Voice Production Changes in Artificial Environments

Pasquale Bottalico[1], Tomas Sierra-Polanco[1]

[1]Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, Champaign, IL, United States of America.
pb81@illinois.edu}

Abstract
The regulation of speech level is affected by the perception of the communication environment and auditory feedback. We examined the effects of room acoustics in an artificial setting on voice production in terms of sound pressure level and the relationship with the perceived vocal comfort and vocal control. Three independent room acoustic parameters were considered: gain (alteration of the sidetone or playback of one's own voice), reverberation time, and background noise. An increase in the sidetone led to a decrease in vocal sound pressure levels, thus increasing vocal comfort and vocal control. This effect was consistent in the different reverberation times considered. Mid-range reverberation times led to a decrease in vocal sound pressure level along with an increase in vocal comfort and vocal control, however, the effect of the reverberation time was smaller than the effect of the gain. The presence of noise amplified the aforementioned effects.

Keywords: room acoustics, gain, speech level, vocal comfort, vocal control.

1 Introduction

Vocal communication involves multiple physiologic (oral and aural) and cognitive systems. From the perspective of production, the regulation of speech level is primarily affected by physiological features of the speaker such as vocal tract size, vocal fold length, and lung capacity [1]. This production regulation is affected by perceived communication demands, such as communication partners or communication environment, sense of vocal comfort, and applied vocal effort. For example, speech level and speech style can be partner-specific such as speaking to a child [2] or to someone with a perceived hearing loss [3]. Another example was presented by [4] where aspects of auditory feedback such as background noise, altered sidetone (amplified playback of one's own voice), hearing loss, and room acoustics were described. The alteration in auditory feedback can modify vocal parameters, such as Sound Pressure Level (SPL), and can modify the talker's perception of vocal comfort and vocal control [5][6]. These parameters may be modified by the implementation of artificial settings delivered by headphones with the goal of increasing vocal comfort and control while decreasing vocal effort in occupational voice users such as teachers and call centre operators. All of these are affected by the relationship between voice production and hearing sensitivity [7] and how the auditory system and auditory feedback play a fundamental role in voice production including the perception of effort and comfort.
The alteration in auditory feedback can also modify the perception of a communication scenario, thus affecting voice production, vocal comfort, and the perception of vocal control. Vocal control can be defined as the capacity to self-regulate vocal production, e.g., SPL, fundamental frequency, and resonance. The sensation of control relates to the ability to adjust the voice consciously. In a communication environment, in general, speakers try to control their voice production in order to increase speech intelligibility. When talking in a noisy

