

Influence of classroom acoustics on the vocal behavior of teachers

*Original*

Influence of classroom acoustics on the vocal behavior of teachers / Astolfi, Arianna; Carullo, Alessio; Vallan, Alberto; Pavese, Lorenzo. - In: PROCEEDINGS OF MEETINGS ON ACOUSTICS. - ISSN 1939-800X. - STAMPA. - 19, 040123:(2013), pp. 1-9. (Intervento presentato al convegno 21st International Congress on Acoustics tenutosi a Montreal (CAN) nel 2-7 June 2013) [10.1121/1.4800427].

*Availability:*

This version is available at: 11583/2507894 since:

*Publisher:*

Acoustical Society of America

*Published*

DOI:10.1121/1.4800427

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



**ICA 2013 Montreal  
Montreal, Canada  
2 - 7 June 2013**

**Noise**

**Session 4pNSa: Effects of Noise on Human Performance and Comfort II**

**4pNSa3. Influence of classroom acoustics on the vocal behavior of teachers**

Arianna Astolfi\*, Alessio Carullo, Alberto Vallan and Lorenzo Pavese

\*Corresponding author's address: Energetics, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin, 10129, Turin, Italy, [arianna.astolfi@polito.it](mailto:arianna.astolfi@polito.it)

Erroneous vocal behavior of teachers and their changes in the voice production due to poor acoustics in classrooms can be investigated through recently developed voice-monitoring devices. These devices are portable analyzers that use a miniature contact-microphone glued to the jugular notch in order to sense the skin acceleration level due to the vibration of the vocal folds. They estimate the Sound Pressure Level (SPL) at a certain distance from the speaker's mouth provided that a preliminary calibration procedure is performed, the fundamental frequency and the time dose. Two different devices are compared in this work: the former is a commercial device, whose phonation sensor is a small accelerometer; the latter, recently developed by the authors, uses an electret condenser microphone to sense the skin acceleration level. SPL and fundamental frequency are estimated over fixed-length frames and the results that refer to a sample of 25 primary school teachers and a university professor are analyzed. The duration of the voice and pause periods is investigated in order to detect the peaks of occurrence and accumulation in different conditions of reverberation. A method for the detection and analysis of the emphatic speech is also proposed.

Published by the Acoustical Society of America through the American Institute of Physics

## INTRODUCTION

Voice disorders of teachers, from the functional dysphonia to the vocal fold nodules, are likely to result from voice abuse, erroneous behaviors in the use of voice and changes in the voice production due to poor acoustics in the classrooms where the voice is used.

According to Astolfi et al. (2012) the higher the voicing time percentage, which teachers exhibit greater than other professionals (Masuda et al., 1993; Bottalico and Astolfi, 2012), the higher the possible category of disease derived by the phoniatric evaluation. Chan (1994) suggested that teachers would be able to improve their voices if they significantly reduced vocal abuses in daily life and practiced specific strategies to maintain classroom order and reduce the use of voice in teaching. Bovo et al. (2007) showed as a short group voice therapy, home-controlled voice exercises, and hygiene, represent cost-effective primary preventions of voice disorders in teachers and Nanjundeswaran et al. (2012) suggested that an individually tailored voice health program may be sufficient to prevent voice problems from teaching in healthy student teachers.

Lyberg Åhlander et al. (2012) discovered that the cause of voice dysfunction in teachers with self-reported voice problems is not found in the vocal apparatus or within the individual, but in the interplay of the individual's behavior and the work environment.

As far as the work environment is concerned, Bottalico and Astolfi (2012) recently investigated vocal doses and parameters for primary school teachers and found an increase in speech level and fundamental frequency with an increase in background noise level, during traditional lessons. Pelegrín-García and Brunskog (2012) demonstrated, under simulated acoustic environments, that teachers in absence of background noise adjust their voice levels under different room acoustic conditions and established optimum room acoustic conditions for speaking, while Lyberg Åhlander et al. (2011) concluded that the teachers with voice problems are more sensitive to room acoustics and to other issues related to the work environment.

The vocal load of teachers can be presently quantified thoroughly and systematically by means of devices for voice monitoring (Hillman and Mehta, 2011). Recent researches have used a miniature accelerometer glued at the base of the neck as a phonation sensor for long-term monitoring of vocal function. The National Center for Voice and Speech (Popolo et al., 2005a; Popolo et al., 2005b; Švec et al., 2005) and the Massachusetts General Hospital (Cheyne, 2003; Hillman et al., 2006; KayPENTAX, 2013), have produced the NCVS dosimeter and APM 3200, respectively, while another device named VoxLog has been recently developed at the Linköpings University of Sweden (Voxlog, 2013; Wirebrand, 2011). The APM and Voxlog are commercial devices, while the NCVS dosimeter is only used for research. These devices estimate the Sound Pressure Level (*SPL*) at a certain distance from the speaker's mouth, provided that a suitable calibration procedure is performed, the fundamental frequency ( $F_0$ ) and the voicing time percentage relative to the total time of monitoring, i.e. the time dose ( $D_t$ ).

