. In the range of typical rooms for speech, the variations in overall voice level that result in a constant autophonic level are of the order of 2.3 dB, and up to 3.4 dB in the 4 kHz band. By comparison of these curves with previous studies, it is shown that talkers use other cues than loudness to adjust their voices when speaking in different rooms.

PACS numbers: 43.55.Hy, 43.70.Mn

I. INTRODUCTION

The sound that a talker perceives from his own voice—auditory feedback or sidetone—is constituted by two main components: direct and indirect auditory feedback. The direct auditory feedback can be separated into two other components: airborne sound and bone-conducted sound. These two last components are of the same order of magnitude[1,2] and are always present for building up the sound of the talker's own voice, as long as the acoustic path between the mouth and the ears is undisturbed and the talker has normal hearing. However, the bone-conducted component is not constant in level and frequency distribution, but varies with different vocalizations.[3] The indirect auditory feedback is essentially airborne and is generated by the reflections of the talker's own voice at the room boundaries, or by a sound reinforcement system when it is used to amplify the talker's voice.

The loudness with which talkers perceive their own voice is called the autophonic rating.[4] The autophonic rating grows at almost twice the rate of the loudness of external sounds, meaning that the change in voice level (in dB) required to double the autophonic rating is half of the amount required for external sounds in order to double the loudness sensation. The differences between the autophonic scale and the loudness (sone) scale are most likely due to the different sensing mechanisms in hearing one's own voice and external sounds. The sensation for external sounds is essentially auditory, whereas for one's own voice, it is also dependent on tactile, proprioceptive, and internal mechanisms.[5]

According to Lane and Tranel,[6] speakers adjust their voices to maintain a speech-to-noise ratio suitable for communication. Some factors affecting the speech-to-noise ratio are linked to the auditory perception, such as noise or alterations in sidetone. Other factors are not linked to the auditory perception, but have a clear influence on the voice levels used, as for example, the distance between the talker and the listener.[7,8]

The variation in voice level due to the presence of noise is known as the Lombard effect (see a review in Lane and Tranel[6]). Lane *et al.*[9] showed that talkers accounted for variations of ambient noise level by varying their voice level at a rate of 0.5 dB/dB (voice/noise). In the same study, Lane *et al.* found an equivalent rate for the so-called sidetone compensation: talkers lowered their voice by 0.5 dB for each additional dB of gain applied to the sidetone, while talking over an interphone. The variations of sidetone can also be due to a temporary hearing loss; Black found a compensation rate of 0.57 dB/dB HL.[10]

In the previous cases, the sidetone was altered by damping the direct auditory feedback, or by reproducing an amplified replica of one's own voice through a monitoring device which had the effect of a single sound reflection with a level high enough to mask the direct auditory feedback components. In rooms, the sidetone is altered in a substantially different way, because the indirect auditory feedback is built up by a number of reflections arriving at different delays, with different amplitudes, and spectral weightings. These reflections may interact with the direct auditory feedback in a different way from a single delay. There are two room acoustic parameters to measure the sidetone variations as caused by a room. The *voice support* (ST_V) is defined as the energy ratio of the indirect (E_I) to the airborne-direct (E_D) auditory feedback.[11] The *room gain* (G_{RG}) is defined as the ratio of the total airborne auditory feedback ($E_I + E_D$) to the airborne-direct auditory feedback,[12]

$$ST_V = 10 \log \frac{E_I}{E_D}, \tag{1}$$

---

[a] Electronic address: `dpg@elektro.dtu.dk`

$$G_{RG} = 10 \log \frac{E_I + E_D}{E_D}. \qquad (2)$$

Some studies have shown an effect of room acoustics on the voice levels. Speakers talk louder in highly damped rooms than they do in more acoustically "live" rooms.[13] Brunskog *et al.* found that the changes in voice level of talkers in classrooms were related to the acoustic parameter room gain at a rate of -13.5 dB/dB.[11,12] The changes in voice level were partially due to the distance between teacher and students, and when the distance factor is removed, the room gain has an effect on voice level of about -3.6 dB/dB.[8] These substantially different rates of change, compared with the sidetone compensation of -0.5 dB/dB, could be due to a contribution of the indirect auditory feedback to the autophonic level different from the contribution from the amplification devices used in previous research on sidetone compensation.

Pick *et al.*[14] experimentally proved that the Lombard effect is systematically present, so is difficult to inhibit. Therefore, variations in background noise, sidetone, or hearing loss are expected to induce similar changes in voice levels. It is of particular interest to apply this knowledge to the teaching situation. Teachers have to use their voice as their primary working tool.[15] The prevalence of voice problems among teachers is much higher than in the rest of the population,[16] around a 13% of them have voice problems,[17] and they have to take absence leave, which is both a social and financial problem. In Poland, voice disorders related to excessive vocal load at work (e.g. for teachers, actors, or singers) are classified as an occupational disease.[18] If the acoustic conditions can effectively induce relevant changes in the voice levels used, occupational health and safety organizations should take actions in supporting and funding initiatives that improve classroom acoustics from the talker's point of view, while granting optimal listening conditions for the students in terms of speech intelligibility.

No previous research that the authors are aware of has related in a quantitative way the room acoustics conditions to sidetone variations and alterations in autophonic level. The present paper investigates the extent to which room acoustics can alter the autophonic level and induce Lombard effect-related changes in voice, by determining the equal autophonic level curves. These are defined as the relative voice levels that keep a constant autophonic level under different room acoustic conditions.

II. METHOD

Fourteen subjects (ten men and four women) with ages between 20 and 30 yr, without any known problems with hearing or voice and without previous instruction in vocal training, took part in the experiment. A reference sound at a constant sound pressure level (SPL) was presented, and the test subjects were asked to produce a vocalization (either /a/, /i/, or /u/) with the same loudness as the reference. Each subject produced a total of 60 vocalizations that were stored and analyzed to extract the results.

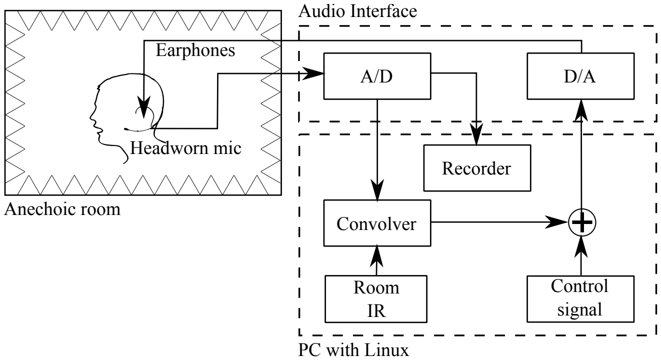

FIG. 1. Experimental setup. The subject was placed inside an anechoic room to remove all the reflections at the boundaries. The different room acoustics conditions were generated by means of software convolution.

A. Experimental setup

The experimental setup is shown in Fig. 1. The experiment took place in an anechoic chamber of dimensions 4.8 m × 4.1 m × 2.9 m in order to remove all reflections from the room. The indirect auditory feedback was generated by picking the voice from the talker, convolving it with a synthetic impulse response, and playing it back via earphones specially designed to minimize the blocking of direct sound and preserve the usual bone conduction path.

The voice of the talker was picked with a microphone DPA (DPA Microphones A/S; Allerød, Denmark) model 4066 located on the cheek at a position 5 cm from the lips' edge in the line between the mouth and the right ear. This signal was sampled at 44.1 kHz with a resolution of 24 bit using an audio interface RME (Audio AG; Haimhausen, Germany) HDSPe Multiface II, which was connected to a computer running the convolution software *jconvolver* under Linux. The convolution system introduced an overall delay of 11.5 ms between the arrival of the direct sound at the ears and the indirect auditory feedback generated in the convolution process. The resulting signal was again converted into the analog domain and reproduced through the two channels (left and right) of the earphones.

These earphones were a customization of the KOSS (KOSS corporation; Milwaukee, WI) model PLUG. The original earphones radiate sound into a short plastic tube and fit into the ear canal with foam pieces. These foam pieces were removed and a bent 3.5 cm silicone tube was attached to the short plastic tubes. At the end of the silicone tube, an Oticon (Oticon A/S; Smørum, Denmark) open dome was placed, so it could fit into the ear canal without modifying the free air transmission and the bone conduction significantly. Figure 2 shows the custom earphones used in the experiment and Fig. 3 shows the insertion loss (IL) introduced by the earphones when used in the ear canal of an artificial ear, B&K (Brüel & Kjær Sound & Vibration Measurement A/S; Nærum, Denmark) type 4159 mounted on a Head and Torso Simulator (HATS) B&K type 4128. The HATS was equipped

FIG. 2. Detail of the earphones with the tubes and the open domes to fit into the ear canal without blocking the direct sound

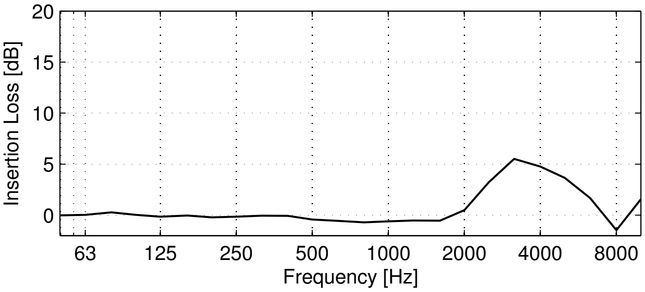

FIG. 3. Insertion loss of the custom earphones, measured in the left ear of a dummy head equipped with a mouth simulator acting as the sound source.

with a mouth simulator which was used as the sound source for the measurements. The peak in IL around 3 kHz and a negative IL value at 8 kHz indicate that the earphones introduce a displacement in the resonance of the ear canal toward higher frequencies, attenuating the resonance peak due to viscous losses. The IL between 63 Hz and 2 kHz is lower than 1 dB, and the maximum attenuation at higher frequencies is 6 dB. These values were assumed to be acceptable for the present application.

With the custom earphones, the frequency response deviates from a flat response (see Fig. 4). Specifically, it has a poor low and mid frequency response, with a roll-off below 2 kHz, and remarkable resonance peaks at high frequencies, between 3 kHz and 8 kHz. A minimum phase FIR filter of 128 samples was used in order to compensate for the frequency response and achieve a relatively flat frequency response, corresponding to the frequency response of the electrostatic headphones STAX (STAX Ltd.; Miyoshi-machi, Japan) model Lambda. This target frequency response was chosen instead of an ideal flat frequency response after realizing—by means of subjective assessment—that the overall sound quality was better in the first case. The FIR filter was preconvolved with the synthetic impulse responses generated for each experimental condition.

A MATLAB program controlled the experiment, changing the synthetic impulse response loaded by *jconvolver* and reproducing different messages to the talker, indicating beginning and the end of vocalization periods, and which vowel should be produced.

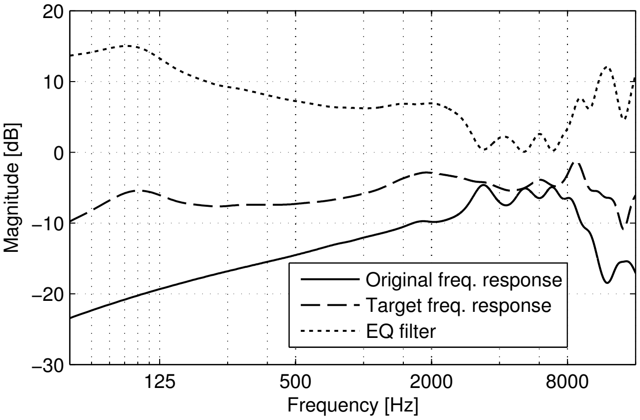

FIG. 4. Equalizer filter applied to the earphones in order to have a magnitude response similar to the one produced by the electrostatic headphones STAX SR Lambda. The magnitude dB reference is arbitrary.

B. Acoustic conditions

There were nine different synthetic impulse responses or conditions C1 to C9 (plus an additional condition C10, namely the absence of simulated reflections), which added the indirect auditory feedback of the talker's voice to the direct sound and the bone conduction. The acoustic properties of the different conditions are summarized in Table I. The synthetic impulse responses were generated artificially, and it was not their goal to replicate the acoustic conditions of actual environments, but to provide well-defined and adjustable experimental conditions. Each synthetic impulse response was obtained in the following manner. First, a white Gaussian noise signal (of 66150 samples at 44.1 kHz) was generated. An exponential decay was applied to the noise signal. The decay constants were chosen so that the reverberation time T of the conditions fell into one of three groups: low (C1 to C3, 0.45 s $\leq T \leq$ 0.55 s), medium (C4 to C6, 0.93 s $\leq T \leq$ 1.12 s) and high (C7 to C9, 1.40 s $\leq T \leq$ 1.65 s). Finally, different gains were applied so that the room gain entered in the categories of low (C1, C4, and C7, 0.07 dB $\leq G_{RG} \leq$ 0.19 dB), medium (C2, C5, and C8, 0.31 dB $\leq G_{RG} \leq$ 1.68 dB), and high (C3, C6, and C9, 2.95 dB $\leq G_{RG} \leq$ 8.63 dB).

The reverberation times were chosen to correspond to usual reverberation times found in rooms for speech (low T: classrooms, medium T: drama theaters, high T: opera houses). The room gain / voice support values were chosen to be representative of real rooms without amplification (-20 dB$\leq ST_V \leq$-5 dB), although higher values were also chosen to explore the possible effects of electroacoustic amplification on the voice production and perception.

For the objective measurements, a HATS B&K type

TABLE I. Experimental conditions and measured acoustic parameters.

Condition	T [s]	G_{RG} [dB]	ST_V [dB]
C1	0.55	0.07	-17.9
C2	0.50	0.31	-11.3
C3	0.45	2.95	-0.12
C4	1.12	0.13	-15.2
C5	1.00	1.03	-5.7
C6	0.93	6.57	5.5
C7	1.65	0.19	-13.5
C8	1.50	1.68	-3.3
C9	1.40	8.63	8.0
C10	0.01	0.04	-20.3

4128 with right ear simulator B&K type 4158 and left ear simulator B&K type 4159 was placed at the talker position in the setup in Fig. 1. The headworn microphone and the earphones were attached to the dummy head as explained in the experimental setup section. The HATS had a mouth simulator and microphones at the ears, so it was possible to measure the impulse response corresponding to the path between the mouth and the ears. The direct sound was generated by direct radiation from the mouth to the ears, whereas the reflections were generated artificially by convolution with a synthetic impulse response and reproduction through the earphones. The mouth-to-ears impulse responses were measured with the MLS module in the 01dB (01dB-Metravib; Limonest Cedex, France) Symphonie system. The backwards-integrated energy-time curves[19] of the measured responses C1 to C9 are shown in Fig. 5. The reverberation time was calculated from the slope of these curves, in a decay of at least 10 dB neither influenced by the noise floor nor the direct sound. The room gain and the voice support were calculated in the way proposed by Pelegrin-Garcia[11]. The corresponding gain introduced by each response on the direct sound, in one-third octave frequency bands between 100 Hz and 4 kHz, is shown in Fig. 6.

C. Vocalizations

Each acoustic condition was repeated three times but using different vowels every time. The three vowels `/a/`, `/i/`, and `/u/` were chosen because they are known to be the so-called corner vowels with the widest spread of the formants.[20] The bone conducted acoustic feedback paths for these vowels are different among them.[3] In this way, the contributions from different bone conduction paths to the autophonic ratings are averaged, and the results are more representative of average speech.

D. Procedure

The experiment was carried out using two different signals as the loudness reference. The first one is called "Voice Level Matching Test" (VLMT) which uses recordings from subjects' own vocalizations as a reference, and the second one is called "Tone Level Matching Test" (TLMT). The reason for this decision was twofold. First, having a human vocalization as the reference could lead to an imitation of the vocal effort and not only to a replication of loudness. Second, using a pure tone could have made the task more difficult because of the mismatch in the perceived sound quality of the reference and the vocalization.

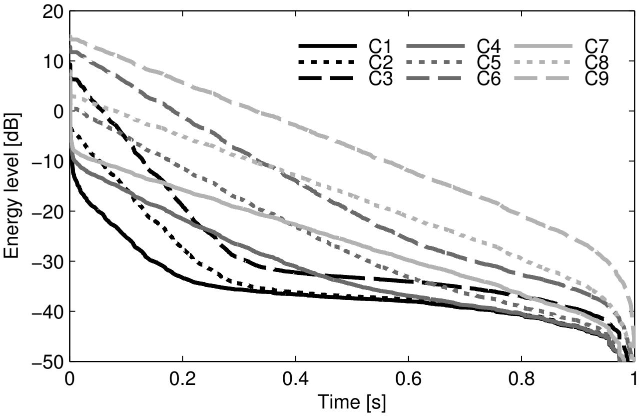

FIG. 5. Backwards-integrated energy-time-curves for the acoustic conditions C1 to C9 presented in the test. The condition C10 (no additional impulse response) is not shown in the figure.

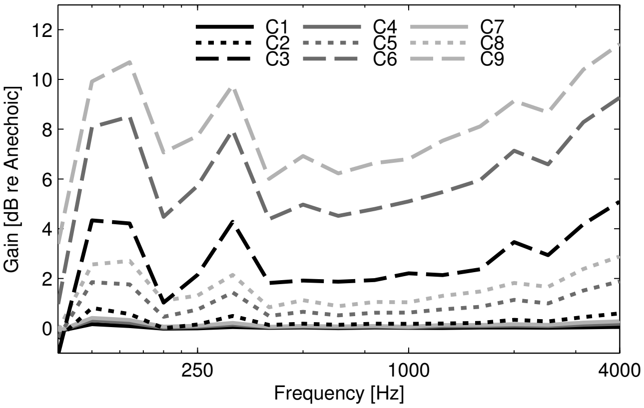

FIG. 6. Gain of the impulse response of each condition C1 to C9 relative to the energy of the impulse response in the anechoic chamber (condition C10), analyzed in one-third octave bands.

The measurements in the VLMT required two steps: (a) recording of references and (b) voice matching test.

a. Recording of references In the beginning of the test, every subject recorded the three vowels `/a/`, `/i/`, and `/u/` with the following protocol (Fig. 7a):

1. A voice played back through the earphones the vowel to utter.

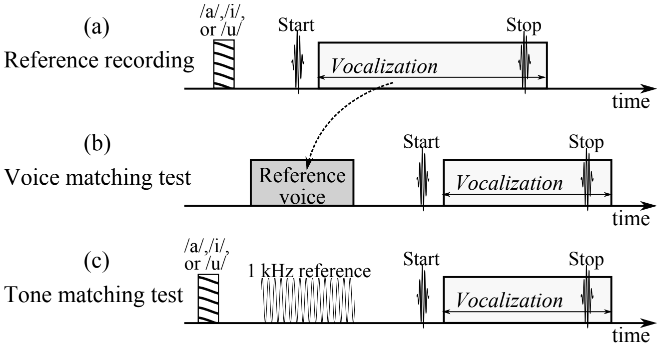

FIG. 7. Procedure followed in the test. Note: The duration of the events and its separation is only approximate

2. After 1.5 s, a beep indicated the beginning of the reference vocalization.

3. The subjects had been instructed to produce a steady vocalization after the beep signal, using a comfortable voice level. The voice was recorded.

4. Another beep, four seconds later, indicated the end of the utterance.

5. The recordings were analyzed to check its steadiness, and they were repeated (from step 1) until the deviation of 200-ms equivalent overall SPL in consecutive, non-overlapping periods, was in a 3 dB range for at least 2 s. The 2 s segment with the lower deviation was chosen as the reference for the given vowel and subject.

6. An equalizer filter was applied to the references recorded with the headworn microphone, so as to later reproduce by the earphones the levels and spectral distributions present at the ears during the original vocalizations.

b. Voice matching test This phase is shown in Fig. 7b.

1. The 3 vowels were selected in random order. The 2-s reference containing the chosen vowel was played back.

2. After 1.5 s, a beep indicated the beginning of the vocalization and, at the same time, the convolver was activated with one of the ten conditions C1 to C10 (in random order).

3. The subjects had been instructed to produce a steady vocalization after the beep signal, with the same vowel and the same loudness as the reference. The voice was recorded.

4. Another beep, three seconds later, indicated the end of the utterance and the deactivation of the convolver.

The measurements with the TLMT (Fig. 7c) were very similar to the voice matching test, but the reference in step 1 was substituted with an audible message of the vowel to produce followed by a 1 kHz sinusoid signal of 2 seconds duration and played back at a level of 75 dB SPL measured at the eardrum of a dummy head. The subjects were explicitly instructed to match the loudness of the pure tone.

At the beginning of the experiment, the subjects made a training run with five conditions and one vowel of the VLMT to get acquainted to the procedure. The results of the training measurements were not used for the posterior analysis. In total, each subject produced 60 vocalizations (10 acoustic conditions, 3 vowels, and 2 references) that were used for the analysis.

E. Post-processing

Each recording was analyzed for a stability criterion, looking for a one-second interval in which the deviation of 200 ms equivalent overall SPL in consecutive, non-overlapping periods, was in a 3 dB range. The one-second interval with the lowest deviation was used in the analysis. The SPL in the one-octave frequency bands between 125 Hz and 4 kHz (L_i), together with the overall unweighted (L_Z) and A-weighted SPL (L_A), were extracted from each of the recordings for building the statistical model. The SPL in condition C10 (anechoic) was used as the reference factor to normalize all the other levels. Therefore, the relative level ΔL_i is defined as

$$\Delta L_{i,j} = L_{i,j} - L_{i,C10}, \tag{3a}$$

$$\Delta L_{Z,j} = L_{Z,j} - L_{Z,C10}, \tag{3b}$$

$$\Delta L_{A,j} = L_{A,j} - L_{A,C10}, \tag{3c}$$

where i is the frequency band and j is one of the conditions C1 to C9.

The spread in SPL among conditions is studied in the frequency domain. For the spectral analysis of the signals, one-third octave band filters are used. Two descriptors are used, one for low frequencies and another one for high frequencies. These are the average rms deviation in the eight one-third octave frequency bands between 100 Hz and 500 Hz, $s_{100-500}$, and the average rms deviation in the nine one-third octave frequency bands between 630 Hz and 4 kHz, s_{630-4k},

$$s_{100-500} = \frac{1}{8}\sum_{i=1}^{8}\sqrt{\frac{1}{9}\sum_{j=1}^{9}\left(\Delta L_{i,j} - \overline{\Delta L_{i,j}}\right)^2} \tag{4a}$$

$$s_{630-4k} = \frac{1}{9}\sum_{i=9}^{17}\sqrt{\frac{1}{9}\sum_{j=1}^{9}\left(\Delta L_{i,j} - \overline{\Delta L_{i,j}}\right)^2} \tag{4b}$$

where

$$\overline{\Delta L_{i,j}} = \frac{1}{9}\sum_{j=1}^{9}\Delta L_{i,j} \qquad i = 1\ldots 17 \tag{5}$$

The subindex i refers to the third-octave band center frequency ($f_{i=1} = 100$ Hz to $f_{i=17} = 4$ kHz), whereas the subindex j refers to one of the acoustic conditions C1 to C9.