environment, people tend to raise the level of their voice in order to maintain understandable communication [8]. The maximization of intelligibility, clarity, vocal comfort and control, and the minimization of vocal effort and fatigue, should be the priority of any professional talker [9].
Growing evidence suggests that there is an association between vocal production level and external auditory feedback. External auditory feedback consists of the external path between mouth and ears and is strongly influenced by the acoustics of the environment where the speaker is speaking. Such environmental effects are room noise, vocal amplification of one's own voice, and, room reverberation.
A commonly experienced external auditory effect that directly affects vocal production level is that of elevated room noise, or the Lombard Reflex or Effect [10], [8]. For example, [11] recorded a monologue passage for twenty-four vocally healthy young adults (12 men and 12 women, aged 19–22 years) using an Ambulatory Phonation Monitor (APM model 3200) under three natural environment conditions in a randomized order. These conditions were: a quiet room (clinic room, mean 35.5 dBA, ranged from 34 to 37 dBA), a room with moderate noise level (clinic corridor, mean 54.5 dBA, ranged from 53 to 56 dBA), and a room with high noise (a pantry room with a noisy exhaust fan, mean 67.5 dBA, ranged from 66 to 69 dBA). The results showed significant increases in mean voice level and self-reported vocal effort in the high-noise environment than in the other two conditions.
Vocal level was shown to be affected by the reverberation time of the room [12], and by the level at which a speaker perceived his/her own voice, as well as the level of the background noise [13]. More recently, studies have added further details to these and other factors such as speaker-listener distance and acoustic characteristics of the room and/or of the communication channel [12], [5] [14] [15][16][17]]. Pelegrín-García *et al.* [17] found that voice level decreased as reverberation time increased; while Black [12] reported that greater vocal intensity was found in less reverberant rooms than in more reverberant rooms. This is common even in extreme reverberation conditions [18].
Furthermore, external auditory feedback can be artificially altered by modifying the playback of one's own voice (i.e. sidetone alteration). In a study of the effect of sidetone alteration on voice levels by increasing the sidetone gain of 20 dB, Siegel and Pick [13] found a ratio of change in the voice level of 0.15 dB/dB. This ratio increased to 0.21 dB/dB, 0.30 dB/dB and 0.34 dB/dB when speech-spectrum noise was added during the experiment at 60 dB, 70 dB and 80 dB, respectively.
Recent investigations on speech adjustments were related to an increase of external auditory feedback [5] and to reverberation times [19]. The above mentioned showed that the effect of reflective panels, placed close to the speaker, had a decrease of about 1 dB in voice level, which was observable in rooms with different reverberation times and in different speech styles.
In summary, previous research suggests that voice level, vocal comfort, and vocal control vary (1) when the gain level of external auditory feedback increases and (2) under different reverberant conditions. These variations could be also affected by the presence of noise. The perceived vocal comfort was lower in rooms with very low or very high reverberation time. Nevertheless, to better understand how speech adjusts to room acoustics, it is necessary to have control of the acoustical parameters. This can be facilitated by creating virtual acoustics scenarios.
To explore this topic, the current study examined the effects of room acoustics in a virtual setting on vocal SPL, and self-reported vocal comfort and control. Three independent room acoustic parameters were considered: gain (alteration of the sidetone), reverberation time (T30), and background noise. This relationship was stated to better understand how these independent and dependent variables relate to each other in simulated scenarios. As we have mentioned, previous studies have been performed in real scenarios, which are not malleable nor changeable, but fixed. By having simulated scenarios, this study proposes a wide range of possibilities that could be infinitely modified, in a simple way, on its initial parameters for independent variables. The main research questions of this study were based on the following statements regarding relationships between:
(1) Voice level variations and participant's gain level of external auditory feedback (sidetone or self-amplification).
(2) Vocal comfort (and control) responses and participant's gain level of external auditory feedback (sidetone or self-amplification).

(3) Voice level variations and different simulated T30 of rooms where participants are speaking.
(4) Vocal comfort (and control) response and different simulated T30 of rooms where participants are speaking.
(5) Finally, if there are such effects:
(5a) Voice level variations and the presence or absence of noise.
(5b) Vocal comfort (and control) and the presence or absence of noise.
Hence, the present work is aimed to provide contributions on how acoustical environments affect voice production in terms of objective measurements such as SPL, but also in terms of perceptual measurements such as self-reported vocal comfort and vocal control.

2 Materials and Methods

The speech of 30 talkers was recorded in fourteen different virtual acoustical scenarios of external auditory feedback, including three gain levels and three T30, each of them with and without the presence of speech-shaped noise. The participants' speech was recorded with a microphone placed at a fixed distance of 15 cm from the mouth. A preliminary calibration procedure of the microphone was performed at the beginning of the recording session per participant. The calibration level was set to 94 dB at 1kHz. The recordings were performed in a soundproof double-walled Whisper Room (interior dimensions: 226 × 287 cm and h = 203 cm). T30 was measured for mid-frequencies to be 0.07 s in the soundproof room and background noise equal to 25 dB(A). The speech signals were processed to calculate SPL.

2.1 Participants and instructions

This study was conducted with approval from and in accordance with the policies of the Office of Protection of Research Subject at the University of Illinois at Urbana Champaign (IRB 18179). Thirty participants (17 females and 13 males) participated in this experiment. All the participants were Native American English-speaking young adults (age 19 to 32 years old; mean age 23 years), with self-reported normal speech and hearing, and no reported or observable upper respiratory infection on the day of the recording. In general, none of them reported hearing conditions. 26 participants reported that their primary ethnicity was "Caucasian", two were "Asian", and two "Hispanic-Latino". Four of them reported being eventual smokers. Five reported voice training in the past, such as singing lessons, and four reported a history of speech or language therapy in their childhood. The participants were instructed to read aloud a standardized text in English under fourteen different virtually simulated acoustic conditions. Each task had a duration of about 27 seconds of reading. Before the measurements, each participant was presented with the printed passage to familiarize themselves with it. The fourteen virtually simulated acoustic conditions were: a reference condition (no gain, no reverberation) and the result of all possible combinations of two gain levels of the external auditory feedback (+5 dB and +10 dB) and three different T30. The six aforementioned conditions were presented with and without speech-shaped noise added. The order of administration of the fourteen scenarios was randomized to provide an equal distribution of any (short-term) vocal discomfort across all the tasks, as well as to control for any unknown confounding variables relating to the task order.

