

CHANGES OF VOICE PRODUCTION IN ARTIFICIAL ACOUSTIC ENVIRONMENTS

BY

TOMÁS ENRIQUE SIERRA POLANCO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Speech and Hearing Science
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Master's Committee:

Assistant Professor Pasquale Bottalico
Professor Fatima Husain
Assistant Professor Keiko Ishikawa

ABSTRACT

The physical production of speech level dynamic range is directly affected by the physiological features of the speaker such as vocal tract size and lung capacity. However, the regulation of these production systems is affected by the perception of the communication environment and auditory feedback. The current study 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 voice comfort and voice 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 voice sound pressure levels, thus increasing voice comfort and voice control. This effect was consistent in different reverberation times considered. Mid-range reverberation times ($T_{30} \approx 1.3$ seconds) led to a decrease in voice sound pressure level along with an increase in voice comfort and voice control, however the effect of the reverberation time was smaller than the effect of the gain. The presence of noise amplified the aforementioned effects for the analyzed variables.

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Pasquale Bottalico, for being always there through this journey replying to my insistent emails and guiding me through the path of Speech and Hearing Science, while analyzing side-by-side data from several experiments. Additionally, I am deeply grateful with my graduate committee, Professors Fatima Husain and Keiko Ishikawa, for revising and commenting this document. Also, Professors Eric Hunter (Michigan State University) and Lady Cantor-Cutiva (Universidad Nacional de Colombia), for helping me out with the details of this study. Furthermore, I 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, who helped me out on assessing participants for data collection. And finally, express my gratitude to the participants involved in the experiment.

To my mom, for being the biggest thing in life. Which I rather write it in Spanish:

¡La cucha es lo más grande!

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: MATERIALS AND METHODS	7
CHAPTER 3: RESULTS.....	13
CHAPTER 4: DISCUSSION.....	28
CHAPTER 5: CONCLUSIONS	33
REFERENCES	35

CHAPTER 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 (Riede & Brown, 2013). This production regulation is affected by perceived communication demands, such as communication partners or communication environment, sense of voice comfort, and applied voice effort. For example, speech level and speech style can be partner specific such as speaking to a child (Rowe, 2008) or to someone with a perceived hearing loss (Krause & Braida, 2004). Another example was presented by Lane & Tranel (1971) 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 voice parameters, such as Sound Pressure Level (SPL), and can modify the talker's perception of voice comfort and voice control (Bottalico et al., 2015; Pelegrín-García & Burnskog, 2012). These parameters may be modified by the implementation of artificial settings delivered by headphones with the goal of increasing voice comfort and control, while decreasing voice effort in occupational voice users such as teachers and call center operators. All of these are affected by the relationship between voice production and hearing sensitivity (Hunter et al., 2006) and how the auditory system and auditory feedback plays a fundamental role in voice production including perception of effort and comfort.

Voice effort has been defined as the perceived exertion of a vocalist to a perceived communication scenario (Hunter et al., 2020). Changes in voice effort have been shown to be correlated with other voice adaptations such as vowel modifications, along with changes in voice

fundamental frequency, dB SPL, spectral tilt, and speech rate (Berardi, 2015; Berardi, 2020; McKenna & Stepp, 2018). Even though voice effort changes associate with a range of voice production parameters, radiated speech level seems to be the primary production parameter related to voice effort (ISO 9921, 2003) even when speech level is being controlled for (McKenna & Stepp, 2018).

Changes in voice effort (as measured using voice production metrics, such as SPL) or communication environment can affect voice comfort. Voice comfort can be defined as a subjective attribute that is directly correlated to the positive evaluation of the room for speech production and to the perceived support. Voice comfort has been shown to be negatively correlated to the feeling of having to raise the voice and to the tiredness after speaking for a long time period in the room (Cipriano et al., 2017; Pelegrín-García et al., 2014). An investigation based on self-reports showed that the voice comfort is more closely related to the perceived noise annoyance than to the perceived room reverberance. Voice comfort is related to all aspects that reduce voice effort (Titze, 1999; Titze, 2000). It appears to decrease with the speaker's perceived fatigue and the sensation of needing to increase the voice level (Pelegrín-García and Brunskog, 2012). Previous research in classroom settings showed that the voice comfort increases with the perception of the classroom as being good to speak in and with the perceived support and enhancement, while it decreases with the perceived exhaustiveness of speaking in a classroom during a lesson and with the sensation of having to increase the voice level (Pelegrín-García & Brunskog, 2012).

The alteration in auditory feedback can also modify the perception of a communication scenario, thus affecting voice production, voice comfort, and the perception of voice control. Voice control can be defined as the capacity to self-regulate voice production, e.g., SPL,

fundamental frequency, and resonance (Bottalico et al., 2016a). 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. For example, while considering a communication partner with hearing limitations, a talker (deliberately or inadvertently) uses “clear speech” (Ferguson & Kewley-Port, 2007; Krause & Braida, 2004). This type of speech has been characterized by a slower speech rate, a wider range of fundamental frequency, and a higher temporal modulation index than conversational speech (Bottalico et al., 2016a). Likewise, when talking in a noisy environment, people tend to raise the level of their voice in order to maintain understandable communication (Lombard, 1911). The maximization of intelligibility, clarity, voice comfort and control, and the minimization of voice effort and fatigue, should be the priority of any professional talker (Bottalico et al., 2016a).

Growing evidence suggests that there is an association between voice production level and external auditory feedback. External auditory feedback consists of the external path between mouth and ears and is strongly influenced by acoustics of the environment where the speaker is speaking. Such environmental effects are room noise, amplification of one’s own voice, and room reverberation.

A commonly experienced external auditory effect which directly impacts voice production level is that of elevated room noise, or the Lombard Reflex or Effect (Junqua, 1993; Lombard, 1911). For example, Yiu et al. (2016) 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 voice effort in the high-noise environment than in the other two conditions.

Voice level was shown to be affected by the reverberation time of the room (Black, 1950), and by the level at which a speaker perceived his/her own voice, as well as the level of the background noise (Siegel & Pick, 1974). 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 (Black, 1950; Bottalico et al., 2015; Bottalico et al., 2016a; Bottalico et al. 2016b; Bottalico, 2017; Bottalico et al. 2017a; Bottalico et al. 2017b; Pelegrín-García et al., 2011). Pelegrín-García et al. (2011) found that voice level decreased as reverberation time increased, while Black (1950) reported that greater voice intensity was found in less reverberant rooms than in more reverberant rooms. This is common even in extreme reverberation conditions (Rollins et al., 2019).

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 (1974) 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 (Bottalico et al., 2015) and to reverberation times (Bottalico et al., 2016b). 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, voice comfort and voice control vary (1) when 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 voice 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 voice SPL, and self-reported voice comfort and control. Three independent room acoustic parameters were considered: gain (alteration of the sidetone), reverberation time (T30), and background noise. This relationship was studied 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) Voice 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) Voice 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) Voice 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 voice comfort and voice control.

