The Influence of Room Acoustics on the Voice

Michael Rollins

A Capstone Project Report submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Bachelor of Science

Timothy Leishman, Advisor

Department of Physics and Astronomy

Brigham Young University

19 August 2016

ABSTRACT

Schoolteachers have a particularly high rate of voice-problem symptoms. Room acoustics could be a significant reason for this prevalence, but more needs to be known about the effects room acoustics on vocal effort. With increased understanding, rooms could be designed to mitigate unhealthy vocal effort, and by extension, voice problems. The present study attempts to measure the influence of room acoustic parameters on vocal changes by comparing the vocal effort of typical talkers in several distinct acoustic environments. Thirty-two participants were recorded completing a battery of speech tasks in eight widely ranging acoustic conditions. Key vocal parameters were derived from associated recordings and the statistical significance of the influence of the room acoustic parameters on each of the vocal parameters was determined using standard one-way ANOVA tests. It was found that changes in EDT, C50, and %ALcons had highly correlated effects on several vocal parameters, notably smoothed cepstral peak prominence, acoustic vocal quality index (AVQI), and pitch strength. As the EDT increased and C50 and %ALcons decreased, these and other vocal parameters tended toward more dysphonic phonation. There were also gender differences in several vocal parameters, including AVQI, pitch strength, and other vocal effort-related parameters, with females tending to exert more vocal effort. These findings begin to objectify the effect of room acoustics on vocal accommodations and provide grounds for developing future talker-oriented room acoustical standards.

Keywords: vocal accommodation, gender differences, EDT, %ALcons, C50

ACKNOWLEDGEMENTS

This capstone project was made possible by the contributions of several people, including student members of the research team: Jennifer Whiting, Mark Berardi, Kieren Smith, and Steven Markham, who made the work a joy; Eric Hunter, who obtained a grant wherein I had the unlikely opportunity to perform quality research about the voice as an undergraduate physics student; my adviser, Timothy Leishman, whose vision and belief in me has been a source of inspiration; and most of all, my wife Clara and daughter Suzanna and their nigh-infinite patience and support. Funding for this research came from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R01DC012315. They would like to remind the reader that the content of this publication is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Table of Contents

List of Abbreviations

%ALcons	Percentage articulation loss of consonants
ANOVA	Analysis of variables
AVQI	Acoustic voice quality index
BYU	Brigham Young University
C50	Clarity; definition; early-to-late sound ratio (50 ms)
CPPs	Smoothed cepstral peak prominence
EDT	Early decay time
F_0	Fundamental frequency
FFT	Fast Fourier transform
GUI	Graphical user interface
HNR	Harmonics-to-noise ratio
LF	Lateral fraction
MLS	Maximum length sequence
RASTI	Rapid speech transmission index
RT, T_{10}, T_{20}, T_{30}	Reverberation time
STI	Speech transmission index
SWIPE	Sawtooth waveform inspired pitch estimator

List of Tables

List of Figures

Chapter 1

Introduction

Schoolteachers, lawyers, tour guides, and many others depend on their voice for their livelihoods. Many such workers use their voices in a way that puts them at risk for developing voice disorders. Angelillo et al. (2009) and Roy et al. (2004) confirmed that in particular, schoolteachers have a higher rate of voice symptoms than nonteachers. Room acoustics could be a significant reason for this prevalence of voice problems, but more needs to be known about the effects room acoustics have on the voice [Kob et al. (2008)]. With an understanding of how room acoustics affect vocal effort, rooms could be later designed to mitigate vocal effort, and by extension, voice problems. The present study attempts to measure the influence of room acoustic parameters on vocal changes by comparing the vocal effort of typical talkers in several distinct acoustic environments.

1.1 Background and Motivations

Because about one third of workers in industrialized societies depend on their own voices as a principal tool of work, vocal health and voice disorders are of great import to both employers and employees [Vilkman (2000)]. Roy et al. (2004) found that schoolteachers in particular ascribed their voice problems to their occupation and indicated that their voices restricted which activities and interactions they could participate in. Furthermore, schoolteachers missed more workdays than nonteachers and were more likely to consider a change of occupation because of their voices.

The effects of classroom acoustics on the listener have been thoroughly studied during the last 50 years, especially with respect to noise levels and speech intelligibility [Hétu et al. (1990), Dreossi and Momensohn-Santos (2005), and Howard and Angus (2001)]. These studies have resulted in listener-focused standards and recommendations [ANSI (2010)]. Studies on the effects of classroom acoustics on the talker are, in comparison, few. Efforts on this front include investigation into the effect of room acoustics on the voice level of talkers in classrooms [Pelegrín-García and Brunskog (2012)], the effect of room characteristics on vocal intensity and rate [Black (1950)], and the vocal dose of schoolteachers [Bottalico and Astolfi (2012)]. The present study is an effort to gain insight into the relationship between room acoustics—specifically early decay time (EDT), definition (C50), percent articulation loss of consonants (%ALcons), and lateral fraction (LF)—and vocal parameters beyond voice level. These insights should be of use to those tasked with designing or treating rooms intended for unamplified single-talker communication, allowing more talker-friendly rooms that mitigate or alleviate unhealthy vocal effort.

1.2 Aims

This research and paper address the following question: To what extent do room EDT, C50, %ALcons, and LF elicit vocal accommodations in talkers? To investigate this question, participants were recorded speaking in several significantly different acoustic environments. Vocal accommodation was analyzed by comparing acoustically derived vocal parameters across the room acoustic parameters in each condition.

The primary goal of this investigation was to determine thresholds (values or ranges) of EDT, C50, %ALcons, and/or LF at which significant shifts in any of the vocal accommodation measurements were evoked. The secondary goal was to determine differences between genders in these vocal accommodations. This research supports the notion that published optimal ranges of room EDT, C50, %ALcons, and LF for the listener are considerate of the voice of the talker [Acoustic Design Ahnert (2009), pp. 647, 650, 656; Long (2006), pp. 585-6]. It assumes no sound reinforcement (microphones, amplifiers, loudspeakers, etc.) and minimal background noise in the acoustic environments. The results should not be extrapolated to conditions violating these assumptions without further investigation.

1.3 Plan of Development

This paper first describes the methods used in designing and carrying out the experiment, including the acoustic environments used, the participant population, the speech elicitation protocol, and data manipulation. Next, the results of the data are presented (in the form of vocal parameters plotted against room acoustic parameters) and discussed. Finally, conclusions are drawn and suggestions for further research are made.

Chapter 2

Methods

To study the influence of room acoustics on vocal accommodations, 32 participants were recorded completing a battery of speech tasks in each of eight acoustically distinct conditions characterized using several architectural parameters. The participants' speech recorded in each condition was analyzed by Praat software to derive key vocal parameters. Statistical significance of the influence of the room acoustic parameters on each parameter was determined using standard one-way ANOVA tests.

2.1 Acoustic Conditions

Eight acoustic conditions were used in the present study: seven in a reverberation chamber and one in an adjacent laboratory room of similar volume ($204m^3$ and $181m^3$, respectively). The latter served to produce a control condition. The acoustics were characterized by four common architectural acoustic parameters: early decay time (EDT), clarity factor for speech (C50), percentage articulation loss of consonants (%ALcons), and lateral energy fraction (LF) [defined respectively in Beranek (2003) pp. 577-8, Long (2006) p. 600, Acoustic Design Ahnert (2009) p. 652, and Beranek (2003) p. 579]. These parameters respectively provide objective measures for a wide range of perceptual room characteristics: reverberance, clarity, degredation of speech signal, and apparent source width or room spaciousness. Reverberation time (T20) information was also measured and is also included in Appendix A.

Characterization measurements were taken using a maximum-length sequence (MLS) signal generated by EASERA. A dodecahedron loudspeaker was used as the source and a GRAS 40AE ½" free-field microphone with a random-incidence corrector (for measuring EDT, C50, and %ALcons) or an Audio Technica AT4050 multi-pattern condenser microphone (for measuring LF) as the receiver (see Fig. 2.1). The loudspeaker and microphone were positioned in the room so as to mimic the locations of the researcher's and participant's heads, respectively (compare Figs. 2.1 and 2.3). They were placed at least 1 m from any wall at a (sitting) height of 120 cm ±2 cm and a source-receiver distance of 185 cm ±2 cm. The source-receiver distance was chosen to be conversational, while taking advantage of the articulation loss in the room. Two people were in the room during the characterization measurements to account for additional absorption from the presence of the researcher and participant (not shown in Fig 2.1).

Between 0 and 32 wedges were introduced incrementally (0, 2, 4, 8, 16, 24, 32) into the chamber and distributed evenly to achieve a wide range of acoustic conditions (see Fig. 2.1). The location of each wedge in each configuration was marked with tape on the floor for

(a) (b)

Figure 2.1. Characterizing the acoustic conditions in the reverberation chamber. (a) GRAS 40AE and 0 wedges. (b) AT4050 and 16 wedges.

consistency during recordings. The background noise in all conditions was well below the maximum recommended for a small classroom setting according to Long [(2006) pp. 606-7]. Two impulse response measurements were taken for each acoustic condition—one with each microphone.

The microphone signals were recorded using the EASERA software interface. Its built-in functions were then used to generate C50 values in octave bands, a RASTI-based %ALcons [Ahnert et al. (2002)], LF values in octave bands, and an impulse response .wav file. From this impulse response, MATLAB code developed at BYU extracted the EDT, T10, T20, and T30 in octave bands. The 500 and 1000 Hz octave bands were averaged to determine a single value for EDT, T10, T20, T30, and C50 in each condition; octave bands 125 through 1000 Hz were averaged to calculate a single LF, according to Beranek [(2003), pp. 577-8]. These values are shown in Figs. 2.2 and A.1. The parameter values measured in the adjacent room as a control condition, while achieved without any wedges, falls on the extrapolated curves very close to what would be expected for 48 wedges. For this reason, the control condition is plotted in Figs. 2.2 and A.1 at the 48 wedges mark along the abscissa and indicated with the subscript con (short for control condition). As can be seen in Fig. 2.2, LF does not exhibit a simple relationship, as do the other room acoustic parameters, so it is not further included in the analysis.

2.2 Participant Population

The participant population was drawn from undergraduate and graduate BYU students (n = 32) with an age range of 18-35. All participants were self-reported to have no hearing aids, hearing disorders, speaking impediments, or vocal disorders. The population was divided into two genders: half (16) of the participants were female and the other half were male. The population was also divided into one of three types: singer, musician (nonsinger) or else. A summary of the

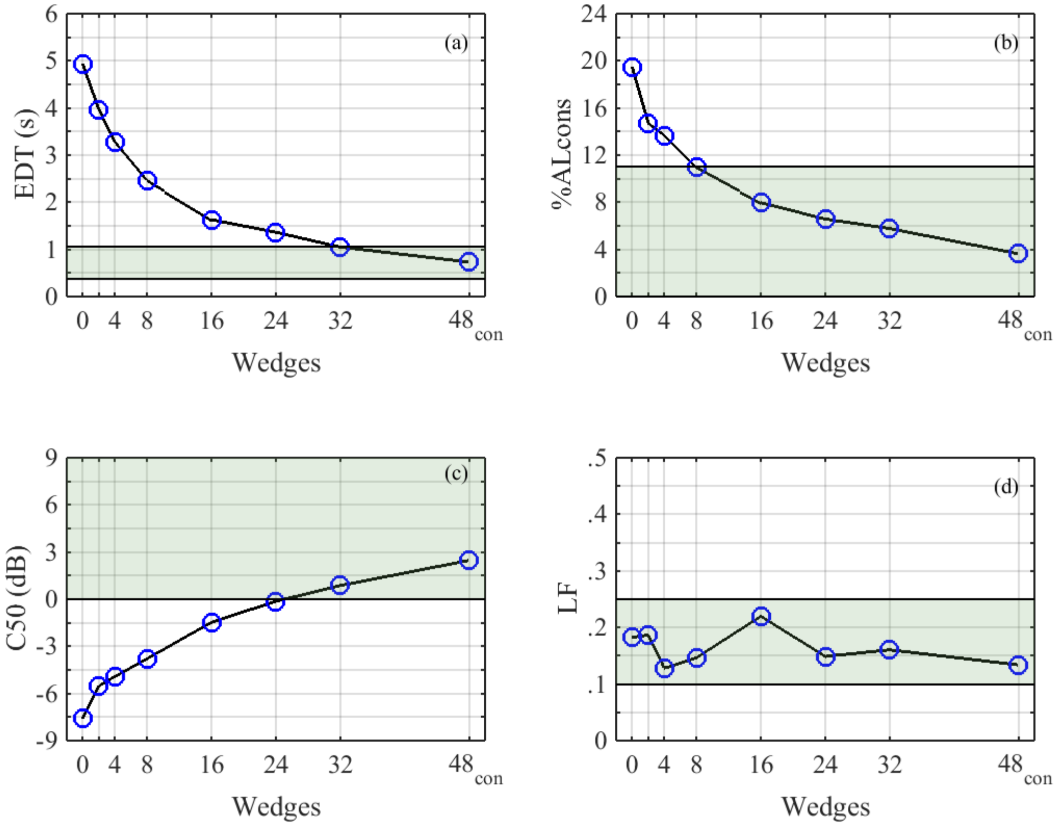

Figure 2.2. Room acoustic parameters in the eight acoustic conditions. Preferred values for unamplified speech, represented by shading in the graphs, are found in the following sources: (a) EDT [Long (2006), pp. 585-6], (b) %ALcons [Acoustic Design Ahnert (2009), p. 650], (c) C50 [ibid., p. 647], and (d) LF [ibid., p. 656]. The control condition is plotted at the 48 wedge mark based on curve extrapolation. The values plotted here are compared in Table C.1.

breakdown is included in Table 2.1. Gender, musicianship, and singing proficiency were self-reported. All study participants were briefed on the purpose and means of the experiment and gave their written consent to full participation.