Two different devices are compared in this work: the former is the APM 3200 and the latter, named Voice-Care, has been recently developed by the authors (Carullo et al., 2013a; Carullo et al., 2013b) and uses an electret condenser microphone to sense the skin acceleration level.

The vocal load of teachers is already proved by several studies concerning the time dose measurement: Masuda et al. (1993) measured a mean phonation time of 21.5% on elementary teachers over 8 h work, Hunter and Titze (2010) found that teachers vocalize an average of 30% over 6 h work compared with 14.5% over 6 h after work, while Bottalico and Astolfi (2012) found a voicing time percentages of 26% on average on primary school female teachers over 4 h work.

A time dose of about 25% over 4 h work figures out 1 h of voicing time, which translate to about  $9 \cdot 10^5$  vocal fold collisions considering an average value of the fundamental frequency for women of 240 Hz (Bottalico and Astolfi, 2012). According to Hunter and Titze (2009) the vocal overuse is cause of physiological vocal fatigue which may be distinguished in laryngeal muscle fatigue and laryngeal tissue fatigue. The former results in soreness, discomfort, and/or muscle tension in the neck region, while the latter likely stems from change or damage to the vocal fold *lamina propria* caused by vibration exposure and results in pain or scratchy voice sensation and/or increased voice breaks, instability and inability to produce soft voice.

If the primary aim in evaluating vocal fatigue is the quantification of the phonation time, an equal remark should be devoted to the recovery time, that can be distinguished in long and short-term recovery (Titze et al., 2007; Hunter and Titze, 2009). Subjective ratings seem to successfully quantify the effect of vocal loading more than objective metrics as proved by Hunter and Titze (2009) which, by means of perceptual ratings, quantified a full long-term recovery time from 12 to 18 h after 2-h oral reading. They hypothesized that with daily use of the voice there is a continual damage of the laryngeal tissue and the healing mechanism is in a state of constant repair, hence recovery time was shown to be similar to a dermal wound healing trajectory.

As far as short recovery time is concerned, the minimum silence period for tissues to experience any degree of recovery is not known, but the distribution of voicing and silence periods during speech can be important to determine which of these periods affects the vocal fatigue (Titze et al., 2007; Hunter, 2010).

To this aim some questions arise: is a monologue of one h more tiring than speaking three h with longer pauses interposed? In a monologue, a typical period of silence exists which allows one to recover more than other periods? Room acoustics and noise have influence on the distribution of the voicing and silence periods?

In the same way as for the long-term recovery, the perceptual rating of vocal fatigue can be used to answer these questions and the results are of interest for health-care providers.

Given the uncertainty in the duration of the recovery process, Titze et al. (2007) began with the investigation of the distributions of voicing and silence periods in a typical teacher's work day. Over two-week monitoring on 31 subjects, they found the greatest accumulation of voicing periods at work in the  $(0.316 \div 1.0)$  s range and the greatest accumulation of silence in the  $(3 \div 10)$  s range.

The present research aim is to detect the distributions of voicing and silence periods for primary school teachers during the workday in classrooms with low and high reverberation time. It is supposed that the length of the voicing periods can increase due to the longer sound tail, with consequently increase in the vocal fatigue.

It is also supposed that teachers rise their voice in classroom to captivate the attention of the pupils. In this case they speak with an exaggerated or emphatic inflection (Titze et al., 2003). A method for the detection of the distribution of voicing and silence durations within intervals of emphatic speech is also proposed to see if specific voice sizes are typical of this inflection.

## DEVICES FOR VOCAL-ACTIVITY MONITORING

Recently, the National Center for Voice and Speech and the Massachusetts General Hospital carried out the most documented researches in this area. They developed two devices which will be hereafter referred as NCVS dosimeter and APM 3200, respectively. Another device, named VoxLog, which has been recently developed at the University of Sweden, is only described in a technical manual and a Master Thesis. These three devices are based on the measurement of the Skin Acceleration Level (*SAL*) due to the vibration of the vocal folds sensed by a contact microphone glued to the neck of the talker through a surgical adhesive or a necklace. The VoxLog also provides a microphone to sense the vocal signal and the background noise. The main characteristics of these devices for daily monitoring of voice use are reported in Table 1, compared to the Voice-Care device developed by the authors and briefly described in the following paragraph.

Analogous calibration procedures are defined for NCVS, APM 3200 and Voice-Care in order to estimate the *SPL*: the person under monitoring wears the contact microphone while emitting a vocal signal in front of a calibrated microphone at a fixed distance from the mouth, then a fitting of the measured data is implemented to extract the calibration function *SPL* vs *SAL*. *SPL* estimation for VoxLog is instead obtained from the signal at the output of the embedded microphone.