CHAPTER 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 each 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

This study was conducted with approval from and in accordance with policies of 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 American Native 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 to be eventual smokers. Five reported voice training in the past, such as singing lessons, and four reported history of speech or language therapy in their childhood.

2.2 INSTRUCTIONS AND CONDITIONS

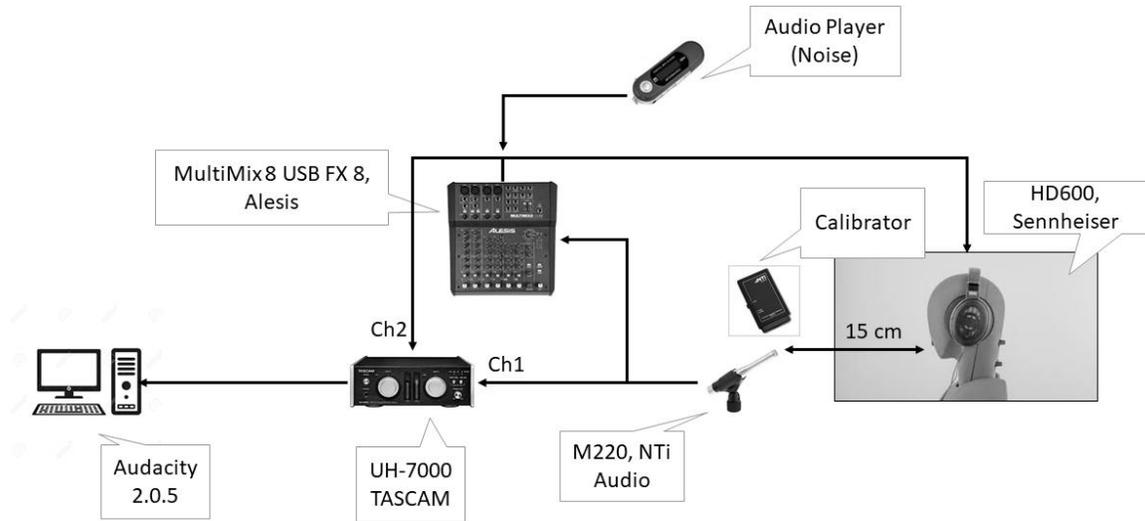
The participants were instructed to read aloud the first 6 sentences of “The Rainbow Passage”, a standardized text in English (Fairbanks, 1960), under fourteen different virtually simulated acoustic conditions. Each task had a duration of about 27 seconds of reading. Before the measurements, each participant 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) voice discomfort across all the tasks, as well as to control for any unknown confounding variables relating to task order.

Participants answered two questions after each task of the experiment: (1) *How comfortable was it to speak in this condition?* And (2) *How well were you able to control your voice in this condition?* These questions were worded in a manner consistent with the relevant ISO standard (ISO 28802, 2012) and administered immediately after the exposure to the noise conditions in each task. Participants responded by making a vertical tick on a continuous horizontal line of 100 mm in length on a visual analogue scale; this scale was provided on paper. The score was measured as the distance of the tick from the left end of the line. The extremes of the lines were ‘not at all’ (left) and ‘extremely’ (right).

2.3 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 each participant using a Class 1 Sound Calibrator NTi Audio (Tigard, OR, USA) with automatic atmospheric pressure compensation (ref: 94 dB \pm 0.2 dB at 1 kHz \pm 1%). The microphone output was split in 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 sound board (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 as noticeable echo (Lezzoum et al., 2016). The disposition of the equipment is depicted in Figure 1.



14

Figure 1: Equipment set-up for experiment measurements.

2.4 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 on the appendix by Pelegrín-García & Brunskog (2012).

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 $L_{Aeq} = 70$ 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
<i>T30 Whisper Room (s)</i>	0.164	0.129	0.079	0.061	0.064	0.054	0.048
<i>T30 Low (s)</i>	0.512	0.821	1.071	1.191	0.922	0.799	0.016
<i>T30 Medium (s)</i>	1.318	1.279	1.383	1.403	1.351	1.270	1.161
<i>T30 High (s)</i>	1.763	1.721	1.965	1.835	1.371	1.163	0.884
<i>Background noise (dB)</i>	34.0	26.7	16.0	14.8	13.4	15.5	16.4
<i>Speech-Shaped Noise (dB)</i>	59.8	62.0	66.6	60.3	64.2	59.7	55.4

2.5 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 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. A random effect is referred to as a factor that may affect the outcome but does not have main relevance. The selection of random effects is based on taking out the variance associated with a specific factor, due to low interest in its effect. Thus, it is used as a random factor to remove variance. Models were selected based on the Akaike information criterion (Akaike, 1998; the model with the lowest value being preferred) and the results of likelihood ratio tests (a significant result indicating that the more complex of the two nested models in the comparison is preferred) and were built using lme4, lmerTest and multcomp packages. 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 p-values for these tests were adjusted using the default single-step method (Hothorn et al., 2008). 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.

CHAPTER 3: RESULTS

Six Linear Mixed Effects (LME) models were run, two LME for each of the three different response variables: SPL, voice comfort, and voice 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, voice comfort, and voice control, and (2) effects of T30 and noise on SPL, voice comfort, and voice control. Table 2 summarizes the outcomes for the 14 conditions.

Table 2: Mean values and standard error (se) for the variable SPL in dB(A), perceived voice comfort and control in %, for the 14 conditions.

