

# Teachers' voicing and silence periods during continuous speech in classrooms with different reverberation times

Pasquale Bottalico,<sup>1,a)</sup> Arianna Astolfi,<sup>2</sup> and Eric J. Hunter<sup>1</sup>

<sup>1</sup>*Voice Biomechanics and Acoustics Laboratory, Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan 48824, USA*

<sup>2</sup>*Department of Energy, Politecnico di Torino, Torino, Italy*  
*pb@msu.edu, arianna.astolfi@polito.it, ejhunter@msu.edu*

**Abstract:** The relationship between reverberation times and the voicing and silence accumulations of continuous speech was quantified in 22 primary-school teachers. Teachers were divided into a high and a low reverberation time groups based on their classroom reverberation time (higher and lower than 0.90 s). Reverberation times higher than 0.90 s implicate higher voicing accumulations and higher accumulations of the silences typical of turn taking in dialogue. These results suggest that vocal load, which can lead to vocal fatigue, is influenced by classroom reverberation time. Therefore, it may be considered a risk factor for occupational voice users.

© 2017 Acoustical Society of America

[AL]

**Date Received:** August 26, 2016     **Date Accepted:** December 8, 2016

## 1. Introduction

Excessive reverberation and noise can be perceived as disturbing to the speaking situation. In the case of voice professionals, such as teachers, whose occupation requires them to be intelligible, adjustment of speech (e.g., lengthened words and pauses) can be a natural adaptive action (Cooke *et al.*, 2014). This behavior is typical of clear speech, which is produced spontaneously when high speech intelligibility is required (Picheny *et al.*, 1986).

While it has been long shown that speech produced in the presence of noise, i.e., Lombard Speech, typically exhibits evidence of word lengthening and the insertion of more and longer pauses (Summers *et al.*, 1988), few studies have dealt with the influence of room acoustics on variations in the speech duration in voice professionals. Black (1950) noted that a reader's mean duration per phrase is longer in larger and more reverberant rooms, while Pelegrín-García *et al.* (2011) found that the phonation time percentage was higher in reverberant room than in anechoic conditions. A tendency to increase the occurrence of longer voicing periods was also observed in voice professionals for higher reverberation times (RT; Astolfi *et al.*, 2015); this tendency resulted in increased phonation time percentages for teachers who taught in more reverberant rooms.

From a clinical point of view, vocal fatigue is used to denote negative vocal adaptation that occurs as a consequence of prolonged voice use or vocal load (Scherer *et al.*, 1991; Welham and Maclagan, 2003). Vocal recovery from prolonged voice use is usually measured on the order of several hours to several days (Hunter and Titze, 2009). However, prolonged voice use does not result from continuous voicing but from short bursts of voicing that accumulate throughout the day and, as suggested by Titze *et al.* (2007), might be related directly to vocal fatigue. Titze *et al.* further discusses the potential benefits of short silence periods (between the voicing) which could be related to short-term vocal recovery. They investigated the distribution of voicing and silence periods for teachers at work and reported the overall voice accumulation of each period by multiplying the number of occurrences by their corresponding duration. Nevertheless, a teacher's voicing and silence periods are likely changing as speech production adjusts to the communication environment (e.g., Lombard, reverberation time). Therefore, a teacher's voicing and silence periods may be influenced by a classroom RT yet such a relation has not been studied and may give insight onto how the RT may affect vocal health.

---

<sup>a)</sup> Author to whom correspondence should be addressed.

Research is needed to associate the change in voicing (e.g., accumulation of short voicing and silence intervals) due to the presence of noise or noise and reverberation, and to correlate them to the perception of uncomfortable speech, vocal fatigue and vocal recovery, for voice professionals. In the present study, we hypothesize that high RT is a risk factor for vocal fatigue in voice professionals because it increases the vocal load with the implication that increased load would require increased recovery time. To test this hypothesis, we quantified the relationship between classroom RT conditions and the voicing and silence accumulations of continuous speech in a sample of vocally healthy primary school teachers.