The fundamental frequency  $F_0$  is estimated using different methods present in literature, while the time dose  $D_t$  is obtained by defining a *SPL* threshold that allows to separate voiced and unvoiced frames.

Uncertainty specifications are not available for all the devices. For the parameter *SPL*, only the calibration uncertainty is provided for NCVS dosimeter (5 dB at 95% of confidence level), while for the APM an average error of 3.2 dB with a standard deviation of 6 dB was estimated.

### The Voice-Care device

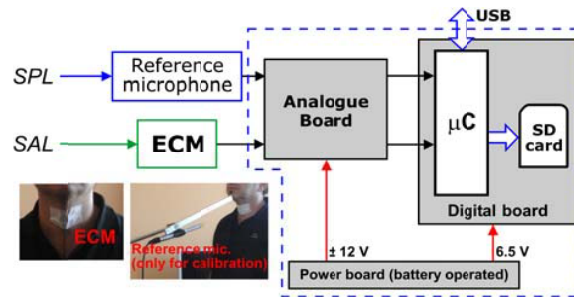
Voice-Care consists of a wearable light-weight battery-operated data-logger based on a low-cost micro-controller board, an Electret Condenser Microphone (ECM Midland MIAE38) used as a contact microphone, and a processing program that allows the vocal parameters to be extracted from the recorded signal. Voice-Care platform has been developed according to the results of a recent authors' work (Carullo et al., 2013a) and is fully described in Carullo et al. (2013b).

Figure 1 shows the architecture of the Voice-Care. The ECM, fixed to the jugular notch of the person under monitoring by means of a surgical band, senses the *SAL* due to the vibration of the vocal folds. The ECM output is amplified and filtered by means of a custom conditioning circuit, which also provides the polarization voltage (phantom power) to the ECM. The voltage output of such a circuit is digitized (sampling frequency 19230 Sa/s) through the Analogue-to-Digital Converter internal to the microcontroller ATMEL ATmega328, which is installed onto a very low-cost Arduino<sup>TM</sup> board. The acquired samples are stored into a micro SD-card, then transferred to a

PC where they are processed in order to estimate the parameters  $SPL$ ,  $F_0$  and  $D_t$ , by subdividing the data stream into frames of 30 ms, which corresponds to the inter-syllabic pauses. A second channel connected to an air reference microphone (Behringer ECM8000), which allows the  $SPL$  to be estimated, is used during a preliminary phase in order to identify the calibration function of the device, i.e. the relationship between  $SAL$  and  $SPL$ .

**TABLE 1.** Main characteristics of recent devices for vocal-activity monitoring compared to Voice-Care.

Name	Sensor	Bandwidth	Frame length	Estimated parameters	Calibration procedure	Algorithm for pitch detection
APM 3200	Accelerometer BU7135 (Knowles Corp.)	2 Hz ÷ 3 kHz Flatness: ± 1.5 dB (50÷1000 Hz)	50 ms	$SPL, F_0, D_t$ and vocal doses defined in Titze et al. (2003)	Defined only for $SPL$	Autocorrelation
NCVS	Accelerometer BU7135 (Knowles Corp.)	2 Hz ÷ 3 kHz Flatness: ± 1.5 dB (50÷1000 Hz)	30 ms	$SPL, F_0$ and the vocal doses defined in Titze et al. (2003)	Defined only for $SPL$	Cepstral analysis
Voxlog	Accelerometer and microphone	Not stated	Not stated	$SPL(A), F_0, D_t$	Obtained for $SPL$ from the microphone	FFT-based processing
Voice-Care	ECM (Midland MIAE38)	10 Hz ÷ 4 kHz	30 ms	$SPL, F_0, D_t$	Defined for $SPL$ and $F_0$	Autocorrelation



**FIGURE 1.** Voice-Care data-logger architecture.

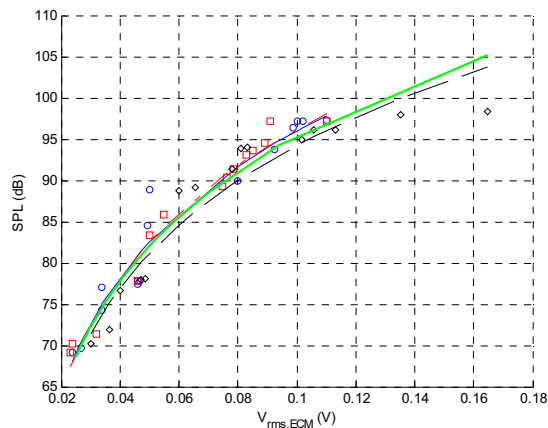
During the calibration procedure, to be performed before each monitoring session, the subject maintains a fixed position in front to the microphone, at 16 cm from the mouth, while sustaining the vowel /a/ at different intensity.