F. Statistical analysis

An analysis of variance (ANOVA) table, including main effects and second order interactions of the acoustic condition (C1 to C9), the gender (male/female), the vowel (/a/, /i/, or /u/), and the reference (TLMT or VLMT), was obtained to calculate their relative contribution to the variations of ΔL_Z and ΔL_A. For the derivation of this table, an additive, fixed-effects model was assumed. ΔL_Z is "a priori" the variable of interest in the study, comparable to other sidetone studies, and ΔL_A is relevant for being a closer indicator of the loudness perception.

From the inspection of the data, the mean values of ΔL_Z, ΔL_A, or all the ΔL_i do not change linearly with the room gain or the voice support. Instead, they follow a non-linear trend of the form

$$\Delta L = A(\mathrm{e}^{-B\times G_{RG}} - 1) - C \qquad (6)$$

as a function of the room gain, or

$$\Delta L = A\left(\left(10^{\frac{ST_V}{10}} + 1\right)^{-\frac{10B}{\ln 10}} - 1\right) - C \qquad (7)$$

as a function of the voice support. A, B, and C are the parameters of the model (identical in the two previous equations) and the relation

$$G_{RG} = 10\log\left(10^{\frac{ST_V}{10}} + 1\right) \qquad (8)$$

has been used.[12]

The fitting of the non-linear function to the measured data, in order to obtain the A, B, and C parameters, was performed with the routine `nls` of the library `stats` of the statistical software `R`.[21]

III. RESULTS

Table II shows the results of the four-way ANOVA for ΔL_Z, considering a fixed-effects, additive model, with the main effects and the two-way interactions. It reveals that there is a significant effect of the acoustic condition ($F(8, 652) = 92.4$, $p < 0.0001$), responsible for almost the 90% of the explained variance. Gender has also a significant effect ($F(1, 652) = 43.2$, $p < 0.0001$), and is responsible for another 5% of the explained variance. The variables reference and vowel do not report significant effects. However, there are significant interactions between reference and vowel ($F(2, 652) = 5.55$, $p = 0.004$) and between vowel and gender ($F(2, 652) = 5.13$, $p = 0.006$), responsible however, for less than 3% of the explained variance. There are no significant interactions between the acoustic condition and any other variable. In the additive model, the average ΔL_Z is -3.3 dB for females, whereas it is -2.2 dB for males.

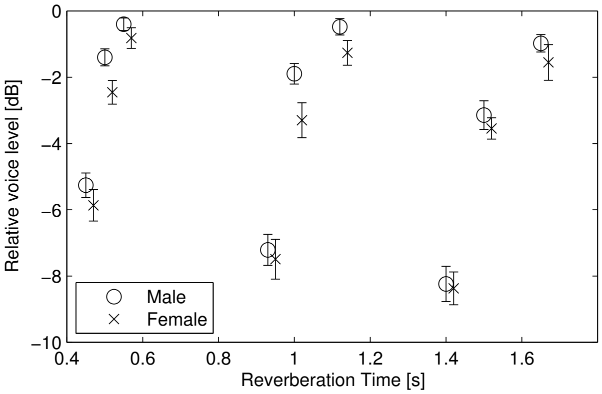

FIG. 8. Relative overall unweighted voice levels as a function of the reverberation time under the different experimental conditions. The bars around the points indicate ±1 standard error.

Table II also shows the results of the four-way ANOVA for ΔL_A. As with ΔL_Z, the most important effect is due to the acoustic condition ($F(8, 652) = 99.6$, $p < 0.0001$) which accounts for 92.8% of the explained variance. This increase in the explained variance is probably due to the closer relationship of the A-weighting to the loudness perception. The gender has also a significant effect ($F(1, 652) = 19.3$, $p < 0.0001$) and accounts for 2.3% of the explained variance. In the additive model, the average ΔL_A is -3.8 dB for females and -2.9 dB for males. The effect of the reference is at the limit of significance ($F(1, 652) = 4.2$, $p = 0.041$) and it accounts for barely a 0.5% of the explained variance. However, a one-way ANOVA model with reference as the only explanatory variable does not pass a significance test. The vowel has no significant effect on ΔL_A. There is also a significant interaction between reference and vowel ($F(2, 652) = 4.7$, $p = 0.009$) accounting for 2.6% of the explained variance, between reference and gender ($F(1, 652) = 4.0$, $p = 0.044$) accounting for 0.47% of the explained variance, and between vowel and gender ($F(1, 652) = 4.9$, $p = 0.008$) accounting for 1.1% of the explained variance.

The values of ΔL_Z are plotted as a function of T in Fig. 8. No trend relating the two variables can be observed from the measurements, because the ΔL_Z are scattered homogeneously.

The average results of ΔL_i in the frequency bands from 125 Hz to 4 kHz, along with the overall unweighted and A-weighted relative SPL values (ΔL_Z and ΔL_A, respectively) are shown in Fig. 9. In the top row, the results are shown for males and females separately. The abscissa shows the room gain parameter. In the bottom row, the same results are shown, but plotted against the voice support. Each data point corresponds to the average of all subjects of one gender, vowels and reference for the same condition. Different symbols correspond to different measures. The bars around the data points indicate ±1 standard error.

It can be seen that the ΔL values are arranged in a non-linear fashion. Observing the data in the room gain

TABLE II. Four-way analysis of variance table with main effects and two-way interactions applied to the relative overall SPL, unweighted (ΔL_Z) and A-weighted (ΔL_A).

	ΔL_Z			ΔL_A		
	F-value	p-value	% Expl. variance	F-value	p-value	% Expl. variance
Main effects						
Reference	1.99	NS	–	4.2	0.041	0.49
Vowel	1.46	NS	–	0.24	NS	–
Gender	43.21	$< 10^{-6}$	5.23	19.3	$5 \cdot 10^{-5}$	2.25
Acoustic condition	92.463	$< 10^{-6}$	89.54	99.6	$< 10^{-6}$	92.8
Two-way interactions						
Reference*Vowel	5.55	0.004	1.34	4.7	0.009	1.10
Reference*Gender	3.058	0.08	0.37	4.0	0.044	0.47
Reference*Acoustic condition	0.48	NS	–	0.64	NS	–
Vowel*Acoustic condition	0.33	NS	–	0.39	NS	–
Gender*Acoustic condition	0.60	NS	–	0.41	NS	–
Vowel*Gender	5.13	0.006	1.24	4.9	0.008	1.13

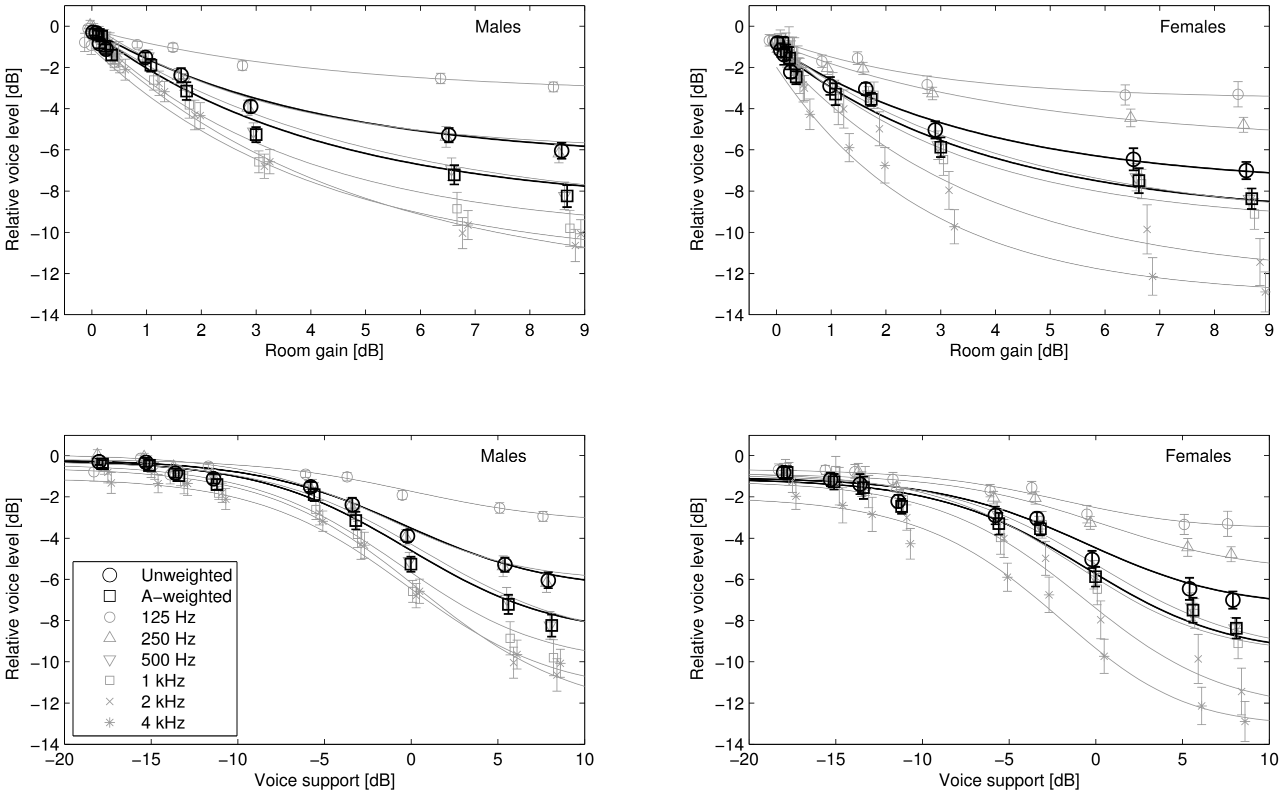

FIG. 9. Relative voice levels as a function of the room gain (top row) and the voice support (bottom row), for male (left column) and female subjects (right column). The reference value for each subject is the voice level produced without simulated reflections. The curves are the best fitting models, Eq. (6) for the top row and Eq. (7) for the bottom row, for each relative voice level descriptor. The bars around the points indicate ±1 standard error.

plots, each level ΔL falls close to a curve given in Eq. (6). This non-linear model indicates that all points converge to a constant level $-C$ for $G_{RG} \to 0$ and that they tend to a limit value $-A - C$ as G_{RG} approaches ∞. The parameter B defines the slope of the curve, together with A. The best fitting curves are overlaid on Fig. 9, and the A, B, and C parameters for all ΔL, separately for males and females, are shown in Table III.

An average model for males and females together, for ΔL_Z and ΔL_A is given by

$$\Delta L_Z = 8.4 \times e^{-0.24 G_{RG}} - 8.9 \text{ [dB]}, \quad (9a)$$

$$\Delta L_A = 6.4 \times e^{-0.25 G_{RG}} - 6.9 \text{ [dB]} \quad (9b)$$

as a function of the room gain, or alternatively, using the

TABLE III. Parameters A, B, and C of the models Eqs. (6) and (7) for the relative voice levels in each of the frequency bands between 125 Hz and 4 kHz, and the overall unweighted and A-weighted relative levels.

Gender	Parameter	ΔL_{125}	ΔL_{250}	ΔL_{500}	ΔL_{1k}	ΔL_{2k}	ΔL_{4k}	ΔL_Z	ΔL_A
Females	A	2.87	4.83	8.73	8.82	11.11	11.12	6.71	8.18
	B	0.35	0.22	0.23	0.29	0.27	0.36	0.26	0.26
	C	0.65	0.87	0.92	0.8	1.22	1.99	1.05	1.11
Males	A	3.11	6.14	8.89	9.70	11.95	10.49	6.31	8.52
	B	0.23	0.30	0.20	0.26	0.21	0.24	0.25	0.24
	C	0.17	-0.07	0.27	0.4	0.58	1.07	0.18	0.22

TABLE IV. Average rms deviations at low frequencies $s_{100-500}$ and high frequencies s_{630-4k} corresponding to the plots in Fig. 10.

$s_{100-500}$	Raw levels		Corrected levels	
	Male	Female	Male	Female
/`a`/	1.5 dB	1.5 dB	1.7 dB	1.7 dB
/`i`/	1.8 dB	2.0 dB	1.5 dB	1.4 dB
/`u`/	1.8 dB	2.1 dB	1.6 dB	1.3 dB
s_{630-4k}				
/`a`/	3.8 dB	3.4 dB	1.4 dB	1.2 dB
/`i`/	3.6 dB	3.4 dB	1.3 dB	1.2 dB
/`u`/	2.6 dB	3.2 dB	1.5 dB	1.1 dB

voice support,

$$\Delta L_Z = 8.4 \times \left(10^{\frac{ST_V}{10}} + 1\right)^{-1.05} - 8.9 \text{ [dB]}, \qquad (10a)$$

$$\Delta L_A = 6.4 \times \left(10^{\frac{ST_V}{10}} + 1\right)^{-1.10} - 6.9 \text{ [dB]}. \qquad (10b)$$

Figure 10a shows the measured spectra in one-third octave bands for the different vowels (/`a`/ on the top row, /`i`/ on the middle row, and /`u`/ on the bottom row), under the different conditions (different line styles), for the female (left column) and male subjects (right column), averaged for the two reference signals and the different subjects for each gender. As shown in Fig. 9, the differences among conditions are greater at high frequencies. This is also reflected in the average rms deviation s in Table IV, which is higher in the frequency bands between 630 Hz and 4 kHz (s_{630-4k} in the range from 2.57 to 3.75 dB) than in the frequency bands between 100 Hz and 500 Hz ($s_{100-500}$ in the range from 1.47 dB to 2.09 dB).

Figure 10b results from adding the gains of each condition in Fig. 6 to the spectra of the vocalizations on those conditions (plotted in Fig. 10a). As can be seen, the deviations among spectra is greatly reduced, in particular at high frequencies, where the average rms deviation s_{630-4k} is now in the range of 1.05 to 1.52 dB, as shown in Table IV. By applying the gain of the IR, the average rms deviation in the low frequency range, $s_{100-500}$, is lower for the vowels /`i`/ and /`u`/, but not for /`a`/, and it ranges from 1.28 to 1.68 dB in all cases. These numbers reflect a uniform spread of the spectra in a broader frequency range for the corrected recordings, which are a closer approximation to the subjects' perceived levels.

IV. DISCUSSION

From the observation of the measured relative voice levels in Figs. 8, 9 and 10a, it is possible to state that different acoustic environments alter the autophonic level for a talker. However, the reverberation time is not a good descriptor of the changes in voice level, as seen in Fig. 8, since it is not directly related to the energy of the indirect auditory feedback. Figure 9 describes the changes in voice level that make the talker's voice sound equally loud at their ears when the indirect acoustic feedback is changed. The curves for ΔL_Z show a constant autophonic level under different room gain conditions (top row), or voice support conditions (bottom row). The A-weighted and the one-octave band values follow the same general trend of the non-linear model in Eq. (6), but with different model parameters. In normal rooms for speech without amplification ($G_{RG} < 1.0$ dB),[22] the variations in voice level to keep a constant autophonic level are within 2.3 dB, according to model Eq. (9a). In the frequency band of 4 kHz, this range increases to about 3.4 dB using the parameters of Table III.

For the three lowest values of voice support(-18.0 dB $\leq ST_V \leq$ -13.5 dB), excluding the anechoic chamber, the range of ΔL_Z is about 0.3 dB, calculated from the model in Eq. (10a). There are consistent voice level variations in a range of less than 0.5 dB, which is considered to be the just noticeable level difference for broadband noise signals.[23] These observations agree with recent findings, which suggest that an auditory motor system controls voice intensity in a non-conscious way and is able to react to level variations below the conscious detectability threshold.[24]

The model in Eq. (9a) shows a varying slope in the dependence of voice level with room gain. It is most negative (or maximum in absolute value) for $G_{RG} \to 0$ with a value of -2.0 dB/dB. In the range observed, the less negative slope is obtained for the highest room gain value ($G_{RG} = 8.6$ dB). In this case, the slope is -0.26 dB/dB. The same equation indicates a saturation effect (zero slope) as $G_{RG} \to \infty$. This could be an indication that the voice levels approach the phonation threshold with the given experimental setup. However, no generalization of the model is intended for values of G_{RG} higher than the studied range.

In a review of different studies of sidetone, Lane *et al.*[9] showed that the sidetone compensation function is linear with slopes varying between -0.4 and -0.6 dB/dB. With

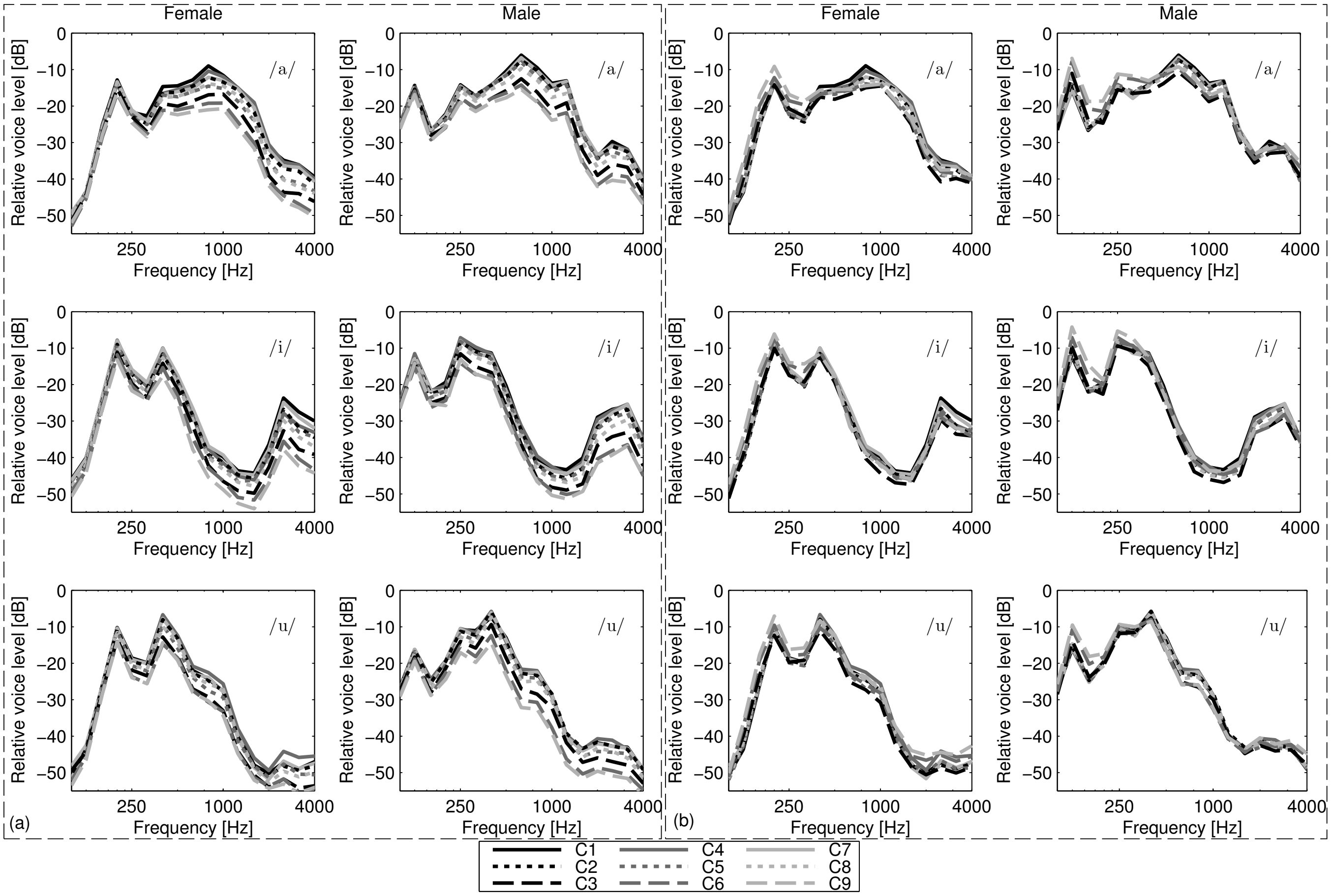

FIG. 10. (a) One-third octave band spectra of the vocalizations averaged for all subjects in one gender and for the two references (tone and voice). The three vowels and the two genders are shown separately. (b) One-third octave band spectra of the vocalizations in (a), where each average vocalization has been corrected with the gain introduced by each condition (in Fig. 6). The dB reference is arbitrary.

the model in Eq. (9a), these slopes are obtained in the range of 5 dB$\leq G_{RG} \leq$ 6.7 dB. Using Eq. (2), a G_{RG} of 5 dB is equivalent to a ratio of indirect-to-direct airborne sound of approximately 2. Several studies have stated that the direct airborne sound and bone conducted sound of one's own voice are of a comparable magnitude.[1–3] A G_{RG} of 5 dB indicates that the reflected sound is of the same importance as the combination of the direct airborne sound and the bone conducted sound of one's own voice. For values of G_{RG} higher than 5 dB, the indirect auditory feedback component is dominating, and the slopes are comparable to those found in traditional sidetone studies.[9]

Lane and Tranel[6] pointed out that the Lombard reflex and the sidetone compensation are two sides of the same coin. In later experiments, Pick *et al.*[14] showed that the Lombard reflex is very difficult to inhibit. Consequently, it is natural that the sidetone compensation is also difficult to inhibit. In the absence of background noise, large values of room gain would make a talker speak softer, as it could happen when using an electroacoustic reinforcement system. From a different perspective, it could be possible to think that a good room for speech has a certain value of room gain. A room of drier acoustics and with a lower room gain would make the talker speak louder. However, in rooms without electroacoustic amplification, the range of room gain is bounded between 0 and approximately 1 dB, which would induce changes in voice level of less than 2 dB. At the first glance, this value seems not to be very significant compared to the dynamic range of the human voice (roughly 30 dB, depending on the person and the fundamental frequency).