## 2. Experimental method

The participating teachers were part of a larger study (Bottalico and Astolfi, 2012). All teachers underwent clinical examinations (Bottalico *et al.*, 2016), and only the subjects without severe voice disorders were selected for the present analysis. The subjects with severe voice disorders were not included because they could have shown a difference vocal behavior (Bottalico *et al.*, 2016; Åhlander *et al.*, 2014). The sub-sample includes 22 teachers at six primary schools in Italy: three schools were built at the end of the nineteenth century, while the others were built in the 1970s. These buildings provided a large range of classroom RTs, from 0.68 to 1.58 s, with the mean value equal to 0.96 s and the median equal to 0.9 s. The teachers were divided in two groups by the RT. If RT in the classroom where the teachers were monitored, was higher than the median (0.9 s), the teacher was inserted in the group high RT, otherwise in the group low RT. The average background noise levels in the high RT and in the low RT groups were 53.6 dB(A) [standard deviation 7.8 dB(A)] and 50.6 dB(A) [standard deviation 5.4 dB(A)]. The difference in the background noise levels between the two groups was not statistically significant ( $t = 1.28$ ,  $df = 23.2$ ,  $p$  value = 0.21). For this reason, the background noise levels were not included in the analysis. More details on the larger study are reported in Bottalico and Astolfi (2012).

The mean age of the 22 teachers was 45 years (range 31–59 years). The teachers were monitored over one or two workdays during 4 hour blocks, the standard working hours for a teacher in Italy. A total of 39 workday samples were collected and included in the analyses. The average duration of the monitoring was 228 min (3.8 h). In the classrooms with high RT the average was 226 min (3.76 h), while in the classrooms with low RT the average was 232 min (3.85 h). The difference of the monitoring times between the two groups was not statistically significant ( $t = -0.73211$ ,  $df = 40.106$ ,  $p$  value = 0.4684). However, the values of the accumulations were reported in seconds/hour in order to compensate for possible difference in the monitoring time. Table 1 reports the main characteristics of the teachers and the number of monitored workdays.

Each teacher was fitted with the Ambulatory Phonation Monitor (APM, model 3200, KayPENTAX, Montvale, NJ). This device consists of an accelerometer, which is positioned at the sternal notch, and an acquisition unit that processes the accelerometer signal. The APM 3200 provides a time-history of voicing information using a frame length of 0.05 s; the voicing information included fundamental frequency,  $f_0$ , and an estimation of the sound pressure level (SPL) at a distance of 15 cm on-axis from the speaker's mouth. Fundamental frequency calculation and estimated SPL were dependent on a subject-specific setup routine, which included a reference microphone in order to correlate the skin acceleration level to the SPL.

Of the information provided by the device, only the presence or absence of detected voice excitation is of interest for the present study. Voiced and unvoiced frames were discriminated by the APM. When the root-mean-square level acquired by the transducer exceeded a preset threshold, the frame was designated as voiced, and for that frame,  $f_0$  and SPL were determined (Cheyne *et al.*, 2003). Otherwise, the output for the frame was equal to 0.

The number of continuous periods of voicing and silence which were between 0.2 to 10 s (in 0.05 s steps) were identified from APM time histories. Subsequently, the time accumulations for each continuous period were calculated by multiplying the number of occurrences by the corresponding duration of the period. The accumulations of voicing and silence periods below 0.2, which were the first three continuous time periods (0.05, 0.10, and 0.15 s), were not considered in this study because, as suggested by Titze *et al.* (2007), least two data points are required to determine the shortest on-off sequence, and considerable "sampling noise" may have contaminated the data in the first three time steps. Furthermore, the threshold commonly used to define a pause in natural speech is 0.25 s (Picheny *et al.*, 1986). As far as the voicing accumulation is concerned, the focus of this analysis was on periods between 0.2 and 1.3 s. The upper limit of 1.3 s was chosen considering the fact that the values of the accumulations for periods longer

Table 1. Characteristics of the investigated teachers, including gender, age, number of monitored working-days and mid-frequencies RT of the classrooms where they had taught.

Subject	Gender	Age	Number of monitored working-days	RT <sub>m,500-2kHz</sub> (s)
1	Female	38	2	1.00
2	Female	43	2	0.99
3	Female	54	2	1.58
4	Female	35	1	1.00
5	Female	39	2	0.99
6	Female	40	2	1.21
7	Female	47	1	1.04
8	Female	31	1	1.09
9	Female	34	2	1.30
10	Female	58	1	1.09
11	Female	34	2	0.83
12	Female	33	2	0.83
13	Male	43	2	0.90
14	Female	49	1	0.83
15	Male	59	2	0.76/1.11 <sup>a</sup>
16	Female	38	2	0.71
17	Female	52	2	0.85
18	Female	55	2	0.73
19	Female	58	2	0.90
20	Female	54	2	0.73
21	Female	56	2	0.68
22	Female	40	2	0.82

<sup>a</sup>The two working-days were monitored in different classrooms.

than 1.3 s were close to 0 seconds/hour and the fact that, according to Italian prosody (C-ORAL-ROM, 2005), the range of the words' duration in Italian language is between 0.2 and 1.3 s. The signal processing was performed with MATLAB R2016a (MathWorks, Natick, MA). Statistical analysis was conducted using R version 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria). For both types of accumulation, nonlinear regression models were fitted combining the function *lm* and *poly* in R.