The parameter  $D_t$  is obtained through the comparison of the estimated  $SPL$  to a threshold value, set to separate voiced and unvoiced frames. This value corresponds to the noise floor during the monitoring session.

The fundamental frequency  $F_0$  for voiced frames is estimated through an autocorrelation algorithm, while a zero value is assigned to unvoiced frames.

### Comparison between Voice-Care and APM 3200

The comparison between Voice-Care and APM 3200 refers to the monitoring of the vocal activity of a female professor during two different university lessons of one hour and forty minutes, in a large classroom of 520 m<sup>3</sup> occupied by about 150 students. As far as Voice-Care is concerned, in order to estimate the calibration repeatability, a multiple calibration session has been done in a short time interval without removing the ECM from the jugular notch of the person. Figure 2 shows the results of three calibrations of the ECM against the reference microphone carried out in a small dead room. The marks represent the experimental values, while the continuous lines are the fitted logarithmic calibration functions in the  $SPL$  range from 68 dB to 105 dB at 16 cm, which corresponds to the range (52 ÷ 89) dB at 1 m. The shape of the curves is almost the same, with a maximum difference of 2.5 dB.



**FIGURE 2.** Example of three calibration results of the Voice-Care device. The *SPL* is estimated from the RMS value of the voltage at the output of the reference microphone,  $V_{rms,MIC}$ . The blue, red and black marks represent the experimental values of the three calibrations while the continuous blue, red and black lines represent the correspondent logarithmic calibration functions in the *SPL* range from 68 dB to 105 dB at 16 cm from the mouth. The green line is the average calibration function used in the monitoring mode to estimate the sound pressure level related to the  $V_{rms,ECM}$  derived from the output of the ECM-based chain.

The sensitivity with respect to the air-borne transmitted sounds has proved to be negligible for Voice-Care, while the body activities of the person under monitoring could affect the output of the Voice-Care chain. Specific tests have been carried out to evaluate such effects by monitoring the body activity of the person that wears the Voice-Care and its exposure to noisy environments while no vocal activity is performed. The effects related to the body activities, which are characterized by very low frequency components, are minimized through the use of a digital high-pass filter with a cut-off frequency of 50 Hz.

The calibration of APM 3200 was carried out using a calibrated reference microphone at 15 cm from the teacher's mouth in a small dead room according to the manufacturer instructions. The reference microphone has been previously calibrated using a pure tone @ 1 kHz against a B&K 2222 sound level meter.

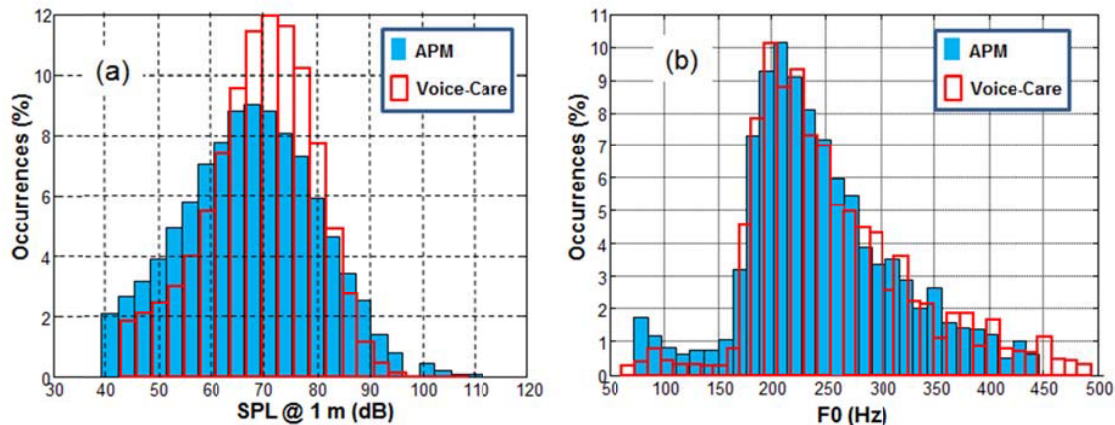
The results of the long-term monitoring with the two devices are summarized in Figure 3. Figure 3(a) shows the histograms of the occurrences of the voiced frames detected by the Voice-Care device and the APM 3200, in terms of *SPL* @ 1 m from the teacher's mouth. For the Voice-Care the histogram starts from a value that corresponds to the electrical noise floor, i.e. 44 dB, while the most occurring value is at 70 dB, that is a reasonable value during a lesson in a large university classroom without wearing a microphone matched to a speech reinforcement system (ISO 9921, 2003). Results by APM are in agreement with those by Voice-Care, with a noise floor at 40 dB and the most occurring value at 68 dB. Changes in the percentage of occurrence between the two lessons can be due to the different typology of lessons hold in two different days.