<i>T30</i>	<i>Gain</i>	<i>Noise</i>	<i>SPL (dB)</i>	<i>se</i>	<i>Comfort (%)</i>	<i>se</i>	<i>Control (%)</i>	<i>se</i>
<i>Whisper Room</i>	<i>0</i>	<i>No Noise</i>	<i>73.3</i>	<i>0.06</i>	<i>79.1</i>	<i>3.90</i>	<i>85.4</i>	<i>2.70</i>
<i>Whisper Room</i>	<i>0</i>	<i>Speech-Shaped</i>	<i>76.7</i>	<i>0.06</i>	<i>54.7</i>	<i>4.22</i>	<i>61.7</i>	<i>4.35</i>
<i>Low</i>	<i>5</i>	<i>No Noise</i>	<i>72.8</i>	<i>0.06</i>	<i>85.0</i>	<i>2.86</i>	<i>86.4</i>	<i>2.67</i>
<i>Low</i>	<i>5</i>	<i>Speech-Shaped</i>	<i>76.0</i>	<i>0.06</i>	<i>65.5</i>	<i>4.56</i>	<i>73.7</i>	<i>4.21</i>
<i>Low</i>	<i>10</i>	<i>No Noise</i>	<i>72.0</i>	<i>0.06</i>	<i>81.2</i>	<i>3.24</i>	<i>82.1</i>	<i>3.17</i>
<i>Low</i>	<i>10</i>	<i>Speech-Shaped</i>	<i>74.4</i>	<i>0.06</i>	<i>69.5</i>	<i>4.37</i>	<i>76.5</i>	<i>3.87</i>
<i>Medium</i>	<i>5</i>	<i>No Noise</i>	<i>72.9</i>	<i>0.06</i>	<i>82.8</i>	<i>3.77</i>	<i>84.6</i>	<i>3.88</i>
<i>Medium</i>	<i>5</i>	<i>Speech-Shaped</i>	<i>75.8</i>	<i>0.06</i>	<i>70.0</i>	<i>3.89</i>	<i>75.8</i>	<i>3.87</i>
<i>Medium</i>	<i>10</i>	<i>No Noise</i>	<i>71.8</i>	<i>0.06</i>	<i>76.7</i>	<i>4.43</i>	<i>84.8</i>	<i>3.12</i>
<i>Medium</i>	<i>10</i>	<i>Speech-Shaped</i>	<i>74.0</i>	<i>0.06</i>	<i>79.0</i>	<i>3.33</i>	<i>83.1</i>	<i>3.20</i>
<i>High</i>	<i>5</i>	<i>No Noise</i>	<i>73.1</i>	<i>0.06</i>	<i>82.3</i>	<i>3.44</i>	<i>85.7</i>	<i>2.84</i>
<i>High</i>	<i>5</i>	<i>Speech-Shaped</i>	<i>76.0</i>	<i>0.06</i>	<i>67.8</i>	<i>3.88</i>	<i>75.4</i>	<i>3.75</i>
<i>High</i>	<i>10</i>	<i>No Noise</i>	<i>71.6</i>	<i>0.06</i>	<i>81.0</i>	<i>3.68</i>	<i>81.9</i>	<i>3.41</i>
<i>High</i>	<i>10</i>	<i>Speech-Shaped</i>	<i>73.9</i>	<i>0.06</i>	<i>77.2</i>	<i>3.83</i>	<i>81.4</i>	<i>3.65</i>

3.1 EFFECTS OF GAIN

3.1.1 Effects of gain and noise on SPL

A Linear Mixed Effects (LME) model was run with the response variable SPL (in dB(A)). This model, reported in Table 3, 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. As shown in Figure 2, 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. Since gender effect was statistically significant, Figure 2 differentiate among females and males where, generically, males were louder than females by 2.93 dB(A).

Table 3: 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. The reference levels were: 0 dB for gain, without speech-shaped noise (No Noise) for noise condition, and female for gender.

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

Signif. codes: '****' < 0.001, '***' < 0.01, '**' < 0.05, '.' < 0.1, ' ' < 1

Post-hoc comparisons are shown on Table 4. Considering the effect of gain and its interaction with noise, 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), $z = -5.87$, $p < 0.001$) and in 10 dB of gain condition (-2.03 dB(A), $z = -21.95$, $p < 0.001$), whereas the difference between 5 dB and 10 dB gain conditions was -1.49 dB(A) ($z = -46.98$, $p < 0.001$).

Table 4: Multiple Comparisons of Means: Tukey Contrasts for the response variable SPL and the interaction effects between (1) gain and (2) noise.

<i>Gain Difference</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>p-value</i>	
<i>Gain 5 – Gain 0</i>	-0.54	0.09	-5.87	<0.001	***
<i>Gain 10 – Gain 0</i>	-2.03	0.09	-21.95	<0.001	***
<i>Gain 10 – Gain 5</i>	-1.49	0.03	-46.98	<0.001	***

Signif. codes: '****' < 0.001, '***' < 0.01, '**' < 0.05, '.' < 0.1, ' ' < 1
(Adjusted p values reported -- single-step method)

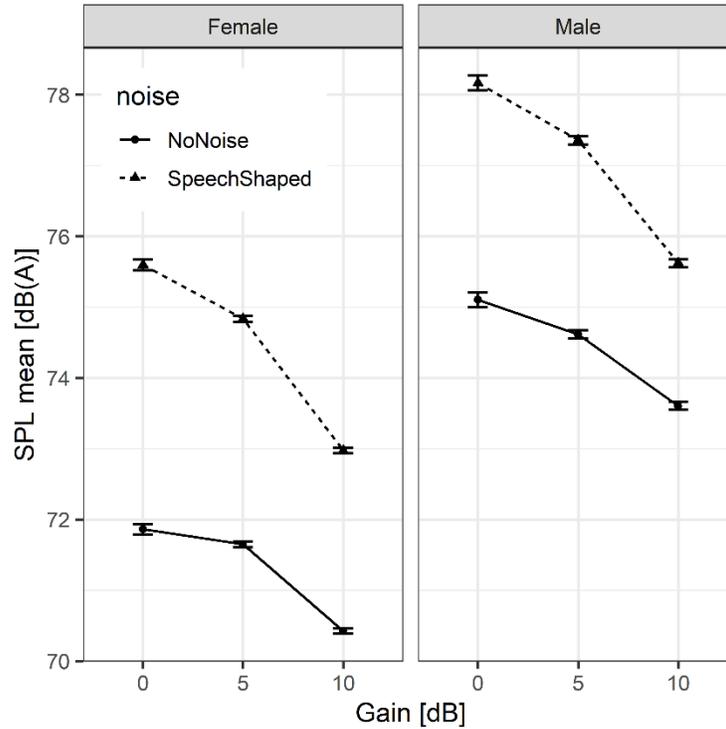


Figure 2: Mean SPL in dB(A) vs room gain regarding noise conditions and gender. The error bars indicate standard error.

3.1.2 Effects of gain and noise on voice comfort

One LME model was run with the response variable self-reported voice 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 5. 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 voice comfort from 0 dB to 5 dB of gain was 3.55%, while it was of 0.27% from 0 dB to 10 dB; in the conditions without noise added. As shown in Figure 3, the voice comfort decreased by -25.31% when the speech-shaped noise was added. For these

conditions, where noise was added, the mean increase in self-reported voice comfort from 0 dB to 5 dB of gain was 13.43%, while it was of 20.77% from 0 dB to 10 dB.

Table 5: 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. The reference levels were: 0 dB for gain, without speech-shaped noise (No Noise) for noise condition, and female for gender.

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

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1

Post-hoc comparisons (in Table 6) confirmed that, overall, the voice 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%, $z = 3.76$, $p < 0.001$), and the condition with 10 dB of gain (10.52%, $z = 4.66$, $p < 0.001$). Furthermore, the voice comfort reported in the condition with 10 dB of gain was 2.02% higher than that in the condition with 5 dB of gain ($z = 1.27$, $p = 0.406$).

Table 6: Multiple Comparisons of Means: Tukey Contrasts for the response variable self-reported voice comfort and the interaction effects between (1) gain and (2) noise.