The equal autophonic level curve for ΔL_Z, described in Eq. (9a), is compared to the results of other two studies (Ref. 12, and Ref. 8) in Fig. 11 (Note: the two studies show variations in voice power level, whereas the equal autophonic level curves are indicated as variations in SPL, so the comparison is approximate). The dataset of Ref. 12 shows the variations in voice level of teachers lecturing in classrooms of different sizes and room gains. The slope of the line that relates voice levels with room gain is -13.5 dB/dB. However, the changes in voice level are not purely due to the perception of room acoustics, but to other aspects of the communication scenario, such as the variation in distance between talker and listeners that occurs naturally in different rooms of different size. At the same time, the smallest room is the one with the largest room gain. Therefore, the dataset of Ref. 12 is

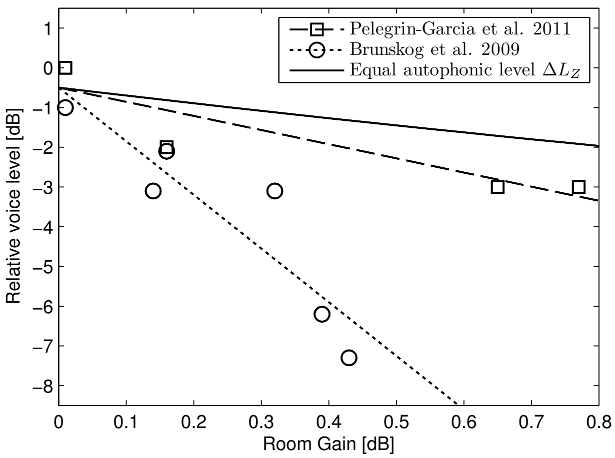

FIG. 11. Comparison of the voice power levels used by teachers in different classrooms (Brunskog et al.[12]), talkers speaking to a listener at 6 m (Pelegrin-Garcia et al.[8]) and voice levels (SPL) to keep a constant autophonic level.

representative of typical voice level variations in rooms without background noise. The dataset of Ref. 8 presents data of a talker addressing a listener at a distance of 6 m in front of him in four different rooms with different room gain. The average voice level varies with the room gain at a rate of -3.6 dB/dB. In the same range of G_{RG}, the equal autophonic level curve approximates a straight line with a slope of -1.8 dB/dB. The talkers in these two experiments did not follow a communication strategy based on keeping the autophonic level constant. In case they did, the voice measurements would have lain on top of the equal autophonic level curve. Talkers apparently "amplify" the effect of the Lombard reflex. This suggests that they make use of other attributes present in the room impulse response than loudness for the adjustment of their voice, probably in combination with other sensory inputs. One explanation for the difference in slope is that the talkers in Ref. 8 adjust their voice level according to some tacit knowledge of sound attenuation with distance, as suggested by Zahorik and Kelly,[25] although do not completely compensate for that. In the experiment of Ref. 8, the sound attenuation at 6 m from the talker differed by more than 15 dB in the two most extreme cases (with $G_{RG} \approx 0$ dB and $G_{RG} \approx 0.8$ dB), whereas the voice variation was only about 3 dB.

The amount of voice level variation to achieve a constant autophonic level is different for the different frequency bands and the two genders. It is less important at low frequencies and more important at higher frequencies and for females. This can be observed in both Figs. 9 and 10a. When applying the frequency-dependent gain introduced by the synthetic IR in Fig. 6 to the voice recordings, they seem to fall on similar curves, as shown in Fig. 10b and in the reduced average rms deviations in Table IV. This means that the subjects kept the resulting sound from their vocalizations constant at their ears, in overall level and in spectral balance of different frequency bands. As a consequence, the parameters A, B, and C of Table III can be used in connection with the models in Eqs. (6) and (7) to describe the amount of compensation expected for the different frequency bands. It may be possible that the compensation at high frequencies is a side-effect of the change in vocalization level, because the spectral slope decreases with increasing vocal effort.[26] The fact that the compensations for the two references (tone and voice) are not significantly different, as shows the ANOVA in Table II, makes this hypothesis likely. The alternative hypothesis is that subjects try to keep the sound quality (loudness and spectral balance) of the vocalizations constant. This hypothesis is reasonable when using the voice as a reference, but not when using the tone. The equivalent results of the test using the two references shows that subjects focused on the loudness cue.

The models in Eqs. (9a) and (10a) can be used to predict the variations in vocal intensity that happen with the use of electroacoustic amplification. As an example, Sapienza *et al.*[27] found that teachers talked on average 2.4 dB softer in classrooms when using a sound reinforcement system. The gain of the system was tuned so that it increased the SPL at a distant listener position by 10 dB. At these positions, the reflected energy dominates over the direct sound energy. Making this consideration, and considering that the amplification system produces a uniform SPL in the room, the amount of non-direct energy E_I increases by 10 dB when the system is turned on, also at the talker position. By Eq.(1), ST_V would increase about 10 dB when the system is turned on. A representative value of ST_V in non-amplified classrooms is -13 dB.[22] By using Eq. (10a), talkers would speak 2.5 dB softer when the system is on ($ST_V = -3$dB), compared with what they would do when the system is off ($ST_V = -13$dB). The good agreement of the measured and predicted variations (2.4 dB and 2.5 dB) are probably due to the fact that the only variable that was changed in the study of Sapienza *et al.* was the sidetone, and not any other variables like the room or the distance to the listeners, and therefore the subjects reacted sympathetically according to the Lombard reflex.

The level of the voice reference recordings was not monitored, and the test subjects received the instruction to produce a vocalization at a "comfortable" level. Since the equal loudness level contours as a function of the frequency in ISO-226:2003[28] are not parallel, it may be possible that the amount of compensation was different at different voice levels. This could have been studied by repeating the test with reference tones at different levels, but this was done only at one level. Since the comfortable and most used voice level changes from subject to subject, the measured equal autophonic level curves are an average indicator of this "most comfortable level". Because the results of the tests using the two references (voice and tone) are similar, as shown by the low significance of the variable "reference" in the ANOVA of Table II, significant differences are not to be expected among different reference levels.

V. CONCLUSIONS

An experiment was conducted to obtain the relative voice levels that kept the autophonic level constant under different room acoustics conditions described by the parameters room gain and voice support. Analyzing the voice levels in one-octave bands and with different frequency weightings, a set of equal autophonic level curves was generated. These curves allow to determine the expected voice level differences in different rooms which are purely related to the Lombard-effect or sidetone compensation. The main conclusions of the study are as follows

1. Voice level variations under different room acoustics conditions are related to the room gain or the voice support, and not to the reverberation time.
2. Typical voice level variations in rooms for speech ($G_{RG} < 1.0$ dB) to keep a constant autophonic level are not higher than 2.3 dB.
3. By comparison with other studies, talkers use other cues than loudness to adjust their voice level in rooms, resulting in larger voice variations than barely keeping the autophonic level constant.

Acknowledgments

This research has been partially funded by the Swedish organization AFA Försäkring as a part of the project "Speaker's comfort and voice health in classrooms". The authors would like to express their gratitude to all the subjects who took part in the tests, and to Søren Laugesen, at Oticon A/S for his unconditional help providing the authors with key elements in the setup.

[1] G. Békésy, "The structure of the middle ear and the hearing of one's own voice by bone conduction", The Journal of the Acoustical Society of America **21**, 217–232 (1949).

[2] C. Pörschmann, "Influences of bone conduction and air conduction on the sound of one's own voice", Acta Acust. united Ac. **86**, 1038–45 (2000).

[3] S. Reinfeldt, P. Ostli, B. Håkansson, and S. Stenfelt, "Hearing one's own voice during phoneme vocalization–transmission by air and bone conduction.", The Journal of the Acoustical Society of America **128**, 751–762 (2010).

[4] H. Lane, A. Catania, and S. Stevens, "Voice level: Autophonic scale, perceived loudness, and effects of sidetone", Journal of the Acoustical Society of America **33**, 160–167 (1961).

[5] L. Raphael, G. Borden, and K. Harris, *Speech science primer: Physiology, acoustics, and perception of speech*, 167–198, fifth edition (Lippincott Williams & Wilkins, Baltimore and Philadelphia) (2007).

[6] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech", Journal of Speech, Language and Hearing Research **14**, 677 (1971).

[7] D. Michael, G. Siegel, and H. Pick Jr, "Effects of distance on vocal intensity", J. Speech Hear. Res. **38**, 1176–83 (1995).

[8] D. Pelegrin-Garcia, B. Smits, J. Brunskog, and C.-H. Jeong, "Vocal effort with changing talker-to-listener distance in different acoustic environments", The Journal of the Acoustical Society of America **129** (2011), accepted for publication.

[9] H. Lane, B. Tranel, and C. Sisson, "Regulation of voice communication by sensory dynamics", The Journal of the Acoustical Society of America **47**, 618–624 (1970).

[10] J. Black, "The effect of noise-induced temporary deafness upon vocal intensity", Communication Monographs **18**, 74–77 (1951).

[11] D. Pelegrin-Garcia, "Comment on "Increase in voice level and speaker comfort in lecture rooms" [J.Acoust.Soc.Am. 125, 2072-2082 (2009)]", The Journal of the Acoustical Society of America **129** (2011), accepted for publication.

[12] J. Brunskog, A. Gade, G. P. Bellester, and L. R. Calbo, "Increase in voice level and speaker comfort in lecture rooms.", The Journal of the Acoustical Society of America **125**, 2072–2082 (2009).

[13] J. Black, "The effect of room characteristics upon vocal intensity and rate", The Journal of the Acoustical Society of America **22**, 174–176 (1950).

[14] H. Pick Jr, G. Siegel, P. Fox, S. Garber, and J. Kearney, "Inhibiting the Lombard effect", The Journal of the Acoustical Society of America **85**, 894–900 (1989).

[15] V. Jónsdóttir, "The voice: an occupational tool. A study of theachers' classroom speech and the effects of amplification", Phd thesis, University of Tampere and University of Oulu, Finland (2003).

[16] N. Roy, R. Merrill, S. Thibeault, R. Parsa, S. Gray, and E. Smith, "Prevalence of voice disorders in teachers and the general population", Journal of Speech, Language and Hearing Research **47**, 281–293 (2004).

[17] V. Lyberg-Åhlander, R. Rydell, and A. Löfqvist, "Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff", J. Voice (2010), available online 25 March 2010.

[18] M. Sliwinska-Kowalska, E. Niebudek-Bogusz, M. Fiszer, T. Los-Spychalska, P. Kotylo, B. Sznurowska-Przygocka, and M. Modrzewska, "The prevalence and risk factors for occupational voice disorders in teachers", Folia Phoniatrica et Logopaedica **58**, 85–101 (2006).

[19] M. Schroeder, "New Method of Measuring Reverberation Time", The Journal of the Acoustical Society of America **37**, 409–12 (1965).

[20] L. Raphael, G. Borden, and K. Harris, *Speech science primer: Physiology, acoustics, and perception of speech*, 105–130, fifth edition (Lippincott Williams & Wilkins, Baltimore and Philadelphia) (2007).

[21] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2010), URL `http://www.R-project.org`, ISBN 3-900051-07-0.

[22] D. Pelegrin-Garcia, V. Lyberg-Ahlander, 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", Proceedings of Meetings on Acoustics **11**, 060001 (2010).

[23] E. Zwicker and H. Fastl, *Psychoacoustics: facts and models*, 175–200, second edition (Springer-Verlag, Berlin Heidelberg New York) (1999).

[24] H. Hafke, "Nonconscious control of voice intensity during vocalization", Archives of Acoustics **34**, 407–414 (2009).

[25] P. Zahorik and J. W. Kelly, "Accurate vocal compensation for sound intensity loss with increasing distance in natural environments.", The Journal of the Acoustical Society of America **122**, EL143–150 (2007).

[26] B. Lindblom and J. Sundberg, "The human voice in speech and singing", in *Springer handbook of acoustics*, edited by

T. D. Rossing, first edition, chapter 16, 175–200 (Springer Science+Business Media, LLC, New York) (2007).

[27] C. Sapienza, C. Crandell, and B. Curtis, “Effects of sound-field frequency modulation amplification on reducing teachers’ sound pressure level in the classroom”, Journal of Voice **13**, 375–381 (1999).

[28] ISO (International Organization for Standardization), *ISO-226:2003: Acoustics - Normal equal-loudness-level contours* (ISO, Geneva, Switzerland) (2003).

Proceedings of Meetings on Acoustics

Volume 11, 2010 http://asa.aip.org

160th Meeting
Acoustical Society of America
Cancun, Mexico
15 - 19 November 2010
Session 5aSC: Speech Communication

5aSC5. Influence of Classroom Acoustics on the Voice Levels of Teachers With and Without Voice Problems: A Field Study

David Pelegrin-Garcia*, Viveka Lyberg-Åhlander, Roland Rydell, Jonas Brunskog and Anders Lofqvist

***Corresponding author's address: Technical University of Denmark, Building 352, Oersteds plads, Kongens Lyngby, DK-2800, Kongens Lyngby, Denmark, dpg@elektro.dtu.dk**

Many teachers suffer from voice problems and classroom acoustics has been considered as one of the potential hazards for this. The present study examines how classroom acoustics interacts with the voices of 14 teachers without voice problems and 13 teachers with voice problems. The assessment of the voice problems was made with a questionnaire and a laryngological examination. During teaching, the sound pressure level at the teacher's position was monitored. The teacher's voice level and the activity noise level were separated using mixed Gaussians. In addition, objective acoustic parameters of Reverberation Time and Voice Support were measured in the 30 empty classrooms of the study. An empirical model shows that the measured voice levels depended on the activity noise levels and the voice support. Teachers with and without voice problems were differently affected by the voice support of the classroom. The results thus suggest that teachers with voice problems are more aware of classroom acoustic conditions than their healthy colleagues and make use of the more supportive rooms to lower their voice levels. This behavior may result from an adaptation process of the teachers with voice problems to preserve their voices. [Work supported by AFA.]

Published by the Acoustical Society of America through the American Institute of Physics

Received 12 Nov 2010; published 15 Dec 2010

INTRODUCTION

Voice is the primary working tool of teachers, and a good voice is essential for communicating with students. Nowadays, many teachers suffer from voice problems. A recent study reported that around 13% of the active school teachers in southern Sweden self-reported voice problems [1]. Voice health problems are a major concern, not only due to the required clinical assistance and the personal consequences in job dissatisfaction and lack of self-esteem, but also due to the financial impact that the teachers' absence produces in the global budget of the country [2]. Investigating possible causes for voice disorders from the testimonies of affected teachers, Vilkman points out "bad classroom acoustics" as one of the hazards for voice health [3].

The present study analyzed the average voice levels used at work by teachers with and without voice problems as a function of relevant environmental acoustic parameters. Two acoustic parameters were considered important: the activity noise level, due to the presence of students and other noise sources during teaching, and the voice support offered by the classroom. Three steps were necessary in the study: first, the choice of teachers and the assessment of voice problems. Second, the monitoring of the teacher's voice levels and the activity noise levels during teaching, and last, the measurement of objective acoustic parameters in the empty classrooms.

METHOD

Choice of teachers

A total of 27 teachers in 5 different schools in the south of Sweden, at educational levels ranging from primary school to high school, were considered for this study. The participants were selected as a follow-up to an epidemiological study[1].

The teachers were classified into two groups: one group (test; $N_T = 13$, 2 male/11 female) containing the teachers with voice problems and another group (control, $N_C = 14$; 2 male/12 female) with those teachers having no remarkable voice problems. The assessment of voice problems was made by means of the VHI-T (Voice Handicap Index with Throat subscale) questionnaire [4] and a laryngological examination.

Measurements during teaching

The teachers were equipped with an IEC 61672-compliant, type 2, sound level meter SVANTEK SV-102. This device measured and stored the A-weighted sound pressure level (SPL), using an exponential averaging with "fast" time constant, sampled at 1 s intervals. The microphone capsule was attached to the teachers' clothing neck, as a lapel microphone, at a distance of about 15 cm from the mouth.

The sound level meter operated for one working day. For each teacher, two SPL sequences were studied. One of them corresponded to a lesson at the beginning of the day and another one to a lesson at the last hour. The duration of the lessons was between 30 and 45 minutes. An example sequence is shown in Fig. 1 and the corresponding histogram is shown as gray bars in Fig. 2.

In these SPL sequences, it was assumed that the SPL from the teacher's voice was several dB higher than the SPL from activity (originated from students, ventilation noise and other external sources), because of the closer placement of the microphone to the teacher's mouth (around 15 cm). The time fraction while the teacher was talking was noted as α. The activity levels were obtained while the teacher was silent, during a time fraction $1-\alpha$.

The teacher's voice (S) and activity noise (N) levels were assumed to be random processes coming from normal distributions, with probability density functions $f_S(L)$ and $f_N(L)$, respectively, where L indicates the A-weighted SPL. The means of these distributions are notated $L_{50,S}$ and $L_{50,N}$ (the symbol L_{50} indicates the level that is exceeded during 50% of the time, also referred to as median level), and their standard deviations σ_S and σ_N. As an example, these distributions are indicated in Fig. 2 with dash-dot and dashed lines, respectively. Thus,

$$S \sim \mathcal{N}(L_{50,S};\sigma_S) \rightarrow f_S(L), \tag{1}$$

$$N \sim \mathcal{N}(L_{50,N};\sigma_N) \rightarrow f_N(L). \tag{2}$$

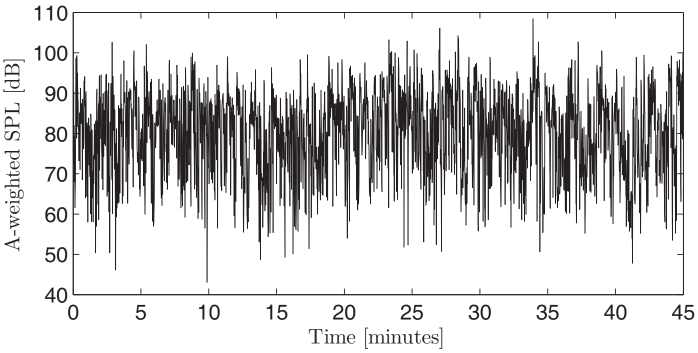

FIGURE 1. A-weighted SPL at the lapel microphone worn by the teacher during one lesson

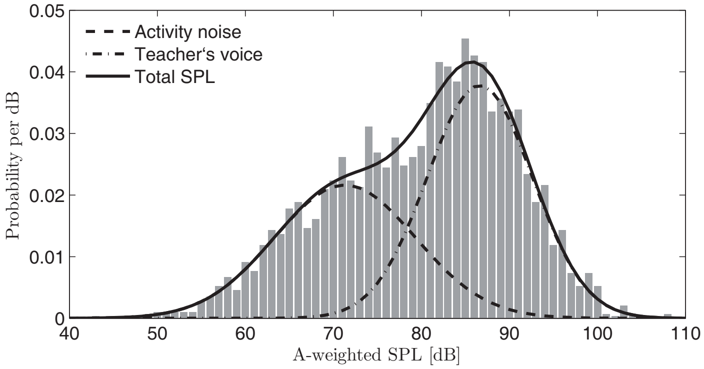

FIGURE 2. In gray, histogram computed from the A-weighted SPL values in Fig. 1. On top, scaled normal probability density functions corresponding to the activity noise (dashed line), the teacher's voice (dash/dot line), and the addition of both processes (solid line).

The joint process corresponding to the observed A-weighted SPL values was regarded as having a probability density function $f_{S+N}(L)$, obtained by overlapping the two normal distributions $f_S(L)$ and $f_N(L)$, scaled by their probability of occurrence in time (α and $1-\alpha$, respectively):

$$f_{S+N}(L) = \alpha f_S(L) + (1+\alpha) f_N(L). \tag{3}$$

According to this principle, a linear combination of two normal distributions was fitted to the A-weighted SPL histogram, by minimizing the squared error with the simplex algorithm implemented in the function `fminsearch` of MATLAB. In this way, there were 5 estimated parameters ($L_{50,S}$, $L_{50,N}$, σ_S, σ_N, and α) for each sequence, although only the A-weighted median levels for the teacher's voice ($L_{50,S}$) and the activity noise ($L_{50,N}$) were used in the analysis. As an example, the probability density function fitted to the measured A-weighted SPL is shown with a solid line in Fig. 2. A similar approach to determine speech and noise levels in classrooms has been previously used [5].

Classroom acoustic measurements

Acoustic measurements were performed in the 30 classrooms where the teachers held their lessons, while they were empty.

Reverberation time. The reverberation time (RT) was calculated according to the standard ISO 3382-2 [6]. The sound source was a B&K Omnisource type 4295, placed at the teacher's position and with the radiating opening at a

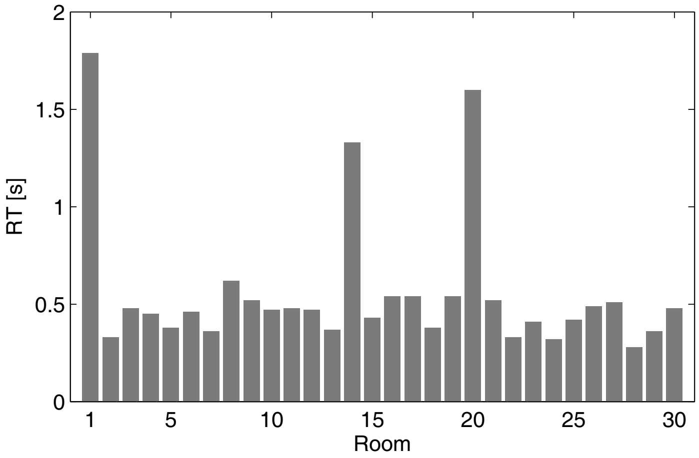

FIGURE 3. Mid-frequency average reverberation time values measured in the classrooms

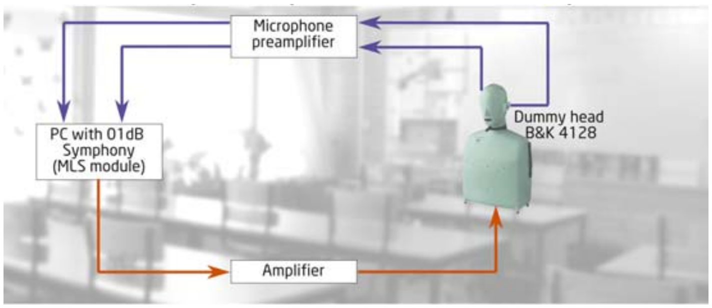

FIGURE 4. Setup used to measure the mouth-to-ears impulse response in the classrooms

height of 1.6 m. Two 1/2" pressure-field microphones B&K type 4192 were used as receivers and were placed close to students' seats at a height of 1.2 m. The 01dB Symphonie system, incorporating the MLS software module, was used to produce the measurement signal and send it to the loudspeaker via a power amplifier, acquire the signal from the microphones, calculate the impulse responses, and derive the RT_{20}. The measured RT values in the classrooms, corresponding to the average of the 500 Hz and 1 kHz octave frequency bands, are shown in Fig. 3. However, the RT was not used in the empirical model due to the lack of normality in the measured values. The three 'outliers' in reverberation time correspond to three sports hall that were used for gymnastics lessons.

Voice support. Instead, the focus in this research was on characterizing the acoustic conditions of classrooms as perceived by the teachers while talking. A parameter called *Voice Support* (ST_V) is introduced in this paper as a measure of how much the sound reflections at the room boundaries amplify the voice of the teacher at his/her ears (NOTE: The exact definition of ST_V is given below).

The voice support is calculated from an impulse response corresponding to the airborne sound transmission between the mouth and the ears (or simply, mouth-to-ears impulse response). For this purpose, a Head and Torso Simulator (HaTS) B&K type 4128 was used. The HaTS included a loudspeaker at its mouth, and microphones at its ears. The HaTS was placed at a representative teaching position, with the mouth at a height of 1.5 m. The 01dB Symphonie system was used to produce the excitation signal and determine the mouth-to-ears impulse response from the measured signal at the microphones. The setup used to measure the mouth-to-ears impulse response is shown in Fig. 4.