The occurrences of the fundamental frequency  $F_0$  in the monitored intervals are shown in Figure 3(b) for the two devices. The histograms only refer to the voiced frames, since according to the described algorithm a zero value is assigned to the parameter  $F_0$  during unvoiced frames. The obtained values are in the range from 75 Hz to about 480 Hz, as expected for a female talker (Titze et al., 2003). A good agreement is shown between the two devices.

The parameter  $D_i$  is 31.8% for Voice-Care and 33.3% for APM: results are very similar for the two devices, and the difference is expectable since the observation times are associated to two different lesson. The values can represent the way of speaking during a conference, with the characteristic of a long-term monologue.

## DISTRIBUTION OF VOICING AND SILENCE PERIODS FOR TEACHERS

The distribution of the voicing and silence periods has been obtained from the same samples of vocal activity of the primary school teachers involved in the work by Bottalico and Astolfi (2012). They investigated the vocal doses (Titze et al., 2003) and the vocal parameters of 35 primary school teachers (32 females and 3 males) in six schools in Italy with the APM 3200 over a 73 working-day samples of 4 h each (66 for females and 7 for males). They estimated the mean value of the *SPL* @ 1 m over the working day, obtaining an average value of 62.1 dB for the females and 57.7 dB for the males.



**FIGURE 3.** (a)  $SPL @ 1\text{ m}$  and (b)  $F_0$  occurrences during a monitoring interval of one hour and forty minutes related to two university lessons using Voice-Care and APM 3200.  $SPL$  and  $F_0$  discretization intervals are 3 dB and 12 Hz respectively.

The full sample of teachers was divided into two groups of three schools, A and B, where A grouped the older school buildings, with an average value of mid-frequency reverberation time in the classrooms of 1.13 s, and B grouped the newer schools, with an average reverberation time of 0.79 s. The average background noise level estimated as the A-weighted percentile level  $L_{A90}$ , measured during traditional lessons, did not differ significantly between the two groups, with average values of 53.2 dB(A) and 50.4 dB(A), respectively.

It was also found that the average values of the vocal doses and parameters did not differ significantly for the two school groups, while the differences in the subjective scores were significant, which involved: the influence of acoustics on teaching, the noise intensity, the noise disturbance, the perceived reverberation, the teacher's vocal effort, the speech comprehension and the acoustical quality satisfaction. The comparison between the average scores of the two groups of schools, A and B, revealed higher negative scores in group A where the classrooms are more reverberant, except for the speech comprehension and the acoustical quality satisfaction scores, which were positively higher in group B. Given the discrepancy between the objective and subjective results related to voice for the two groups of schools, it has been supposed that the objective increase in the vocal fatigue due to reverberation could be detected in another way: the hypothesis is that the length of the voicing periods can increase due to the longer sound tail, with a consequently increase in the vocal fatigue.

A small number of working-day samples are excluded from the analysis since they are not characteristic of group A and group B, because the classrooms have been acoustically renovated or because the lessons were held in atypical way. A reduced group of 23 females and 2 males were selected, for a total of 18 and 24 working-day samples of 4 h each for the school A and B, respectively. This set of teachers is comparable to the group of subjects used by Titze et al. (2007) for the same research, and the results can be considered statistically significant to draw preliminary conclusion about vocalization of teachers.

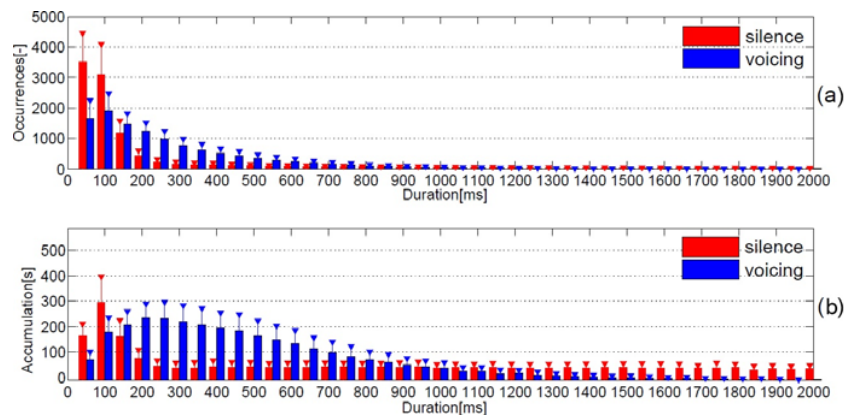
### Distributions of voicing and silence periods at school

Figure 4 and 5 show the ensemble averages of the histograms for voice and silence occurrences (a) over 4 h, for specific durations, and the voice and silence accumulation in seconds (b) for the same occurrences, related to the teachers in group A (18 samples) and B (24 samples) respectively.