Table 2.1. Breakdown of the participant population by gender and type.

	Singer	Musician	Else	Row Total
Female	7	5	4	16
Male	6	5	5	16
Column Total	13	10	9	32

2.3 Speech Elicitation

In each acoustic condition, an interviewer (researcher) prompted the participant through several speech tasks. These included reading the first paragraph of the Rainbow Passage in a clear and conversational manner [Fairbanks (1960)], sustaining the vowel /ɑ/ for five seconds three times, describing a cartoon image (from the Diapix image inventory), and answering an open-ended prompt (e.g. “Tell me about your favorite city.” “What’s your favorite dessert?” etc.). At the end of the battery of tasks, the participant was asked to rate their level of vocal effort and vocal fatigue in the acoustic condition, and estimate what their vocal fatigue would be after speaking for 20 minutes in the same condition. An iteration of all speech tasks in a given acoustic condition constituted one trial.

Each participant went through nine trials. The first, which took place in the control condition, was used for instructing the participant. Its data was not included in the analysis. After it was completed, the participant and researcher relocated to the reverberation chamber for the next seven trials (see Fig. 2.3) before returning again to the control condition for the ninth trial. The order of acoustic conditions in the reverberation chamber (trials two through eight)

(a) (b)

Figure 2.3. Recording participants in (a) 0-wedge condition and (b) 16-wedge condition.

was randomized for each participant. Between each trial, research assistants entered the reverberation chamber to add or remove wedges, during which the participant's voice was given a chance to rest.

The participant wore a DPA 4060 pre-polarized condenser microphone mounted to a headband with behind-the-ear grips (see Fig. 2.4). The microphone was positioned close to the corner of the mouth. A GRAS ½" free-field microphone with a random-incidence corrector was placed halfway between the participant and interviewer to capture the speech from both talkers as affected by the changing room acoustics.

2.4 Data Exploration

Recordings were taken using Reaper digital audio workstation and exported as .wav audio files. Metadata was stored in the filename via use of a strict file naming protocol (included in Appendix B). Using a MATLAB GUI (see Fig. 2.5), the raw recordings were segmented into separate .wav files for each task.

Two of the tasks—sentences 2 and 3 of the Rainbow Passage and 3 seconds of the sustained /ɑ/—were concatenated into a single audio file from which a MATLAB script

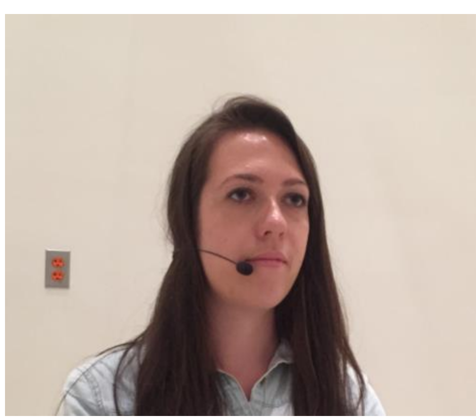

Figure 2.4. The DPA 4060 microphone worn by participants to record their speech.

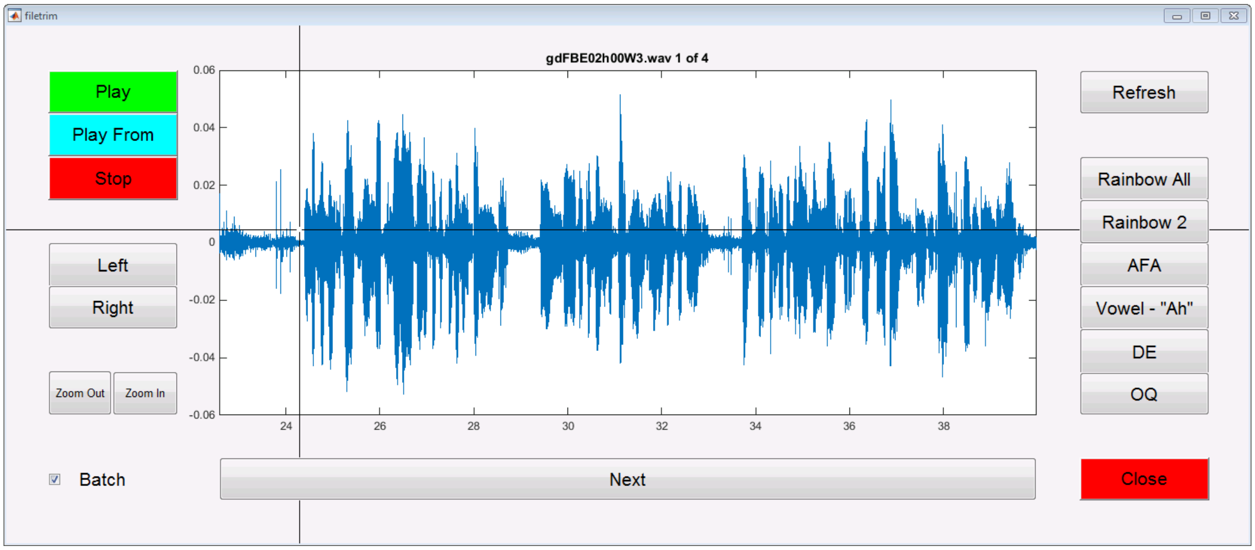

Figure 2.5. The MATLAB GUI developed to segment multi-channel audio files in batches.

controlled Praat to extract several useful vocal parameters. These vocal parameters were compared against several factors using several one-way ANOVA tests. Vocal parameters that were found to be significantly influenced by any factor are presented in the results of Chapter 3.

The vocal parameters investigated included mean fundamental frequency (F_0); mean loudness of voiced speech in dB; alpha ratio, a spectral balance calculated as a ratio between spectral energy below and above 1 kHz; spectral slope above F_0; speaking rate; FFT total energy; pitch strength, an objective measure meant to capture how salient the presence of pitch is; harmonics-to-noise ratio (HNR), which is a component of pitch strength but more widely used; smoothed cepstral peak prominence (CPPs), which is highly correlated with dysphonia severity [Maryn et al. (2015)]; shimmer dB, which measures local change in amplitude; and acoustic voice quality index (AVQI). The AVQI, developed by Maryn et al. (2010, 2015), is calculated from a weighted combination of CPPs, HNR, local shimmer, shimmer dB, slope, and tilt. These values were obtained using a special Praat script included by Maryn et al. in their 2015 paper.

The mean pitch strength was determined via MATLAB implementation of Aud-SWIPE-P, based on the SWIPE' (sawtooth waveform inspired pitch estimator) script provided by Camacho [(2007), pp. 99-101].

Chapter 3

Results

This chapter presents the results of the data analysis in three manners: a table listing vocal parameters that were significantly influenced, plots showing full and subpopulation means of vocal parameters across acoustic conditions, and plots showing selections from the previous two sets of plots normalized to the control condition.

3.1 Vocal Parameters Influenced by Factors

Each vocal parameter was subjected to a one-way ANOVA test considering the following factors: each room acoustic parameter, gender, type, and trial. In this study, significance was established by having a p-value of $p < 0.001$. As shown in Table 3.1, several vocal parameters were influenced by room acoustic parameters. Because of the high correlation of the latter, the vocal parameters were found to be influenced almost identically by them. As a result, the plots in this chapter all feature the same parameter (EDT) along the abscissa. A comparison of room acoustic parameter values in each condition can be found in Appendix C. Plots involving other parameters along the abscissa are shown in Appendix D.

Gender was found to influence many vocal parameters, including most of those influenced by room acoustic parameters. Some, such as F_0, were expected, while others, such as

Table 3.1. Vocal parameters influenced by factors ($p < 0.001$). Parameters with an asterisk are some of the six total parameters used in calculating the AVQI [Maryn et al. (2010)].

Vocal parameter	**Influential Factors**		
	Room Acoustic Parameters	**Gender**	**Type**
Acoustic voice quality index (AVQI)	Yes	Yes	
Harmonics-to-noise ratio (HNR)*	Yes	Yes	
Pitch strength mean	Yes	Yes	
Shimmer dB*	Yes	Yes	
Smoothed cepstral peak prominence (CPPs)*	Yes		
Speaking rate	Yes		
Alpha ratio		Yes	
F_0 mean		Yes	
FFT total energy		Yes	Yes
Mean voiced dB level		Yes	Yes
Spectral slope > F_0		Yes	

AVQI, were more surprising. Type was only found to significantly influence two vocal parameters: total energy and mean voiced dB level. Trial only influenced the speaking rate. This could be the result of increasing familiarity with the Rainbow Passage with each successive trial leading to increasing reading speeds. However, it should be noted that the room acoustic parameters also had a significant influence on the speaking rate.

3.2 Plots of Total Population Means

3.2.1 Vocal Parameters Influenced by Room Acoustic Parameters

Following are plots of the six vocal parameters listed in Table 3.1 as influenced by room acoustic parameters. The plotted values were averaged across the entire population ($n = 32$) into a single value per acoustic condition. Vocal parameters calculated as a component of the AVQI have a subscript AVQI after the parameter name [Maryn et al. (2015)].

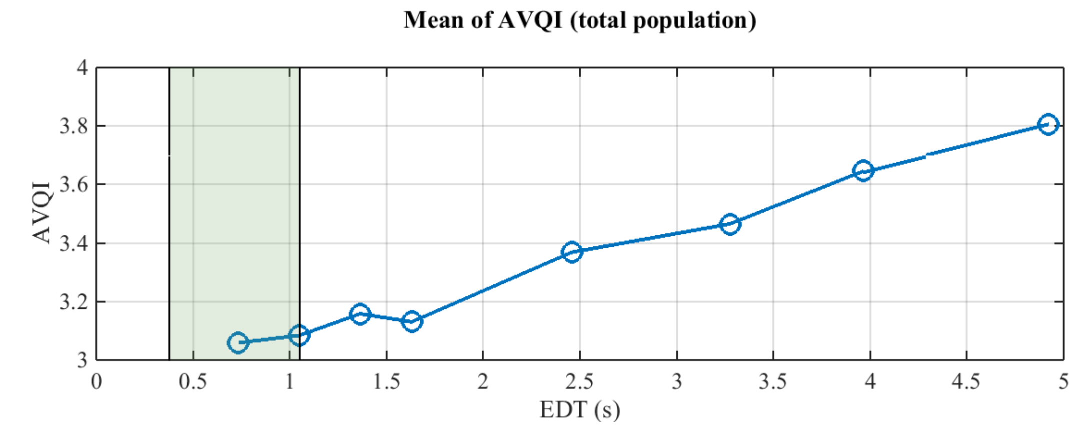

Figure 3.1. Plot of total population mean AVQI against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to higher AVQI.

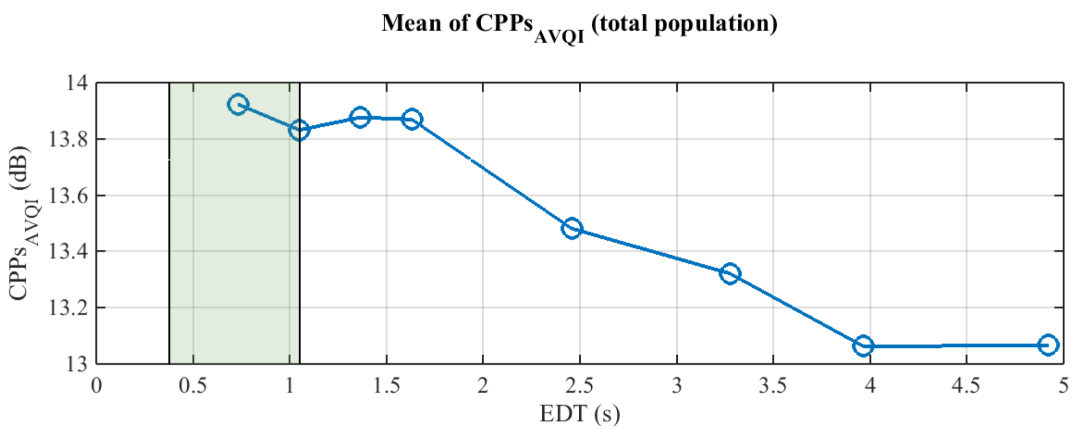

Figure 3.2. Plot of total population mean CPPs against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to lower CPPs.