2.2 Equipment

The speech material was recorded by a frequency response Class 1 microphone placed at a fixed distance of 15 cm from the mouth (M2211, NTi Audio, Tigard, OR, USA). The microphone was calibrated at the beginning of the recording session per participant using a Class 1 Sound Calibrator NTi Audio (Tigard, OR, USA) with automatic atmospheric pressure compensation (ref: 94 dB ± 0.2 dB at 1 kHz ± 1%). The microphone output was split into two lines: the first for direct recording and the second for creating the virtual acoustic environment. The direct digital recording sampled at 44.1 kHz was recorded using an external soundboard (UH-7000 TASCAM, Teac Corporation, Montebello, CA, USA) connected to a personal computer (PC) running Audacity 2.0.5 (SourceForge, La Jolla, CA). For the virtual environment, the direct microphone output was combined, in half of the conditions, with speech-shaped noise using a digital mixer (MultiMix 8 USB FX 8, Alesis, Cumberland, RI, USA). The voice signal was digitally processed to add reverberation using a real-time effect processor of the digital mixer and played back to the participant using open headphones (HD600, Sennheiser, Wedemark, Germany). The delay between the uttered voice and its transmission through the processing loop (i.e. Alesis digital mixer) and back to the participant's headphones was measured to be lower than 5 ms. This value is below the range between 16 and 26 ms threshold which is considered a noticeable echo (Lezzoum et al., 2016).

2.3 Room acoustic parameters

Room acoustic T30 conditions (ISO 3382-2, 2008) of the virtual scenarios were obtained from impulse responses (IRs) calculated with the convolution method. An exponential sweep signal was emitted by the mouth of a Head and Torso Simulator (HATS, GRAS 45BB KEMAR). The sweep was captured by the microphone, real-time processed, played back with open headphones, and finally recorded by the ears of the HATS. The recorded sweep was deconvolved with the emitted sweep inverted on the time axes, obtaining the IR, as exposed in the appendix by Pelegrín-García and Brunskog [6].
The average T30 for combined 500 Hz and 1 kHz octave bands, were determined for the Whisper Room and each of the 3 simulated environments (ISO 3382-2, 2008). It was 0.07 s in Whisper Room T30 condition, 1.13 s in Low T30 condition, 1.39 s in Medium T30 condition, and 1.90 s in High T30 condition. The measured values of T30 for the Whisper Room and the three simulated conditions between 125 and 8k Hz are given in Table 1. To manipulate the level of external auditory feedback, three different gain factors were introduced in the real-time processor. These gain factors were chosen with the goal of obtaining a difference between the voice level measured at the ears in the air (with no sidetone modification) and the voice level measured at the ears position after the real-time processor, equal to 0 dB, 5 dB, and 10 dB.
In 7 out of the 14 tasks performed by each participant, speech-shaped noise was added to the real-time processor with the same power. The power level was set to obtain an A-weighted equivalent level averaging both ears of about LAeq = 70 dB(A) at the ears of the talker (measured with the HATS). This level was chosen among the one used by Siegel and Pick (1974) to stimulate the variation in the voice level with the sidetone alteration without excessive noise exposure for the participants. The values per octave band for background noise conditions, with and without speech-shaped noise, are reported in Table 1.

Table 1 – T30 measured in Whisper Room conditions and 3 simulated environments (Low, Medium, and High) per octave band. Background noise conditions with and without speech-shaped noise spectrum per octave band. The measurements were performed with the HATS.