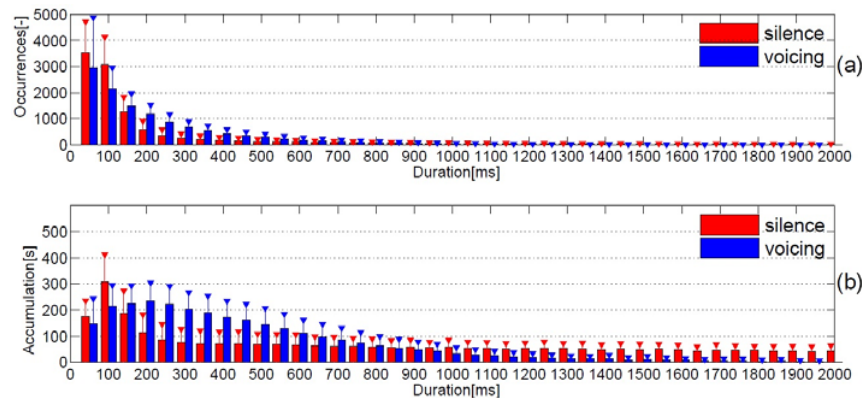
Since the acquired data has been subdivided into frames of 50 ms, the voicing or silence durations (or periods) are multiple of this interval and are represented on a linear scale. This scale has been chosen for detecting differences between the two groups with a greater detail. The accumulation for each period is obtained multiplying the number of occurrences by the corresponding duration.

For teachers in group A, that speak in schools with longer reverberation time, the greatest occurrence of voicing is in the 100 ms voicing period and the greatest occurrence of silence is in the 50 ms silence period. For teachers in group B, that speak in schools with lower reverberation time (in the acceptable range for speech intelligibility according to Yang and Bradley, 2009), the greatest occurrences of voicing and silence are in the same 50 ms period.

As far as the maximum accumulation is concerned, the two groups provide similar results with voicing periods between 200 ms and 250 ms and silence period of 100 ms.



**FIGURE 4.** Ensemble averages of the histograms for (a) voice and silence occurrences over 4 hours, for specific durations, and (b) the voice and silence accumulation in seconds, for the same occurrences, for the teachers in group A (18 samples). Error bars indicate standard deviation across samples.



**FIGURE 5.** Ensemble averages of the histograms for (a) voice and silence occurrences over 4 hours, for specific durations, and (b) the voice and silence accumulation in seconds, for the same occurrences, for the teachers in group B (24 samples).

Titze et al. (2007) represented the occurrences of voicing and silence periods taking into account the frame lengths typical of the speech rhythms and pauses, that led to the adoption of a logarithmic scale with bin duration of a half decade of logarithmic time in the range from  $0.0316\text{ s}$  to  $10^4\text{ s}$ . Over two weeks monitoring on 31 teachers using the NCVS dosimeter, they found two equal peaks of occurrences for voicing at work in the bins  $(0.0316 \div 0.1)\text{ s}$  and  $(0.10 \div 0.316)\text{ s}$ , and for silence in the bin  $(0.0316 \div 0.10)\text{ s}$ . The greatest accumulation of voicing periods has been found in the bin  $(0.316 \div 1.0)\text{ s}$  and the greatest accumulation of silence periods in the bin  $(3.16 \div 10)\text{ s}$ . The same results by Titze et al. (2007) have been found for the two groups of teachers of our work by clustering data in 5 bins of a half decade of logarithmic time in the range from  $0.0316\text{ s}$  to  $10\text{ s}$ . No significant difference has been found for the two groups of schools.

The shortest bin  $(0.0316 \div 0.10)\text{ s}$  includes voicing and silence periods below and up to the phonemic segmental level, the bin  $(0.10 \div 0.316)\text{ s}$  includes voicing and silence periods at the phonemic and syllabic level, the bin  $(0.316 \div 1.0)\text{ s}$  includes voicing and silence periods at the word and sentence level, the bin  $(1.0 \div 3.16)\text{ s}$  includes all-voiced sentences and pauses between sentences, and the bin  $(3.16 \div 10)\text{ s}$  includes sustained phonations and pauses between sentences (Titze et al., 2007; Zellner, 1994).

In conclusion, from the analysis of the distributions of the voicing and silence periods for the two groups of schools A and B, any significant difference does not emerge between the schools if the voicing and silence periods are clustered in 5 logarithmic bins related to speech rhythms and pauses, as proposed in Titze et al. (2007). As found in their research, the greatest occurrence of voicing is in the speech intervals below the phonemic segmental level up to the syllabic level, and the greatest occurrence of silence corresponds almost to the same interval.



When the periods are clustered in intervals multiple of 50 ms and represented on a linear scale, it is noticed that for teachers in group A the greatest occurrence of voicing increases of 50 ms with respect to the greatest occurrence of voicing in group B. Although the differences in voicing durations can be due to intra-speaker variability, other various effects can influence their durations (Van Heerden and Barnard, 2008), and one of the cause could be the longer sound tail. This result also supports the difference in subjective scores for the two groups of schools, which revealed higher negative scores in group A, whose classrooms are more reverberant.