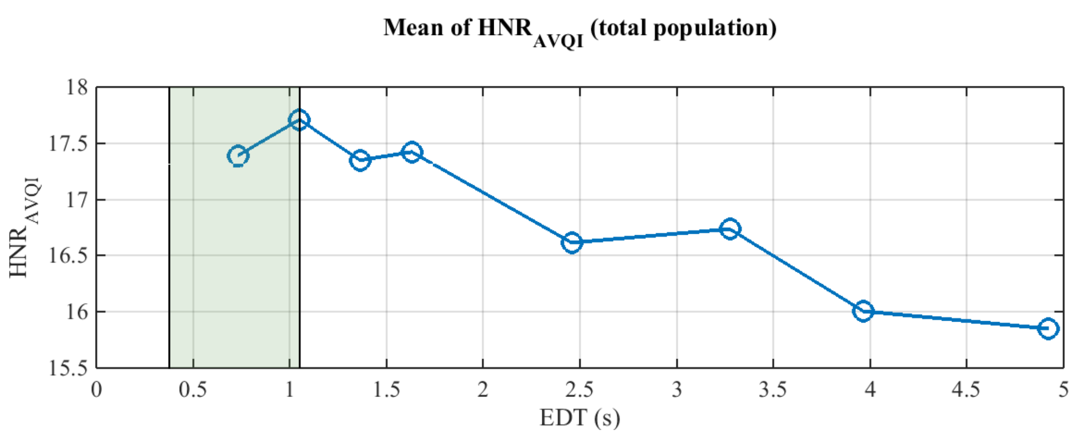

Figure 3.3. Plot of total population mean HNR against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to a lower HNR.

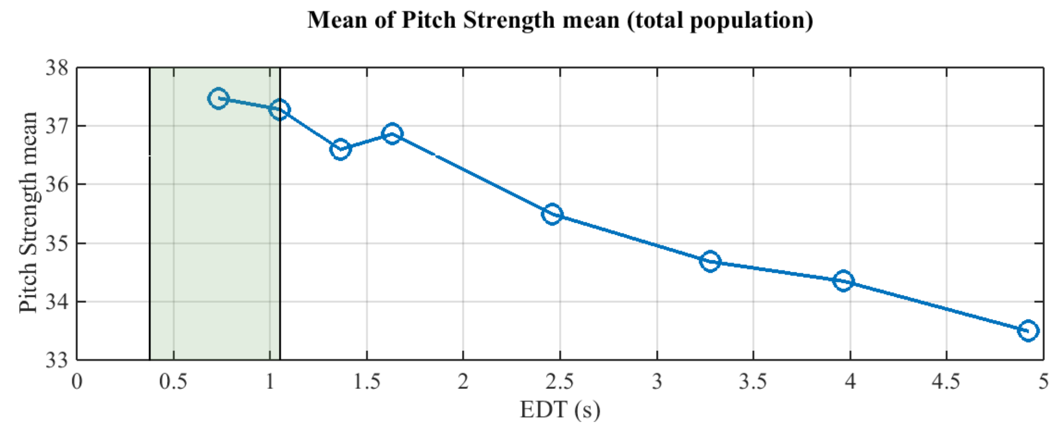

Figure 3.4. Plot of total population mean pitch strength mean against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to a lower pitch strength.

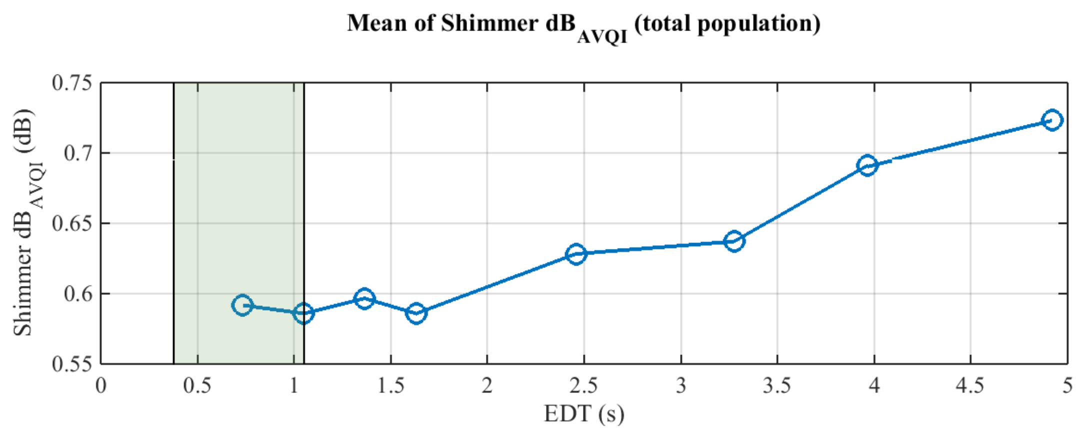

Figure 3.5. Plot of total population mean shimmer dB against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to higher shimmer dB.

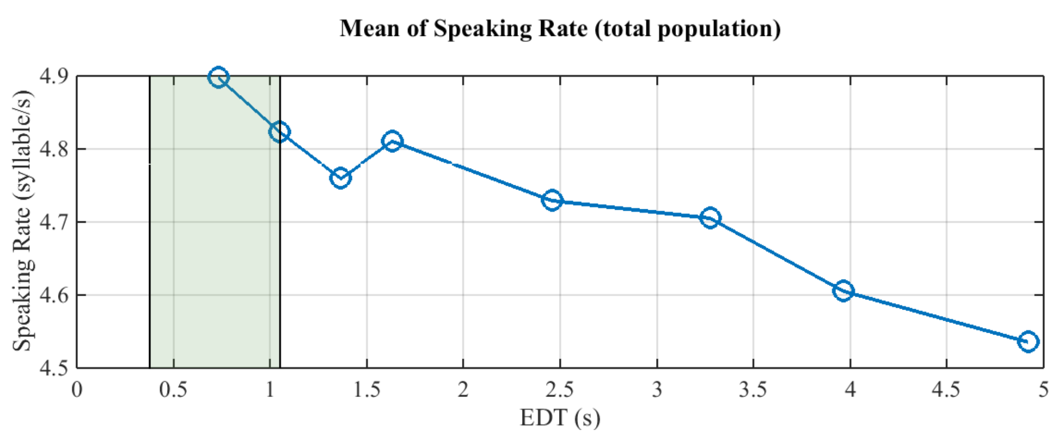

Figure 3.6. Plot of total population mean speaking rate against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to slower speaking rate.

3.2.2 Vocal Parameter Influenced by Trial Number

The only vocal parameter significantly influenced by trial number—speaking rate—is shown in Fig. 3.7.

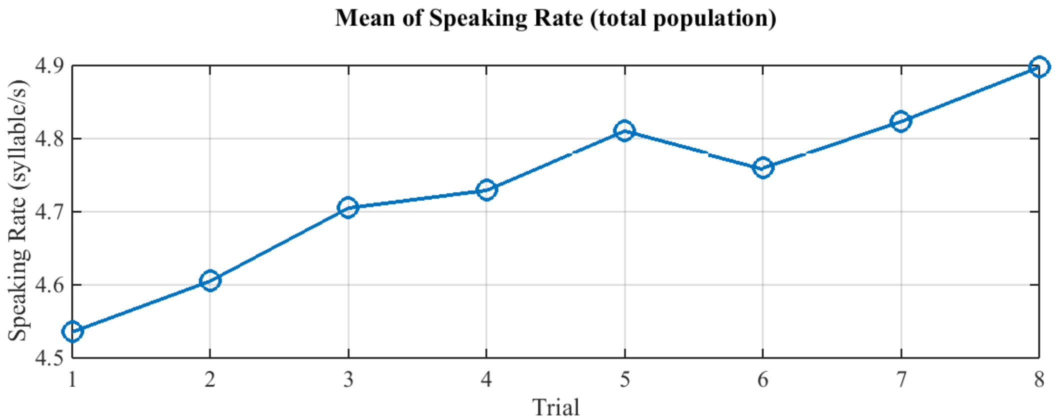

Figure 3.7. Plot of total population mean speaking rate against trial number. The speaking rate increases as the trials advance.

3.3 Plots Comparing Subpopulations

3.3.1 Vocal Parameters Influenced by Gender

Following are plots of the vocal parameters in Table 3.1 that were significantly influenced by gender. The vocal parameters in Figs. 3.8 through 3.11 are influenced by both room acoustic parameters and gender, while those in Figs. 3.12 through 3.16 are only influenced by gender.

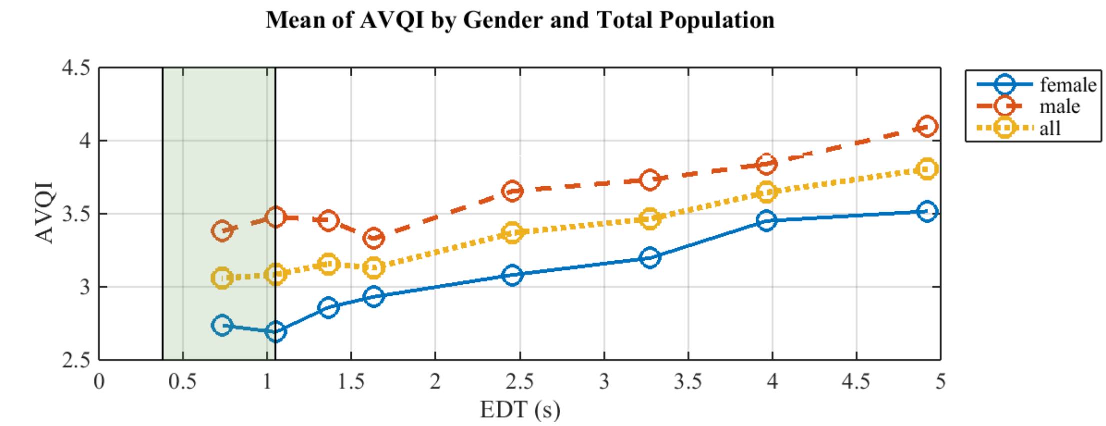

Figure 3.8. Plot of AVQI means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to higher AVQI and females have a lower AVQI than males.

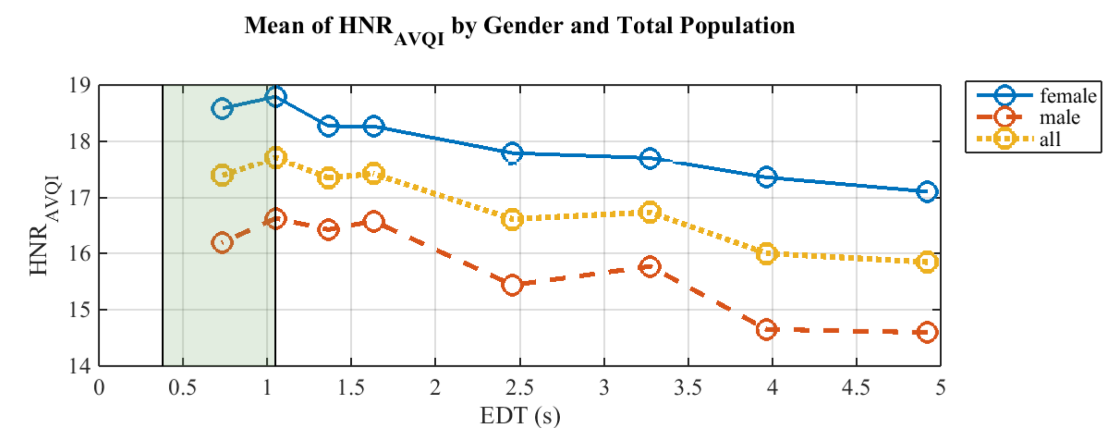

Figure 3.9. Plot of HNR means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to lower HNR rate and females have a higher HNR than males.

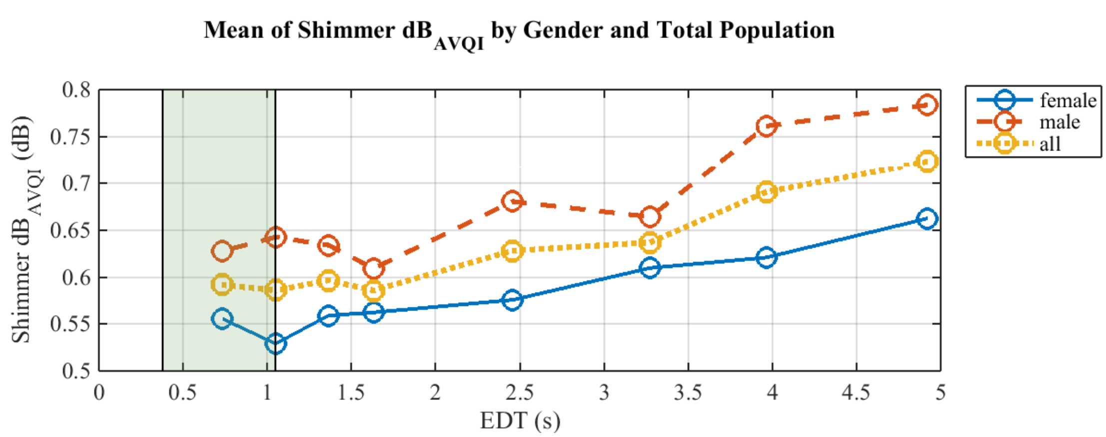

Figure 3.10. Plot of shimmer dB means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to higher shimmer dB and females have a lower shimmer dB than males.