	125 Hz	250 Hz	500 Hz	1 KHz	2 KHz	4 KHz	8 KHz
T30 Whisper Room (s)	0.164	0.129	0.079	0.061	0.064	0.054	0.048
T30 Low (s)	0.512	0.821	1.071	1.191	0.922	0.799	0.016
T30 Medium (s)	1.318	1.279	1.383	1.403	1.351	1.270	1.161
T30 High (s)	1.763	1.721	1.965	1.835	1.371	1.163	0.884
Background noise (dB)	34.0	26.7	16.0	14.8	13.4	15.5	16.4
Speech-Shaped Noise (dB)	59.8	62.0	66.6	60.3	64.2	59.7	55.4

2.4 Voice processing and statistical analysis

Analysis of the speech parameters was performed with Matlab R2017a (MathWorks, Natick, MA, USA). For each of the 14 tasks, a time history of A-weighted SPL was obtained from recorded speech. The time information associated with time histories (which typically ranged from 0 to 30 seconds within a task) was be retained for inclusion in the statistical analysis.

Statistical analysis was conducted using R Studio (version 1.2.5033). Linear Mixed-Effects (LME) models were fitted by restricted maximum likelihood (REML). Random effects terms were chosen based on variance explained. Models were selected based on the Akaike information criterion and the results of likelihood ratio tests. Tukey's post-hoc pair-wise comparisons (Multiple Comparisons of Means: Tukey Contrasts) were performed to examine the differences between all levels of the fixed factors of interest. These are pair-wise z tests, where the z statistic represents the difference between an observed statistic and its hypothesized population parameter in units of the standard deviation. The LME output includes the estimates of the fixed effects coefficients, the standard error associated with the estimate, the degrees of freedom (df), the test statistic (t), and the p-value. The Satterthwaite method is used to approximate degrees of freedom and calculate p-values.

3 Results

Six Linear Mixed Effects (LME) models were run, two LME for each of the three different response variables: SPL, vocal comfort, and vocal control. The first of the two sets of LME models focused on gain as a fixed effect, while the second on T30. Both of them considered the effect of noise and gender as a fixed factor. The results section is divided into two subsections: (1) effects of gain and noise on SPL, vocal comfort, and vocal control, and (2) effects of T30 and noise on SPL, vocal comfort, and vocal control.

3.1 Effect of gain

A Linear Mixed Effects (LME) model was run with the response variable SPL (in dB(A)). This model, reported in Table 2, has the following fixed factors (1) gain, (2) noise, (3) gender, and (4) the interaction of gain and noise. The random effects were (1) T30, (2) chronological task order, (3) time (where time was measured in ms for each participant overall assessment), and (4) identification number of each participant. The reference levels used in the models were: 0 dB for gain, background without speech-shaped noise (No Noise) for noise condition, and female for gender. The estimates of standard deviation for time as a random effect was 1.26 dB(A), for participant identification number was 2.78 dB(A), for order was 0.20 dB(A), and for T30 was 0.07, whereas the residual standard deviation was 6.49 dB(A). The mean variation in SPL from 0 dB to 5 dB of gain for no noise added condition, was -0.31 dB(A), while it was -1.41 dB(A) from 0 dB to 10 dB. When the speech-shaped noise is added, overall, the voice SPL increases 3.49 dB(A). When noise was added, the differences from 5 dB and 10 dB to the reference level (0 dB), were -0.78 dB(A) and -2.65 dB(A), respectively. Post-hoc comparisons were made considering the effect of gain and its interaction with noise. These comparisons confirmed that, overall, SPL measured in 0 dB of gain condition was higher than both, that in the condition with 5 dB of gain (-0.54 dB(A), SE=0.09, $z = -5.87$, $p<0.001$) and in 10 dB of gain condition (-2.03 dB(A), SE=0.09, $z = -21.95$, $p <0.001$), whereas the difference between 5 dB and 10 dB gain conditions was -1.49 dB(A) (SE=0.03, $z = -46.98$, $p<0.001$).

One LME model was run with the response variable self-reported vocal comfort in % (0 = 'not at all comfortable', 100 = 'extremely comfortable') and the fixed factors (1) gain, (2) noise, (3) gender, and (4) the interaction between gain and noise. The random effects were (1) T30, (2) chronological task order, and (3) participant. The output of this model is reported in Table 3. The estimate of standard deviation for participant as a random effect was 14.58%, for order was 2.56% and for T30 was 0.00 %, whereas the residual standard deviation was 14.74%. The mean increase in self-reported vocal comfort from 0 dB to 5 dB of gain was 3.55%, while it was 0.27% from 0 dB to 10 dB; in the conditions without noise added. The vocal comfort decreased by -25.31% when the speech-shaped noise was added. For these conditions, when the noise was added, the mean increase in self-reported vocal comfort from 0 dB to 5 dB of gain was 13.43%, while it was 20.77% from 0 dB to 10 dB. Post-hoc comparisons confirmed that, overall, the vocal comfort measured in the condition with 0 dB of gain was lower than that in both the condition with 5 dB of gain (8.50%, SE=2.26, $z = 3.76$, $p<0.001$), and the condition with 10 dB of gain (10.52%, SE=2.26, $z = 4.66$, $p<0.001$). Furthermore, the vocal comfort reported in the condition with 10 dB of gain was 2.02% higher than that in the condition with 5 dB of gain (SE=1.59, $z =1.27$, $p =0.406$).