The greatest accumulation of voicing periods can be directly related to vocal fatigue and the greatest accumulation of silence periods to short-term vocal recovery. In this work and in the similar work by Titze et al. (2007), the  $(0.316 \pm 1.0)$  s range, which includes voicing periods at the word and sentence level, represents the greatest accumulation of voicing periods, while the  $(3.16 \pm 10)$  s range, which includes pauses between sentences, the greatest accumulation of silence periods. Further analyses are needed to associate these accumulation intervals to the subjective perception of comfortable speaking.

### Distribution of voicing and silence periods within intervals of emphatic speech

It is supposed that teachers rise their voice in classroom to captivate the attention of the pupils, speaking with an emphatic (or exaggerated) inflection. A method for the detection of the distribution of voicing and silence durations within intervals of emphatic speech is proposed to detect if specific voice sizes are typical of this inflection.

The starting point is the invariability of  $D_t$  in monotone, normal and exaggerated speech (Titze et al., 2003). In our work  $D_t$  was of about 25% for both the two school groups A and B, so it is assumed the same  $D_t$  value in the emphatic speech intervals. Moreover, the emphatic speech has been identified as the speech for which the  $SPL$  is higher than the speech percentile level  $L_{v,10}$ , which represents the speech level exceeded for the 10% of the phonation time. The proposed method is able to automatically select time windows of emphatic speech of variable width that are characterized by a  $D_t$  of 25%. Figure 6 shows the ensemble averages of the histograms for voice and silence occurrences over 4 h, for specific durations, related to the emphatic speech for teachers in group A (18 samples) (a) and B (24 samples) (b) respectively.

For both the groups of teachers the greatest occurrence of voicing is in the 50 ms voicing period and the greatest occurrence of silence is in the 150 ms silence period. For teachers in group A a less number of occurrences and the presence of longer voicing and silence periods are noticed.

The less number of occurrences in group A is due to the presence of a less number of temporal windows which satisfies the  $D_t$  criterion of 25%. Group A is characterized by emphatic intervals longer than group B, not detectable with a fixed  $D_t$ . The presence of longer voicing periods can be related to the longer reverberation time, as already indicated in the analysis of the full speech sample.

As conclusion, for the reasons described above, the main result of this analysis on emphatic speech should be focused only on the peaks of durations of the voicing and silence periods, which are the same for the two groups. Anyway, further research on this topic are advisable for better detecting differences for emphatic speech in different room acoustics conditions.

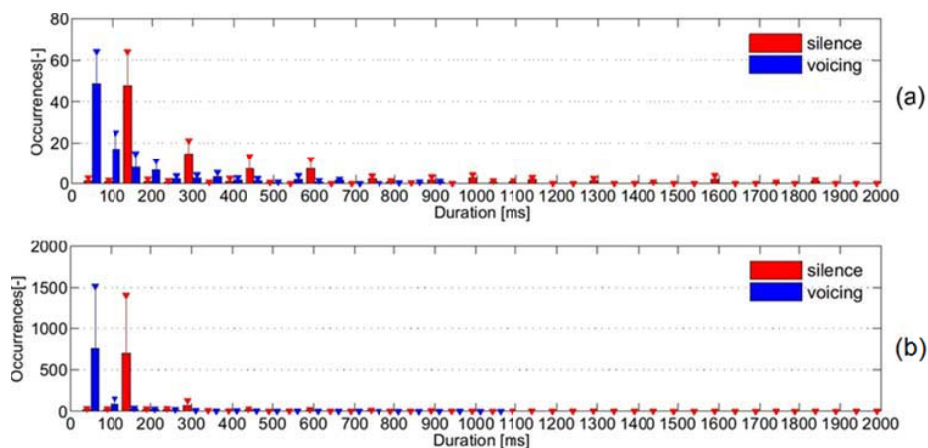


FIGURE 6. Ensemble averages of the histograms for voice and silence occurrences over 4 hours, for specific durations, related to emphatic speech for the teachers in group A (18 samples) (a) and in group B (24 samples) (b).