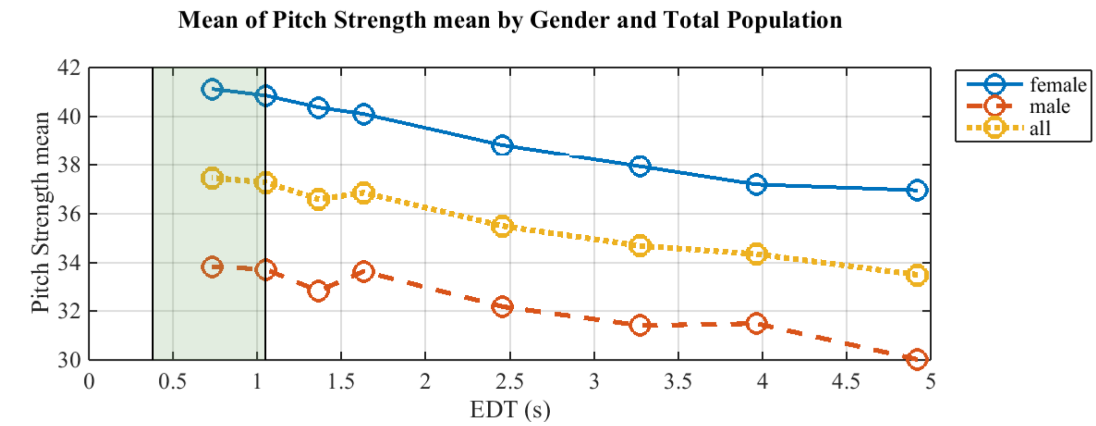

Figure 3.11. Plot of pitch strength means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Longer EDT corresponds to lower pitch strength and females have a higher pitch strength than males.

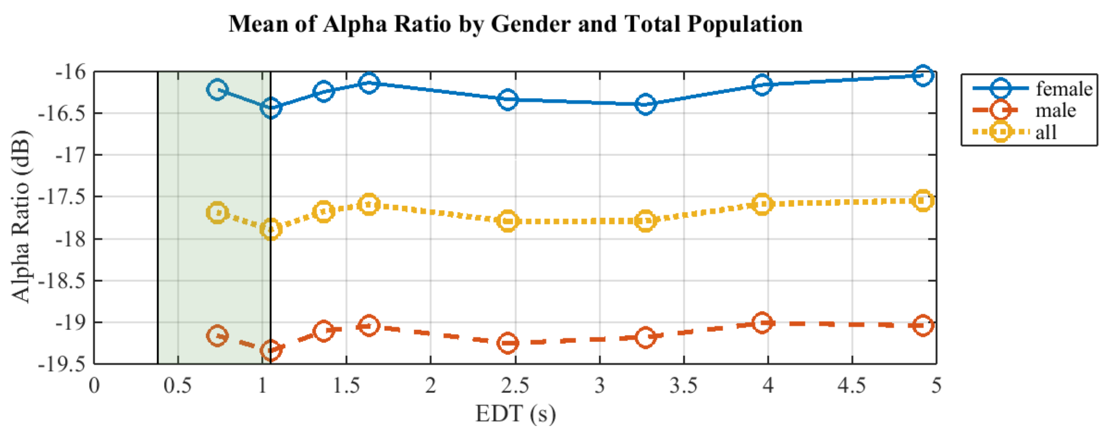

Figure 3.12. Plot of alpha ratio means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Females have a higher alpha ratio than males.

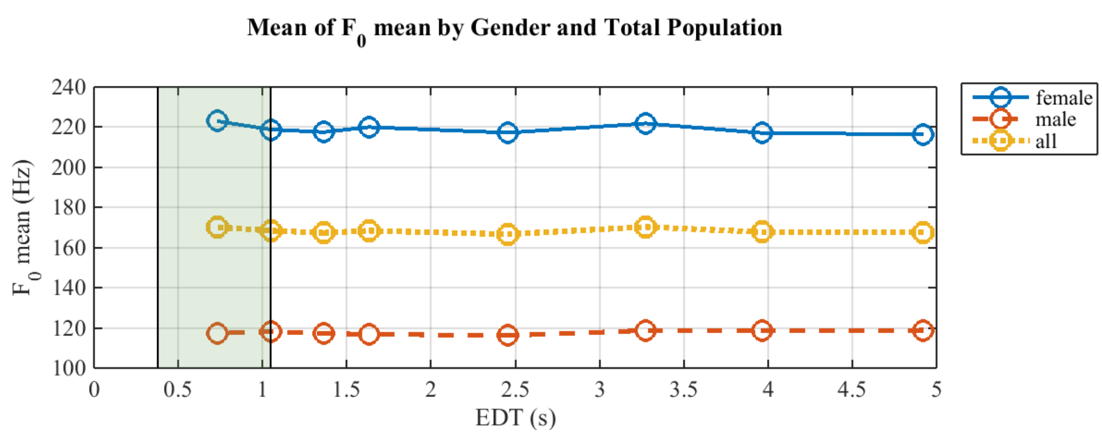

Figure 3.13. Plot of F_0 means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Females have a higher F_0 than males (as is well established).

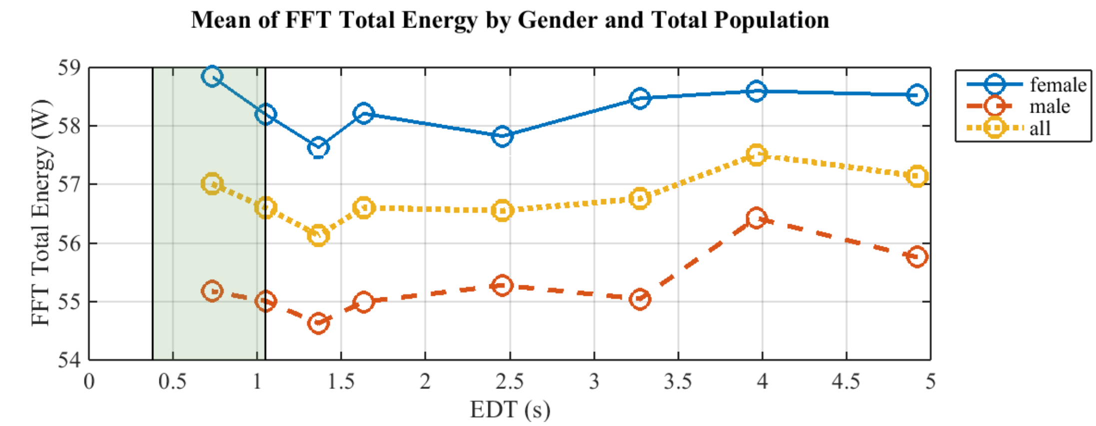

Figure 3.14. Plot of FFT energy means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Females have more FFT energy than males.

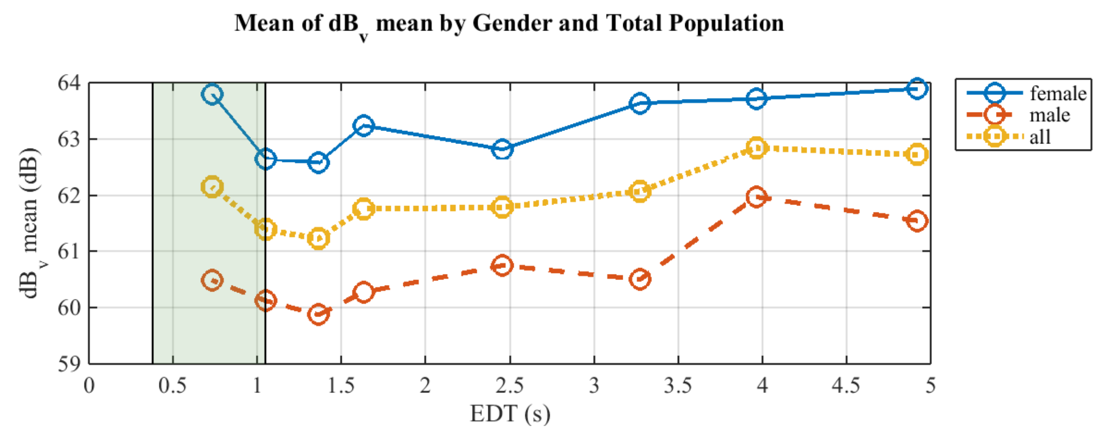

Figure 3.15. Plot of voiced dB means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Females have a higher voiced dB level than males.

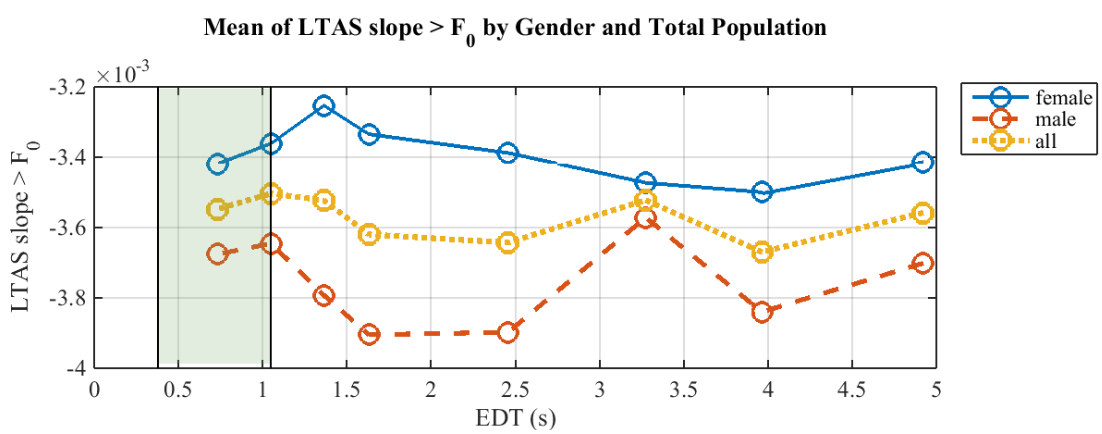

Figure 3.16. Plot of spectral slope means by gender against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Females have a higher spectral slope than males.

3.3.2 Vocal Parameters Influenced by Type

Following are plots of the vocal parameters in Table 3.1 that were significantly influenced by type. These two vocal parameters were not influenced by room acoustic parameters, but they were both influenced by gender (see Figs. 3.14 and 3.15).

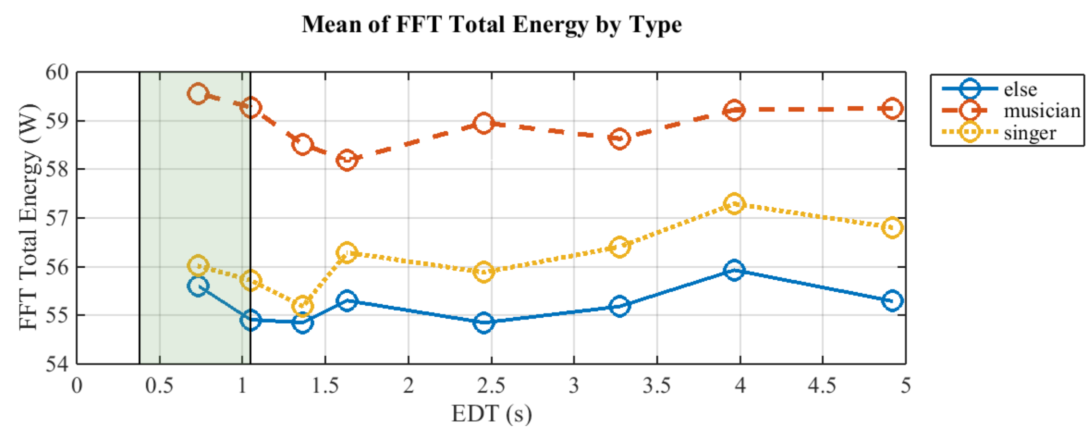

Figure 3.17. Plot of FFT energy means by type against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Musicians have more FFT energy than the rest of the population.

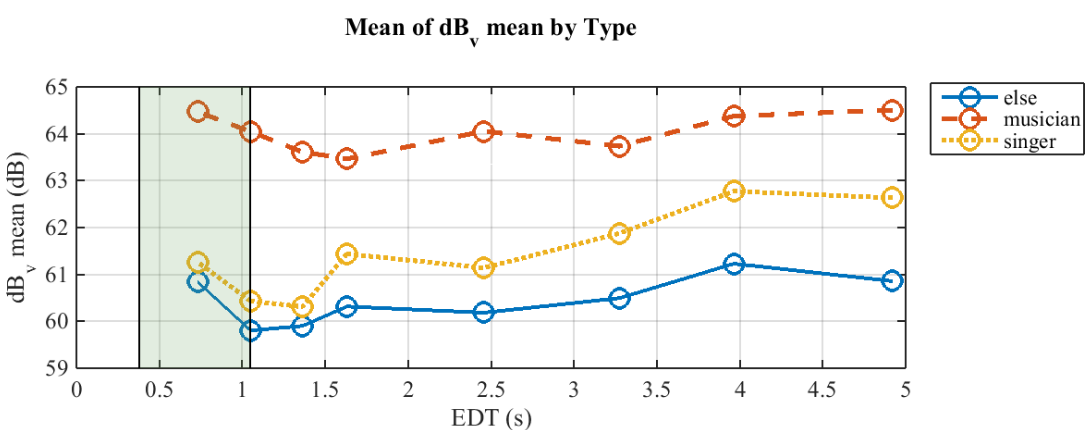

Figure 3.18. Plot of voiced dB means by type against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Musicians have a higher voiced dB level than the rest of the population.