From the measured mouth-to-ears impulse response $h(t)$ (example shown in Fig. 5), the direct sound $h_d(t)$ is obtained by applying a window $w(t)$ to the measured impulse response $h(t)$,

$$h_d(t) = h(t) \times w(t), \tag{4}$$

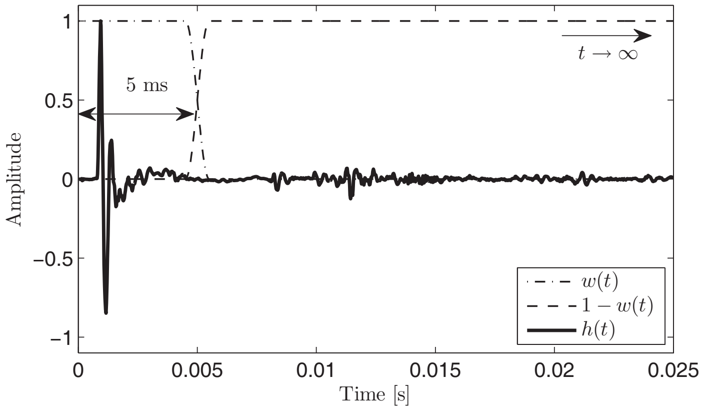

FIGURE 5. Example of a measured mouth-to-ears impulse response, with the windowing applied in order to calculate the direct and the reflected airborne sound components of one's own voice.

where $w(t)$ is

$$w(t) = \begin{cases} 1 & t < 4.5 \text{ ms} \\ 0.5 + 0.5\cos\left(2\pi(t-t_0)/T\right) & 4.5 \text{ ms} < t < 5.5 \text{ ms} \\ 0 & t > 5.5 \text{ ms} \end{cases} \tag{5}$$

with $t_0 = 4.5$ ms and $T = 2$ ms. The reflected sound $h_r(t)$ is the complementary signal

$$h_r(t) = h(t) \times (1 - w(t)) = h(t) - h_d(t) \tag{6}$$

From the above signals, the energy levels corresponding to the direct sound ($L_{E,d}$) and the reflected sound ($L_{E,r}$) are calculated as

$$L_{E,d} = 10\log\frac{\int_0^\infty h_d^2(t)\,\mathrm{d}t}{E_0}, \tag{7}$$

$$L_{E,r} = 10\log\frac{\int_0^\infty h_r^2(t)\,\mathrm{d}t}{E_0}. \tag{8}$$

From these two equations, the *voice support* ST_V, in analogy to Gade's objective support [7], is defined as the difference between the reflected sound and the direct sound from the mouth-to-ears impulse response,

$$ST_V = L_{E,r} - L_{E,d}, \tag{9}$$

The ST_V values measured in the 30 classrooms of the study, averaged for two HaTS positions and the two ears, without applying any filtering, are shown in Fig. 6. The average value is indicated with a solid line, whereas one standard deviation above and below the mean is indicated with dashed lines.

Statistical method

We used a multiple regression to analyze the combined influence of the covariates *voice support* (ST_V) and *median activity noise* ($L_{50,N}$) on the teachers' median voice levels ($L_{50,S}$). The two covariates ST_V and $L_{50,N}$ were fairly uncorrelated ($\rho = -0.07$). Additionally, we accounted for possible differences in voice use between the teachers of the test and control groups (with and without voice problems) by including a binary variable named Test/Control which indicated which group the teacher belonged to.

Since we considered the effect of ST_V and $L_{50,N}$ to be potentially different for the teachers of the test and control groups, we included also the interaction between the Test/Control variable and the two covariates. Nevertheless, the

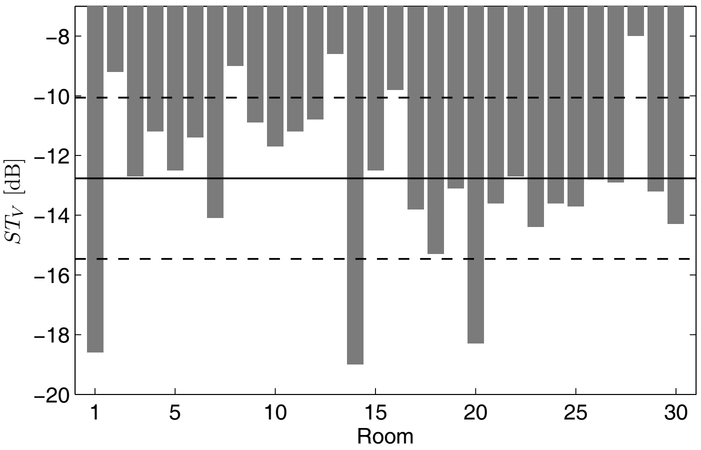

FIGURE 6. Voice support values measured in the 30 classrooms, by averaging the results of two positions and the two ears in each room.

interaction between $L_{50,N}$ and Test/Control was found to be non-significant ($F_{1,48} = 0.15, P = 0.70$) and was left out from the final model.

We fitted the model in R [8] using the function `lm`. Prior to running the model, we applied the square root, affine transformation to the activity noise levels $\sqrt{75 - L_{50,N}}$, in order to obtain an approximately normal distribution of the observed values of the covariate. None of the measured noise levels was higher than 75 dB.

This transformed variable, and ST_V, which already presented an absence of outliers and skew, were further z-transformed. We checked various diagnostics of model validity and stability (Cook's distance, dfits, distribution of residuals, residuals plotted against predicted values) and none of these indicated obvious influential cases or outliers, nor obvious deviations from the assumptions of normality and homogeneity of residuals [9]. The significance of each variable in the model was assessed by means of F-tests resulting from an analysis of variance.

RESULTS

Overall, the median voice levels were clearly influenced by the combination of predictor variables in the proposed statistical model ($R^2 = 0.69$, $F_{4,49} = 27.8$, $p < 0.001$):

$$L_{50,S}(\text{test}) = 81.3 - 3.87 \times \sqrt{75 - L_{50,N}} - 0.72 \times ST_V \text{ [dB]}, \tag{10a}$$

$$L_{50,S}(\text{control}) = 102.9 - 3.87 \times \sqrt{75 - L_{50,N}} + 0.84 \times ST_V \text{ [dB]}. \tag{10b}$$

The effect of the transformed noise levels on the voice levels ($F_{1,49} = 92.2$, $p < 0.001$) was highly significant. The overall effect of the covariate voice support ST_V ($F_{1,49} = 0.65$, $p = 0.43$) and the factor Test/Control ($F_{1,49} = 2.12$, $p = 0.15$) were not significant at the 5% level. However, the interaction between the ST_V and the Test/Control variable was found to be highly significant ($F_{1,49} = 16.5$, $p < 0.001$).

The measured $L_{50,S}$ values as a function of ST_V are shown in Fig. 7. For the average observed noise levels ($L_{50,N} = \overline{L}_{50,N}$), the model (10) is:

$$L_{50,S}(\text{test}) = 69.8 - 0.72 \times ST_V \text{ [dB]}, \tag{11a}$$

$$L_{50,S}(\text{control}) = 91.4 + 0.84 \times ST_V \text{ [dB]}. \tag{11b}$$

For teachers without voice problems (control group), the median voice levels increased with the measured voice support at a rate of 0.8 dB/dB. On the other hand, teachers with voice problems (test group) lowered their voice levels the higher the voice support, at a rate of -0.7 dB/dB.

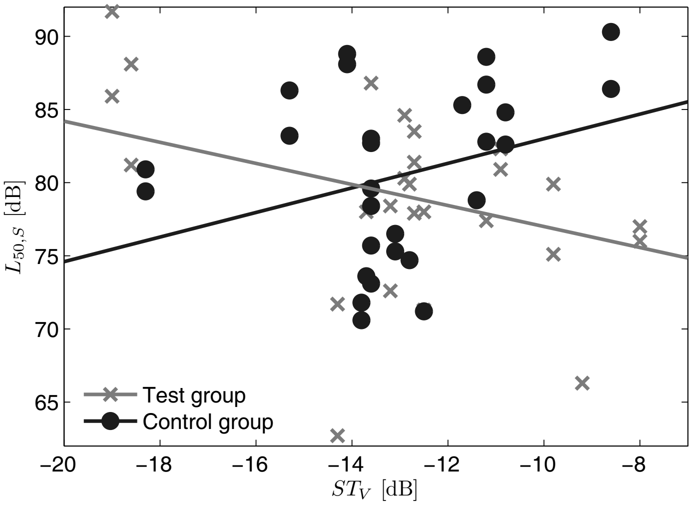

FIGURE 7. Median voice SPL used by teachers versus voice support measured in the empty classrooms. The solid lines show the regression model in (11). The two teacher groups make use of the voice support in significantly different ways.

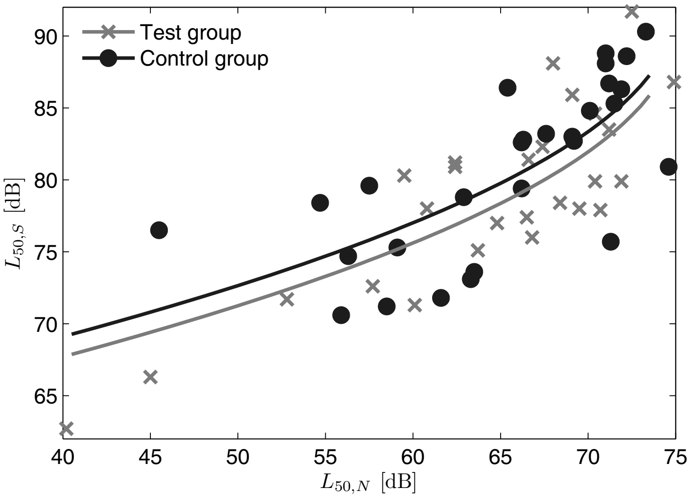

FIGURE 8. Median voice SPL used by teachers versus median activity noise SPL. The solid lines show the regression model (12). As a consequence of the Lombard effect, the voice levels increase with the noise levels, equally for teachers with and without voice problems. However, teachers in the control group use higher voice levels than those in the test group.

The measured $L_{50,S}$ values as a function of $L_{50,N}$ are shown in Fig. 8. For the average observed voice support ($ST_V = \overline{ST}_V$), the model (10) is:

$$L_{50,S}(\text{test}) = 90.6 - 3.87 \times \sqrt{75 - L_{50,N}}\ [\text{dB}], \tag{12a}$$

$$L_{50,S}(\text{control}) = 92.0 - 3.87 \times \sqrt{75 - L_{50,N}}\ [\text{dB}]. \tag{12b}$$

For all teachers, There was an increase of median voice level with the activity noise present during teaching. This increase was non-linear in the observed range of levels, being more relevant for the highest noise levels. Additionally, the teachers from the test group talked 1.4 dB on average softer than the teachers in the control group. However, this difference was not statistically significant with the number of teachers considered in this study.

DISCUSSION

Teachers from the test group (with voice problems) decreased their voice levels with increasing voice support (-0.7 dB/dB) in the classrooms, as opposed to the control group (without voice problems, 0.8 dB/dB). The behavior of the test group would be desirable for the prevention of voice problems. The measurements suggest that teachers from the test group made good use of the voice support as an adaptive mechanism to preserve their vocal health. This finding supports the results from a study by Kob *et al.* [10], who found that teachers with voice problems were more affected by poor classroom acoustics than their healthy colleagues. The behavior of the teachers in the test group follows the results of Brunskog *et al.* [11], who found that teachers lowered their voice levels as a function of the amplification offered by the room to their own voice. However, the behavior of teachers in the control group does not follow a logical pattern. A hypothetical answer would be that the voice support increases in rooms with sound reflecting boundaries, and the activity noise levels would increase in this case. Due to the Lombard effect, the talkers (students and teacher) would perceive increased noise levels and automatically raise their voices. However, the lack of correlation between voice support and activity noise invalidates this hypothesis.

Teachers from the test and control groups were equally affected by noise. Both groups increased their vocal intensity with increasing activity noise, in accordance with the Lombard effect. If the curves are approximated by straight lines for $L_{50,N}$ above 55 dB, the slope is 0.6 dB/dB, in good agreement with the literature (for example, Lazarus reports slopes between 0.5 dB/dB and 0.7 dB/dB [12]). The teachers from the test group talked on average 1.4 dB softer than the control group, although this difference was not significant. Nevertheless, this might be an additional indication that teachers with voice problems tried to limit their vocal effort in terms of vocal intensity.

CONCLUSIONS

The main conclusions from the field study are the following:

- Teachers with voice problems make a more efficient use of the voice support in classrooms than their healthy colleagues, probably as an adaptive mechanism to preserve their voice health.
- Teachers with and without voice problems react identically to changes in activity noise, according to the Lombard effect.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to all the teachers and staff of schools participating in the measurements. This research has been funded by the Swedish company AFA Försäkring as a part of the project "Speaker's comfort and voice health in classrooms".

REFERENCES

1. V. Lyberg-Åhlander, R. Rydell, and A. Löfqvist, "Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff," *J. Voice*, 2010. Available online 25 March 2010.
2. K. Verdolini and L. Ramig, "Review: occupational risks for voice problems," *Logopedics Phonatrics Vocology*, vol. 26, no. 1, pp. 37–46, 2001.
3. E. Vilkman, "Voice problems at work: A challenge for occupational safety and health arrangement.," *Folia Phoniatr. Logo.*, vol. 52, pp. 120–25, 2000.
4. V. Lyberg-Åhlander, R. Rydell, J. Eriksson, and L. Schalén, "Throat related symptoms and voice: development of an instrument for self assessment of throat-problems," *BMC Ear, Nose and Throat Disorders*, vol. 10, no. 5, 2010. Online edition.
5. M. Hodgson, R. Rempel, and S. Kennedy, "Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures," *J. Acoust. Soc. Am.*, vol. 105, pp. 226–233, 1999.
6. International Organization for Standardization, "EN ISO 3382-2:2008, Acoustics - Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms," 2008. Geneva.
7. A. Gade, "Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and laboratory experiments," *Acustica*, vol. 69, pp. 193–203, 1989.
8. R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. (date last viewed 10/16/10). ISBN 3-900051-07-0.

9. G. Quinn and M. Keough, *Experimental design and data analysis for biologists*, ch. Multiple regression and correlation. Cambridge: Cambridge Univ Pr, 2002.
10. 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,” *Acoust. Sci. Tech.*, vol. 29, pp. 86–94, 2008.
11. J. Brunskog, A. Gade, G. Payá-Ballester, and L. Reig-Calbo, “Increase in voice level and speaker comfort in lecture rooms.,” *J. Acoust. Soc. Am.*, vol. 125, pp. 2072–82, 2009.
12. H. Lazarus, “Prediction of Verbal Communication in Noise–A review: Part 1,” *Applied Acoustics*, vol. 19, no. 6, pp. 439–464, 1986.

Teacher's Voice Use in Teaching Environments:
A Field Study Using Ambulatory Phonation Monitor (APM)

Viveka Lyberg Åhlander, David Pelegrín García[1], Roland Rydell*, and Anders Löfqvist

Dept. Logopedics, Phoniatrics and Audiology, Clinical Sciences, Lund, Lund University, Lund, Sweden,
[1]Acoustic Technology, Dept. Electrical Engineering, Technical University of Denmark, Lyngby, Denmark,
*also ENT Department, Lund University Hospital, Lund, Sweden

ABSTRACT

Purpose: To investigate the vocal behavior and voice use in teachers with self-estimated voice problems and their age-, gender and school-matched colleagues without voice problems. The main hypothesis was that teachers with and without voice problems act differently with respect to classroom acoustics and air-quality, and that the vocal doses would separate the groups.

Method*:* This study is a case-control designed field study. Teachers with self-estimated voice-problems from three schools were matched for age- and gender to voice-healthy school-colleagues, n=14 pairs, 12 F/2M. The subjects had been examined for laryngeal-, vocal-, hearing- and psychosocial aspects. The teachers' fundamental frequency, Sound Pressure Level, and phonation-time were recorded with an Ambulatory Phonation Monitor during one workday and they also reported their activities in a structured diary. The ambient noise level was simultaneously recorded with a dosimeter; the room temperature and air quality were also measured. The acoustic properties of the classrooms were measured without any students present.

Results*:* The results showed that teachers with voice problems behaved differently from their voice healthy peers, in particular during teaching sessions. The time and cycle doses were significantly higher in the group with voice problems. Also the F0 pattern, related to both voice level and room acoustics differed between the groups. The results thus suggest a higher vocal load with fewer possibilities for vocal recovery during teaching in subjects with subjective voice problems.

With this paper we intended to investigate teachers' voice use in their work-environment. We aimed at exploring the vocal behavior in a group of teachers with self-assessed voice problems and to compare them to a group of teachers with self-assessed voice health. This is a study within the project "speakers' comfort".

One of the most important aspects of teaching is for the teacher to make her- or himself heard. The demands on a teacher's voice are varied. The voice is needed to communicate, instruct and clarify. The teaching tasks at elementary and middle school levels can vary from soft, facilitating talk during morning assembly, to singing, reading loud, lecturing, and teaching in the sports hall. The most important trait of a teacher's voice is thus to be flexible. However, with high levels of background noise and unfavorable room acoustics, this might be an effortful task that may be detrimental to the voice. Recently published data suggest that very few teachers in Swedish schools have undergone any voice training and that voice amplification is very rare, even in the schools' sports halls (Lyberg Åhlander, Rydell, & Löfqvist, 2010). Thus, teachers are at risk of developing voice problems and there is a high prevalence of voice disorders in teaching staff also when compared to other occupations with vocal demands (Fritzell, 1996; Roy, Merrill, Gray, & Smith, 2005; Smith, Lemke, Taylor, Kirchner, & Hoffman, 1998; Titze, Lemke, & Montequin, 1997). This is also shown by data on sick-leave due to voice problems. In a group of teachers who assess themselves as suffering from voice problems, 35% compared to 9% in a group of voice healthy teachers reported sick-leave due to voice problems

(Lyberg Åhlander, et al. 2010). According to Sapir, Keidar, & Mathers-Schmidt (1993), 0% of a group with no occupational vocal demands reported sick-leave for this reason.

Recent results from comparisons between a group of teachers with self-assessed voice problems and their voice-healthy colleagues (Lyberg Åhlander, et al, submitted} indicated that there were no major differences between the groups in vocal, laryngeal, hearing or psychosocial aspects. This leads us to conclude that, apart from differences in the time needed to recover from voice problems, the differences between teachers with and without voice problems might be found in the their daily voice use, possibly in combination with the teaching environment.

During the last decades, a number of research groups have tried to understand teachers' daily voice use, based on the hypothesis that this behavior might differ from what can be seen in laboratory or clinical settings. Jonsdottir, Laukkanen &Vilkman (2002), Lindström, Ohlsson, Sjöholm, & Waye (2010), Ohlsson, Järvholm, & Löfqvist (1987), Rantala, Paavola, Korkko, & Vilkman (1998). Rantala & Vilkman (1999), Rantala, Vilkman, & Bloigu (2002), and Södersten, Granqvist, Hammarberg, & Szabo (2002), among others, studied the vocal behavior of subjects at their work place. Hunter and Titze (2010) also studied non-occupational time. Parameters that have been used are fundamental frequency, sound pressure level, and phonation (or speaking) time.

The change in fundamental frequency during a workday has been identified as a sign of voice load. Rantala et al. (2002) studied teachers' vocal behavior during a workday, and found a tendency for teachers with few voice complaints to increase their F0 level more than their colleagues with many complaints. Moreover, Jonsdottír, Laukkanen &Vilkman (2002) found a greater F0 and SPL increase in teachers with voice complaints when they used voice amplification compared to when they did not. Laukkanen, Ilomaki, Leppanen, & Vilkman (2008) described the rise of F0 as an increase in muscular activity, most likely an adaptation to vocal loading during a day at work. In addition, they described that the voice changes during vocal loading include a rise of the sound pressure level (SPL) and a decrease of jitter and shimmer. Jonsdottír et al. (2002) suggests that an F0 increase is a healthy reaction to voice load, prone to promote effective voice function.

Several different methods to study the vocal behavior in vivo have been developed during the years (Airo, Olkinuora, &Sala, 2000; Buekers, Bierens, Kingma, & Marres, 1995; Cheyne, Hanson, & Genereux, 2003; Granqvist, 2003; Lindström et al. 2010; Masuda, Ikeda, Manako, & Komiyama, 1993; Ohlsson, Brink, & Löfqvist, 1989; Popolo, Svec, & Titze, 2005; Svec, Popolo, & Titze, 2003; Szabo, Hammarberg, Granqvist, &Södersten, 2003). These devices have used various techniques. The ones in use today are based on accelerometers that record fundamental frequency and sound pressure level from skin vibrations. Using this technique, it is possible to track the speaker's voice also in noisy environments without recording the background noise.

Description of the APM and definition of parameters

The APM, made by Kay-Pentax, is a microprocessor based system, estimating the fundamental frequency, SPL, phonation time, and the number of vibratory cycles from skin vibrations captured by an accelerometer that is attached to the front of the subject's neck, above the sternal notch (Cheyne et al. 2003; Hillman, Heaton, Masaki, Zeitels, & Cheyne, 2006). The software calculates the mode and average of the fundamental frequency in Hz and the mode and average sound pressure level in dB. Based on these measures, Titze, Svec, and Popolo (2003) have defined vocal dose measures. *The time dose* is defined as the total duration of phonation, i.e., the total cumulated time and the percentage of this time spent phonating (see Cheyne et al. [2003] for further information about the APM microprocessor's identification of phonation); t*he cycle dose* is the total number of vibratory cycles during a period of time (Svec et al. 2003). The cycle dose was originally introduced by Rantala & Vilkman (1999) as Vocal Loading Index.

The teacher does not act alone in the classroom. Results from an earlier study showed that 92% of the teachers found the noise from the students to be disturbing (Lyberg Åhlander, Rydell, & Löfqvist

2010). Thus, it is important to consider the effect of the background noise on the teacher's voice. The Lombard effect (Lane & Tranel, 1971) describes the influence of noise on the voice. The speaker automatically raises sound pressure level and changes the spectral contents and the of the voice signal as the noise level increases. The background noise level in classrooms is usually high, also during instruction, as shown by Pekkarinen & Viljanen (1991). Södersten, Ternström, & Bohman (2005) investigated the rise of F0 and SPL due to background noise in healthy subjects. They showed that the speaker increases the SPL and F0 and prolongs the phonation time when exposed to noise, especially continuous noise. In that study, female talkers also reported less success in making themselves heard and greater effort to do so. Ternström, Bohman , & Södersten (2006) measured the spectral balance, the ratio of energy in the frequency bands 2-6 kHz and .1-1 kHz, and found it to be less negative as a function of increasing background noise level and voice SPL. Moreover, Lindström, Persson Waye, Södersten, McAllister, & Ternström (in press) showed that there is a large variation in vocal behavior due to noise exposure. Thus, it is important to study voice use in real life to further understand the vocal behavior and detect possible individual differences in voice use and in the management of vocal load.