The analysis of vocal control was similar to vocal comfort. One LME model was run with the response variable self-reported vocal control in % (0 = 'not at all controlled', 100 = 'extremely controlled') and the fixed factors (1) gain, (2) noise, (3) gender, and (4) the interaction between gain and noise. The random effects were (1) T30, (2) chronological task order, and (3) participant. The output of the model is reported in Table 4. The estimate of standard deviation for participants as a random effect was 13.47%, for order was 2.71%, and 0.00% for T30, whereas the residual standard deviation was 13.20%. The mean decrease in self-reported vocal control from 0 dB to 5 dB of gain was 0.45%, while it was 3.01%, from 0 dB to 10 dB; in the conditions without noise added. The vocal control decreased by 24.19% when the speech-shaped noise was added. For these conditions, when the noise was added, the mean increase in self-reported vocal control from 0 dB to 5 dB of gain was 13.28%, while it was 18.87% from 0 dB to 10 dB. Post-hoc comparisons regarding the interactions between gain and noise confirmed that, overall, the vocal control measured in the condition with 0 dB of gain was lower than that in both the conditions with 5 dB of gain (6.42%, SE=2.03, z = 3.17, p=0.004) and the condition with 10 dB of gain (7.93%, SE=2.02, z = 3.92, p<0.001). Furthermore, the vocal control reported in the condition with 10 dB of gain was 1.51% higher than that in the condition with 5 dB of gain (SE=1.43, z =1.06, p =0.533).

Table 2 – LME models fit by REML for the response variable SPL and the fixed factors (1) gain, (2) noise, (3) gender, and the interaction between gain and noise.

	Estimate	Std. Error	df	t value	p-value	
(Intercept)	71.76	0.68	31.7	104.89	<0.001	***
Gain 5	-0.31	0.10	5.9	-3.00	0.024	*
Gain 10	-1.41	0.10	5.9	-13.66	<0.001	***
Noise Speech-Shaped	3.49	0.08	176742.7	45.12	<0.001	***
Gender Male	2.93	1.02	30.0	2.86	0.008	**
Gain 5: Noise Speech-Shaped	-0.47	0.09	184158.1	-5.23	<0.001	***
Gain 10: Noise Speech-Shaped	-1.24	0.09	186146.1	-13.93	<0.001	***

Table 3 – LME models fit by REML for the response variable self-reported comfort and the fixed factors (1) gain, (2) noise, (3) gender, and (4) the interaction between gain and noise.

	Estimate	Std. Error	df	t value	p-value	
(Intercept)	80.40	4.67	61.2	17.23	<0.001	***
Gain 5	3.55	3.18	367.4	1.12	0.265	
Gain 10	0.27	3.18	367.1	0.08	0.993	
Noise Speech-Shaped	-25.32	3.91	369.4	-6.48	<0.001	***
Gender Male	-1.75	5.64	27.0	-0.31	0.759	
Gain 5: Noise Speech-Shaped	9.88	4.51	368.7	2.19	0.029	*
Gain 10: Noise Speech-Shaped	20.50	4.50	367.4	4.55	<0.001	***

Table 4 – LME models fit by REML for the response variable Control and the fixed factors (1) gain, (2) noise, (3) gender, (4) the interaction between gain and noise.