<i>Gain Difference</i>	<i>Estimate (%)</i>	<i>Std. Error (%)</i>	<i>z value</i>	<i>p-value</i>	
<i>Gain 5 – Gain 0</i>	8.50	2.26	3.76	<0.001	***
<i>Gain 10 – Gain 0</i>	10.52	2.26	4.66	<0.001	***
<i>Gain 10 – Gain 5</i>	2.02	1.59	1.27	0.406	

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1
(Adjusted p values reported -- single-step method)

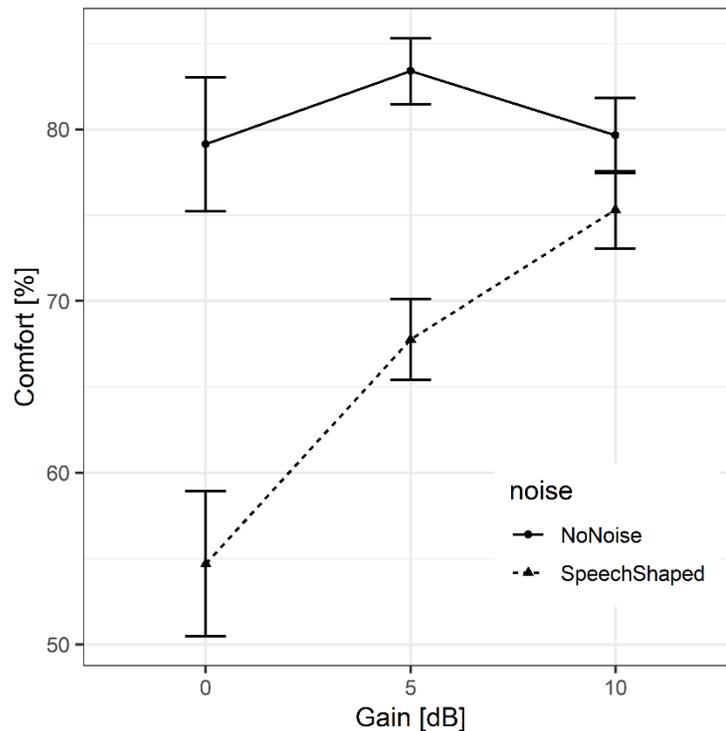


Figure 3: Mean self-reported voice comfort in % (0 = ‘not at all comfortable’, 100 = ‘extremely comfortable’) across participants per gain level. The error bars indicate \pm standard error.

3.1.3 Effects of gain and noise on voice control

The analysis of voice control was similar to voice comfort. One LME model was run with the response variable self-reported voice 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 7. 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 voice control from 0 dB to 5 dB of gain was 0.45%, while it was of 3.01%, from 0 dB to

10 dB; in the conditions without noise added. As shown in Figure 4, the voice control decreased by 24.19% when the speech-shaped noise was added. For these conditions, where noise was added, the mean increase in self-reported voice control from 0 dB to 5 dB of gain was 13.28%, while it was of 18.87% from 0 dB to 10 dB.

Table 7: 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. The reference levels were: 0 dB for gain, without speech-shaped noise (No Noise) for noise condition, and female for gender.

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

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1

Post-hoc comparisons regarding the interactions between gain and noise confirmed that, overall, the voice 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%, $z = 3.17$, $p=0.004$) and the condition with 10 dB of gain (7.93%, $z = 3.92$, $p<0.001$). Furthermore, the voice control reported in the condition with 10 dB of gain was 1.51% higher than that in the condition with 5 dB of gain ($z = 1.06$, $p = 0.533$).

Table 8: Multiple Comparisons of Means: Tukey Contrasts for the response variable self-reported voice control and the interaction between (1) gain and (2) noise.

<i>Gain Difference</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>p-value</i>	
<i>Gain 5 – Gain 0</i>	6.42	2.03	3.17	0.004	**
<i>Gain 10 – Gain 0</i>	7.93	2.02	3.92	<0.001	***
<i>Gain 10 – Gain 5</i>	1.51	1.43	1.06	0.533	

Signif. codes: ‘****’ < 0.001, ‘***’ < 0.01, ‘**’ < 0.05, ‘.’ < 0.1, ‘ ’ < 1
(Adjusted p values reported -- single-step method)

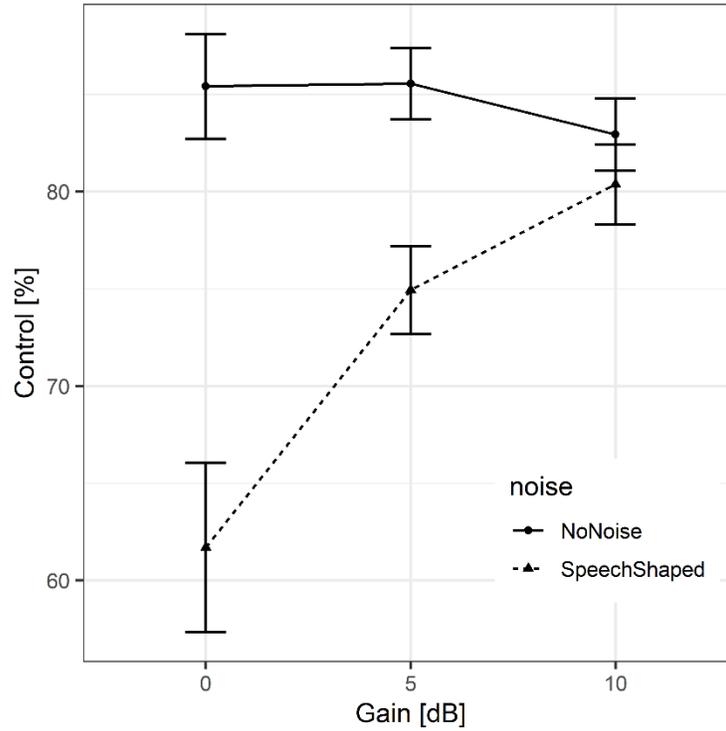


Figure 4: Mean self-reported voice control in % across participants per gain level. The error bars indicate \pm standard error.

3.2 EFFECTS OF REVERBERATION TIME (T30)

3.2.1 Effects of reverberation time and noise on SPL

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 9. 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). As shown in Figure 5, 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).

Table 9: 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. The reference levels were: Low T30, without speech-shaped noise (No Noise) for noise condition, and female for gender.

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

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1

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), $z = -4.28$, $p < 0.001$) and in the condition with High T30 ($z = -3.01$, $p = 0.007$), whereas the difference between the condition with Medium T30 and High T30 was 0.05 dB(A) ($z = 1.22$, $p = 0.443$), as shown in Table 10.

Table 10: Multiple Comparisons of Means: Tukey Contrasts for the response variable SPL and the interaction between (1) T30 and (2) noise.