3.4 Selected Normalized Plots

Following are plots of the AVQI, pitch strength mean, and voiced dB mean normalized to the control condition. This serves to demonstrate the extent of departure from the control condition of the population for these three vocal parameters.

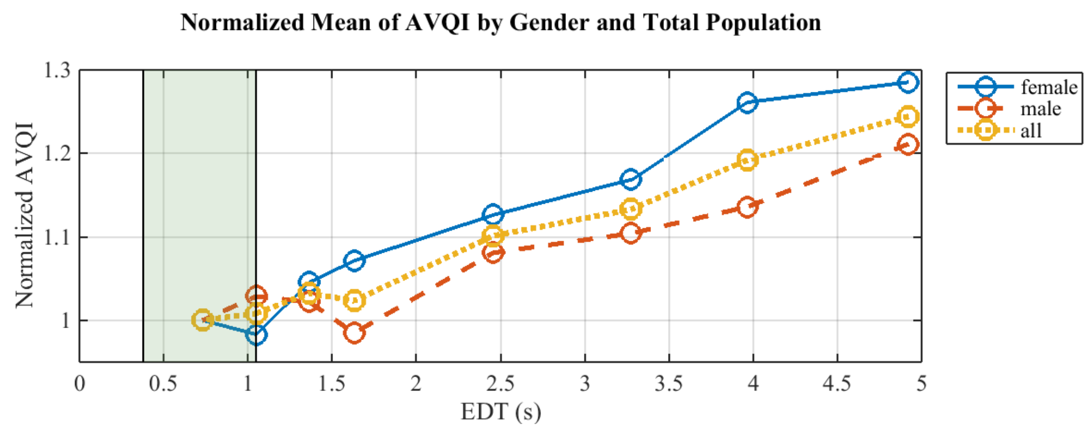

Figure 3.19. Plot of normalized AVQI means against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Females tended to increase their AVQI by a higher proportion than males with increasing EDT.

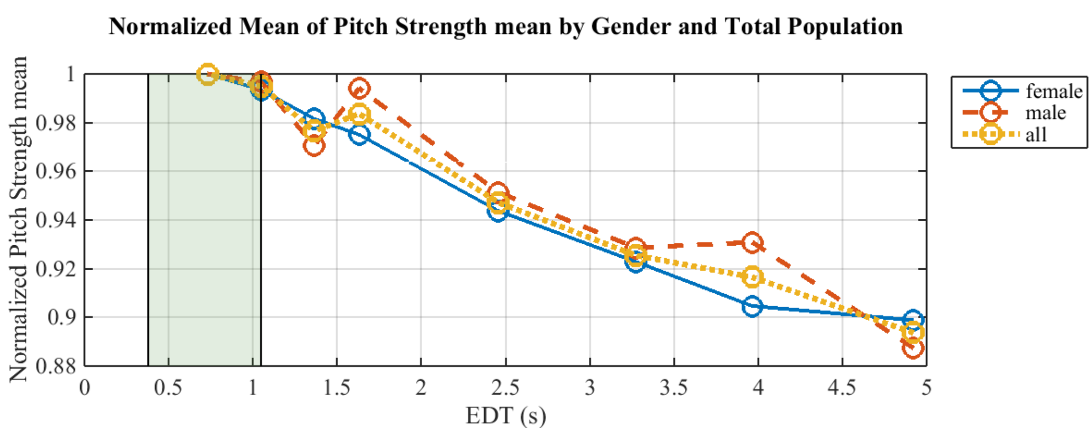

Figure 3.20. Plot of normalized pitch strength means against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Males and females changed their pitch strength by almost the same proportions with increasing EDT.

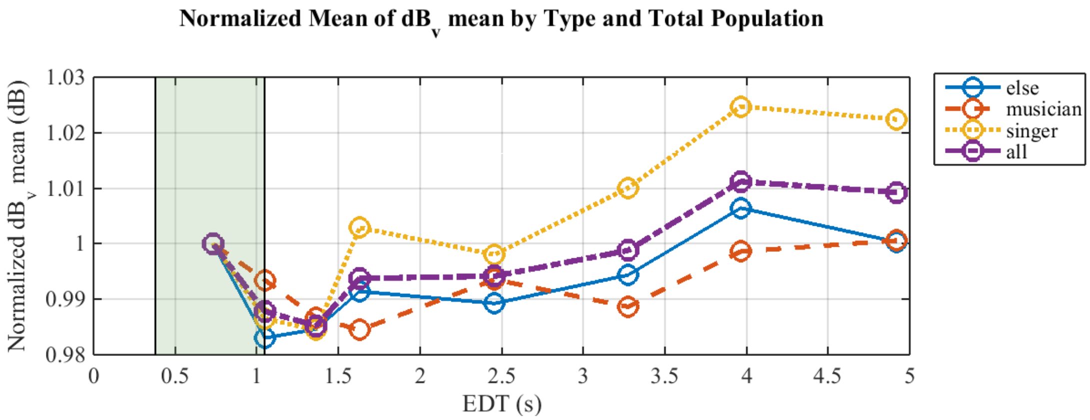

Figure 3.21. Plot of normalized voiced dB means against EDT. The shaded area designates the optimal EDT range for speech intelligibility. Singers increased their voiced dB level by a greater proportion than any other type above an EDT of 1.5 s.

Chapter 4

Analysis and Discussion

4.1 Factor Influence

4.1.1 Room Acoustic Parameters

Several vocal parameter means were significantly influenced by the room acoustic parameters: AVQI, CPPs, HNR, shimmer dB, pitch strength mean, and speaking rate. The trend of each of these is toward values more typical of dysphonic speech as the EDT increases [Maryn et al. (2010)]. This suggests that extremely reverberant acoustic conditions encourage more dysphonic-like use of the voice.

4.1.2 Gender

Additionally, gender affected several vocal parameters: AVQI, HNR, shimmer dB, pitch strength mean, alpha ratio, F_0, total FFT energy, loudness during voiced speech, and spectral slope above F_0. Females consistently exhibited a greater average F_0, FFT energy, HNR, loudness during voiced speech, pitch strength, and spectral slope than males, which suggests a higher level of vocal effort in females in each acoustic condition. Interestingly, the AVQI and shimmer dB were slightly lower for females than for males, which suggests that the females in this study were producing slightly less-dysphonic speech than their male counterparts, even with their increased vocal effort.

4.1.3 Type

Type also seemed to exert an influence on vocal parameters, albeit only loudness of voiced speech and total FFT energy. It is interesting that the biggest group difference for both parameters was between musician and else, with singers averaging between the other two. The influence of type on vocal parameters was an auxiliary pursuit, so not finding a strong difference between types does not lessen the import of the other findings of the study.

4.2 Extent of Change

4.2.1 Room Acoustic Parameters

In addition to the direction of change, it is informative to examine the magnitude of the change in vocal parameters corresponding to the room acoustics. For AVQI, a cut-off score to discriminate between dysphonic and normal voices has been set at 3.46 [Reynolds et al. (2012)]. This value is reached or surpassed by the mean of the total population for all $\mathrm{EDT} \geq 3.3$ s; by females when $\mathrm{EDT} \geq 1.1$ s (except for the condition where $\mathrm{EDT} = 1.6$ s); and by males for when $\mathrm{EDT} \geq 4.0$ s.

CPPs as calculated by Praat (as it was in the present study) is relatively new and does not yet have established normative values for typical or dysphonic speech [Maryn et al. (2015)]. Pitch strength of the voice (as opposed to noise, tones, or a mixture thereof) has also only been evaluated recently [Eddins et al. (2016)], and there is not yet an established cut-off value for dysphonic speech, and speakers with dysphonia vary in pitch strength [Shrivastav et al. (2012)]. The discovery of such a cut-off for these vocal parameters would be valuable in determining optimal room acoustics for talkers. The difference in CPPs between maximum and minimum EDT conditions being less than 1 dB is underwhelming, but the difference for pitch strength mean—4 points for males, females, and the total population—is not insignificant.

While some individuals varied in HNR by up to 7.5 dB, these were generally in nearly adjacent EDT conditions. The total population mean only varied by 1.5 dB, and females and males each by 1.7 dB, and therefore while statistically significant, the overall HNR is seen to not vary very much with the room acoustics.

Shimmer dB has a history of inconsistency across different computational algorithms, and indeed it seems to be more informative in many cases to measure its change than its absolute value [Bielamowicz et al (1996)]. As such, no critical value is used in this study to classify a corresponding critical EDT value. The largest difference in shimmer dB is 0.14 dB for the mean of the total population; 0.13 dB for females; and 0.17 dB for males, all of which are significant changes, considering the absolute values between 0.5 and 0.8 dB.

The difference between maximum and minimum speaking rate is moderate: 0.36 syll/s for the whole population, 0.40 syll/s for females, and 0.32 syll/s for males. However, it should be remembered that trial number also had a statistically significant effect on speaking rate (see Fig. 3.7), which indicates that a more nuanced treatment is necessary to understand the effect.

4.2.2 Gender

Gender affected a steady difference for several vocal parameters: females tend to have an HNR 2.2 ± 0.35 dB higher than males; a pitch strength mean 6.8 ± 0.57 points higher than males; an alpha ratio 2.9 ± 0.064 points higher than males; a F_0 mean 101 ± 2.6 Hz higher than males; 3 ± 0.49 W more energy than males; and a voiced level 2.6 ± 0.54 dB higher than males.

Gender elicits a looser difference in other vocal parameters: females tend to have an AVQI score 0.56 ± 0.13 points lower than males; a shimmer of 0.091 ± 0.034 dB higher than males; and an LTAS slope above F_0 $3.6 \times 10^{-4} \pm 1.6 \times 10^{-4}$ s higher than males.

Error bounds given in this section are ± 1 standard deviation from the mean. Of all the gender differences, the AVQI score, pitch strength, alpha ratio, F_0, and voiced dB level are significant in magnitude.

4.3 Implications

These trends in the data suggest that established listener-oriented standards bode well for mitigating extra vocal exertion, in the absence of noise. In fact, some of the vocal parameters do not seem to change significantly until optimal values of C50 or EDT are far exceeded. However, further investigation measuring the voice in varying levels of noise is needed to validate or challenge the effectiveness of current standards for the vocal health of talkers in noisy environments, such as school classrooms.

Of course, many questions remain, including the potential difference in accommodation between males and females (compare, for instance, the slopes of the different genders in Figs. 3.8 and 3.9). Another point of interest is an effective room acoustic parameter threshold at which several vocal parameters start to shift. From the data of this study, it appears that such a threshold may exist for EDT between 1 and 3 s. Because the present study ranged widely in EDT values, this question needs to be answered via finer EDT intervals. This may even be possible to investigate using a virtual real-time convolution system in lieu of physical rooms [Whiting (2014)].

4.4 Errors and Limitations

While convincing, the results drawn from this study ought not to be taken out of context. This experiment recorded only native talkers without voice or hearing problems. The speech elicitation was also conducted at a single talker-listener distance, sitting down, with minimal background noise and without sound reinforcement. Ambient conditions, such as humidity and temperature, were not measured nor controlled for. Any extrapolations of the findings of this study into any such circumstances should be first substantiated with further experimentation. Many of these issues have been researched to some degree, and all would benefit from further investigation.

A slight practical inconvenience in the study may have also affected the data. Specifically, the required room change between the control condition and the rest of the conditions may have had an effect on the research participants' voices. This room change not only signified the last trial, but took about two minutes and may have incurred a slight psychological shift in some participants.

Chapter 5

Conclusions

5.1 Summary of Findings

The present study aimed to discover objective relationships between the EDT, C50, %ALcons, or LF of a room and vocal accommodations. Based on the analysis, it seems safe to conclude that the LF within optimal range does not have a noticeable impact on vocal accommodation. On the other hand, changes in EDT, C50, and %ALcons were very highly correlated in the study's conditions and had indistinguishably similar effects on several vocal parameters, notably AVQI, pitch strength, and CPPs. As the reverberance increased and clarity and %ALcons decreased, these and other vocal parameters tended toward more dysphonic production. There were gender differences in several vocal parameters, including AVQI, pitch strength, and other vocal effort-related parameters, with females tending to exert more vocal effort in every acoustic condition.

5.2 Further Research

Several courses for further inquiry were presented in the previous chapter. The most fruitful of these might include repeating these experiments with finer EDT (and accompanying RT) incremental values to find a more distinctive threshold range, or to determine effects of EDT values less than 0.4 s. Investigating these same vocal accommodations with the introduction of noise to simulate a schoolteacher working environments would also be valuable. Using the existing data from this study, further explorations could be made into interaction between factors, the potential nonlinear influence of LF on certain vocal parameters, and differences in gender accommodations. Eventually, the findings presented and suggested here could lead to improved building standards that incorporate both talker and listener needs.

References

Acoustic Design Ahnert (**2009**). *Ease 4.3 Users Manual*. Berlin, Germany: Acoustic Design Ahnert, pp. 638-656.

Ahnert, W., Feistel, S., and Schmitz, O. (**2002**). "Implementation of intelligibility algorithms into EASE 4.0," Audio Engineering Society Conference: 21st International Conference: Architectural Acoustics and Sound Reinforcement (pp. 92-103). Audio Engineering Society E-Library. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=11200

Angelillo, M., Maio, G. D., Costa, G., Angelillo, N., and Barillari, U. (**2009**). "Prevalence of occupational voice disorders in teachers," J Prev Med Hyg, **50**, 26-32. doi:http://dx.doi.org/10.15167/2421-4248/jpmh2009.50.1.152

American National Standards Inst. (**2002**). ANSI S12.60 PART 1 (R2010) American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools, Part 1: Permanent Schools (American National Standards Institute, New York).