	Estimate	Std. Error	df	t value	p-value	
(Intercept)	86.39	4.28	59.9	20.19	<0.001	***
Gain 5	-0.45	2.85	366.3	-0.16	0.874	
Gain 10	-3.01	2.85	366.1	-1.06	0.291	
Noise Speech-Shaped	-24.19	3.51	368.4	-6.90	<0.001	***
Gender Male	-1.19	5.20	27.0	-0.23	0.821	
Gain 5: Noise Speech-Shaped	13.73	4.04	367.6	3.40	<0.001	***
Gain 10: Noise Speech-Shaped	21.88	4.04	366.3	5.42	<0.001	***

3.2 Effects of reverberation time (T30)

One LME model was run with the response variable SPL (in dB(A)). This model has (1) T30, (2) noise, (3) gender, and (4) the interaction between T30 and noise as fixed factors, as reported in Table 5. The random effects were (1) gain, (2) chronological task order, (3) time (where time was measured in ms for each participant overall assessment), and (4) identification number of each participant. The reference levels used in this model were: Low T30, background without speech-shaped noise (No Noise) for noise conditions, and female for gender. The differences among T30 conditions were more pronounced in the noise added conditions. The estimates of standard deviation for time as a random effect was 1.25 dB(A), for participant was 2.82 dB(A), for order was 0.26 dB(A), and for gain was 0.85 dB(A), whereas the residual standard deviation was 6.43 dB(A). The mean variation in SPL from the Low T30 to Medium T30 was -0.08 dB(A), and a variation of -0.01 dB(A) from Low T30 to High T30, without noise added. When artificial speech-shaped noise was present, voice SPL increases 2.80 dB(A). For noise added conditions, the differences were -0.25 dB(A) and -0.23 dB(A) for Low T30 versus Medium T30 and Low T30 versus High T30, respectively. Generically, males were louder than females by 2.93 dB(A). Post-hoc comparisons including interaction between T30 and noise confirmed that, overall, SPL measured in Low T30 condition was higher than that in the condition with Medium T30 (-0.17 dB(A), SE=0.04, z = -4.28, p<0.001) and in the condition with High T30 (-0.12 dB(A), SE=0.04, z = -3.01, p=0.007), whereas the difference between the condition with Medium T30 and High T30 was 0.05 dB(A) (SE=0.04, z = 1.22, p=0.443).

To analyse the effects of T30 on vocal comfort, another LME model was run with the response variable self-reported vocal comfort (in %) and the fixed factors (1) T30, (2) noise,(3) gender, and (4) the interaction between T30 and noise. The random effects were (1) gain, (2) chronological task order, and (3) participant. The output of this model is reported in Table 6. The estimate of standard deviation for participant as a random effect was 15.08%, for order was 2.66%, and for gain was 0.85%, whereas the residual standard deviation was 14.05%. The mean decrease in self-reported vocal comfort, without noise added, from Low T30 to Medium T30 was 3.47%, while it was 1.15% from Low T30 to High T30. When the artificial speech-shaped noise was present, the vocal comfort decreased by 16.01%. For noise added conditions, there was an increase of comfort when T30 factors were higher than Low T30, 7.67% for Medium T30, and 5.48% for High T30. Generically, males' comfort was lower than females by 2.78%, with no statistical significance. Post-hoc comparisons regarding interaction between T30 and noise confirmed that, overall, the vocal comfort measured in Low T30 condition was lower than that in Medium T30 condition (2.10%, SE=1.86, z = 1.13, p=0.498) and High T30 condition (2.16%, SE=1.88, z = 1.15, p=0.483), whereas the difference between Medium T30 and High T30 was 0.06 % (SE=1.86, z = 0.03, p=0.999). None of these comparisons were statically significant.

To analyse the effects of T30 on vocal control, a final LME model was run with the response variable self-reported vocal control in % (0 = 'not at all controlled', 100 = 'extremely controlled') and the fixed factors (1) T30, (2) noise, (3) gender and (4) the interaction between T30 and noise. The random effects were (1) gain, (2) chronological task order, and (3) participant. The output of this model is reported in Table 7. The estimate of standard deviation for participant as a random effect was 13.83%, for order was 2.54%, and for gain was 0.00%, whereas the residual standard deviation was 12.79%. The mean decrease in self-reported vocal control, without noise added, was 0.16% from Low T30 to Medium T30, while it was 0.56% from Low T30 to High T30. When the artificial speech-shaped noise was present, the vocal comfort decreased by 9.17%. For noise added conditions, there was an increase of control when T30 factors were higher than Low T30, 4.37% for Medium T30, and 3.07% for High T30. Generically, males' control was lower than females by 2.06%, with no statistical significance. Post-hoc comparisons regarding interaction between T30 and noise confirmed that, overall, the vocal control measured in the condition Low T30 was lower than that in Medium T30 condition (2.10%, SE=1.70, z = 1.24, p=0.430) and High T30 condition (1.25%, SE=1.71, z = 0.73, p=0.745), whereas the control was 0.85% lower in High T30 than Medium T30 (SE=1.69, z = -0.50, p=0.871). None of these comparisons were statically significant.