### 3. Results

Figure 1 shows the average values of voicing accumulations in seconds per hour for continuous voicing period durations between 0.2 and 1.3 s (word level) for all the

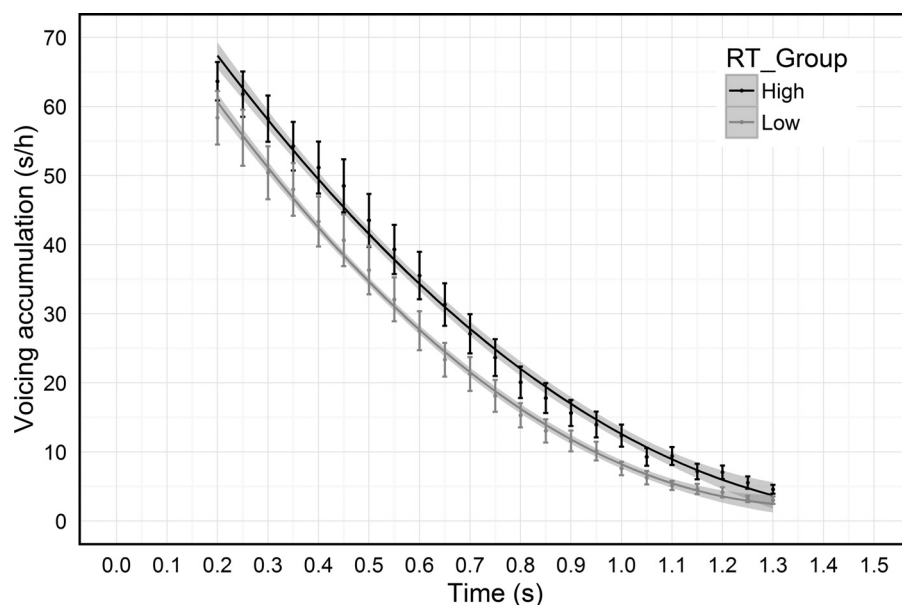


Fig. 1. Mean values of voicing accumulations in seconds per hour, for time step duration between 0.2 and 1.3 s for all the subjects, per group of RT (high for RT higher than 0.9 s in grey, low for RT lower than 0.9 in black), with SE indicated by error bars. The curves represent the polynomial regression models in the two conditions (high and low RT), together with the 95% confidence bands.

Table 2. Polynomial models (2 degree) for response variables voicing and silence accumulations considering as predictor for the interaction between time (first and the second degree terms) and RT Group (high and low).

Voicing accumulations	Estimate	Standard Error	<i>t</i> value	<i>p</i> value <sup>a</sup>
(Intercept)	85.40	1.98	43.02	<0.001
Time: high RT	-103.16	6.16	-16.75	<0.001
Time <sup>2</sup> : high RT	30.65	4.22	7.26	<0.001
Time: low RT	-125.27	6.25	-20.05	<0.001
Time <sup>2</sup> : low RT	47.746	4.34	11.00	<0.001
Silence accumulations				
(Intercept)	17.01	0.15	111.65	<0.001
Time: high RT	-2.82	0.07	-37.58	<0.001
Time <sup>2</sup> : high RT	0.15	0.01	19.79	<0.001
Time: low RT	-3.19	0.08	-41.36	<0.001
Time <sup>2</sup> : low RT	0.19	0.01	23.93	<0.001

<sup>a</sup>All *p* values were significant at the <0.001 levels.

subjects. Because a higher number of voicing occurrences is present in shorter voicing periods, the voicing accumulation in time (seconds/hour) from these short voicing periods is higher than the accumulation from longer voice periods. The grey points represent the mean voicing accumulation in classrooms with low RT ( $0.58 \text{ s} < \text{RT} < 0.90 \text{ s}$ ), while the black points represent the voicing accumulation in classrooms with high RT ( $0.90 \text{ s} \leq \text{RT} < 1.58 \text{ s}$ ). Eighteen working days were monitored in classroom with low RT, while 21 in classrooms with high RT. The error bars represent the standard errors. The curves represent the polynomial regression models in the two conditions (high and low RT) together with the 95% confidence bands. The model results are shown in Table 2 and include the intercept and the interaction between time (first and the second degree terms) and RT group (high and low). The residual standard error was 11.36 with 892 degrees of freedom. The  $R^2$  adjusted for the complexity of the model was 0.74 and *F* statistic was 628; while the *p* value was lower than 0.0001. As it can be noticed from Fig. 1, the values of voicing accumulation in low RT classrooms are significantly shorter than the accumulation in high RT classrooms in almost all the range of Time (0.2–1.3 s). The difference seems to be negligible for voice frames durations longer than 1.2 s, where the confidence bands of the two curves overlap.