Dry air is often mentioned by patients at voice clinics to affect their voices. The dryness of air has been proved to affect vocal prerequisites in laboratory settings (Vintturi, Alku, Sala, Sihvo, & Vilkman, 2003; Leydon, Sivasankar, Falciglia, Atkins, & Fisher, 2009). However, no field studies seem to have been made where the effects of air quality and temperature on voice problems have been examined.

One of the factors that is often mentioned, but seldom studied in relation to the development of voice disorders, is the influence of the room acoustics on the teacher's voice. Pekkarinen &Viljanen (1991) concluded that many Finnish classrooms were too reverberant with a resulting reduced intelligibility, which may cause the speaker use more effort when speaking. Kob, Behler, Kamprolf, Goldschmidt, & Neuschaefer-Rube (2008) studied teachers with different voice status acting in different rooms and concluded that teachers with voice problems were more affected by the acoustic properties of the room than their voice healthy colleagues. For examining the effect of room acoustics on voice use, one useful starting point is the results by Brunskog, Gade, Bellester, & ReigCalbo (2008) on the preferred acoustical properties of a room for a good speaker's comfort. Lacking a measure describing the speaker's perception of the room acoustics, earlier investigations have used measures that focus on the listeners' perspective, like the reverberation time or the speech transmission index (see Rossing, Dunn, Hartmann, Murray Campbell, & Fletcher, 2007). Brunskog & Pelegrín Garcia (2010) introduced a measure of *voice support*. It is a measure based on the two properties of the impulse response defining the airborne acoustic path between the mouth and the ears. These are the direct sound from the mouth to the ears, and the indirect sound from the reflection at the boundaries of the room (Brunskog & Pelegrín García, 2010). Thus, the voice support is the ratio between the energy of the reflected sound (E_r) and the energy of the direct sound (E_d), Equation 1.

$$ST\mathrm{v} = 10\log\frac{E_r}{E_d} \qquad (1)$$

This paper presents the results of the measurements of the teachers' voices with the aim of exploring the vocal behavior in a group of teachers with self-assessed voice problems and to compare them to a group of teachers with self-assessed voice health. A second aim was to relate the vocal behavior to the conditions of the room acoustics, background noise, and air quality in the teaching environment. Some of the results are presented in Pelegrín García, Lyberg Åhlander, Rydell, Löfqvist, & Brunskog (2010). The study is a field study with case-control design.

METHODS

Subjects and schools

The subject group, 28 teachers were recruited among the participants in an earlier study (Lyberg Åhlander, Rydell, & Löfqvist, 2010). Based on self-ratings of the voice, the teachers with self-assessed voice problems were age and gender matched to voice healthy colleagues to form 14 pairs (12 female and 2 male), all non-smokers. For the present study, the schools that had the highest frequency of matched pairs were asked to participate. The teachers worked at three different schools in Lund, teaching primary to high school levels. The schools were approximately equivalent to each other in size, the number of students and staff and were built during the same decade (mid 1960 to mid 1970). Two schools taught all levels from primary to junior high, and one junior high level. All participants underwent an examinations of the larynx, voice, hearing and psychosocial aspects.

The subjects were contacted by phone, were informed about the project, and asked if the still wanted to participate. Written information was sent by e-mail after the contact was established. Both the teachers and the headmasters of the three schools gave their written consent to participate. The teachers were further asked to identify a “typical” workday on which the APM measurement could be performed. The demographics of the teachers are shown in Table I. There were no differences between the groups in age or time in occupation as shown by a t-test.

Table I. *Demographics of n=28 teachers. Group I = teachers with voice disorders. Group II = teachers without voice disorders.*

Group	Gender F/M	Age Median (range)	Time in occupation, Median (range)
Group I, n=14	12/2	41 (24-62)	13 (2-40)
Group II, n=14	12/2	43 (28-57)	18 (2-28)

Materials

The data was collected the Ambulatory Phonation Monitor 3200 vers. 1.04 (APM) (KayPentax New Jersey, USA). The APM measures the phonation time, when the phonation occurred and estimates the teacher’s SPL, and F0 (KayPentax 2009). Based on a pilot study, a diary was developed for the teacher to complete during the day, to track the activities of the teacher. The diary had two sections. The first consisted of nine questions on general information: the number and grade of the students taught, the teaching activities, the distance to and noise-level of students along with one question on voice hygiene (intake of water during the lesson). There was also a 100 mm VA-scale for continuous voice self-assessment where the subjects rated their current voice status (no voice problems - maximum voice problems). The second part consisted of nine questions on voice aspects and one on stress, rated on a categorical scale (not at all; partly; moderately and very much). The voice questions were modified from the VHI-T (Lyberg Åhlander, Rydell, Eriksson, & Schalén, 2010).

Simultaneously with the APM recordings, the noise and voice levels at the teacher’s position were measured with a sound level meter Svantek, mod. SV-102. The signals were picked with a lapel microphone at a distance of 15 cm from the teacher’s mouth. The sound level meter was placed in the same waist-bag as the APM box. The acoustic properties of the classrooms were evaluated with the following acoustic parameters *background noise level*, *reverberation time*, *speech transmission index*, *sound strength* and *room support* while the classrooms were empty, due to logistics. Additionally, the geometrical dimensions of the room were measured. The air humidity, room temperature and the carbon dioxide (CO_2) contents of the air were simultaneously measured during the work-hours with an indoor air quality measuring device: Q-Trak IAQ Monitor Model 8550, TSI Inc, USA, analyzed with Trak Pro Data Analysis Software.

APM-procedures

To avoid possible pollen allergy, data was collected during late January through March 2010. Before the workday started, the APM accelerometer was glued to the subject's neck, just above the sternal notch. The cable, connecting the accelerometer and the APM device, was taped to the back of the neck and thread under the clothes, exiting the garments at waist level. The APM was calibrated following the manufacturer's recommendations. The teacher stood or sat in front of the calibration microphone, with the distance guide (15 cm) resting on the upper lip. The subject was then instructed to phonate on the vowel /a/ from the softest to the loudest phonation possible. The APM device was thereafter put in a waist-bag. The APM was worn by the teacher during the workday and preferably also after work hours. Moreover, the subject was instructed how to complete the diary, which was supposed to be filled out after each lesson together with the VA-scale on current voice status. The voice part was completed on three occasions: after the first lesson, after lunch and just after the removal of the accelerometer.

Statistics and ethical considerations

The statistical analyses were computed using SPSS 18.1. For most continuous variables, paired samples t-tests were used. Chi-Square tests were used when parameters where categorical. One way ANOVA was used to compare variations between activities. The alpha level for all statistical analyses was set to .05. The study has been approved by the Institutional Review Board at Lund University (#248/2008).

RESULTS

Activities and duration-times

The activities during the day were classified into preparation, teaching, lunch (incl. private lunch and lunch with the students), meeting, after work, and exercising. The data were analyzed by comparisons between work time, and time after work and teaching (lessons). There were no differences between the groups for the type of activity during work time, as shown by a chi-square test for independence, nor were there any differences between the groups in duration of each activity, shown by a paired samples t-test. No comparisons between males and females were made due to the small group of men. Furthermore, no comparisons were made for time off-work, due to a too small number of recordings.

Phonation-time

There were significant differences between the groups for percent of voicing during the work-day as shown by a paired t-test: Group I (M=20.9 SD=8.1) and Group II (M=15.5 SD=8.0) t(87)=4.870, p=.0006. For teaching, there were differences between the groups for percent of voicing, Group I (M=23.6 SD=7.1) and Group II (M=17.3 SD=9.0) t(50)=3.929, p=.0003. A one-way ANOVA showed significant differences for percent of voicing between activities for Group I: F(1720, 5116)= 6593, p=.0002. Post-hoc comparisons using the Tukey HSD test indicated that the mean scores for all activities except for "meeting" differed significantly from "teaching": "teaching: (M=23.6 SD=7.1) "preparation/break" (M=18.5 SD=7.2), "lunch" (M=13.5 SD=5.4) "after work" (M=11.9 SD= 5.6). The percent of voicing also differenced significantly between some of the activities in Group II F(1302, 5842)=4.192, p=.001. The post hoc test indicated significant differences between "teaching" (M=17.3, SD=9.0) and the activities "preparation/break" (M=13.4 SD=7.2) and "after work" (M=8.7 SD=5.0). The phonation time during "teaching", thus ranges between 17 and 24% (SD=9/7) in this material. Further, the number of cycles during work-time differed significantly between the groups as shown by a paired t-test: Group I: M=169 921 SD=162 931 and Group II; M=118 946 SD= 101 247 t(93)=2.875, p=.005. A one-way ANOVA showed significant differences between activities for both groups: Group I F(5, 98)=9.623, p=.0001 and Group II: F(6,113)=10.131 p=.0006. Post-hoc comparisons using the Tukey HSD test for Group I indicated that the mean scores for "teaching" (M=202 823, SD=117 202) differed significantly from those of "preparation/break" (M=65 252, SD=46 842) and "after-work" (M=383 158, SD=332 327). The post-hoc test for Group II only showed a significant difference between "teaching" (M=169 829, SD= 93 543) and "preparation/break" (M=47 228, SD=52 955).

F0 and SPL, changes during the day

The variations of F0 and SPL between the different activities were estimated with a one way ANOVA. No significant differences were found for any group. When comparing occupational versus non-occupational time (after work) no significant differences were found for any group for F0 or SPL. The F0 and SPL values are shown in Tables II and III.

***Table II**. Mean values of F0 and SPL for activities during a day for two groups of female teachers: Group I, n=12 teachers with voice problems and Group II n=12 teachers with healthy voices. N denotes number of measured sessions.*

Activity	F0 Hz (Sd)	SPL dB (Sd)
Break/Planning		
Group I (n=21)	226 (17)	70 (5)
Group II (n=24)	235 (23)	73 (8)
Teaching		
Group I (n=46)	237 (25)	72 (5)
Group II (n=41)	245 (29)	75 (7)
Meeting		
Group I (n=5)	224 (16)	66 (2)
Group II (n=11)	233 (31)	73 (8)
Lunch		
Group I (n=7)	241 (35)	71 (7)
Group II (n=13)	235 (22)	75 (9)

***Table III**. Mean values of F0 and SPL for activities during a day for four male teachers from two groups: Group I, n=2 teachers with voice problems and Group II n=2 teachers with healthy voices. N denotes number of measured sessions.*

Activity	F0 Hz (Sd)	SPL dB (Sd)
Break/Planning		
Group I (n=5)	170 (17)	87 (10)
Group II (n=2)	136 (6)	70 (4)
Teaching		
Group I (n=7)	194 (25)	95(9)
Group II (n=5)	169 (20)	76 (5)
Meeting		
Group I (n=1)	-	-
Group II (n=2)	153 (25)	76 (8)
Lunch		
Group I (n=1)	-	-
Group II (n=1)	-	-

Fundamental frequency and sound pressure level

There was a difference between the groups in the direction of the correlation coefficients, correlating F0 to SPL during teaching. Group I: r= -.379 whereas Group II: r= .295. As shown in Figure 1, this indicates that the group with voice problems decreases the F0 when increasing the SPL, but the voice healthy group increases the F0 when increasing the SPL.

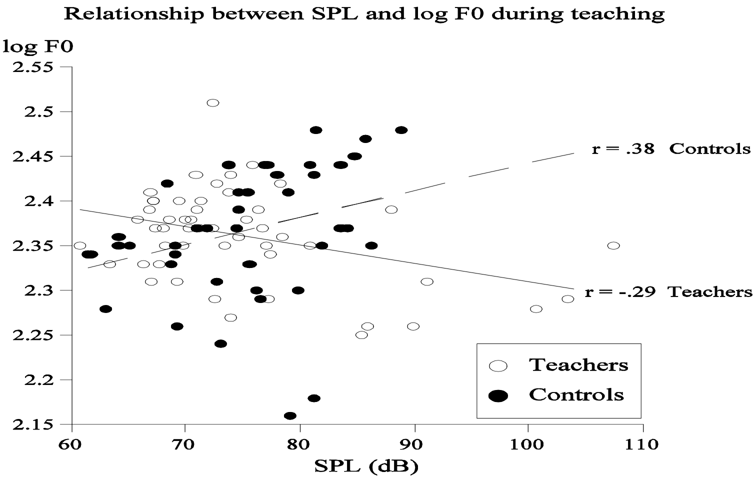

***Figure I.** Plot of sound pressure level and fundamental frequency during teaching.*

Air quality parameters

There were no differences in temperature between the classrooms of the groups, the temperature ranging from 17,3°-25,1°C (Median 22°C). Similarly, there were no significant differences between the rooms for the aspects of air-quality or humidity. The mean for air-quality,CO^2 measured in particles/million (ppm) = 670 (SD =139) and for Humidity, in % = 26% (SD=5,3). A standard multiple regression-model did not find any parameter to be significantly contributing to differences in F0 or

SPL in any group. For a proper comparison of these measurements the back-ground noise needs to be taken into consideration and thus, these results are to be presented elsewhere.

Diary and VAS estimates

There were significant differences between the groups for the following questions as shown by a chi-square test for independence: voice-fatigue; throat-clearing; throat-ache; tenseness of throat; hoarseness; air-loss and stress-level. There were no differences for voice changes during speech, difficulties in making oneself heard, or coughing. The distribution of the answers along with χ^2 and p-values emerge from Table IV a & b.

The estimations of voice problems during the day on a VA scale showed significant differences between the groups according to a paired t-test: Group I (M=32,3 SD=20,8) and Group II (M=11,2 SD=11,8) t(19)=3.441, *p*=.003. An ANOVA showed no differences between teaching-sessions.

Table IV. *The result of Chi-square test for independence of the diary-questions in two groups of teachers: Group I: teachers with voice problems (n=14), Group II: teachers without voice problems (n=14).. Distribution are presented in % . Chi-square values, degrees of freedom and p-values are also provided. Number of answers: Group I: n=42, Group II: n=43.*

a)

Question	**No** (%)	**Partly** (%)	**Moderately** (%)	**Much** (%)	χ^2 **(Df)**	***p***
Do you perceive voice fatigue?					12,245 (3)	0.007
Group I	29	48	19	5		
Group II	58	40	2	0		
Does your voice break or tire?					5,757 (3)	0,12
Group I	64	29	5	2		
Group II	86	12	2	0		
Do you have difficulties in making yourself heard?					0.770 (2)	0,68
Group I	67	29	5	0		
Group II	74	23	2	0		
Do you have a need to clear your throat?					9,647 (3)	0.02
Group I	31	36	17	17		
Group II	44	46	9	0		
Do you have a need to cough?					5,684 (3)	0,128
Group I	57	26	12	5		
Group II	79	14	7	0		
Does your throat ache?					9,088 (3)	0,03
Group I	52	33	9	5		
Group II	81	16	2	0		
Is your throat tense?					10,951 (3)	0,01
Group I	38	43	17	2		
Group II	70	28	2	0		
Do you have a hoarse voice?					6,443 (2)	0,04
Group I	67	14	19	0		
Group II	77	21	2	0		

b)

Question	*(%)*	*(%)*	*(%)*	*(%)*	χ^2 **(Df)**	***p***
Do you have enough air when you talk?	Always	Nearly always	Almost never	Never	9,907(2)	0,007
Group I	52	45	2	0		
Group II	84	16	0	0		
Stress-level	Low	Rel. low	Rel. high	High	8,522 (3)	0,04
Group I	45	31	19	5		
Group II	35	58	7	0		

DISCUSSION

There was a clear difference between the teachers with voice problems and their voice healthy colleagues in their own assessments of vocal aspects during the day. Also the VAS estimations were significantly higher in the group with voice problems. Moreover, the groups seemed to use different strategies for F0 control in relation to increasing sound pressure levels: the group with voice problems did not raise their F0 with increasing SPL whereas the voice healthy group raised the F0 with the SPL-increase. Interestingly, the voice-problem group either kept the F0 stable or decreased it, see Figure 1. According to Gramming, Sundberg, Ternström, Leanderson, and Perkins (1988) a ”normal” increase of F0 is 0.5 semitone/dB.

The rise of the fundamental frequency as a reaction to voice load has been found in other similar studies. Laukkanen et al. (2008), and Jonsdottir, Laukkanen & Vilkman (2002) showed this rise to be more evident after voice amplification. Rantala &Vilkman (1999) found F0 changes to be significant in a group of teachers with few complaints of voice problems but not in the group with many complaints. The difference between the groups in the present study confirms this line of reasoning. We suggest that a rise of F0 should be occurring as a healthy adaption to voice load. A non-rise of F0 might either contribute to or be the consequence of voice problems. We may speculate that the non-occurring F0 rise might reflect a loss of tissue flexibility, a difference in behavior or of neuro-muscular capacity. The measurement of F0 in relation to SPL during a work day is thus an important measure for the detection and objective verification of voice problems.

One of the two questions in the diary that separated the groups the most was, not surprisingly, the question about perceived vocal fatigue. The answers to the question of loss of air while talking (do you have enough air when you talk?) might of course reflect an insufficient breathing technique. We may speculate that loss of air during talking also can mirror an underlying functional incompetency of the vocal apparatus or a compensatory behavior to reduce the phonatory effort.

The measured F0 and SPL values depend also on the teachers’ vocal behavior in relation to the room acoustics and the background noise. These aspects were examined for all the subjects in this study by Pelegrín García, Lyberg Åhlander, Rydell, Brunskog, & Löfqvist (2010). The results showed that both groups are equally affected by noise and behave in accordance with the Lombard effect (e.g., Lane & Tranel, 1971), increasing their voice intensity with increasing background noise. The teachers in Group I talked 1.4 dB softer than their voice healthy colleagues. An interesting finding is the behavior in relation to the room support, where the two groups showed opposite trends. The teachers in Group I decrease the SPL of the voice with increasing support in the classrooms, but the teachers in Group II increase it. The results thus show that different individuals make different use of the room acoustics.

The room support is a measure of the help that the talker gets from the room (Brunskog, Gade, Bellester, & Reig-Calbo (2009). These results are intriguing. One possibility is that teachers experiencing voice problems are more attentive to the room acoustics, because they have to, in order to preserve their vocal health. Another possibility is that they are too affected by their voice problems to be able to behave differently. The combined results of their behavior in relation to the room support and their decrease of F0 in relation to increasing SPL of their own voice might indicate a loss of vocal flexibility, since healthy subjects would raise their F0 with increasing SPL.

We don't know, however, if wearing the APM influenced the teachers' vocal behavior during the day. Apart from some comments on the cables being in the way at some points during the day, the device caused no problems for the subjects. Nor do we know if the measurements influenced the students' behavior. Some of the smaller children reacted when they saw the accelerometer and even told their peers: "better keep the voice low today, miss xxx has a sore throat". However, the teachers did not note any differences in the behavior of the children. Perhaps more importantly, the measurements were made during a single day. The day was chosen by the subjects for being a "normal" day at work. However, Masuda, Ikeda, Manako, and Komiyama (1993), concluded that their measurements of 29 subjects over several days did not vary between days within the one and same subject. Based on these findings, the chosen day is most likely representative of the subjects' daily pattern of voice use.

Cycle and time doses

The time and cycle doses are two of the potentially most useful measures for understanding the stresses on the tissue of the vocal folds during phonation (Svec, Popolo, & Titze, 2003; Titze, Svec, & Popolo, 2003). There were significant differences between the groups for both doses in our material. The percentage of phonation was significantly greater in the voice problem group (24% vs 17%) which confirms the findings by Rantala & Vilkman (1999). Our result of 17-24% of percent of voicing is in line with the findings of others for the teachers' time at work. Earlier studies have reported of phonation-time in teachers. Masuda et al. (1993) reported a phonation-time of 20%, Titze, Hunter, & Svec (2007) a phonation-time of 23% and in a recent study Hunter & Titze (2010) reported phonation-times as high as 30%, +/-11%. Södersten et al. (2002) reported a phonation-time of 16.9% in pre-school teachers. When making comparisons of phonation-time between countries it is, however, necessary to take into account possible differences in teaching methods. In Sweden today, co-teaching is rather common, especially at the elementary and middle school levels. Moreover, there is a general paradigm shift towards a more student-focused teaching style. To teach under such circumstances means less need to lecture in a traditional manner and probably also means better possibilities for vocal rest.

Also the cycle dose differed significantly between the groups for the time at work, with the higher dose in the group of teachers with voice problems. This measure was originally introduced as Vocal Loading Index by Rantala & Vilkman (1999) and had a moderate correlation with the voice complaints in their subjects. That is, the more voice complaints, the higher the VLI values. The selection of the teachers for the two groups in the present study was based on the teachers' own assessment of voice problems (Lyberg Åhlander, Rydell & Löfqvist, 2010). A higher cycle dose in the group with voice problems may thus indicate the usefulness of the cycle dose as measure of vocal load.

Recovery

The importance of pauses, both long and short, has been identified in relation to voice recovery after vocal load (Carroll, Nix, Hunter, Emerich, Titze & Albaza, 2006; Hunter & Titze, 2009; Titze, Hunter & Svec, 2007; Vintturi, Alku, Lauri, Sala, Shivo & Vilkman, 2001). Short pauses occur during breathing and swallowing. An earlier study by Lyberg-Åhlander, Rydell & Löfqvist (2010) concluded that the subjects with voice problems, also included in the present study, reported significantly longer times for recovery after voice load than their voice-healthy colleagues. In the present study, the difference in both vocal- and time dose between the groups indicate a difference in vocal load. This may also reflect a difference in the pausing during phonation. Carroll, Nix, Hunter, Emerich, Titze, & Abaza (2006) found that singers rated their vocal effort and inability to produce soft voice harsher 24-72 hrs after a peak of voce load. The authors speculate that this time lag might represent injury and healing of the lamina propria. Our results of the higher vocal dose in the teachers with voice problems and the difference in estimated recovery times between the vocally affected and the voice healthy groups might indicate that there are changes taking place at a micro-level.

The lunch-break may be considered to be a time for pause and recovery. Lindström et al. (2010) observed a decrease in F0 during lunch time. We did not find such a decrease. Instead there were peaks for both SPL and F0 during lunch-time, probably due to a number of teachers having lunch with the children in so called "pedagogic lunches".

Air quality and temperature

Albeit, there were no differences in temperature between the classrooms of the groups, with temperature ranging from 17,3°-25,1°C, the high temperature in some classrooms is still worth to comment on. The present measurements were made during the winter, which means that the temperature comes from indoor heating. There is evidence that mild heat might be subduing (Hygge, 1991), making the children sleepy and un-focused. Thus, most probably the higher temperatures may cause the students to be noisier, due to their need to stay alert which contribute to the voice-load.

The mean CO2 levels were below the Swedish regulation for indoor work, 1000 ppm (AFS, 2009), but, in a few rooms the CO2 level exceeded the stipulated maximum value. The air-humidity measures are more complicated compared to the other measures, as it is not possible to set a limit value (Swedish occupational safety and health administration). The mean humidity estimate was low, 26%, which is normal during the winter in Sweden (AFS, 2009). However, Sivasankar & Fisher (2003) and Sivasankar, Erickson, & Schneider (2008) conclude that oral breathing increases the phonation threshold which in turn, results in increased vocal effort. Thus, a dryer indoor climate may increase the vocal load.