Table 5 – LME models fit by REML for the response variable SPL and the fixed factors (1) T30, (2) noise, (3) gender, and (4) the interaction between T30 and noise.

	Estimate	Std. Error	df	t value	p-value	
(Intercept)	70.96	0.92	8.6	77.20	<0.001	***
T30 Medium	-0.08	0.06	147858.3	-1.40	0.162	
T30 High	-0.01	0.06	141383.0	-0.11	0.911	
Noise Speech-Shaped	2.80	0.06	165453.1	50.77	<0.001	***
Gender Male	2.93	1.04	29.6	2.82	0.008	**
T30 Medium: Noise Speech-Shaped	-0.17	0.08	148129.7	-2.21	0.026	*
T30 High: Noise Speech-Shaped	-0.22	0.08	161267.1	-2.89	0.004	**

Table 6 – LME models fit by REML for the response variable self-reported vocal comfort and the fixed factors (1) T30, (2) noise, (3) gender, and (4) the interaction between T30 and noise.

	Estimate	Std. Error	df	t value	p-value	
(Intercept)	84.24	4.36	37.7	19.31	<0.001	***
T30 Medium	-3.47	2.65	311.7	-1.31	0.191	
T30 High	-1.15	2.66	311.9	-0.43	0.666	
Noise Speech-Shaped	-16.01	2.64	308.5	-6.05	<0.001	***
Gender Male	-2.78	5.83	27.0	-0.48	0.637	
T30 Medium: Noise Speech-Shaped	11.13	3.75	311.7	2.97	0.003	**
T30 High: Noise Speech-Shaped	6.62	3.73	308.8	1.77	0.077	.

Table 7 – LME models fit by REML for the response variable self-reported vocal control and the fixed factors (1) T30, (2) noise, (3) gender, and (4) the interaction between T30 and noise.

	Estimate	Std. Error	df	t value	p-value	
(Intercept)	85.32	3.96	40.1	21.53	<0.001	***
T30 Medium	-0.16	2.41	312.6	-0.07	0.946	
T30 High	-0.56	2.43	312.9	-0.23	0.816	
Noise Speech-Shaped	-9.18	2.41	309.5	-3.81	<0.001	***
Gender Male	-2.07	5.35	27.0	-0.39	0.702	
T30 Medium: Noise Speech-Shaped	4.53	3.42	312.5	1.32	0.186	
T30 High: Noise Speech-Shaped	3.63	3.40	309.8	1.07	0.285	

4 Conclusions

The aim of this study was to evaluate the effect of external auditory feedback, such as reverberation time, altered sidetone (i.e. gain level), and background noise. The external auditory feedback was modified by changing the sidetone with three levels of gain (0 dB, 5 dB, and 10 dB); these changes showed that an increase in the sidetone led to a decrease of SPL and an increase in self-perception of voice comfort and control. This information is important because it can guide vocal health promotion actions helping to decrease the occurrence of voice disorders and improve speakers´ voice-related quality of life. For instance, among occupational voice users, such as teachers and call centre operators, considering their high risk of developing voice disorders associated with their working conditions [20][14][15] [9][21][6], it is determinant to identify specific elements that can help to improve "healthy" occupational voice use. Therefore, knowing that sidetone may help to decrease SPL and increase self-perceived voice comfort and control, speech and language pathologists at the workplaces may train occupational voice users using sidetone to strengthen voice comfort and control and reduce occupational voice misuse. In addition, results on Medium T30 being associated with the highest voice comfort and control (along with lowest SPL), when speech-shaped noise was added are also interesting. At the workplaces, professionals from Safe and Health at Work may consider these results for designing "safe" workplaces (classrooms, call centre rooms, schools) for "healthy" occupational voice use. In this way, the intervention actions would start in the environment and not in the workers, which is suggested in the hierarchy of controls. Finally, all the experiments conducted in this study were based on simulated acoustical environments, which represents a great step forward in the development of alternative techniques to performs research on voice production and sound propagation.