<i>T30 Differences</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>p-value</i>	
<i>Medium - Low</i>	-0.17	0.04	-4.28	<0.001	***
<i>High - Low</i>	-0.12	0.04	-3.01	0.007	**
<i>High - Medium</i>	0.05	0.04	1.22	0.443	

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1
(Adjusted p values reported -- single-step method)

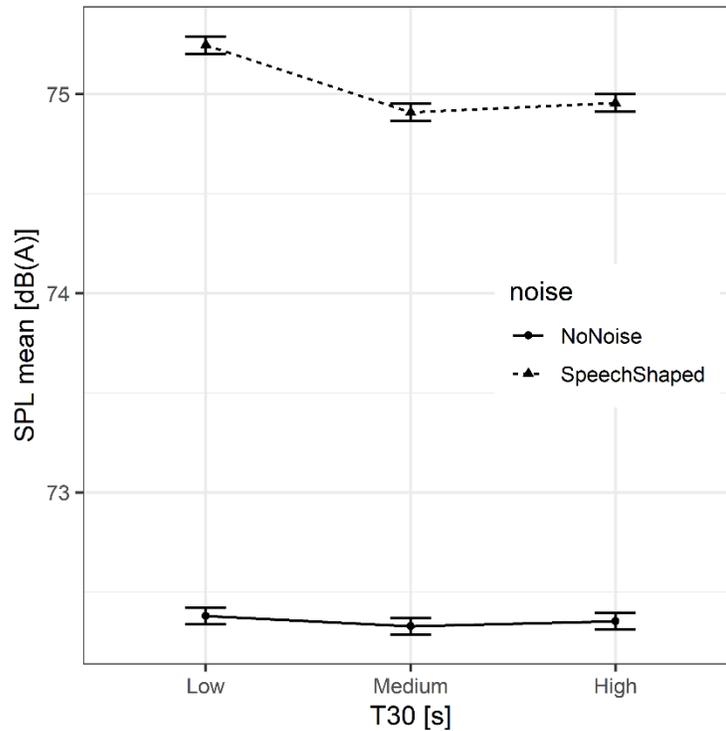


Figure 5: Mean SPL in dB(A) vs T30 regarding noise conditions. The error bars indicate \pm standard error.

3.2.2 Effect of reverberation time and noise on voice comfort

To analyze the effects of T30 on voice comfort, another LME model was run with the response variable self-reported voice 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 11. 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 voice comfort, without noise added, from Low T30 to Medium T30 was 3.47%, while it was of 1.15% from Low T30 to High T30. As shown in Figure 6, when the artificial speech-shaped noise was present, the voice 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.

Table 11: LME models fit by REML for the response variable self-reported voice comfort and the fixed factors (1) T30, (2) noise, (3) gender, and (4) the interaction between T30 and noise.

The reference levels were: Low T30, without speech-shaped noise (No Noise) for noise condition, and female for gender.

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

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1

Post-hoc comparisons regarding interaction between T30 and noise confirmed that, overall, the voice comfort measured in Low T30 condition was lower than that in Medium T30 condition (2.10%, $z = 1.13$, $p=0.498$) and High T30 condition (2.16%, $z = 1.15$, $p=0.483$), whereas the difference between Medium T30 and High T30 was 0.06 % ($z = 0.03$, $p=0.999$), as shown in Table 12. None of these comparisons were statically significant.

Table 12: Multiple Comparisons of Means: Tukey Contrasts for the response variable self-reported voice comfort and the interaction between (1) T30 and (2) noise.

<i>T30 Differences</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>p-value</i>
<i>Medium - Low</i>	2.10	1.86	1.13	0.498
<i>High - Low</i>	2.16	1.88	1.15	0.483
<i>High - Medium</i>	0.06	1.86	0.03	0.999

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1
(Adjusted p values reported -- single-step method)

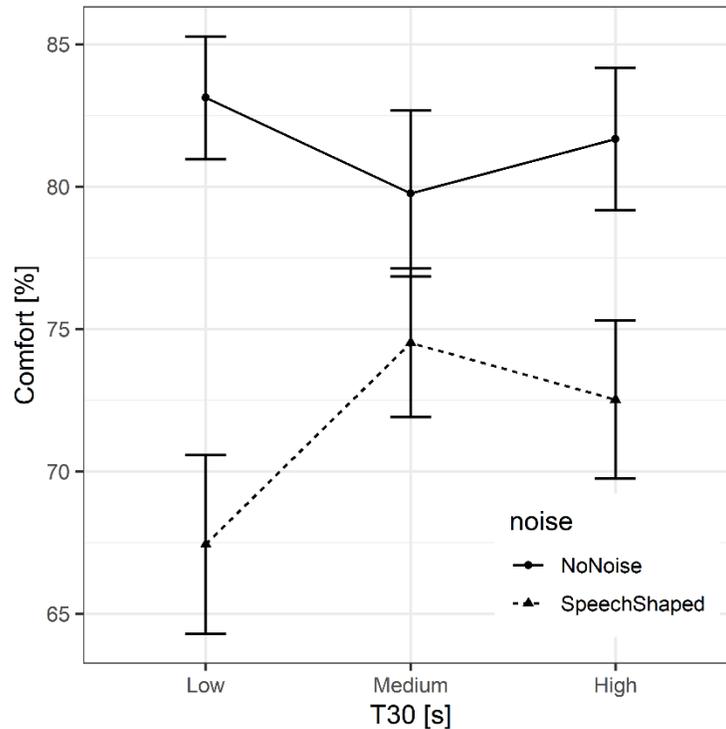


Figure 6: Mean self-reported voice comfort in % (0 = ‘not at all comfortable’, 100 = ‘extremely comfortable’) across participants per T30. The error bars indicate \pm standard error.

3.2.3 Effect of reverberation time and noise on voice control

To analyze the effects of T30 on voice control, a final LME model was run with the response variable self-reported voice 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 13. 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 voice control, without noise added, was 0.16% from Low T30 to Medium T30, while it was of 0.56% from Low T30 to High T30. As shown in Figure 7, when the artificial speech-shaped noise was

present, the voice 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.

Table 13: LME models fit by REML for the response variable self-reported voice control and the fixed factors (1) T30, (2) noise, (3) gender, and (4) the interaction between T30 and noise. The reference levels were: Low T30, without speech-shaped noise (No Noise) for noise condition, and female for gender.

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

Signif. codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1, ' ' < 1

Post-hoc comparisons regarding interaction between T30 and noise confirmed that, overall, the voice control measured in the condition Low T30 was lower than that in Medium T30 condition (2.10%, $z = 1.24$, $p=0.430$) and High T30 condition (1.25%, $z = 0.73$, $p=0.745$), whereas the control was 0.85% lower in High T30 than Medium T30 ($z = -0.50$, $p=0.871$), as shown in Table 14. None of these comparisons were statically significant.

Table 14: Multiple Comparisons of Means: Tukey Contrasts for the response variable self-reported voice control and the interaction between (1) T30 and (2) noise.