To closer investigate the contributions from the indoor-climate, the results of the measurements of air-quality and temperature in this study may be correlated to changes in F0 and SPL levels. However, the results need to be discussed in relation to the back-ground noise. This is beside the scope of this paper and, such correlations thus will be discussed elsewhere.

Conclusion

The APM measurements of two groups of teachers showed that teachers with self-estimated voice problems differed from their age-, gender- and school-matched voice healthy peers in several aspects of voice use in particular during teaching sessions. The time- and cycle doses were both significantly higher in the group with voice problems. This suggests a higher vocal load with fewer opportunities for vocal recovery during teaching. Moreover, the pattern of F0 changes in relation to both room acoustics and the SPL of the voice differed between the groups, possibly indicating a reduced vocal flexibility in the group with voice problems.

Acknowledgement

The authors wish to express their gratitude to reg. speech pathologist Susanna Whiting who through her master thesis helped designing the procedure for the collection of data. This work was supported by AFA.

REFERENCES

Airo, E., Olkinuora, P., & Sala, E. (2000). A method to measure speaking time and speech sound pressure level. *Folia Phoniatrica et Logopaedica, 52*, 275-288.

AFS 2009:02 Arbetsplatsens utformning, [work environment design] (2009).

Brunskog, J., Gade, A. C., & Payá Bellester, G., & **Reig-Calbo, L.** (2009). Increase in voice level and speaker comfort in lecture rooms. *Journal of the Acoustical Society of America, 125*, 2072-2083.

Brunskog, J., & Pelegrín García, D. (2010). Speaking comfort and voice use of teachers in classrooms. *Italian Journal of Acoustics, 34, 51-56.*

Buekers, R., Bierens, E., Kingma, H., & Marres, E. H. (1995). Vocal load as measured by the voice accumulator. *Folia Phoniatrica et Logopaedica, 47*, 252-261.

Carroll, T., Nix, J., Hunter, E., Emerich, K., Titze, I., & Abaza, M. (2006). Objective measurement of vocal fatigue in classical singers: A vocal dosimetry pilot study. *Otolaryngology Head and Neck Surgery, 135*, 595-602.

Cheyne, H. A., Hanson, H. M., & Genereux, R. P. (2003). Development and testing of a portable vocal accumulator. *Journal of Speech, Language & Hearing Research, 46*, 1457-1468.

Fritzell, B. (1996). Voice disorders and occupations. *Logopedics Phoniatrics Vocology, 21*, 7 - 12.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice, 2*, 118-126.

Granqvist, S. (2003). The self-to-other ratio applied as a phonation detector for voice accumulation. *Logopedics Phoniatrics Vocology, 28*, 71-80.

Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., & Cheyne, H. A. (2006). Ambulatory monitoring of disordered voices. *Annals of Otology, Rhinology, and Laryngology, 115*, 795-801.

Hunter, E. J., & Titze, I. R. (2009). Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading exercise. *Annals of Otology, Rhinology, and Laryngology, 118*, 449-461.

Hunter, E. J., & Titze, I. R. (2010). Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings. *Journal of Speech Language, and Hearing Research, 53*, 862-875.

Hygge, S. (1991). The interaction of noise and mild heat on cognitive performance and serial reaction time. *Environment International, 17*, 239-244.

Jonsdottir, V., Laukkanen AM., & Vilkman, E. (2002). Changes in teachers' speech during a working day with and without electric sound amplification. *Folia Phoniatica et Logopaedica, 54*, 282-287.

KayPENTAX (2009) Software instruction manual, Ambulatory Phonation Monitor, mod 3200. New Jersey, USA.

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

Lane, H., & Tranel, B. (1971**).** The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14*, 677-709.

Laukkanen, A. M., Ilomaki, I., Leppanen, K., & Vilkman, E. (2008). Acoustic measures and self-reports of vocal fatigue by female teachers. *Journal of Voice, 22*, 283-289.

Leydon, C., Sivasankar, M., Falciglia, D. L., Atkins, C., & Fisher, K. V. (2009). Vocal fold surface hydration: A review. *Journal of Voice, 23*, 658-665.

Lindström, F., Ohlsson, A. C., Sjöholm, J., & Persson Waye, K. (2010). Mean F0 values obtained through standard phrase pronunciation compared with values obtained from the normal work environment: A study on teacher and child voices performed in a preschool environment. *Journal of Voice, 24*, 319-323.

Lindström, F., Persson Waye, K., Södersten, M., McAllister, A., & Ternström, S. Observations of the relationship between noise exposure and preschool teacher voice usage in day-care center environments. *Journal of Voice,* in Press.

Lyberg Åhlander, V., Rydell, R., & Löfqvist, A. (submitted). *How do teachers with self-reported voice problems differ from their peers with self-reported voice health?* Manuscript submitted for publication.

Lyberg Åhlander, V., Rydell, R., & Löfqvist, A. (2010). Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff. *Journal of Voice*, in press.

Lyberg Åhlander, V., Rydell, R., Eriksson, J., & Schalén, L. (2010). Throat related symptoms and voice: Development of an instrument for self assessment of throat-problems. *BMC Ear, Nose and Throat Disorders, 10*, 5.

Masuda, T., Ikeda, Y., Manako, H., & Komiyama, S. (1993). Analysis of vocal abuse: Fluctuations in phonation time and intensity in 4 groups of speakers. *Acta Otolaryngologica, 113*, 547-552.

Ohlsson, A.-C., Brink, O., & Löfqvist, A. (1989). A voice accumulator - validation and application. *Journal of Speech and Hearing Research, 32*, 451-457.

Ohlsson AC, Järvholm B, & Löfqvist, A (1987). Vocal symptoms and vocal behaviour in teachers. *Scandinavian Journal of Logopedics and Phoniatrics., 12*, 61-69.

Pekkarinen, E., & Viljanen, V. (1991). Acoustic conditions for speech communication in classrooms. *Scandinavian Audiology, 20*, 257 - 263.

Pelegrín García, D., Lyberg Åhlander, V., Rydell, R., Löfqvist, A., & Brunskog, J. (2010). Influence of classroom acoustics on the voice levels of teachers with and without voice problems: A field study. *Proceedings of Meetings on Acoustics, Vol. 11*. (available at: http://scitation.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=PMARCW000011000001060001000001&idtype=cvips, last visited 12/16.10.)

Popolo, P., Svec, J. G., & Titze, I. R. (2005). Adaptation of a pocket pc for use as a wearable voice dosimeter. *Journal of Speech, Language & Hearing Research, 48*, 780-792.

Rantala, L., Paavola, L., Korkko, P., & Vilkman, E. (1998). Working-day effects on the spectral characteristics of teaching voice. *Folia Phoniatrica et Logopaedica, 50*, 205-211.

Rantala, L., & Vilkman, E. (1999). Relationship between subjective voice complaints and acoustic parameters in female teachers' voices. *Journal of Voice, 13*, 484-495.

Rantala, L., Vilkman, E., & Bloigu, R. (2002). Voice changes during work - subjective complaints and objective measurements for female primary and secondary schoolteachers. *Journal of Voice, 16*, 344-355.

Rossing, T. D., Dunn, F., Hartmann, W. M., Murray Campbell, D., & Fletcher, N. H. (2007). *Springer Handbook of Acoustic.s* New York: Springer-Verlag.

Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *Laryngoscope, 115*, 1988-1995.

Sapir, S., Keidar, A., & Mathers-Schmidt, B. (1993). Vocal attrition in teachers: Survey findings. *European Journal of Disordered Communication, 28*, 177 - 185.

Sivasankar, M., & Fisher, K. V. (2003). Oral breathing challenge in participants with vocal attrition. *Journal of Speech Language, and Hearing Research, 46*, 1416-1427.

Sivasankar, M., Erickson, E., & Schneider, S. (2008). Phonatory effects of airway dehydration: preliminary evidence for impaired compensation to oral breathing in individuals with a history of vocal fatigue. *Journal of Speech, Language, and Hearing Research, 51*, 1494-1507.

Smith, E., Lemke, J., Taylor, M., Kirchner, H. L., & Hoffman, H. (1998). Frequency of voice problems among teachers and other occupations. *Journal of Voice, 12*, 480-488.

Svec, J. G., Popolo, P. S., & Titze, I. R. (2003). Measurement of vocal doses in speech: Experimental procedure and signal processing. *Logopedics, Phoniatrics, Vocology, 28*, 181-192.

Szabo, A., Hammarberg, B., Granqvist, S., & Sodersten, M. (2003). Methods to study pre-school teachers' voice at work: Simultaneous recordings with a voice accumulator and a DAT recorder. *Logopedics, Phoniatrics, Vocology, 28*, 29-39.

Södersten, M., Granqvist, S., Hammarberg, B., & Szabo, A. (2002). Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural dat recordings. *Journal of Voice, 16*, 356-371.

Södersten M, Ternström S, & Bohman M. (2005). Loud speech in realistic environmental noise: Phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers. *Journal of Voice, 19*, 29–46.

Ternström, S., Bohman, M., & Södersten, M. (2006). Loud speech over noise: Some spectral attributes, with gender differences. *Journal of the Acoustical Society of America, 119,* 1648-1665.

Titze, I. R., Lemke, J., & Montequin, D. (1997). Populations in the U.S. workforce who rely on voice as a primary tool of trade: A preliminary report. *Journal of Voice, 11*, 254-259.

Titze, I. R., Svec, J. G., & Popolo, P. S. (2003). Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *Journal of Speech, Language, and Hearing Research, 46*, 919-932.

Titze, I. R., Hunter, E. J., & Svec, J. G. (2007). Voicing and silence periods in daily and weekly vocalizations of teachers. *The Journal of the Acoustical Society of America, 121*, 469-478.

Vintturi, J., Alku, P., Lauri, E. R., Sala, E., Sihvo, M., & Vilkman, E. (2001). The effects of post-loading rest on acoustic parameters with special reference to gender and ergonomic factors. *Folia Phoniatrica et Logopaedica, 53*, 338-350.

Vintturi, J., Alku, P., Sala, E., Sihvo, M., & Vilkman, E. (2003). Loading-related subjective symptoms during a vocal loading test with special reference to gender and some ergonomic factors. *Folia Phoniatrica et Logopaedica, 55*, 55-69.

Measurement and prediction of acoustic conditions for a talker in school classrooms

David Pelegrín-García[a] and Jonas Brunskog
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark

Viveka Lyberg-Åhlander and Anders Lofqvist
Dept. Logopedics, Phoniatrics & Audiology, Lund University, Lund, Sweden

(Dated: March 29, 2011)

Objective acoustic parameters were measured in 30 school classrooms. These parameters include usual descriptors of the acoustic quality from the listeners' standpoint, such as reverberation time, speech transmission index, sound strength, and background noise levels, and two descriptors of the acoustic properties for a talker: voice support and room gain. The paper describes the measurement method for these two parameters and presents a prediction model for voice support and room gain derived from the diffuse field theory. The voice support for the medium-sized classrooms with volumes between 100 and 250 m^3 lies in the range between -14 and -9 dB, whereas the room gain is in the range between 0.2 and 0.5 dB. The prediction model for voice support describes the measurements in the classrooms with a coefficient of determination of 0.84 and a standard deviation of 1.4 dB. Regression models for early and late sound strength show that the level of early reflections decrease with distance as previously reported in the literature. However, the level of late reflections after 50 ms is apparently uniform at all the positions in the room.

PACS numbers: 43.55.Hy, 43.70.Mn

I. INTRODUCTION

Learning spaces or classrooms are environments where people spend a large amount of their lifetime, mainly dedicated to acoustic communication tasks. Students spend the early part of their lives listening to the teacher in classrooms in order to learn, and also need to communicate efficiently with other fellow students and the teacher. The success in communication is fundamental to develop the full potential of every student. At the same time, school classrooms are the working place of teachers, who represent an important percentage of the working population. Acoustical conditions have to be evaluated both for teachers and students.

Most of past research in classroom acoustics has been devoted to the acoustic design for students. The negative effects of noise on children perception and performance have been observed,[1–3] the effect of reverberation on speech intelligibility has been quantified,[4,5] and the combination of noise and reverberation has been object of a number of studies.[6–9] Different magnitudes are used to predict speech intelligibility: signal-to-noise ratios, useful-to-detrimental ratios, and speech transmission index (STI).[10,11]

Acoustic conditions are also important for teachers. Teachers suffer from voice disorders in a higher proportion than in the rest of the population[12] (around 13% in Sweden[13] and a similar proportion in the US[12]), which is most likely due to the high vocal requirements that the teaching occupation demands. Noise and bad classroom acoustics are often reported risk factors for voice disorders.[14] Talking in the presence of high noise levels results in the use of higher voice power levels than required to talk in soft noise conditions. This is known as the Lombard effect,[15] and it is estimated that for each dB of noise, the talker raises his voice power level between 0.5 and 0.7 dB. In the presence of low background noise, talkers still modify their voice power under different room acoustic conditions,[16] even when the distance between talker and listener is kept constant.[17]

The magnitude *room gain*[16–18] has a negative correlation with the voice power levels used by talkers in different rooms. The room gain is defined as the gain applied by the room to the voice of the talker at his own ears, relative to free-field. However, this magnitude has a low dynamic range, and the use of *voice support* seems more appropriated in room acoustics.[18] The voice support is conceptually equivalent to Gade's objective support[19] used in the assessment of the acoustic conditions for musicians in concert halls.

A number of surveys have analyzed the acoustic conditions of school classrooms based on measurements of reverberation time and background noise,[20] many times reporting measures of speech intelligibility.[7,21,22] The decay of speech level with distance has received some attention,[23,24] because it is important for the prediction of the signal-to-noise ratio at different positions in the classroom.

Despite the importance of assessing the acoustic conditions for a talker, there are no studies that report room gain, voice support or other talker-related parameters in school classrooms. Some studies advice about possible detrimental effects of poor acoustic conditions on vocal loading.[22,25] The present paper aims at providing some

---

[a] Electronic address: `dpg@elektro.dtu.dk`

information in this respect, giving some reference values for voice support and room gain in typical school classrooms. In addition, the two parameters are explained in more detail than previously reported[18] and a prediction model based on the diffuse-field theory is presented.

II. THEORY

A. Definition and calculation of voice support and room gain

Brunskog *et al.*[16] introduced the parameters room gain and voice support, and Pelegrin-Garcia[18] suggested an alternative method for the calculation of these two parameters from a single impulse response. The procedure followed in the present paper is based on the latter approach, although it is refined regarding the frequency weighting. Given the IR (impulse response) measured with a dummy head between the mouth and the ears $h_{\mathrm{ME}}(t)$, the room gain G_{RG} is defined as difference between the total energy level of the IR L_t and the energy level of the direct sound L_d. The voice support ST_V is defined as the difference between the energy level of the reflections coming back from the boundaries L_r and the energy level of the direct sound,

$$G_{\mathrm{RG}} = L_t - L_d, \tag{1}$$

$$ST_V = L_r - L_d. \tag{2}$$

Assuming that the total energy is the sum of the direct and reflected energies,

$$G_{\mathrm{RG}} = 10\log\left(10^{\frac{ST_V}{10}} + 1\right). \tag{3}$$

The practical calculation of the voice support from the IR is illustrated in the diagram of Fig. 1.

The IR measured between mouth and ears is split into two branches: the top one is multiplied by a window $w_d(t)$ to extract the direct sound $h_d(t)$ and the low branch is multiplied by a window $w_r(t)$ to extract the reflected sound $h_r(t)$. The two window functions are defined as

$$w_d(t) = \begin{cases} 1 & t < 4.5 \text{ ms} \\ 0.5 + 0.5\cos\left(2\pi(t-t_0)/T_W\right) & 4.5 \text{ ms} < t < 5.5 \text{ ms} \\ 0 & t > 5.5 \text{ ms} \end{cases} \tag{4}$$

$$w_r(t) = 1 - w_d(t) \tag{5}$$

with $T_W = 1$ ms. To separate the direct and the reflected components, the mouth/source and the ears/receivers must be located at least one meter away from reflecting surfaces or scattering objects other than the dummy head and the mounting elements. The window for the reflected sound is intended to include all the decaying energy of the IR, because all of it contributes to increase the loudness of one's own voice.

TABLE I. Relevant frequency-dependent quantities used in the prediction model of voice support.

Band i	1	2	3	4	5	6
Frequency [Hz]	125	250	500	1000	2000	4000
Typical speech SPL on-axis at 1 m						
$L_{d,1m}$ [dB]	44.9	57.3	61.8	58.2	53.7	48.9
Difference with SPL at eardrum						
$L_d - L_{d,1m}$ [dB]	13.1	11.8	11.7	13.5	15.3	14.1
Typical speech levels at the eardrum						
$L_{\mathrm{ref,ears}}$ [dB]	58.0	69.1	73.5	71.7	69.0	63.0
Relation between L_W and on-axis SPL at 1 m						
$L_{d,1m} - L_W$ [dB]	-9.5	-8.1	-9.2	-9.5	-7.0	-6.0
Constant K for model Eq. (18)						
K [dB]	3.6	3.7	2.5	4.0	8.3	8.1
Directivity of human speech on downward direction						
Q^* [dB]	0.95	0.78	0.79	0.60	0.21	0.25
Diffuse field HRTF						
ΔL_{HRTF} [dB]	0	0	2	4	11	13

The next stage in the diagram of Fig. 1 is the spectral analysis. The direct sound IR $h_r(t)$ is decomposed into narrow band components $h_{r,i}(t)$ by using a filterbank composed of six one-octave band filters with the standardised center frequencies between 125 Hz ($i = 1$) and 4 kHz ($i = 6$). The energies $E_{d,i}$ and energy levels $L_{d,i}$ are calculated for each band. The same spectral analysis is applied to the reflected sound. The energy levels for the direct sound $L_{d,i}$ are subtracted from the reflected sound $L_{r,i}$, obtaining the values of voice support in one-octave bands $ST_{V,i}$. These values are weighted with a typical speech spectrum at ears $L_{\mathrm{ref,ears}}$ shown in Table I. These levels have been determined from typical anechoic speech levels on-axis at 1 m[26] and the relation between the SPL on-axis at 1 m $L_{d,1m}$ and the SPL at the eardrum measured in anechoic chamber L_d. The overall weighted reference direct sound level $\tilde{L}_d$ and reflected sound level $\tilde{L}_r$ are

$$\tilde{L}_d = 10\log\left(\sum_{i=1}^{6} 10^{\frac{L_{\mathrm{ref,ears},i}}{10}}\right) \tag{6}$$

$$\tilde{L}_r = 10\log\left(\sum_{i=1}^{6} 10^{\frac{L_{\mathrm{ref,ears},i}+ST_{V,i}}{10}}\right) \tag{7}$$

From which the overall speech-weighted voice support (or simply, voice support) is finally calculated as

$$ST_V = \tilde{L}_r - \tilde{L}_d = 10\log\frac{\sum_{i=1}^{6} 10^{\frac{L_{\mathrm{ref,ears},i}+ST_{V,i}}{10}}}{\sum_{i=1}^{6} 10^{\frac{L_{\mathrm{ref,ears},i}}{10}}}. \tag{8}$$

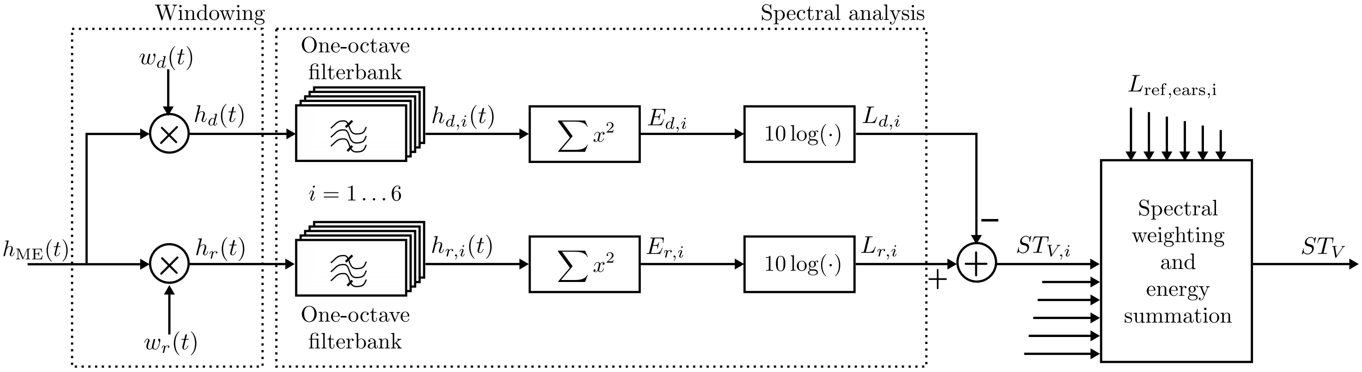

FIG. 1. Block diagram for the calculation of the parameter voice support.

B. Prediction model

Using the definition of ST_V in Eq. (2), a prediction model must account for the relation between the direct and the reflected sound at the ears, when the mouth acts as a source. To build this model, it is assumed that the measurement equipment is a HATS (head and torso simulator) B&K (Brüel & Kjær Sound & Vibration Measurement A/S; Nærum, Denmark) type 4128.

In general, the sound pressure level (SPL) caused by a point source with sound power level L_W at a distance r in free-field (direct sound level, L_d) is

$$L_d = L_W + 10\log\left(\frac{Q}{4\pi r^2}\right), \tag{9}$$

where Q is the directivity of the source. If the source is radiating into half-space (e.g. due to the presence of a reflective plane, like a typical floor) Q becomes 2. When this source is placed in a room, the SPL increases due to sound reflections at the boundaries. The SPL in a room L_p becomes

$$L_p = L_W + 10\log\left(\frac{Q}{4\pi r^2} + \frac{4}{R}\right), \tag{10}$$

where $R = S\bar{\alpha}/(1-\bar{\alpha})$ is sometimes called "room constant", S is the total surface area of the room and $\bar{\alpha}$ is the mean absorption coefficient, which is derived from the volume V and the reverberation time T measurements through Sabine's formula $\bar{\alpha} = 0.161V/(S \times T)$. The reflected sound level, L_r, due to the reflections alone is

$$L_r = L_W + 10\log\left(\frac{4}{R}\right). \tag{11}$$

For predicting ST_V, it would be enough to substitute Eqs. (9) and (11) into (2). However, there are three factors that make the calculation of ST_V slightly different:

1. Modeling of the direct sound

To account for the special propagation between mouth and ears, instead of using Eq. (9), L_d is related to L_W through

$$L_d = L_W + K, \tag{12}$$

where

$$K = (L_d - L_{d,1m}) + (L_{d,1m} - L_W). \tag{13}$$

By introducing this pair of terms, the value of K is decomposed in two quantities. The first quantity, $(L_d - L_{d,1m})$, is determined by the simultaneous SPL measurement at the ears and one meter in front of the mouth of a HATS B&K 4128 reproducing pink noise in an anechoic chamber. The second quantity, $(L_{d,1m} - L_W)$, is determined from the speech directivity patterns measured by Chu and Warnock.[27] The values of the two quantities and K in the different frequency bands are shown in Table I.