Acknowledgements

The authors would like to thank for their assistance, Ella Marzolf and Lydia Fleming, members of the Speech Accommodation to Acoustics Laboratory, University of Illinois at Urbana-Champaign. Additionally, they would like to express their gratitude to the subjects involved in the experiment.

References

[1] Riede, T.; Brown, C. Body size, vocal fold length, and fundamental frequency–implications for mammal vocal communication. *Nova Acta Leopoldina NF*, 111(380), 2013, pp 1-20.

[2] Rowe, M. L. Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, 35(1), 2008, pp 185–205.

[3] Krause, J. C.; Braida, L. D. Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1), 2004, pp 362–378.

[4] Lane, H.; Tranel, B. The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14(4), 1971, pp 677-709.

[5] Bottalico, P.; Graetzer, S.; Hunter, E. J. Effects of voice style, noise level, and acoustic feedback on objective and subjective voice evaluations. *The Journal of the Acoustical Society of America*, 138(6), 2015, pp EL498-EL503.

[6] Pelegrín-García, D.; Brunskog, J. Speakers' comfort and voice level variation in classrooms: Laboratory research. *The Journal of the Acoustical Society of America*, 132(1), 2012, pp 249-260.

[7] Hunter, E. J.; Švec, J. G.; Titze, I. R. Comparison of the produced and perceived voice range profiles in untrained and trained classical singers. *Journal of Voice*, 20(4), 2006, pp 513-526

[8] Lombard, E. "Le signe de l'elevation de la voix," [The sign of voice rise] in *Annales des Maladies de L'Oreille et du Larynx*, 37 (2), 1911, pp 101-119.

[9] Bottalico, P.; Graetzer, S.; Hunter, E. J. Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *The Journal of the Acoustical Society of America*, 139(5), 2016, pp 2870-2879.

[10] Junqua, J. C. The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, 93(1), 1993, pp 510–524.

[11] Yiu, E. M.; Yip, P. P. Effect of noise on vocal loudness and pitch in natural environments: an accelerometer (ambulatory phonation monitor) study. *Journal of Voice*, 30(4), 2016, pp 389-393.

[12] Black, J. W. The effect of room characteristics upon vocal intensity and rate. *The Journal of the Acoustical Society of America*, 22(2), 1950, pp 174-176.

[13] Siegel, G. M.; Pick Jr, H. L. Auditory feedback in the regulation of voice. *The Journal of the Acoustical Society of America*, 56(5), 1974, pp 1618-1624.

[14] Bottalico, P. Speech adjustments for room acoustics and their effects on vocal effort. *Journal of Voice*, 31(3), 2017, pp 392.e1-392.e12.

[15] Bottalico, P.; Astolfi, A.; Hunter, E. J. Teachers' voicing and silence periods during continuous speech in classrooms with different reverberation times. *The Journal of the Acoustical Society of America*, 141(1), 2017, pp EL26-EL31.

[16] Bottalico, P.; Passione, I. I.; Graetzer, S.; Hunter, E. J. Evaluation of the starting point of the Lombard effect. *Acta Acustica United with Acustica*, 103(1), 2017, pp 169-172.

[17] Pelegrín-García, D.; Smits, B.; Brunskog, J.; Jeong, C. H. Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America*, 129(4), 2011, pp 1981-1990.

[18] Rollins, M. K.; Leishman, T. W.; Whiting, J. K.; Hunter, E. J.; Eggett, D. L. Effects of added absorption on the vocal exertions of talkers in a reverberant room. *The Journal of the Acoustical Society of America*, 145(2), 2019, pp 775–783.

[19] Bottalico, P.; Graetzer, S.; Hunter, E. J. Effect of reverberation time on vocal fatigue. *International Conference on Speech Prosody 2016*, Boston, USA, May 31-June 3, 2016, pp 494-497.

[20] Banks, R. E.; Bottalico, P.; Hunter, E. J. The effect of classroom capacity on vocal fatigue as quantified by the vocal fatigue index. *Folia Phoniatrica et Logopaedica*, 69(3), 2017, pp 85-93.

[21] Cipriano, M.; Astolfi, A.; Pelegrín-García, D. Combined effect of noise and room acoustics on vocal effort in simulated classrooms. *The Journal of the Acoustical Society of America*, 141(1), 2017, pp EL51-EL56.