Beranek, L. (**2003**). *Concert halls and opera houses: music, acoustics, and architecture* (2nd ed.). New York: Springer, pp. 570-580.

Bielamowicz, S., Kreiman, J., Gerratt, B., Dauer, M., and Berke, G. (**1996**). "Comparison of Voice Analysis Systems for Perturbation Measurement," J. of Speech, Lang., and Hear. Research, **39**, 126-134. doi:http://dx.doi.org/10.1044/jshr.3901.126

Black, J. W. (**1950**). "The Effect of Room Characteristics upon Vocal Intensity and Rate," J. Acoust. Soc. Am., **22**, 174. doi: http://dx.doi.org/10.1121/1.1906585

Bottalico, P., and Astolfi, A. (**2012**). "Investigations into vocal doses and aparameters pertaining to primary school teachers in classrooms," J. Acoust. Soc. Am., **131**, 2817-2827. doi:http://dx.doi.org/10.1121/1.3689549

Camacho, A. (**2007**). *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*. Ph.D. dissertation, U. of FL, Gainesville, FL: University of Florida. Retrieved from http://www.cise.ufl.edu/~acamacho/publications/dissertation.pdf

Dreossi, R. C., and Momensohn-Santos, T. M. (**2005**). "Noise and its interference over students in a classroom environment: literature review," Pró-Fono R. Atual. Cient., **17**(2) doi:http://dx.doi.org/10.1590/S0104-56872005000200014

Fairbanks, G. (**1960**). *Voice and articulation drillbook* (2nd ed.). New York: Harper & Row.

Eddins, D., Anand, S., Camacho, A., and Shrivastav, R. (**2016**). "Modeling of Breathy Voice Quality Using Pitch-strength Estimates," J. Voice. doi:http://dx.doi.org/10.1016/j.jvoice.2015.11.016

Hétu, R., Truchon-Gagnon, C., and Bilodeau, S. A. (**1990**). "Problems of Noise in School Settings: A Review of Literature and the Results of an Exploratory Study," J. Speech Lang. Pathol. Audiol., **14**, 31-38.

Howard, D. M., and Angus, J. A. (**2001**). "Room Acoustics: How they affect vocal production and perception," in P. Dejonckere (Ed.), Occupational Voice: Care and Cure (pp. 29-46). The Hague, The Netherlands: Kugler Publications.

Kob, M., Kamprolf, A., Neuschaefer-Rube, C., Goldschmidt, O., and Behler, G. (**2008**). “Experimental investigations of the influence of room acoustics on the teacher's voice,” Acoust. Sci. and Tech., **20**, 86-94. doi:http://dx.doi.org/10.1250/ast.29.86

Long, M. (**2006**). *Architectural Acoustics*. Burlington, MA: Elsevier Academic Press, pp. 85-88, 585-586, 600, 606-607.

Maryn, Y., and Weenink, D. (**2015**). “Objective Dysphonia Measures in the Program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index,” J. Voice, **29**(1), 35-43. doi:http://dx.doi.org/10.1016/j.jvoice.2014.06.015

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., and De Bodt, M. (**2010**). “Toward Improved Ecological Validity in the Acoustic Measurement of Overall Voice Quality: Combining Continuous Speech and Sustained Vowels,” Journal of Voice, 540-555. doi:http://dx.doi.org/10.1016/j.jvoice.2008.12.014

Pelegrín-García, D., and Brunskog, J. (**2012**). “Speakers' comfort and voice level variation in classrooms: Laboratory research,” J. Acoust. Soc. Am., **132**, 249. doi:http://dx.doi.org/10.1121/1.4728212

Reynolds, V., Buckland, A., Bailey, J., Lipscombe, J., Nathan, E., Vijayasekaran, S., Kelly, R., Maryn, Y., and French, N. (**2012**). “Objective Assessment of Pediatric Voice Disorders with the Acoustic Voice Quality Index,” J. Voice, **26**(5), 627.e1-627.e7. doi:http://dx.doi.org/10.1016/j.jvoice.2012.02.002

Roy, N., Merrill, R. M., Thibeault, S., Gray, S. D., and Smith, E. M. (**2004**). “Voice Disorders in Teachers and the General Population: Effects on Work Performance, Attendance, and Future Career Choices,” J. Speech Lang. Hear. R., **47**, 542-551. doi:http://dx.doi.org/10.1044/1092-4388(2004/042)

Shrivastav, R., Eddins, D., and Anand, S. (**2012**). "Pitch strength of normal and dysphonic voices," J. Acoust. Soc. Am., **131**(3), 2261-2269. doi:http://dx.doi.org/10.1121/1.3681937

Vilkman, E. (**2000**). "Voice problems at work: A challenge for occupational safety and health arrangement," Folia Phoniatr Logop, **52**(1-3), 120-125. doi:http://dx.doi.org/10.1159/000021519

Whiting, J. (**2014**). *On the Use of EASE and a Real-Time Convolution System for the Creation and Use of Virtual Acoustic Environments*. Unpublished bachelor thesis, Brigham Young U., Provo, UT.

Appendix A

RT Information

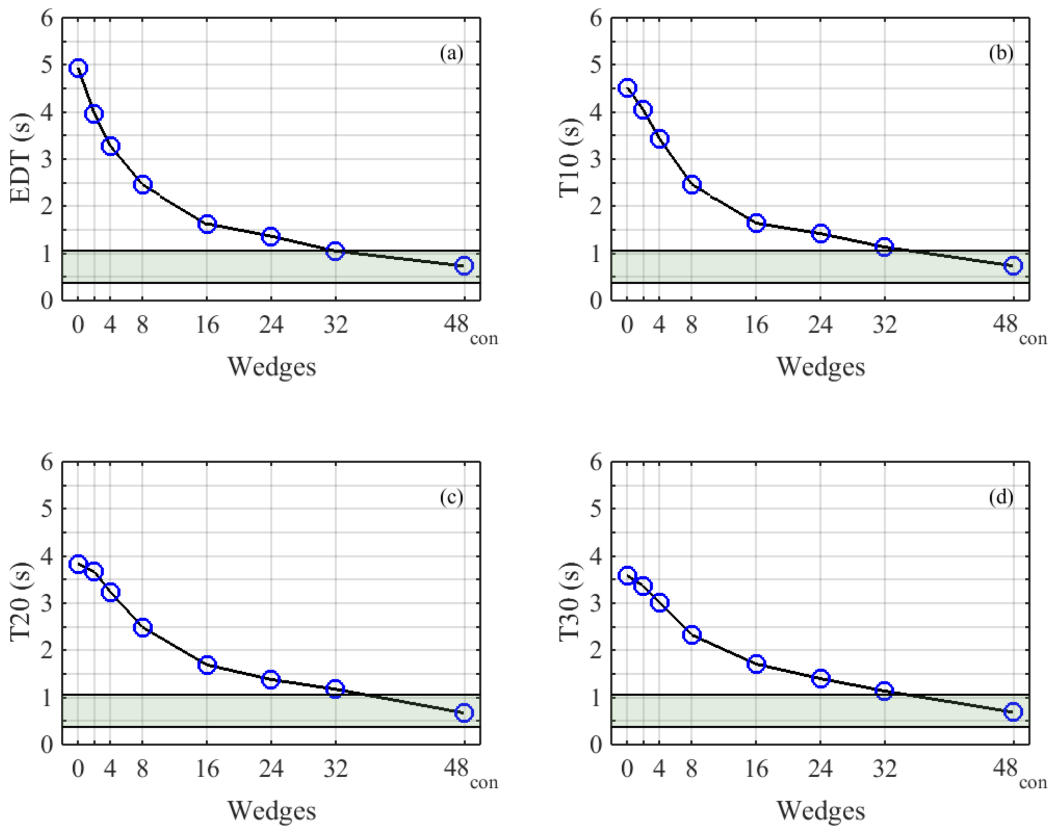

Figure A.1. Values for the (a) EDT, (b) T10, (c) T20, and (d) T30 values for the eight acoustic conditions. Shaded areas represent preferred values for unamplified speech. Preferred values are from Long (2006) pp. 585-6. The control condition is plotted at the 48 wedge mark based on curve extrapolation.

Appendix B

File Naming Protocol

Table B.1. Description of file-naming protocol for recordings. Positions 13-14 are only used for trimmed recordings.

Position	**Rule**	**Specifies**
1-2	gd	grant
3	M *or* F	gender
4	B	facility used
5	S *or* M *or* E	type (singer / musician / else)
6-7	##	participant ID number, e.g. 01, 32
8	h *or* r	microphone (head-worn / room)
9-11	##W *or* 99C	acoustic condition (number of wedges / control condition)
12	#	trial number, between 1 and 9
13-14	AH *or* R2	specify task (sustained /ɑ/ / Rainbow Passage sentences 2-3)

Examples of raw recording filename:
gdFBM01r04W4.wav

Example of trimmed recording filename:
gdMBE01h12W6R2.wav

Appendix C

Comparison of Room Acoustic Parameters

Table C.1. Correspondence of seven room acoustic parameters in the eight acoustic conditions used in the study. The EDT, T10, T20, T30, and C50 values are averages from the 500 and 1000 Hz octave bands. The %ALcons values are based on the RASTI calculation, which is in turn based on the 500 and 2000 Hz octave bands. The LF values are averages of all octave bands from 125 to 1000 Hz. The control condition is listed at the 48 wedge mark based on curve extrapolation. Plots of these values can be seen in Figs. 2.2 and A.1.

Wedges	EDT (s)	T10 (s)	T20 (s)	T30 (s)	C50 (dB)	%ALcons	LF
0	4.92	4.52	3.83	3.58	-7.57	19	0.182
2	3.96	4.04	3.67	3.36	-5.52	15	0.186
4	3.27	3.43	3.24	3.01	-4.92	14	0.128
8	2.46	2.47	2.49	2.33	-3.78	11	0.146
16	1.63	1.64	1.69	1.71	-1.47	8	0.219
24	1.36	1.42	1.38	1.40	-0.13	7	0.148
32	1.05	1.13	1.17	1.14	0.87	6	0.160
(48) control condition	0.73	0.73	0.67	0.69	2.47	4	0.133

Appendix D

Plots of Vocal Parameters against More Room Acoustic Parameters

This appendix includes plots associated with Chapter 3, but with %ALcons or C50 along the abscissa instead of EDT.

D.1 Plots of Vocal Parameters against %ALcons

D.1.1 Vocal Parameters Influenced by %ALcons against %ALcons

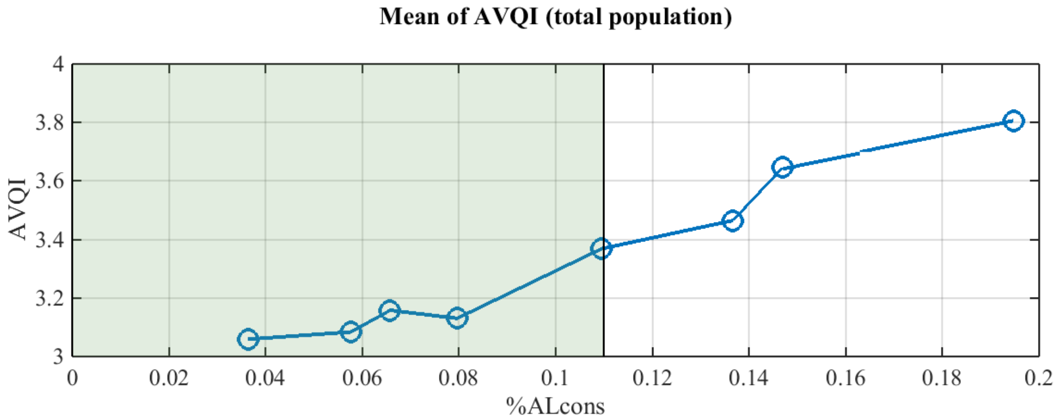

Figure D.1. Plot of total population mean AVQI against %ALcons. The shaded area designates the optimal %ALcons range for speech intelligibility. Higher %ALcons corresponds to higher AVQI.

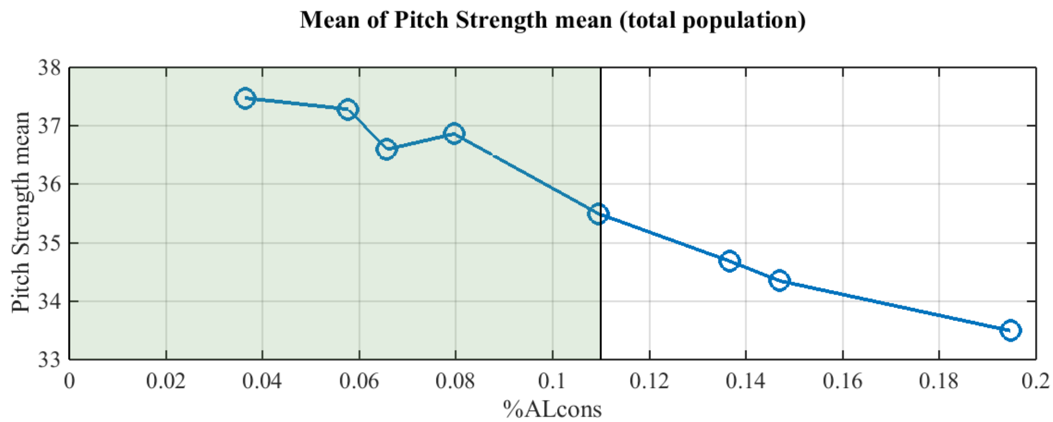

Figure D.2. Plot of total population mean pitch strength against %ALcons. The shaded area designates the optimal %ALcons range for speech intelligibility. Higher %ALcons corresponds to higher pitch strength.