2. Ground reflection

The level of a sound reflection from the ground L_{refl} would be

$$L_{\mathrm{refl}} = L_W + 10\log\left(\frac{Q^*}{4\pi(2d)^2}\right) \tag{14}$$

at the position of the source, which is at a height d from the ground. Q^* is the directivity factor of speech in the downward direction (derived from Chu and Warnock[27]) and its frequency-dependent values are shown in Table I. The height d can be regarded as 1.5 m, which corresponds to the mouth position of a standing female talker.

In these conditions, the expected amount of reflected sound at the position of the dummy head (without it disturbing the sound field) would be

$$L_r = L_W + 10\log\left(\frac{4}{R} + \frac{Q^*}{4\pi(2d)^2}\right). \tag{15}$$

3. HRTF correction

Actually, the artificial head used for measurements disturbs the sound field. Therefore, it is necessary to apply a correction term that relates the SPL at the measurement

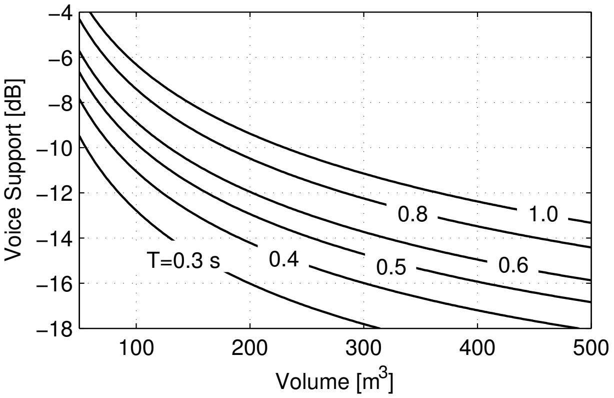

FIG. 2. Voice support versus room volume for a room of proportions 28:16:10 according to the predictions of the model in Eq. (18), for different values of reverberation time.

position when the equipment is present to the SPL at the same position in the absence of the equipment. In the case of the HATS, this correction corresponds to the definition of the *head related transfer function* (HRTF) and is notated as ΔL_{HRTF}. This magnitude is usually direction dependent. As the reflected sound can arrive from many different directions, a direction averaged quantity—the diffuse field ΔL_{HRTF}—given by the manufacturer[28] is used (see Table I).

Therefore, the reflected sound measured with the HATS is

$$L_r = L_W + 10\log\left(\frac{4}{R} + \frac{Q^*}{4\pi(2d)^2}\right) + \Delta L_{\mathrm{HRTF}}. \quad (16)$$

Finally, combining Eqs. (12) and (16) into (2), the frequency-dependent model for voice support is

$$ST_V = 10\log\left(\frac{4}{R} + \frac{Q^*}{4\pi(2d)^2}\right) + \Delta L_{\mathrm{HRTF}} - K, \quad (17)$$

or in terms of directly measurable variables

$$ST_V = 10\log\left(\frac{24.8T}{V} - \frac{4}{S} + \frac{Q^*}{4\pi(2d)^2}\right) + \Delta L_{\mathrm{HRTF}} - K. \quad (18)$$

The results from the individual bands should be weighted to obtain a single value by means of Eq. (8). Figure 2 shows an example set of curves for calculating ST_V from V and T, assuming that the room has proportions 28:16:10 and the reverberation time has a flat frequency characteristic.

III. MATERIAL AND METHOD

Acoustic measurements of the objective parameters T, sound strength G, STI, ST_V, G_{RG}, and background noise level were performed in 30 unoccupied but totally furnished school classrooms. The physical dimensions of the rooms are shown in Table II. According to the volume, the rooms were classified into three groups: small ($V < 100\mathrm{m}^3$), medium ($100 < V < 500\mathrm{m}^3$), and large ($V > 3500\mathrm{m}^3$) classrooms. The rooms in this last group were sports halls where gymnastic lessons took place.

TABLE II. Average (standard deviation) dimensions and volumes of the measured school classrooms

Group size	W (m)	L (m)	H (m)	Volume (m^3)
Small	4.5 (0.8)	3.3 (0.8)	2.7 (0.2)	40.6 (18.3)
Medium	8.9 (1.7)	7.0 (1.0)	2.8 (0.2)	180.2 (61.6)
Large	23.6 (2.4)	20.8 (0.2)	7.4 (0.7)	3614.3 (77.0)

A. Background noise level measurements

The A-weighted, 10-second equivalent background noise levels ($L_{N,Aeq}$) were measured in the empty classrooms using the 01dB Symphonie system with two microphones B&K type 4192. For each classroom, the measurements across four positions were averaged.

B. Measurements with an omnidirectional sound source

1. Reverberation time and speech transmission index

T and STI were derived from the measurements of the room IR $h_{\mathrm{RIR}}(t)$ using an omnidirectional sound source B&K type 4295 "Omnisource". The source was placed at two different teaching positions and with the radiating opening at a height of 1.6 m pointing upwards. Two 1/2" pressure-field microphones B&K type 4192 were used as receivers and were placed close to student seats at a height of 1.2 m. The 01dB (01dB-Metravib; Limonest Cedex, France) Symphonie system, incorporating the MLS software module, was used to produce the measurement signal and send it to the loudspeaker via a power amplifier, acquire the signal from the microphones, calculate the IR, and derive the parameters T and index STI. T was obtained by evaluating the backwards integrated curve[29] of the room IR in the decay interval from -5 to -25 dB. A single value descriptor corresponding to the average on the frequency bands between 500 Hz and 2 kHz T_{500-2k} is given. The average (SD) values of the signal-to-noise ratio of the IR measurements in the different classroom groups were 52 dB (4.1 dB) in small classrooms, 46 dB (5.4 dB) in middle classrooms, and 34 dB (5.4 dB) in sports halls.

2. Sound strength

The sound strength G^{30} was derived from the same IR measurements as T and STI. However, the amplifier levels were not kept constant during the measurements and therefore it was necessary to post-process the room IR files. An auxiliary IR was used: the free-field response of the loudspeaker $h_{\mathrm{dir}}(t)$, measured at a distance of 10 m, with the direct sound peak placed at $t = 0$, and the peak amplitude normalized to 1. The diagram to extract the sound strength due to the early reflections arriving in the first 50 ms G_{er} and the late reflections arriving after 50 ms G_{late} is shown in Fig. 3.

First, $h_{\mathrm{RIR}}(t)$ was analyzed to extract the time location t_0 of the peak corresponding to the direct sound, which had an amplitude p_{peak}. This value of t_0 was used to normalize the amplitude of the whole signal so that the direct sound peak had an amplitude $1/(c \cdot t_0)$, where c is the speed of sound in air at normal condition. In addition, $h_{\mathrm{dir}}(t)$ was scaled to a peak amplitude of $1/(c \cdot t_0)$ and delayed so that the peak was located at t_0. The scaled-delayed version of the direct sound was subtracted from the scaled room IR to obtain an IR corresponding to the reflections $h_{\mathrm{refl}}(t)$,

$$h_{\mathrm{refl}}(t) = \frac{1}{p_{\mathrm{peak}} \cdot c \cdot t_0} h_{\mathrm{RIR}}(t) - \frac{1}{c \cdot t_0} h_{\mathrm{dir}}(t - t_0). \quad (19)$$

The choice of this method (instead of windowing) for separating the reflected sound from the direct sound is due to the fact that, at long distances, the path difference between direct and reflected sound is too small to separate the two components by windowing. It is assumed that the direct sound preserves the same shape at all distances and all rooms with only a change in amplitude and delay. This is not completely true, due to air absorption and atmospheric differences among measurement conditions, but its effect is regarded as small.

The direct sound at 10 m, $1/10 \times h_{\mathrm{dir}}(t)$ was analyzed in the six one-octave bands between 125 Hz and 4 kHz, and the resulting energy levels (re 1) L_{ref} were used as the reference levels for the computation of G. The early reflections were determined from $h_{\mathrm{refl}}(t)$ by applying a window function $w_{\mathrm{er}}(t)$, which is one from 0 to $t_0 + 50$ ms and zero afterwards. The late reflections were determined by applying the complementary window function $w_{\mathrm{late}}(t) = 1 - w_{\mathrm{er}}(t)$ to $h_{\mathrm{refl}}(t)$. The energy levels for early (L_{er}) and late (L_{late}) reflections were calculated. G_{er} and G_{late} were calculated by subtracting L_{ref} from L_{er} and L_{late},

$$G_{\mathrm{er}} = L_{\mathrm{er}} - L_{\mathrm{ref}} \quad (20)$$

$$G_{\mathrm{late}} = L_{\mathrm{late}} - L_{\mathrm{ref}}. \quad (21)$$

The sound strength for the reflected sound G_{refl} is

$$G_{\mathrm{refl}} = 10 \log \left(10^{G_{\mathrm{er}}/10} + 10^{G_{\mathrm{late}}/10} \right) \quad (22)$$

The sound strength for the direct sound G_{dir} is

$$G_{\mathrm{dir}} = 10 \log(100/d^2). \quad (23)$$

From the previous quantities, G is calculated as

$$G = 10 \log \left(10^{G_{\mathrm{dir}}/10} + 10^{G_{\mathrm{refl}}/10} \right) \quad (24)$$

Usually, G is described with a single-value, namely the average of the values at 500 Hz, 1 kHz, and 2 kHz (G_{500-2k}).

C. Measurements with a dummy head

ST_V was calculated from the measurement of an IR corresponding to the airborne sound transmission between the mouth and the ears in the empty classrooms. For this purpose, a HATS B&K type 4128 was used. The HATS included a loudspeaker at its mouth, and microphones at its ears. The HATS was placed at a representative teaching position, with the mouth at a height of 1.5 m, and more than 1 m away from reflecting surfaces. The 01dB Symphonie system was used to produce the excitation signal and determine the mouth-to-ears impulse response from the measured signal at the microphones. For each classroom, the ST_V values of the two ears at two different positions were averaged. G_{RG} was calculated by applying Eq. (3) on the ST_V values.

D. Prediction model for voice support

The prediction model for ST_V in Eq. (18) was evaluated in one-octave frequency bands by using the frequency-dependent measured values of T, along with the volume and total surface area of the classrooms. In addition, a broadband value (speech-weighted ST_V) was calculated from the frequency-band values using Eq. (8).

The prediction model was assessed by comparing the measured and the predicted ST_V values. In each frequency band (or overall speech-weighted), a regression line of the type $\overline{ST}_{V,\mathrm{pred}} = a \cdot ST_{V,\mathrm{meas}} + b$, where $\overline{ST}_{V,\mathrm{pred}}$ is the regressor for the predicted values of voice support (notated as $ST_{V,\mathrm{pred}}$), $ST_{V,\mathrm{meas}}$ are the measured values, and a and b are the coefficients of the regression line. Ideally, a perfect model would result if the predicted and the measured values were equal ($ST_{V,\mathrm{pred}} = ST_{V,\mathrm{meas}}$). Nevertheless, an unbiased model would result if $a = 1$ and $b = 0$, i.e. $\overline{ST}_{V,\mathrm{pred}} = ST_{V,\mathrm{meas}}$.

The goodness of fit of the prediction model was evaluated with three parameters: a) the R^2 of the linear regression model for the measured versus predicted values, b) the residual deviation σ_ϵ of the predicted values from this regression line, and c) the deviation σ_T of the predicted values from an unbiased prediction, which is a measure of the bias in the prediction.

$$\sigma_\epsilon^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(ST_{V,\mathrm{pred}} - \overline{ST}_{V,\mathrm{pred}} \right)^2 \quad (25)$$

$$\sigma_T^2 = \frac{1}{N} \sum_{i=1}^{N} \left(ST_{V,\mathrm{pred}} - ST_{V,\mathrm{meas}} \right)^2 \quad (26)$$

IV. RESULTS

A. Correlation between parameters

The correlation coefficients between the measurements of the magnitudes V, $\log(V)$, $L_{N,Aeq}$, T_{500-2k}, G_{500-2k}, STI, G_{RG}, and ST_V are shown in Table III. $L_{N,Aeq}$ has

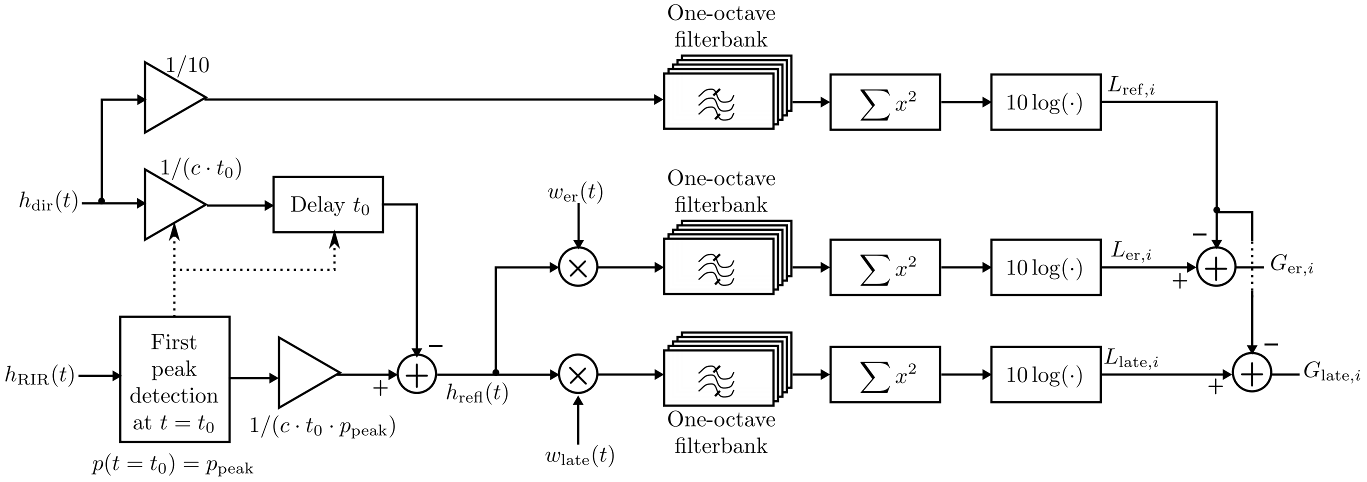

FIG. 3. Block diagram of the post-processing applied to the the impulse responses to extract the sound strength of early and late reflections

very low correlation with all the other parameters, because it is not determined from physical properties of the room, but depends on different noise sources from installations inside the room, and from other external noise sources (traffic noise, students in neighboring classrooms, corridors, or playground). The reverberation time is correlated to the volume and negatively correlated to the STI. The voice support is strongly correlated to the logarithm of the volume, as expected from the prediction model in Eq. (18), and to the sound strength. Both parameters are descriptors of the amplification of sound due to reflections. The presence of some high correlation coefficients is largely caused by the large measured range of volume and most of the other parameters of the classrooms.

B. Background noise levels

The mean and maximum background noise levels (A-weighted and in one octave frequency bands) are shown in Table IV. Although it is not explicitly shown, 73.3% of the classrooms had $L_{N,Aeq}$ lower than 35 dB, another 13.3% between 35 and 40 dB, and the remaining 13.3% of the measurements were between 40 and 45 dB. In most of the cases, the noise sources corresponded to the ventilation systems, although in a few cases, the background noise was affected by external sources, such as neighboring activities, playground, and traffic. The background noise levels were similar for all room sizes, although the overall level in the large rooms was slightly higher than in smaller rooms. In all cases, low frequency noise was markedly dominating. This is an indication that the sources, in most of the cases, were in fact machinery of the ventilation systems, or external noise that breaks in the room due to the usually low insulation performance of walls, doors, and windows at low frequencies.

C. Reverberation time

The mean reverberation times (in one octave frequency bands and 500 Hz-2 kHz average) and their standard deviation are shown in Table V. 81.5% (22 out of 27) of the small and medium classrooms had reverberation times lower than 0.5 s, and the remaining 18.5% were between 0.5 and 0.6 s. In the sports halls, T was between 1.4 s and 1.8 s.

D. Speech transmission index

The average (standard deviation) measured STI was 0.80 (0.02) in small classrooms, 0.75 (0.03) in medium classrooms and 0.63 (0.02) in large classrooms. The spread of STI among rooms, indicated by the standard deviation, was similar in all of the three classroom groups. The small classrooms had the highest STI, which falls in the category of 'excellent'. The medium classrooms had an average STI rating which is between 'good' and 'excellent', and the sports halls had an STI rating of 'good' which is likely to decrease in the presence of activity noise.

E. Sound strength

The numerical values of the mean and standard deviation of G among different classroom groups, in one octave bands from 125 Hz to 4 kHz, along with the average on the 500 Hz, 1 kHz, and 2 kHz bands, are shown in Table VI. G depends on the volume and it has highest values for the smallest rooms. There are two factors contributing to the increased G values. First, the energy density increases as the volume decreases and, second, the students are closer to the teacher than in larger rooms, receiving an important contribution from the direct sound. The spread of G among rooms, indicated by the standard deviation, does not show any clear depen-

TABLE III. Correlation coefficient indicating the strength of the linear dependence between pairs of variables.

	V	$\log V$	$L_{N,Aeq}$	T_{500-2k}	STI	G_{RG}	ST_V	G_{500-2k}
V	1.00	**0.91**	0.29	**0.97**	**-0.81**	-0.50	-0.75	-0.78
$\log V$		1.00	0.27	**0.91**	-0.57	-0.75	**-0.87**	**-0.90**
$L_{N,Aeq}$			1.00	0.24	-0.14	-0.10	-0.19	-0.27
T_{500-2k}				1.00	**-0.89**	-0.47	-0.70	-0.74
STI					1.00	0.55	0.68	0.68
G_{RG}						1.00	**0.94**	**0.81**
ST_V							1.00	**0.89**
G_{500-2k}								1.00

TABLE IV. Frequency band values and overall A-weighted background noise levels (BNL) measured in the classrooms.

Octave band center frequency (Hz)	125	250	500	1000	2000	4000	A-weighted
Small classrooms							
Mean BNL (dB)	38.3	32.4	28.2	26.1	22.3	19.4	32.3
Maximum BNL (dB)	48.8	39.3	34.5	32.6	27.5	21.3	38.5
Medium classrooms							
Mean BNL (dB)	40.2	33.7	27.8	24.4	22.7	19.9	32.7
Maximum BNL (dB)	53.4	43.6	43.7	40.1	37.3	32.4	43.5
Large classrooms							
Mean BNL (dB)	45.1	37.9	33.5	32.0	28.3	21.9	37.6
Maximum BNL (dB)	51.5	46.2	41.1	37.4	30.1	23.2	43.5

dence on the frequency.

It is important to know G at different positions in classrooms to calculate the level of the speech signal and predict speech intelligibility in classrooms based on signal-to-noise ratios. The dependence of G with distance is hidden in Table VI. The contribution of G_{dir}, G_{er}, and G_{late} to G at different distances in the 3 classroom groups is shown in Fig. 4. Regression lines were obtained for G_{er} and G_{late} in each group of classrooms. The corresponding equations (with dB units) are

$$G_{\mathrm{er}} = \begin{cases} 21.5 - 1.1d & \text{small } (R^2 = 0.36; p = 0.11) \\ 16.5 - 0.6d & \text{medium } (R^2 = 0.18; p < 0.001) \\ 7.9 - 0.3d & \text{large } (R^2 = 0.61; p = 0.003) \end{cases} \tag{27}$$

$$G_{\mathrm{late}} = \begin{cases} 9.2 + 0.2d & \text{small } (R^2 = 0.01; p = 0.79) \\ 8.6 - 0.04d & \text{medium } (R^2 = 4 \cdot 10^{-4}; p = 0.86) \\ 5.9 - 0.02d & \text{large } (R^2 = 0.02; p = 0.69) \end{cases} \tag{28}$$

where d is the source-receiver distance. There is a large spread in the data, which is reflected in the low R^2 values in the regression lines. However, the regression lines for G_{er} in medium and large classrooms are significant at the 5% level, and in small classrooms the p-value is 0.11, indicating a likely trend which could have been significant with a larger sample of measured classrooms. In all classroom sizes, G_{er} decreases with distance, and the slope is less steep the larger the room. The low R^2 values and non-significant effects for the regression lines of G_{late} indicate that, most likely, G_{late} is uniform at the different positions in the room.

F. Voice support and room gain

1. Measurements

The mean and standard deviation of ST_V and G_{RG} in the one-octave frequency bands between 125 Hz and 4 kHz measured in the classrooms is shown in Table VII. The frequency characteristics of ST_V and G_{RG} are similar for small and medium classrooms, with a remarkable increase at high frequencies. The only difference between the two classroom groups is that the small classrooms have a slightly higher overall value. The large classrooms (sports halls) have an overall lower value and, in addition, the frequency characteristic is qualitatively different, because the low frequencies are predominant. This indicates that these large rooms do not reflect efficiently the high frequencies of a talker. The spread of ST_V among rooms does not depend on the frequency band, since the standard deviation does not present a frequency-dependent pattern in the different classroom groups. However, the standard deviation of G_{RG} is proportional to its absolute value.

2. Prediction model

The values of V and S of each classroom, together with the frequency-dependent average measurements of T, were used in connection with Eq. (18) to predict the ST_V values. The comparison between the measured and the predicted values of ST_V in the one-octave frequency bands between 125 Hz and 4 kHz is shown in Fig. 5. The most accurate predictions are found in the most important bands for speech (between 500 Hz and 2 kHz). In

TABLE V. Reverberation times (T) measured in the classrooms.

Octave band center frequency (Hz)	125	250	500	1000	2000	4000	Average 500–2000
Small classrooms							
Mean T (s)	0.59	0.39	0.32	0.34	0.35	0.34	0.33
s.d.	0.42	0.14	0.04	0.05	0.05	0.02	0.05
Medium classrooms							
Mean T (s)	0.72	0.53	0.45	0.47	0.47	0.44	0.46
s.d.	0.33	0.17	0.08	0.08	0.07	0.07	0.08
Large classrooms							
Mean T (s)	1.46	1.58	1.59	1.55	1.35	1.04	1.57
s.d.	0.24	0.35	0.29	0.18	0.07	0.07	0.23

TABLE VI. Frequency band values and overall speech-weighted sound strength (G) measured in the classrooms averaged for four distances in each room.

Octave band center frequency (Hz)	125	250	500	1000	2000	4000	Average 500–2000
Small classrooms							
Mean G (dB)	21.7	21.4	18.1	20.7	21.5	22.9	20.1
s.d.	3.6	2.0	0.8	1.1	1.3	1.4	1.1
Medium classrooms							
Mean G (dB)	19.4	18.2	13.8	15.8	16.8	17.7	16.0
s.d.	2.6	2.1	1.9	1.5	1.6	1.7	1.6
Large classrooms							
Mean G (dB)	13.4	12.9	6.8	8.8	9.2	8.9	9.4
s.d.	1.2	0.3	1.0	1.1	1.4	2.1	0.7

these bands, R^2 was at least 0.8, the residual deviation was not higher than 1.2 dB, and the bias or deviation from the unbiased prediction was lower than 2 dB. The prediction for the 125 Hz band had a large uncertainty, shown by the low value of R^2 (0.18), and large residual deviation (3.3 dB) and bias (4.3 dB).