## REFERENCES

- Astolfi, A., Bottalico, P., Accornero, A., Garzaro, M., Nadalin, J., and Giordano, C. (2012). "Relationship between vocal doses and voice disorders on primary school teachers," *Proc. Euronoise 9*, 55-60.
- Bottalico, P. and Astolfi, A. (2012). "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms," *J. Acoust. Soc. Am.* **131**(4), 2817-2827.
- Bovo, R., Galceran, M., Petruccelli, J., Hatzopoulos, S. (2007). "Vocal problems among teachers: Evaluation of a preventive voice program," *J. Voice* **21**(6), 705-22.
- Carullo, A., Vallan, A., and Astolfi, A. (2013a). "Design issues for a portable vocal analyzer," *IEEE Trans. on IM*, in press, pp. 10, DOI: 10.1109/TIM.2012.2236724.
- Carullo, A., Vallan, A., and Astolfi, A. (2013b). "A low-cost platform for voice monitoring," *Proc. IEEE I2MTC 2013*, in press.
- Chan, R.W. (1994). "Does the voice improve with vocal hygiene education? A study of some instrumental voice measures in a group of kindergarten teachers," *J. Voice* **8**, 279-291.
- Cheyne, H.A., Hanson H.M., Genereux, R.P., Stevens, K.N., and Hillman, R.E. (2003). "Development and testing of a portable vocal accumulator," *J. Speech Lang. Hear. Res.* **46**(6), 1457-1467.
- Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., and Cheyne, H. A. (2006). "Ambulatory monitoring of disordered voices," *Ann. Otol. Rhinol. Laryngol.* **115**(11), 795-801.
- Hillman, R.E., and Mehta, D.D. (2011). "Ambulatory Monitoring of Daily Voice Use," *ASHA Journals Perspectives on Voice and Voice Disorders*, 56-61.
- Hunter, E., and Titze, I. R. (2009). "Quantifying vocal fatigue recovery: dynamic vocal recovery trajectories after a vocal loading exercise," *Ann. Otol. Rhinol. Laryngol.* **118**, 449-460.
- Hunter, E., and Titze, I. R. (2010). "Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings," *J. Speech Lang. Hear. Res.* **53**, 862-875.
- Hunter, E. (2010). "Vocal dose measures: general rationale and recent efforts made by the national center for voice and speech," *Rivista Italiana di Acustica* **34** (3), 37-42.
- International Organization for Standardization (2003). ISO 9921 (2003), *Ergonomics - Assessment of Speech Communication* (Genève).
- KayPENTAX (2013). "Ambulatory Phonation Monitor (APM), Model 3200," [www.kaypentax.com](http://www.kaypentax.com) (Last viewed 18 Jan. 2013.)
- Lyberg Åhlander, V., Rydell, R., and Löfqvist, A. (2011). "Speaker's comfort in teaching environments: voice problems in swedish teaching staff," *J. Voice* **25**(4), 430-440.
- Lyberg Åhlander, V., Rydell, R., and Löfqvist, A. (2012). "How do teachers with self-reported voice problems differ from their peers with self-reported voice health?," *J. Voice* **26**(4), e149-e161.
- Masuda, T., Ikeda, Y., Manako, H., and Komiyama, S. (1993). "Analysis of vocal abuse: Fluctuations in phonation time and intensity in 4 groups of speakers," *Acta Otolaryngol.* **113**, 547-552.
- Nanjundeswaran, C., Li, N.Y.K., Chan, K.M.K., Wong, R.K.S., Yiu, E.M.L., and Verdolini-Abbott, K. (2012). "Preliminary data on prevention and treatment of voice problems in student teachers," *J. Voice* **26**(6), 816.e1-816.e12.
- Pelegrín-García, D., and Brunskog, J. (2012). "Speakers' comfort and voice level variation in classrooms: laboratory research," *J. Acoust. Soc. Am.* **132**(1), 249-260.
- Popolo, P. S., Rogge, M. K., Švec, J. G., and Titze, I. R. (2005a). "Technical considerations in the design of a wearable voice dosimeter," NCSV, Salt Lake City, UT, [www.ncsv.org](http://www.ncsv.org) (Last viewed 19 Jan. 2013.)
- Popolo, P. S., Švec, J. G., and Titze, I. R. (2005b). "Adaptation of a pocket PC for use as a wearable voice dosimeter," *J. Speech Lang. Hear. Res.* **48**(4), 780-791.
- Švec, J. G., Titze, I. R., and Popolo, P. S. (2005). "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," *J. Acoust. Soc. Amer.* **117**(3), 1386-1394.
- Titze, I.R., Švec, J.G., and Popolo, P.S. (2003). "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," *J. Speech Lang. Hear. Res.* **46**, 919-932.
- Titze, I.R., Hunter, E.J., and Švec, J.G. (2007). "Voicing and silence periods in daily and weekly vocalizations of teachers," *J. Acoust. Soc. Am.* **121**(1), 469-478.
- Van Heerden, C. J., Barnard, E. (2008). "Speaker-specific variability of phoneme durations," *South Afr. Comput. J.* **40**, 44-50.
- VoxLog Portable Voice Meter (2013). <http://www.sonvox.com/voxlog>. (Last viewed 20 Jan. 2013.)
- Wirebrand, M. (2011). *Real-time monitoring of voice characteristics using accelerometer and microphone measurements*, Master thesis, Dept. Elect. Eng., Linköping Univ., Linköping, Sweden.
- 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.
- Zellner, B. (1994). "Pauses and the temporal structure of speech," in E. Keller (Ed.) *Fundamentals of speech synthesis and speech recognition*, Chichester: John Wiley, pp. 41-62.