<i>T30 Differences</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z value</i>	<i>p-value</i>
<i>Medium - Low</i>	2.10	1.70	1.24	0.430
<i>High - Low</i>	1.25	1.71	0.73	0.745
<i>High - Medium</i>	-0.85	1.69	-0.50	0.871

Signif. codes: '****' < 0.001, '***' < 0.01, '**' < 0.05, '.' < 0.1, ' ' < 1
(Adjusted p values reported -- single-step method)

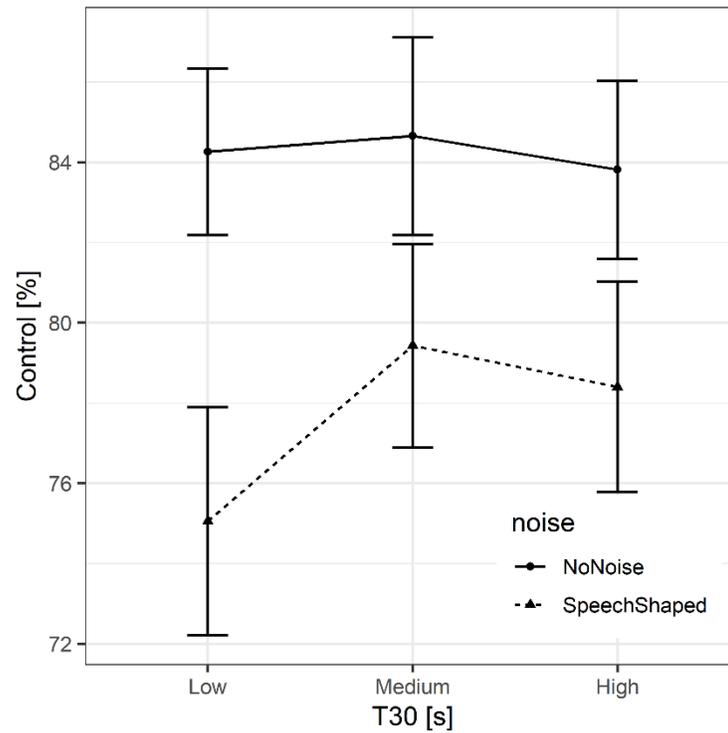


Figure 7: Mean self-reported voice control in % across participants per T30. The error bars indicate \pm standard error.

CHAPTER 4: DISCUSSION

In this study, several acoustics scenarios have been virtually created by modifying the external gain, as well as reverberation time and by adding speech-shaped noise on the overall external auditory feedback. The speech adjustments in terms of SPL, and self-reported voice comfort and control were measured in the aforementioned virtual scenarios.

4.1 EFFECT OF NOISE AND GENDER

Overall, the mean SPL at 15 cm from the mouth was measured as 75.26 dB(A) and 72.48 dB(A) for conditions with and without speech-shaped noise, respectively. The equivalent level of the speech-shaped noise was 70 dB(A), while the background noise in the whisper room was 25 dB(A) when no noise was added. The increase in SPL when artificial noise was added is consistent with the Lombard effect (Lombard, 1911), which refers to the tendency of speakers to raise their voice in order to be understood in noisy environments. As a result of adding background noise, the perceived voice comfort and control decreased by 12.06% and 9.05%, respectively. This decrease in self-reported voice comfort and control, when noise was added, confirmed the tendency showed by Bottalico et al. (2015) in real rooms. Bottalico et al. (2015) showed that the differences for comfort and control on normal voice production were estimated to be 11.1% and 9.4% lower when noise was added. Even if gender was a statistically significant factor in the regulation of voice SPL (i.e., males were louder than females by 2.93 dB(A)), voice comfort and control were not statistically different between gender in the two sets of conditions with and without noise.

4.2 EFFECT OF GAIN

Regarding the effect of sidetone alteration (an alteration of the level at which a person is perceiving his/her own voice), when gain factor was increased, participants decreased their voice SPL while reporting a higher level of voice comfort and control. When the sidetone was increased, by 5 dB and 10 dB, participants showed a statistically significant decrease in their voice SPL of 0.54 dB(A) and 2.03 dB(A), respectively, confirming the results of Siegel and Pick (1974). Siegel and Pick (1974) conducted four different experiments which concluded that when the sidetone is increased by 10 dB there was a decrease of voice SPL within the range of 1.0 to 5.8 dB, with a mean estimate of 3.5 dB. Regarding voice comfort and control, differences were statistically significant only in the conditions in which speech-shaped noise was added. In the different sidetone conditions related to this study, the results showed that the lowest levels of voice comfort and control were reported in the condition without alteration of the sidetone (i.e., 0 dB of gain). Voice comfort increased by 8.50% and 10.52%, for gain 5 dB and 10 dB, while control increased by 6.42% and 7.93%, for gain 5 dB and 10 dB, respectively, setting 0 dB as gain reference. This could have important implications for professional voice users. Specifically, many of these professionals who use electroacoustic systems for the playback of their own voice (like singers, broadcasters, or call center operators) may benefit from increasing the level of their monitors/headphones for an increase in the perceived voice comfort and control and a decrease in voice effort. However, it is necessary to be careful not to increase the feedback level over the limit that may induce hearing loss.

4.3 EFFECT OF REVERBERATION TIME

According to the results presented in this study and comparing them with other studies cited in this discussion, we hypothesized that there is a trend for individuals to react differently while speaking in “middle range” reverberation times, considering “middle range” within the values that are explicitly cited on each study conditions. This does not pretend to concretely give an assessment on specific quantitative values, or ranges of values (high or low), but a relationship within three or more different reverberation times (ordered by levels) when compared in the same experiment. Our opinion pertains on how “middle range” reverberation times guide participants to improve their own comfort and control (along with the decrease of SPL). Nonetheless, it is important to remind that the reverberation times for this study were measured from the oral-binaural impulse response recorded by the HATS, rather than using the standardized method following the ISO 3381.

Following the former, these differences were 0.17 dB(A) and 0.12 dB(A) lower for Medium T30 and High T30, as compared to Low T30, respectively. These results are similar from trends shown in previous studies (Bottalico & Astolfi, 2012; Durup et al., 2015; Puglisi et al., 2017; Puglisi et al., 2014), where voice SPL was presented to lower down in medium ranges of reverberation times. The medium range conditions for reverberation times on those experiments were within 0.7 and 0.9 s, which indicate that there is a trend for higher comfort and control when reverberation conditions are in-between a range of values, i.e., by comparing higher and/or lower values of reverberation times with medium range conditions. Whether for gain conditions the variations in SPL were substantial, the effect of variations of T30 were rather small.