The speech-weighted ST_V predictions are plotted in Fig. 6 as a function of the measured ST_V values. The regression line relating measurements and predictions had a slope of 0.95 (whereas ideally, it should be 1). The R^2 was 0.84, the residual error was 1.1 dB and the bias was 1.4 dB.

V. DISCUSSION

The acoustic properties of school classrooms described in the results section correspond to typical primary and secondary schools in southern Sweden built during the decade of 1970's. The background noise levels in almost three fourths of the small and medium sized classrooms were below 35 dBA—which is the maximum acceptable value on different guidelines, e.g. the standard ANSI S12.60-2002[31] in the US, the Building Bulletin 93[32] in the UK, or the guidelines from the World Health Organisation (WHO).[33] The background noise levels in the remaining fourth of classrooms, were one half below 40 dB and the other half between 40 and 45 dB. The average value was 32.6 dB, which is lower than the 45 to 48 dB reported by Shield and Dockrell[1] in their review from several surveys on empty classrooms (without acoustical treatment).

In small and medium-sized classrooms, T did not exceed 0.6 s, in fulfilment of different guidelines of classroom acoustic design[31–33]. Reverberation times and background noise levels are within the recommended values in most of the cases. This seems to be reflected in the non-problematic perception of classroom acoustics by teachers without voice problems in schools of the same region in Sweden.[13] The Swedish standard for acoustic conditions in classrooms[34] is more strict, requiring reverberation times below 0.5 s for the octave frequency bands above 250 Hz and below 0.6 s at 125 Hz, which only a few of the classrooms fulfill.

The STI measured in classrooms with high signal-to-noise ratios was higher than 0.6 in all cases, even in the sports halls. However, the subjective speech intelligibility with ongoing activity, specially in the sports halls, will be lower than predicted, due to an actual lower signal-to-noise ratio under these conditions. Unfortunately, none of the current guidelines specify the signal-to-noise ratio that should be used for the assessment of STI.

The sound strength can be used to calculate the speech SPL at a given position in the classroom. The effect of the room, ignoring the contribution from the direct sound, can be evaluated with G_{refl}. According to the diffuse field theory, this value is constant across the room and is

$$G_{\mathrm{refl,dif}} = 10 \log \left(\frac{31220T}{V} - \frac{5026}{S} \right). \qquad (29)$$

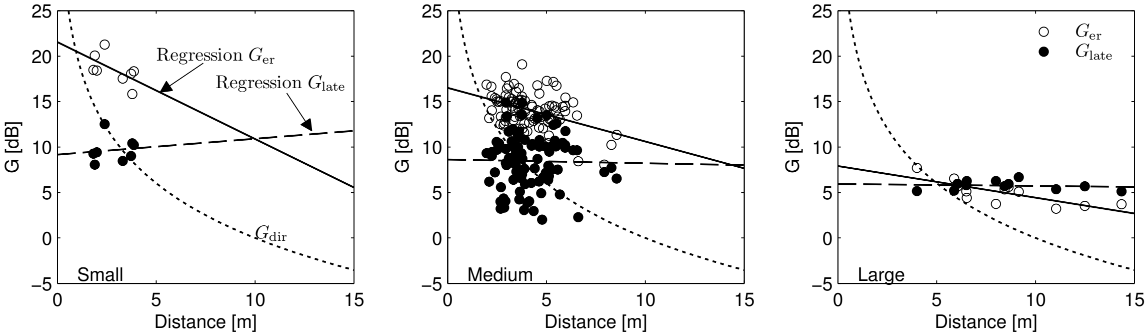

FIG. 4. Measurements of sound strength due to direct sound (G_{dir}), early reflections (G_{er}), and late reflections (G_{late}), as a function of the distance to the source. The regression lines for the two latter components are shown. Left: small rooms. Middle: medium rooms. Right: large rooms.

TABLE VII. Frequency band values and overall speech-weighted voice support (ST_V) and room gain (G_{RG}) measured in the classrooms.

Octave band center frequency (Hz)	125	250	500	1000	2000	4000	Speech-weighted
Small classrooms							
Mean ST_V (dB)	-9.4	-11.1	-9.5	-7.6	-6.4	-4.6	-8.3
s.d.	0.46	0.81	0.91	0.38	0.72	1.04	0.50
Mean G_{RG} (dB)	0.50	0.33	0.47	0.70	0.91	1.31	0.60
s.d.	0.02	0.06	0.09	0.06	0.13	0.25	0.07
Medium classrooms							
Mean ST_V (dB)	-12.1	-13.9	-13.5	-11.6	-10.9	-9.1	-12.2
s.d.	1.46	1.27	1.43	1.68	1.75	1.52	1.43
Mean G_{RG} (dB)	0.28	0.18	0.20	0.32	0.37	0.54	0.26
s.d.	0.10	0.06	0.07	0.13	0.16	0.19	0.09
Large classrooms							
Mean ST_V (dB)	-10.8	-16.0	-18.2	-19.1	-19.5	-19.4	-17.9
s.d.	1.56	1.91	0.92	1.31	1.40	1.31	0.51
Mean G_{RG} (dB)	0.36	0.12	0.07	0.06	0.05	0.06	0.07
s.d.	0.14	0.06	0.01	0.02	0.02	0.02	0.01

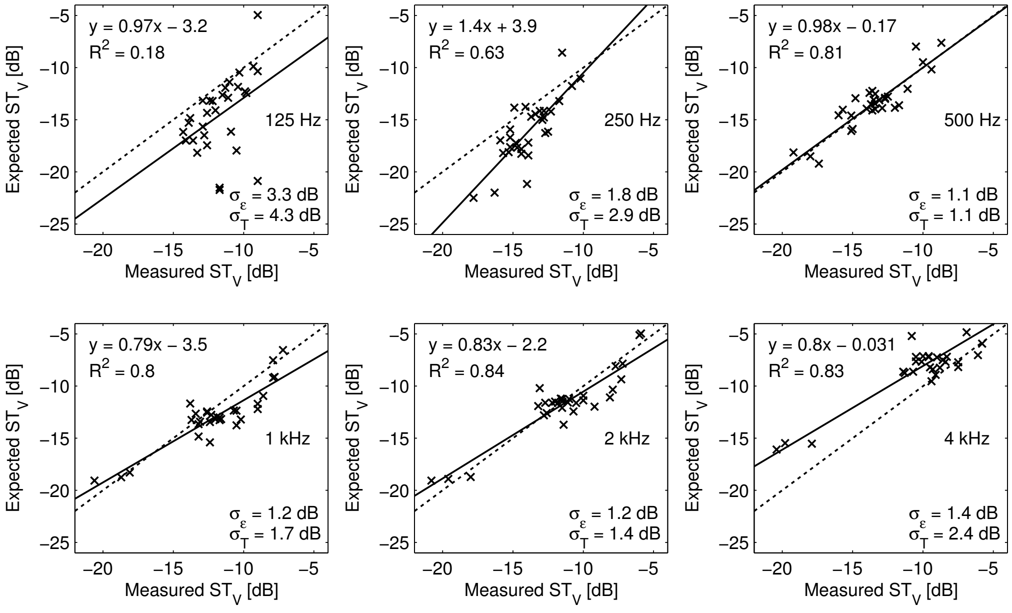

FIG. 5. Expected versus measured values of voice support in frequency bands. The solid lines show the regression lines for the predictions and the dotted lines indicate the ideal and unbiased prediction lines.

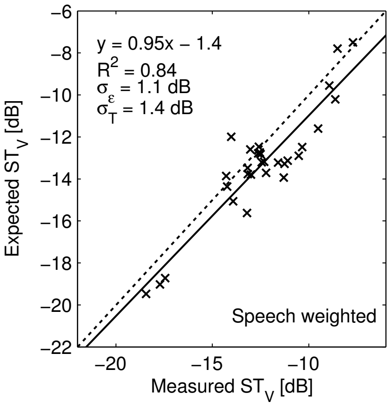

FIG. 6. Expected versus measured speech-weighted overall values of voice support. The solid lines show the regression lines for the predictions and the dotted lines indicate the ideal and unbiased prediction lines.

The values of $G_{\mathrm{refl,dif}}$ are shown as solid lines in Fig. 7 for each of the 3 classroom groups. The quantities V and S were derived from the average dimensions in Table II, assuming a box geometry, which is the case for nearly every room. V and S were 41 m^3 and 72 m^2 for small classrooms, 180 m^3 and 218 m^2 for medium-sized classrooms, and 3614 m^3 and 1640 m^2 for large classrooms. The values used for T are the ones on Table V: 0.33 s for small classrooms, 0.46 s for medium classrooms, and 1.57 s for large classrooms.

According to the revised theory of Barron and Lee,[35] the reflected energy (early+late) at the source position is equal to the energy predicted by the diffuse field theory and decays exponentially with the distance from the source according to

$$G_{\mathrm{refl,BL}} = 10\log\left[\left(\frac{31220T}{V} - \frac{5026}{S}\right)\mathrm{e}^{-\frac{0.04d}{T}}\right]. \qquad (30)$$

The predictions for $G_{\mathrm{refl,BL}}$ according to Barron and Lee's theory for each group of classrooms are shown as dashed lines in Fig. 7. The slopes of these lines are -0.53 dB/m for small classrooms, -0.38 dB/m for midsize classrooms, and -0.11 dB/m for large classrooms. These values are approximately one half of the slopes from the regression lines in Eq. (27). This result is in good agreement with the findings of Sato and Bradley,[24] who found that the experimental decrease of the level of the early reflected sound was twice as derived by Barron and Lee.[35] The sound strength of the reflected sound, $G_{\mathrm{refl,meas}}$ obtained by the combination of the regression lines for the early reflected sound in Eq. (27) and the late reflected sound in Eq. (28) through Eq. (24) are shown as dash-dot curves in Fig. 7. This figure shows that G_{late} has an influence on $G_{\mathrm{refl,meas}}$ only at long distances. Therefore, the average deviation of $G_{\mathrm{refl,meas}}$ from $G_{\mathrm{refl,BL}}$ are lower than considering a linear decay of G_{late} according to Sato and Bradley.

An important observation from Fig. 7 is that $G_{\mathrm{refl,meas}}$ matches the predictions from the diffuse field theory at the source position, and also Barron and Lee's prediction, with a deviation of less than 1 dB. This finding validates the use of the diffuse field theory to obtain the level of the reflected sound at the ears of a talker in the prediction model of ST_V.

According to the measurements, G_{late} is nearly uniform across the room. This is yet another indication that the late part of an impulse response has statistical properties related to the room and not to the placement of source and receiver. The early and the late reflections have been separated as those arriving before and after 50 ms from the arrival of the direct sound, respectively. However, the present results suggest that the transition time between early and late reflections could be defined as the minimum time in the impulse response after the arrival of the direct sound for which the statistics of the late reflections are independent of the position in the room.

The prediction model for ST_V has been derived theoretically and it has been assessed by comparing its predictions with actual ST_V measurements. There is a bias in the prediction, as the regression line of measured versus predicted ST_V is not $\overline{ST}_{V,\mathrm{pred}} = ST_{V,\mathrm{meas}}$ but $\overline{ST}_{V,\mathrm{pred}} = 0.95 \cdot ST_{V,\mathrm{meas}} - 1.4$ (see Fig. 6). This bias results in a deviation of 1.4 dB from the actual values, slightly higher than the residual deviation (1.1 dB). Taking into account that the measurement dataset has not been used to derive the model, the predictions are reasonably accurate.

In the range of medium-sized classrooms (with volumes $100 < V < 250$ m^3), G_{RG} is in the range between 0.2 and 0.5 dB, whereas ST_V is in the range between -14 and -9 dB. There is some spread of data in this range, as seen in Fig. 6. Measured ST_V values can deviate as much as 3 dB from the predicted value. ST_V is influenced by the early reflections which can not be accurately represented with a statistical model such as the one in Eq. (18).

The voice support, analogously to the objective support in concert halls, is not a stand-alone parameter to design classroom acoustics. It is a magnitude related to the additional vocal load that teachers experience while speaking in a classroom due to the acoustic conditions. Other magnitudes, like T, G, STI, and background noise levels should be taken into account as well. There is not enough scientific evidence to establish a definite range of recommended values of ST_V, but the range between -14 and -9 dB obtained in most of the medium-sized classrooms seems adequate, since T and STI fulfilled the recommendations without the rooms being too damped. Using the graph in Fig. 2, for a room of 100 m^3, the range of $-14 < ST_V < -9$ dB corresponds to reverberation times in the range $0.25 < T < 0.6$ s. For a room of 300 m^3, the same range of ST_V corresponds to the range $0.55 < T < 1.4$ s. In this last case, the design criteria should be to aim at the highest reverberation time that does not compromise speech intelligibility, because too high values of reverberation are detrimental to speech intelligibility. For the same reason, it is not adviceable to aim at values of ST_V higher than -9 dB. However, in very small classrooms, ST_V may be higher than -9 dB without compromising speech intelligibility.

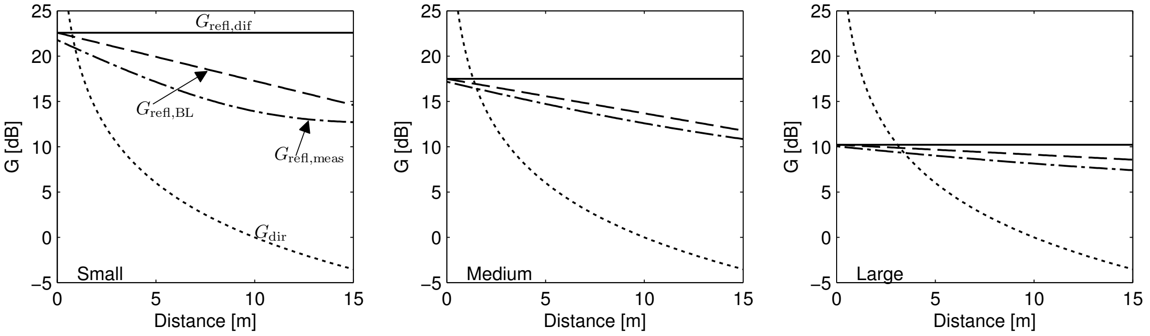

FIG. 7. Predicted sound strength as a function of the distance to the source for the direct and reflected components, according to the diffuse-field theory ($G_{\mathrm{refl,dif}}$, solid line), Barron and Lee's revised theory ($G_{\mathrm{refl,BL}}$, dashed line), and the combination of the regression lines on Eqs. (27) and (28) ($G_{\mathrm{refl,meas}}$, dash-dot line). Left: small rooms. Middle: medium rooms. Right: large rooms.

VI. CONCLUDING REMARKS

The acoustic design of classrooms and the objective assessment of the acoustic performance is traditionally based on measures related to speech intelligibility: reverberation time, speech and background noise levels, and STI. These measures, however, do not offer a direct picture of the acoustic conditions for a talker. Recently developed parameters, *voice support* and *room gain*—which are different ways of expressing the same magnitude—, link variations in vocal effort to room acoustics conditions. The present study provided a reference set of values for these two parameters, in addition to more traditional measures. Most of the 30 school classrooms of the study complied with main classroom acoustic guidelines in terms of reverberation time, background noise and STI. In classrooms with volumes between 100 and 250 m^3, the measured voice support is in the range between -14 and -9 dB, whereas the room gain is in the range between 0.2 and 0.5 dB.

The dependence of sound strength with distance in classrooms has been examined. Empirical regression models derived from the measured data confirm the reflected energy levels predicted with diffuse-field theory at the source position. Additional regression models for early and late sound strength show that the level of early reflections decrease with distance as previously reported in the literature. However, the level of late reflections after 50 ms is apparently uniform in all the positions in the room.

A prediction model has been developed from the diffuse-field theory to predict average values of voice support in classrooms, based on geometrical room properties of volume and total surface area and reverberation time. This model describes the present voice support measurements in classrooms with a coefficient of determination of 0.84 and a standard deviation of 1.4 dB. The prediction model can also be used to assess the acoustic conditions for a talker in classrooms during the design phase. However, considerations about speech intelligibility should not be left aside.

Acknowledgments

This research has been funded by the Swedish organization AFA Försäkring as a part of the project "Speaker's comfort and voice health in classrooms". The authors would like to express their gratitude to all the schools and staff who made these measurements possible.

[1] B. Shield and J. Dockrell, "The effects of noise on children at school: a review", Build. Acoust. **10**, 97–116 (2003).

[2] M. Klatte, M. Meis, H. Sukowski, and A. Schick, "Effects of irrelevant speech and traffic noise on speech perception and cognitive performance in elementary school children", Noise Health **9**, 64–73 (2007).

[3] B. Shield and J. Dockrell, "The effects of environmental and classroom noise on the academic attainments of primary school children", J. Acoust. Soc. Am. **123**, 133–144 (2008).

[4] A. Neuman and I. Hochberg, "Children's perception of speech in reverberation", J. Acoust. Soc. Am **73**, 2145–2149 (1983).

[5] J. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms", J.Acoust.Soc.Am. **113**, 3233–3244 (2003).

[6] A. Neuman, M. Wroblewski, J. Hajicek, and A. Rubinstein, "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults", Ear Hear. **31**, 336–344 (2010).

[7] J. Bradley, "Speech intelligibility studies in classrooms", J.Acoust.Soc.Am. **80**, 846–854 (1986).

[8] J. Bradley, R. Reich, and S. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility", J.Acoust.Soc.Am. **106**, 1820–1828 (1999).

[9] M. Klatte, T. Lachmann, and M. Meis, "Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting", Noise Health **12**, 270–282 (2010).

[10] J. Bradley, "Predictors of speech intelligibility in rooms", J.Acoust.Soc.Am. **80**, 837–845 (1986).

[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**, 861–875 (2000).

[12] N. Roy, R. Merrill, S. Thibeault, R. Parsa, S. Gray, and E. Smith, "Prevalence of voice disorders in teachers and the

general population", J.Speech Lang.Hear.Res. **47**, 281–293 (2004).

[13] V. Lyberg-Åhlander, R. Rydell, and A. Löfqvist, "Speaker's comfort in teaching environments: Voice problems in Swedish teaching staff", J. Voice (2010), available online 26 March 2010.

[14] E. Vilkman, "Voice problems at work: A challenge for occupational safety and health arrangement.", Folia Phoniatr. Logo. **52**, 120–125 (2000).

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

[16] 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).

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

[18] D. Pelegrin-Garcia, "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).

[19] A. Gade, "Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and laboratory experiments", Acustica **69**, 193–203 (1989).

[20] H. Knecht, P. Nelson, G. Whitelaw, and L. Feth, "Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements", Am.J.Audiol. **11**, 65–71 (2002).

[21] E. Pekkarinen and V. Viljanen, "Acoustic conditions for speech communication in classrooms", Scand.Audiol. **20**, 257–263 (1991).

[22] M. Hodgson, "Rating, ranking, and undestanding acoustical quality in university classrooms", J.Acoust.Soc.Am. **112**, 568–575 (2002).

[23] M. Hodgson, "Empirical prediction of speech levels and reverberation in classrooms", Build. Acoust. **8**, 1–14 (2001).

[24] H. Sato and J. Bradley, "Evaluation of acoustical conditions for speech communication in working elementary school classrooms", J.Acoust.Soc.Am. **123**, 2064–2077 (2008).

[25] L. Nijs and M. Rychtarikova, "Calculating the Optimum Reverberation Time and Absorption Coefficient for Good Speech Intelligibility in Classroom Design Using U50", Acta Acust. united Ac. **97**, 93–102 (2011).

[26] W. Olsen, "Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report", Am.J.Audiol. **7**, 21–25 (1998).

[27] W. Chu and A. Warnock, *Detailed directivity of sound fields around human talkers*, Institute for Research in Construction, National Research Council Canada, Tech. Rep, Canada (2002).

[28] Brüel & Kjær, *Head and Torso Simulator Type 4128 - Product data*, Nærum, Denmark (2009).

[29] M. Schroeder, "New Method of Measuring Reverberation Time", J.Acoust.Soc.Am. **37**, 409–412 (1965).

[30] International Organization for Standardization, *ISO-3382:2009, Acoustics - Measurement of room acoustic parameters – Part 1: Performance spaces*, ISO, Geneva (2009).

[31] American National Standards Institute, *ANSI S12.60-2002: American National Standard Acoustical performance criteria, design requirements and guidelines for schools*, Melville, NY (2002).

[32] C. Hopkins, *Building Bulletin 93. Acoustic design of schools, a design guide*, Department of Education and Skills, London, UK (2004).

[33] B. Berglund, T. Lindvall, and D. Schwela, "Guidelines for community noise", World Health Organization, Geneva (1999).

[34] SIS-Building and construction, *SS 25268:2007 Building acoustics - Sound classification of spaces in buildings - health care facilities, educational facilities, day-care centers, offices and hotels*, Swedish Standard, Stockholm, Sweden (2007).

[35] M. Barron and L. Lee, "Energy relations in concert auditoriums. I", J.Acoust.Soc.Am. **84**, 618–628 (1988).

Final report of the project

Speakers comfort and voice disorders in classrooms

An overall aim of this research project has been to investigate the voice use of teachers in relation to the acoustic properties of the classroom, and to study whether speakers take into account auditory cues to regulate their voice levels, even in the absence of background noise. The most common means of communication in a classroom is speaking and listening. The teacher's voice is thus the tool for communicating with the students. The room acoustics in the classroom is the communication channel from the speaker to the listener. It affects the quality of the speech signal and thus the ability to understand what the teacher says.

One of the core concepts in this project is "speakers' comfort" that is tied to the voice use and the speaker's subjective perception of the voice. It is defined as the subjective impression that talkers have when they feel that their vocal message reaches the listener effectively [with no or low vocal effort]. In this subjective impression, experienced while hearing and perceiving one's own voice, some attributes play important roles: the voice-support provided by the room and the speech intelligibility along with the sensory-motor feedback from the phonatory apparatus.

Among the conclusions drawn from the project:
Voice problems in teachers arise from the interplay of the individual and the environment. Teachers with voice problems are more affected by factors in the work environment than their voice healthy colleagues. The differences between a group of teachers with selfassessed voice problems and their voice healthy colleagues were most clearly shown during field-measurements of the voice during a typical school day, while the findings from the clinical examinations of larynx and voice did not differ between the groups.

The results from the prevalence study show that 13% of the teachers suffer from voice problems frequently or always. Most teachers however, reported occurrence of symptoms of vocal disturbances. Voice-related absence from work was common in both teachers with and without voice problems.

LUND UNIVERSITY

Sound Environment Center at Lund University, Sweden
Ljudmiljöcentrum vid Lunds universitet

Box 117, 221 00 Lund
Telefon +46 (0)46222 09 46

www.ljudcentrum.lu.se

Print: Media-Tryck, Lund 2011