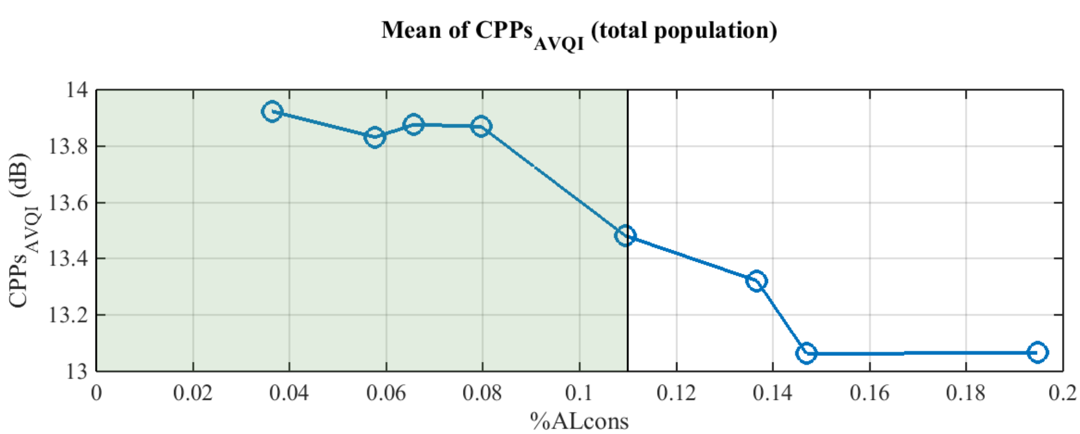

Figure D.3. Plot of total population mean CPPs against %ALcons. The shaded area designates the optimal %ALcons range for speech intelligibility. Higher %ALcons corresponds to lower CPPs.

D.1.2 Vocal Parameters Influenced by Gender against %ALcons

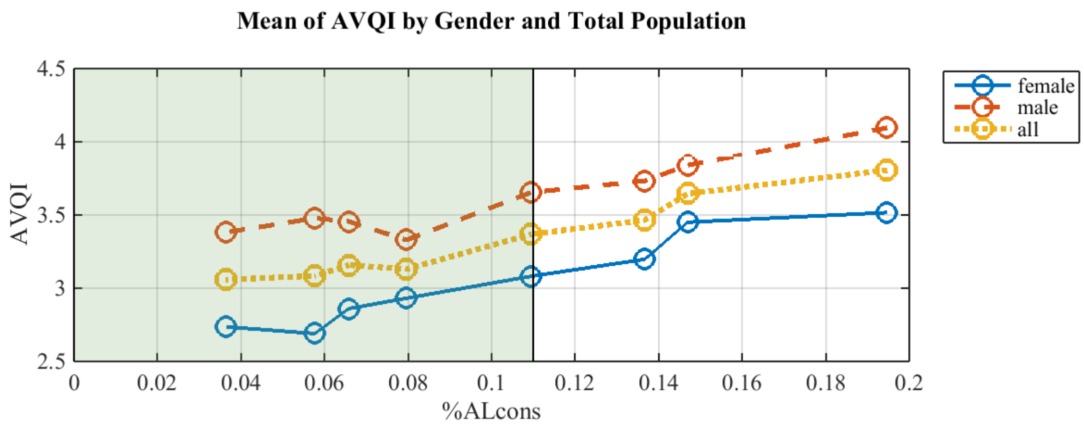

Figure D.4. Plot of AVQI means by gender against %ALcons. The shaded area designates the optimal %ALcons range for speech intelligibility. Higher %ALcons corresponds to higher AVQI, and females have a lower AVQI than males.

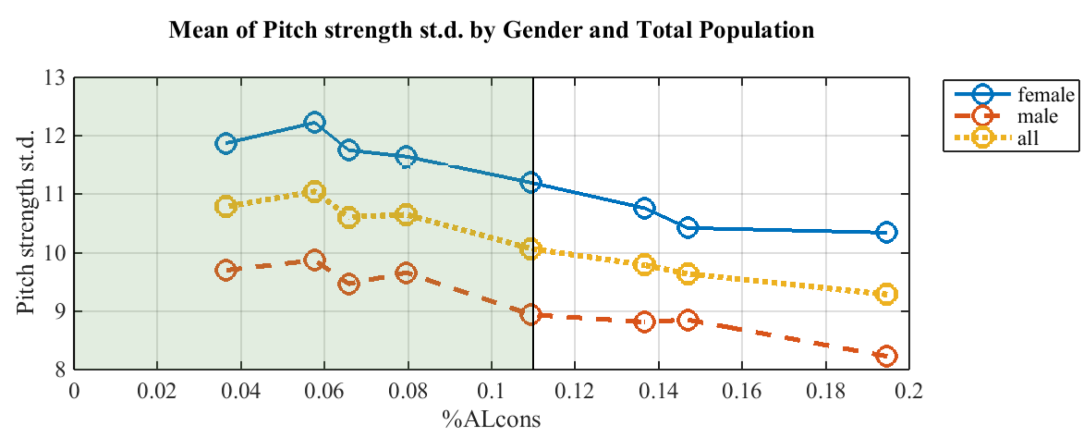

Figure D.5. Plot of pitch strength means by gender against %ALcons. The shaded area designates the optimal %ALcons range for speech intelligibility. Higher %ALcons corresponds to lower pitch strength, and females have higher pitch strength than males.

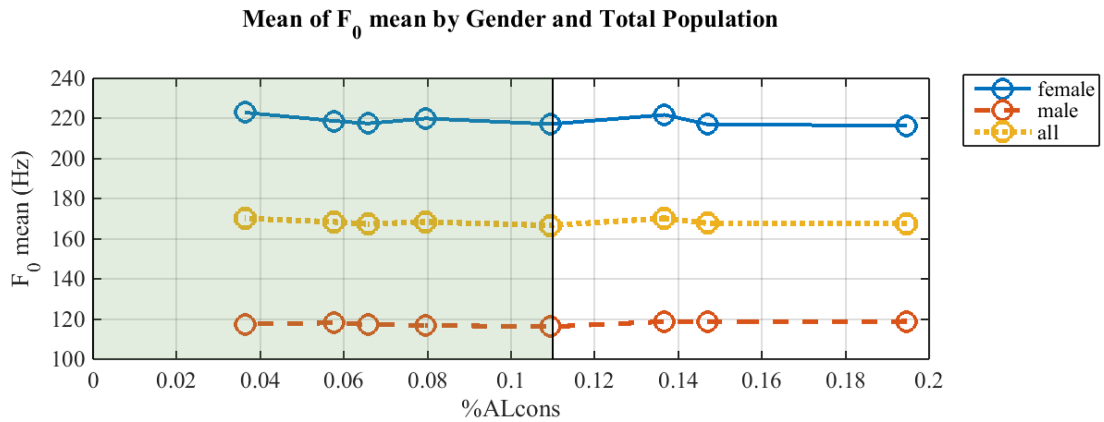

Figure D.6. Plot of mean F_0 means by gender against %ALcons. The shaded area designates the optimal %ALcons range for speech intelligibility. Females have a higher F_0 than males.

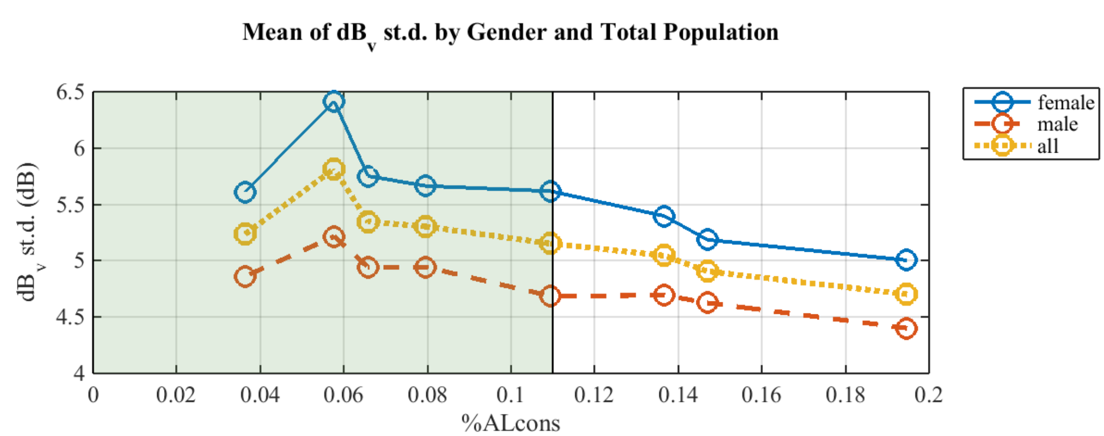

Figure D.7. Plot of voiced dB means by gender against %ALcons. The shaded area designates the optimal %ALcons range for speech intelligibility. Females have a higher voiced dB than males.

D.2 Plots of Vocal Parameters against C50

D.2.1 Vocal Parameters Influenced by C50 against C50

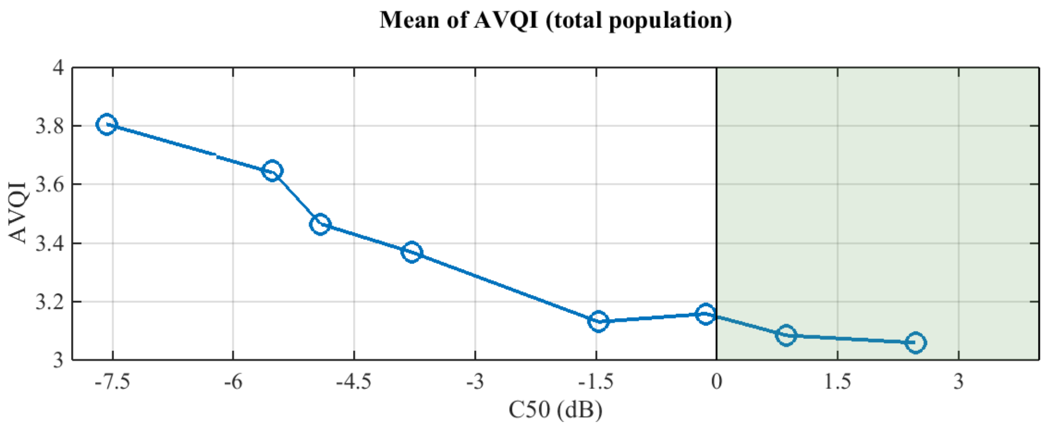

Figure D.8. Plot of total population mean AVQI against C50. The shaded area designates the optimal C50 range for speech intelligibility. Higher C50 corresponds to lower AVQI.

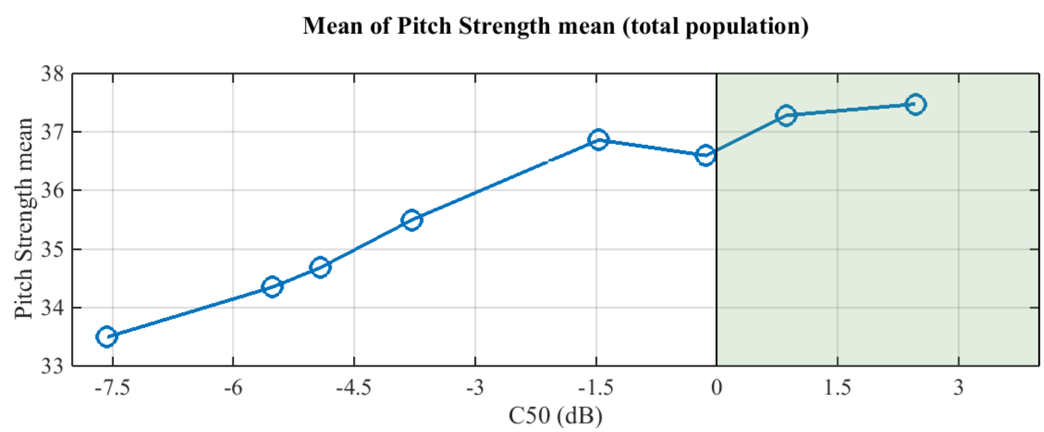

Figure D.9. Plot of total population mean pitch strength against C50. The shaded area designates the optimal C50 range for speech intelligibility. Higher C50 corresponds to higher pitch strength.