The effect of reverberation time in self-reported voice comfort had a similar trend, regarding Medium T30. Comfort was 2.10% and 2.16% higher for Medium T30 and High T30 than that from Low T30. Also, there was an increase on control for Medium T30 and High T30 of 2.10% and 1.25% higher with respect to Low T30. This might indicate a greater comfort and control for middle range reverberation times, as opposed to Bottalico et al (2016a) study, where an increase of comfort is shown in an anechoic and a reverberant room, as opposed to a semi-reverberant room. Bottalico et al reported 3.4% higher comfort for anechoic and 0.8% higher comfort for a reverberant room, both compared with the semi-reverberant room. A similar behavior was shown in that same study for control, where it was 4.5% higher for anechoic and 3.9% for reverberant than that from the semi-reverberant. It is important to point out that in Bottalico et al (2016a), the authors presented two voice styles (normal and loud) and calculate averages among those two voice styles to give estimates on Δ SPL magnitudes. These findings could lead to a misunderstanding on self-reported voice parameters because in the loud style, the voice intensity was higher and, consequently, the reflected sound was more intense.

More investigation is expected on this topic to build up a better understanding on how T30 is affecting (or even if it is affecting at all) voice production in a meaningful way. However, the variations in SPL suggest that lower voice demands were experienced by talkers in Medium T30 conditions ($T30 = 1.39$ s).

4.4 LIMITATIONS AND FUTURE DIRECTIONS

Some limitations of this study were the use of standardized reading material instead of spontaneous speech, avoids variations in phonation time, which may represent a limitation on evaluating self-reported voice comfort and control. Furthermore, due to participants being

American native English speakers, hinders the generalization of these results to speakers in other languages and/or in other forms of spoken English.

In the future, studies on simulated environments could have an increase in the levels of reverberation time, gain, and background noise. By broadening the range of reverberation times, gain levels, and noise conditions might show up further recommendations about acoustical conditions that would maximize voice comfort and control, while minimizing SPL and voice effort. Finally, it is important to point out that adding other acoustical objective measurements would be useful for better understanding the variations on voice comfort and control, such as speed rate of speech and frequency of utterances, which are directly related with the movement of the vocal folds, thus with voice effort and fatigue.

CHAPTER 5: 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 center operators, considering their high risk of developing voice disorders associated with their working conditions (Banks et al., 2017; Bottalico et al., 2015; Bottalico et al., 2016a; Cipriano et al., 2017; Pelegrín-García & Brunskog, 2012; Cantor-Cutiva et al., 2013; Cantor-Cutiva & Burdorf, 2016; Carrillo-Gonzalez et al., 2019; Cantor-Cutiva et al., 2019), 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 strength 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 center 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 (Castro, 2003).

Finally, all the experiments conducted in this study were based on simulated acoustical environments, which represents a great step forward the development of alternative techniques to performs research on voice production and sound propagation.

REFERENCES

- Akaike, H. (1998). "Information theory and an extension of the maximum likelihood principle," in Selected papers of Hirotugu Akaike (New York, NY: Springer), 199-213.
- Banks, R. E., Bottalico, P., Hunter, E. J. (2017). The effect of classroom capacity on vocal fatigue as quantified by the vocal fatigue index. *Folia Phoniatrica et Logopaedica*, 69(3), 85-93. doi: [10.1159/000484558](https://doi.org/10.1159/000484558)
- Berardi, M. L. (2015). Quantification of Vocal Effort in Varying Acoustical Environments [Honors]. [Provo, UT]: Brigham Young University.
- Berardi, M. L. (2020). Validation and Application of Experimental Framework for the Study of Vocal Fatigue. [Doctoral dissertation]. [East Lansing, MI]: Michigan State University. doi: [10.25335/epf5-c186](https://doi.org/10.25335/epf5-c186)
- Black, J. W. (1950). The effect of room characteristics upon vocal intensity and rate. *The Journal of the Acoustical Society of America*, 22(2), 174-176. doi: [10.1121/1.1906585](https://doi.org/10.1121/1.1906585)
- Bottalico, P., and Astolfi, A. (2012). Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *The Journal of the Acoustical Society of America*, 131(4), 2817-2827. doi: [10.1121/1.3689549](https://doi.org/10.1121/1.3689549)
- Bottalico, P., Graetzer, S., Hunter, E. J. (2015). 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), EL498-EL503. doi: [10.1121/1.4936643](https://doi.org/10.1121/1.4936643)
- Bottalico, P., Graetzer, S., Hunter, E. J. (2016a). Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *The Journal of the Acoustical Society of America*, 139(5), 2870-2879. doi: [10.1121/1.4950812](https://doi.org/10.1121/1.4950812)

- Bottalico, P., Graetzer, S., Hunter, E. J. (2016b). Effect of reverberation time on vocal fatigue. In Proceedings of the International Conference on Speech Prosody (Vol. 2016, pp. 494-497). doi: [10.21437/SpeechProsody.2016-101](https://doi.org/10.21437/SpeechProsody.2016-101)
- Bottalico, P. (2017). Speech adjustments for room acoustics and their effects on vocal effort. *Journal of Voice*, 31(3), 392-e1. doi: [10.1016/j.jvoice.2016.10.001](https://doi.org/10.1016/j.jvoice.2016.10.001)
- Bottalico, P., Astolfi, A., Hunter, E. J. (2017a). Teachers' voicing and silence periods during continuous speech in classrooms with different reverberation times. *The Journal of the Acoustical Society of America*, 141(1), EL26-EL31. doi: [10.1121/1.4973312](https://doi.org/10.1121/1.4973312)
- Bottalico, P., Passione, I. I., Graetzer, S., Hunter, E. J. (2017b). Evaluation of the starting point of the Lombard effect. *Acta Acustica United with Acustica*, 103(1), 169-172. doi: [10.3813/AAA.919043](https://doi.org/10.3813/AAA.919043)
- Cipriano, M., Astolfi, A., Pelegrín-García, D. (2017). Combined effect of noise and room acoustics on vocal effort in simulated classrooms. *The Journal of the Acoustical Society of America*, 141(1), EL51-EL56. doi: [10.1121/1.4973849](https://doi.org/10.1121/1.4973849)
- Cantor-Cutiva, L. C., Vogel, I., Burdorf, A. (2013). Voice disorders in teachers and their associations with work-related factors: a systematic review. *Journal of Communication Disorders*, 46(2), 143-155. doi: [10.1016/j.jcomdis.2013.01.001](https://doi.org/10.1016/j.jcomdis.2013.01.001)
- Cantor-Cutiva, L. C., & Burdorf, A. (2016). Work-related determinants of voice complaints among school workers: An eleven-month follow-up study. *American journal of speech-language pathology*, 25(4), 590-597. doi: [10.1044/2016_AJSLP-14-0191](https://doi.org/10.1044/2016_AJSLP-14-0191)
- Cantor-Cutiva, L. C., Bottalico, P., & Hunter, E. (2019). Work-related communicative profile of radio broadcasters: a case study. *Logopedics Phoniatrics Vocology*, 44(4), 178-191. doi: [10.1080/14015439.2018.1504983](https://doi.org/10.1080/14015439.2018.1504983)