Figure 2 shows the average values of continuous silence accumulations in seconds per hour for duration periods between 0.2 and 10 s for all the subjects. Because a higher number of silence occurrences is present in shorter silence periods, the silence

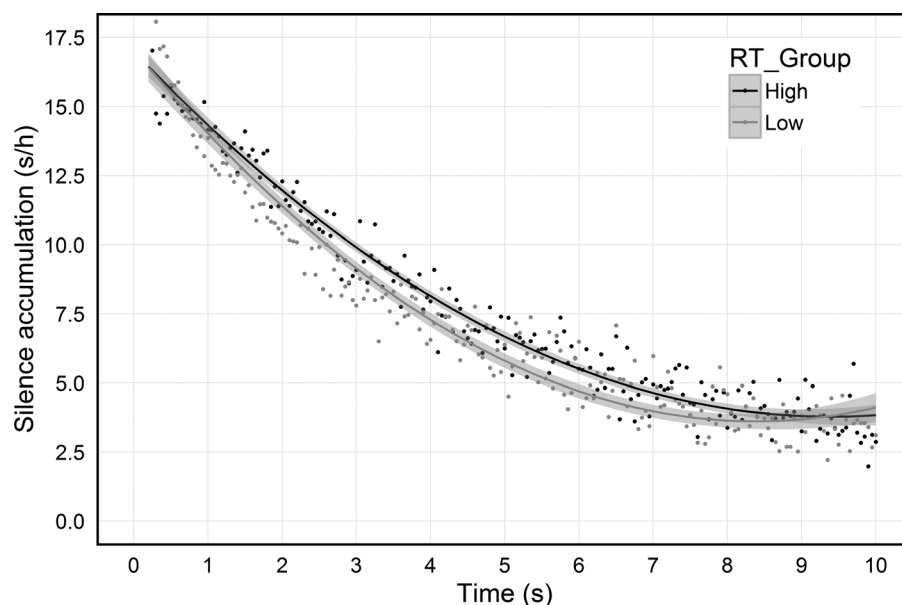


Fig. 2. Mean values of silence accumulations in seconds per hour, for time step duration between 0.2 and 10 s for all the subjects, per group of RT (high for  $\text{RT} \geq 0.9 \text{ s}$  in grey, low for  $\text{RT} < 0.9$  in black). Error bars have not been included for a better data visualization. The curves represent the polynomial regression models in the two conditions (high and low RT), together with the 95% confidence bands.

accumulation in time (seconds/hour) from these short periods is higher than the accumulation from longer silence periods. The grey points represent the silence accumulation in classroom with low RT ( $0.58\text{ s} < \text{RT} < 0.90\text{ s}$ ), while the black points represent the silence accumulation in classroom with high RT ( $0.90\text{ s} \leq \text{RT} < 1.58\text{ s}$ ). The curves represent the polynomial regression models in the two conditions (high and low RT) together with the 95% confidence bands. The model results are shown in Table 2 and include the intercept and the interaction between time (first and the second degree terms) and RT Group (high and low). The residual standard error was 4.15 with 7678 degrees of freedom. The  $R^2$  adjusted for the complexity of the model was 0.46 and  $F$  statistic was 1634; while the  $p$  value was lower than 0.0001. As it can be noticed from Fig. 2, the values of silence accumulation in low RT classrooms are significantly shorter than the accumulation in high RT classrooms in almost all the ranges of continuous time periods (0.2–10 s). The difference seems to be negligible for time periods shorter than 1.5 s and longer than 8 s, where the confidence bands of the two curves are overlapped.

#### 4. Conclusions

The determination of the distribution of short voicing and silence periods represents the first step in the quantification of the amount of vibration exposure that teachers have by talking. Bottalico *et al.* (2016) assessed the relationship between the silence and voicing accumulations of primary school teachers and the teachers' clinical status. The aim of that study was to determine whether more voicing accumulations periods and fewer silence accumulations were measurable for the vocally unhealthy subjects than for the healthy subjects, which would imply more vocal loading and fewer short-term recovery moments. The authors concluded that the teachers with structural voice disorders accumulated more voicing occurrences in intervals between 0.1 and 3.15 s than teachers without structural voice disorders.