- Carrillo-Gonzalez, A., Camargo-Mendoza, M., & Cantor-Cutiva, L. C. (2019). Relationship between sleep quality and stress with voice functioning among college professors: a systematic review and meta-analysis. *Journal of Voice*. doi: [10.1016/j.jvoice.2019.11.001](https://doi.org/10.1016/j.jvoice.2019.11.001)
- Castro, A. B. (2003). 'Hierarchy of Controls': Providing a framework for addressing workplace hazards. *AJN The American Journal of Nursing*, 103(12), 104.
- Durup, N., Shield, B., Dance, S., Sullivan, R. (2015). An investigation into relationships between classroom acoustic measurements and voice parameters of teachers. *Building Acoustics*, 22(3-4), 225-241. doi: [10.1260/1351-010X.22.3-4.225](https://doi.org/10.1260/1351-010X.22.3-4.225)
- Fairbanks, G. (1960). "The rainbow passage," in *Voice and Articulation Drillbook*, 2nd Edition (New York: Harper & Row), 124-139.
- Ferguson, S. H., and Kewley-Port, D. (2007). Talker Differences in Clear and Conversational Speech: Acoustic Characteristics of Vowels. *Journal of Speech, Language, and Hearing Research*, 50(5), 1241–1255. doi: [10.1044/1092-4388\(2007/087\)](https://doi.org/10.1044/1092-4388(2007/087))
- Hothorn, T., Bretz, F., Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(3), 346-363. doi: [10.1002/bimj.200810425](https://doi.org/10.1002/bimj.200810425)
- Hunter, E. J., Švec, J. G., Titze, I. R. (2006). Comparison of the produced and perceived voice range profiles in untrained and trained classical singers. *Journal of Voice*, 20(4), 513-526. doi: [10.1016/j.jvoice.2005.08.009](https://doi.org/10.1016/j.jvoice.2005.08.009)
- Hunter, E. J., Cantor-Cutiva, L. C., Van Leer, E., Van Mersbergen, M., Nanjundeswaran, C. D., Bottalico, P., Sandage, M., Whitling, S. (2020). Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *Journal of Speech, Language, and Hearing Research*, 63(2), 509-532. doi: [10.1044/2019_JSLHR-19-00057](https://doi.org/10.1044/2019_JSLHR-19-00057)

- International Organization for Standardization. (2003). Ergonomics: Assessment of Speech Communication. ISO 9921:2002(E).
- ISO, U. (2008). 3382-2: Acústica, Medición de parámetros acústicos en recintos, Parte 2: Tiempo de reverberación en recintos ordinarios. *Asociación Española de Normalización y Certificación AENOR*.
- International Organization for Standardization, 28802-2012(E), Ergonomics of the physical environment – Assessment of environments by means of an environmental survey involving physical measurements of the environment and subjective responses of people. International Organization for Standardization, Genève, 2012.
- Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, 93(1), 510–524. doi: [10.1121/1.405631](https://doi.org/10.1121/1.405631)
- Krause, J. C., and Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1), 362–378. doi: [10.1121/1.1635842](https://doi.org/10.1121/1.1635842)
- Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14(4), 677-709. doi: [10.1044/jshr.1404.677](https://doi.org/10.1044/jshr.1404.677)
- Lezzoum, N., Gagnon, G., & Voix, J. (2016). Echo threshold between passive and electro-acoustic transmission paths in digital hearing protection devices. *International Journal of Industrial Ergonomics*, 53, 372-379. doi: [10.1016/j.ergon.2016.04.004](https://doi.org/10.1016/j.ergon.2016.04.004)
- Lombard, E. (1911). “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): 101-119.

- McKenna, V. S., and Stepp, C. E. (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America*, 144(3), 1643–1658. doi: [10.1121/1.5055234](https://doi.org/10.1121/1.5055234)
- Pelegrín-García, D., Smits, B., Brunskog, J., Jeong, C. H. (2011). Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America*, 129(4), 1981-1990. doi: [10.1121/1.3552881](https://doi.org/10.1121/1.3552881)
- Pelegrín-García, D., and Brunskog, J. (2012). Speakers' comfort and voice level variation in classrooms: Laboratory research. *The Journal of the Acoustical Society of America*, 132(1), 249-260. doi: [10.1121/1.4728212](https://doi.org/10.1121/1.4728212)
- Pelegrín-García, D., Brunskog, J., Rasmussen, B. (2014). Speaker-oriented classroom acoustics design guidelines in the context of current regulations in European countries. *Acta Acustica united with Acustica*, 100(6), 1073-1089. doi: [10.3813/AAA.918787](https://doi.org/10.3813/AAA.918787)
- Puglisi, G. E., Astolfi, A., Pavese, L., Carullo, A. (2014). Long-term vocal parameters in semi-anechoic and reverberant rooms. *Forum Acusticum 2014 Krakow 7-12.09*.
- Puglisi, G. E., Astolfi, A., Cantor-Cutiva, L. C., Carullo, A. (2017). Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. *The Journal of the Acoustical Society of America*, 141(1), 441-452. doi: [10.1121/1.4973805](https://doi.org/10.1121/1.4973805)
- Riede, T., & Brown, C. (2013). Body size, vocal fold length, and fundamental frequency—implications for mammal vocal communication. *Nova Acta Leopoldina NF*, 111(380), 1-20.

- Rollins, M. K., Leishman, T. W., Whiting, J. K., Hunter, E. J., Eggett, D. L. (2019). Effects of added absorption on the vocal exertions of talkers in a reverberant room. *The Journal of the Acoustical Society of America*, 145(2), 775–783. doi: [10.1121/1.5089891](https://doi.org/10.1121/1.5089891)
- Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, 35(1), 185–205. doi: [10.1017/S0305000907008343](https://doi.org/10.1017/S0305000907008343)
- Siegel, G. M., & Pick Jr, H. L. (1974). Auditory feedback in the regulation of voice. *The Journal of the Acoustical Society of America*, 56(5), 1618-1624. doi: [10.1121/1.1903486](https://doi.org/10.1121/1.1903486)
- Titze, I. R. (1999). Toward occupational safety criteria for vocalization. *Logopedics Phoniatrics Vocology*. 24, 49–54. doi: [10.1080/140154399435110](https://doi.org/10.1080/140154399435110)
- Titze, I. R. (2000). *Principles of voice production* (second printing). Iowa City, IA: National Center for Voice and Speech. pp. 229–233, 361–366.
- Yiu, E. M., and Yip, P. P. (2016). Effect of noise on vocal loudness and pitch in natural environments: an accelerometer (ambulatory phonation monitor) study. *Journal of Voice*, 30(4), 389-393. doi: [10.1016/j.jvoice.2015.05.016](https://doi.org/10.1016/j.jvoice.2015.05.016)