In the current study, only teachers without structural voice disorders were considered with the goal to find possible associations between the vocal behavior, in terms of voicing and silence accumulations, and the RT in the classrooms where they were teaching.

Higher voicing accumulations in all the time intervals were found in higher RT. This finding supports the results of Black (1950), Pelegrín-García *et al.* (2011), Astolfi *et al.* (2015), and Bottalico (2016). All these studies found higher phonation times in more reverberant conditions. Titze *et al.* (2007) and Bottalico *et al.* (2016) suggested that the elevated accumulation of voicing periods might be related directly to vocal fatigue and lead to a voice disorder.

Regarding the silence accumulations, the silence accumulated by the low and high RT teachers was similar for intervals shorter than 1.5 s (associated to short pauses between words typical of monologue communication) and for intervals longer than 8 s. The pauses longer than 8 s represent the greatest accumulation of vocal rest during the workday for teachers. The silence accumulated between 1.5 and 8 s were higher in classrooms with high RT. This length of silences is typical of pauses between sentences, as perhaps waiting for a response from a student (Titze *et al.*, 2007). The results suggest that teachers are adjusting their teaching strategies, spending more time waiting for the students to answer in classroom with higher RT, probably with the goal of improving intelligibility.

Future research will consider the effect on voicing and silence accumulations of RT and other acoustical parameters using a larger sample size. Additionally, future research will consider how such acoustic effects would impact the subjective perception of vocal effort and vocal fatigue in real classrooms. The results of this study provide insight into how RT influences vocal load, which is directly related to vocal fatigue (Scherer *et al.*, 1991), and how bad acoustics can be considered a risk factor for occupational voice users.

#### Acknowledgments

The kind cooperation of the teachers has made this work possible. This research was supported by the Italian National Institute for Occupational Safety and by the National Institute on Deafness and Other Communication Disorders of National Institutes of Health under Award Number R01DC012315. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References and links

- Åhlander, V. L., Pelegrín-García, D., Whitting, S., Rydell, R., and Löfqvist, A. (2014). "Teachers' voice use in teaching environments: A field study using ambulatory phonation monitor," *J. Voice* **28**(6), 841.e5–841.e15.
- Astolfi, A., Carullo, A., Pavese, L., and Puglisi, G. E. (2015). "Duration of voicing and silence periods of continuous speech in different acoustic environments," *J. Acoust. Soc. Am.* **137**(2), 565–579.
- Black, J. (1950). "The effect of room characteristics upon vocal intensity and rate," *J. Acoust. Soc. Am.* **22**, 174–176.
- Bottalico, P. (2016). "Speech adjustments for room acoustics and their effects on vocal effort," *J. Voice*, DOI: 10.1016/j.jvoice.2016.10.001.
- 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.
- Bottalico, P., Graetzer, S., Astolfi, A., Eric, J., and Hunter, E. J. (2016). "Voicing and silence accumulations in Italian primary school teachers with and without voice disorders," *J. Voice*, DOI: 10.1016/j.jvoice.2016.05.009.
- 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.
- Cooke, M., King, S., Garnier, M., and Aubanel, V. (2014). "The listening talker: A review of human and algorithmic context-induced modifications of speech," *Comput. Speech Lang.* **28**, 543–571.
- C-ORAL-ROM (2005). *Integrated Reference Corpora for Spoken Romance Languages*, edited by E. Cresti and M. Moneglia (John Benjamins Publishing Company, Amsterdam, the Netherlands), 304 pp.
- Hunter, E. J., and Titze, I. R. (2009). "Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading exercise," *Ann. Otol. Rhinol. Laryngol.* **118**(6), 449–460.
- Pelegrín-García, D., Smits, B., Brunskog, J., and Jeong, C. (2011). "Vocal effort with changing talker-to-listener distance in different acoustic environments," *J. Acoust. Soc. Am.* **129**(4), 1981–1990.
- Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Lang. Hear. Res.* **29**(4), 434–446.
- Scherer, R. C., Titze, I. R., Raphael, B. N., Wood, R. P., Ramig, L. A., and Blager, R. F. (1991). "Vocal fatigue in a trained and an untrained voice user," in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, and K. Harris (Singular Publishing Group, San Diego, CA), pp. 533–555.
- Summers, W. V., Pisoni, D. P., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). "Effects of noise on speech production: Acoustic and perceptual analyses," *J. Acoust. Soc. Am.* **84**(3), 917–928.
- 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.
- Welham, N. V., and Maclagan, M. A. (2003). "Vocal fatigue: Current knowledge and future directions," *J. Voice* **17**(1), 21–30.