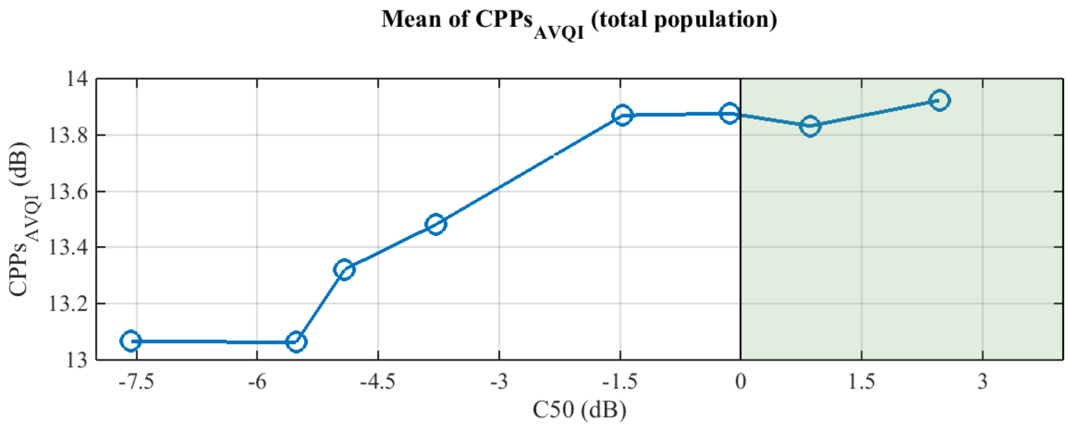

Figure D.10. Plot of total population mean CPPs against C50. The shaded area designates the optimal C50 range for speech intelligibility. Higher C50 corresponds to higher CPPs.

D.2.2 Vocal Parameters Influenced by Gender against C50

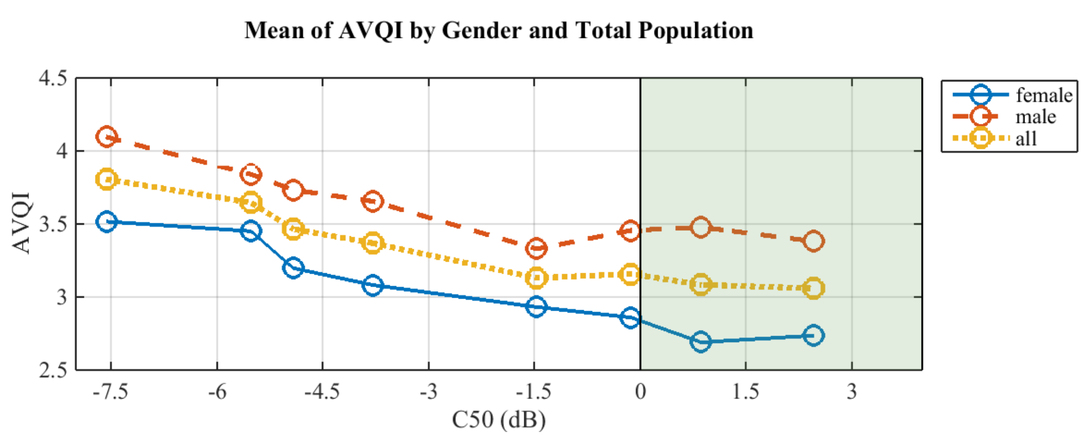

Figure D.11. Plot of AVQI means by gender against C50. The shaded area designates the optimal C50 range for speech intelligibility. Higher C50 corresponds to lower AVQI, and females have a lower AVQI than males.

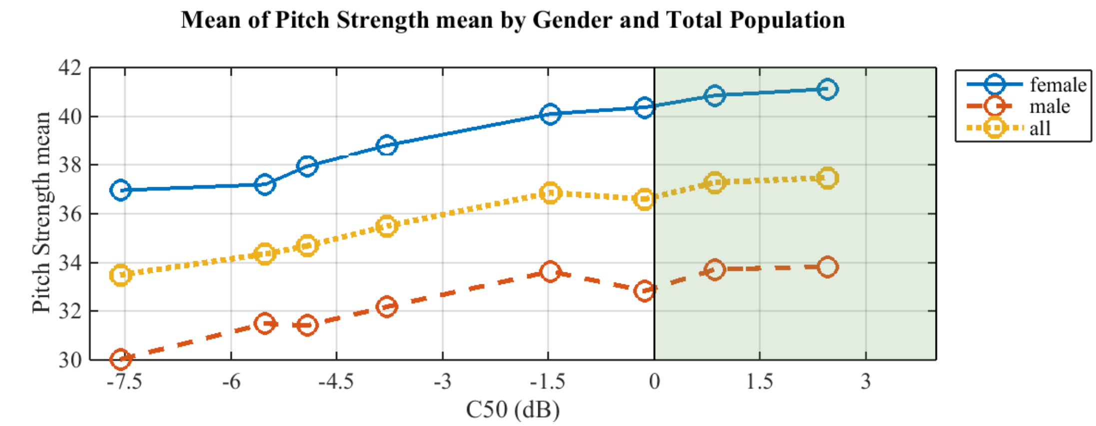

Figure D.12. Plot of pitch strength means by gender against C50. The shaded area designates the optimal C50 range for speech intelligibility. Higher C50 corresponds to higher pitch strength, and females have a higher pitch strength than males.

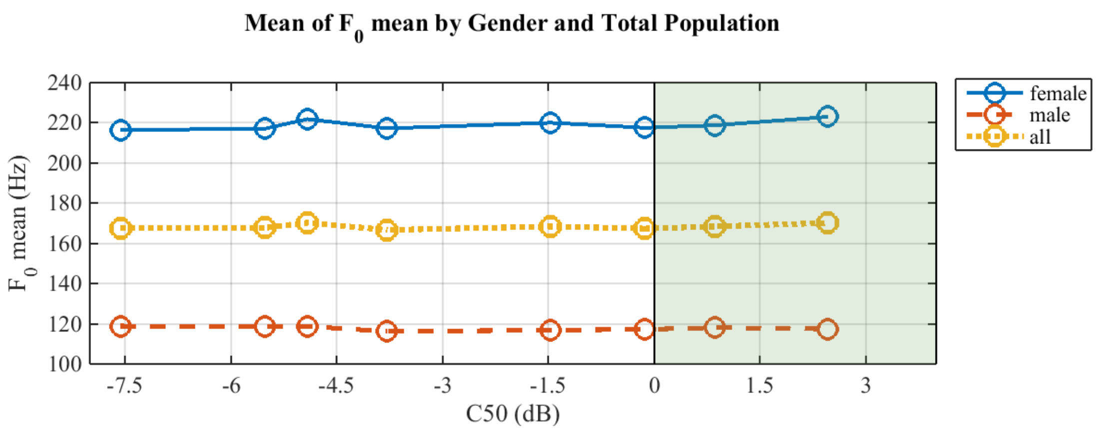

Figure D.13. Plot of F_0 means by gender against C50. The shaded area designates the optimal C50 range for speech intelligibility. Females have a higher F_0 than males.

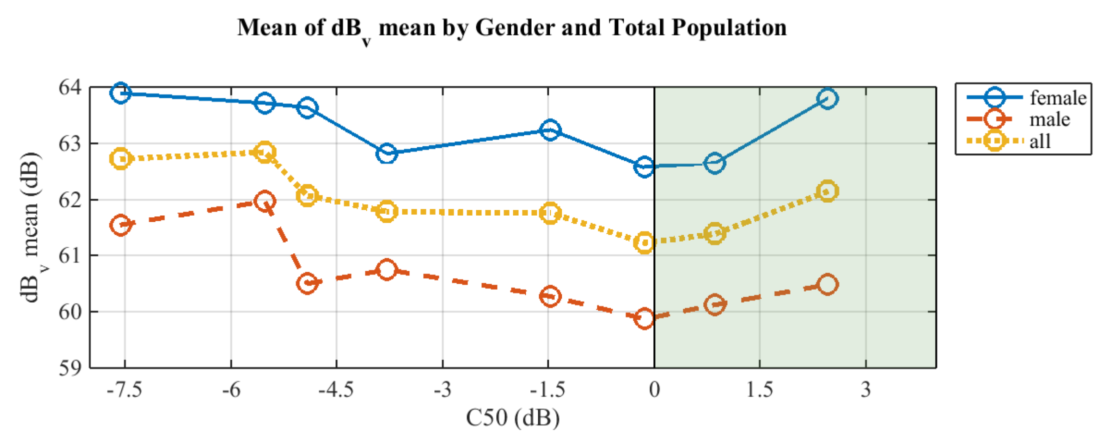

Figure D.14. Plot of voiced dB means by gender against C50. The shaded area designates the optimal C50 range for speech intelligibility. Females have a higher voiced dB than males.

D.3 Plots of Vocal Parameters against T20

D.3.1 Vocal Parameters Influenced by T20 against T20

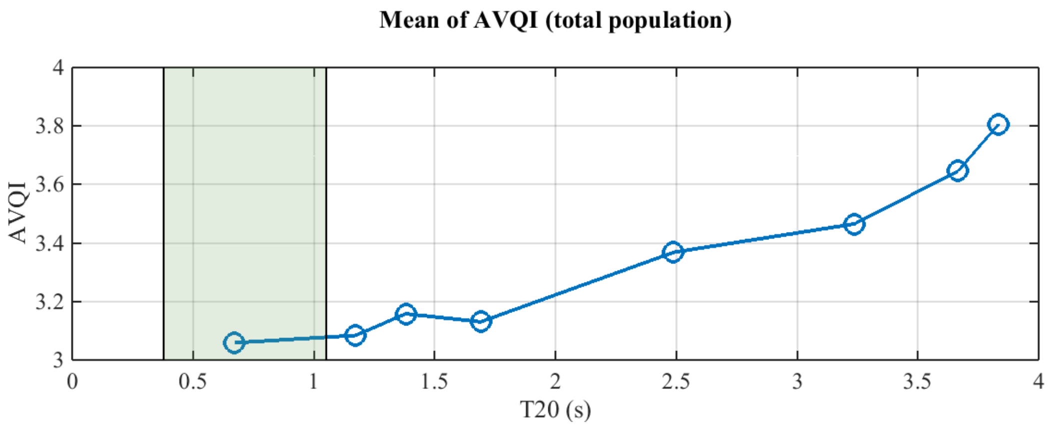

Figure D.15. Plot of total population mean AVQI against T20. The shaded area designates the optimal T20 range for speech intelligibility. Higher T20 corresponds to higher AVQI.

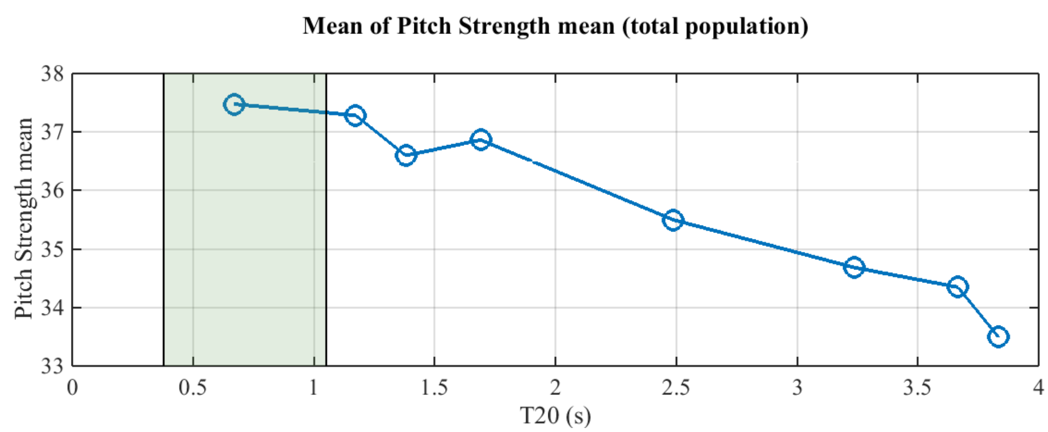

Figure D.16. Plot of total population mean pitch strength against T20. The shaded area designates the optimal T20 range for speech intelligibility. Higher T20 corresponds to lower pitch strength.

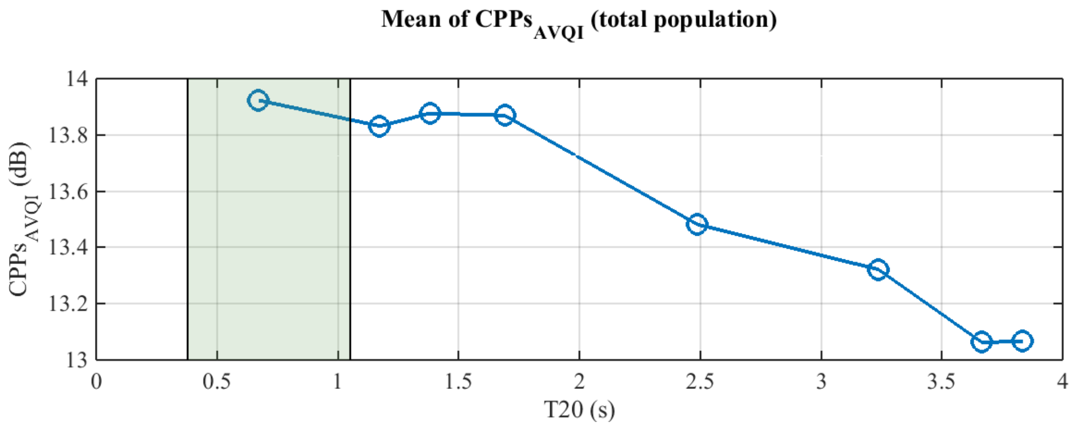

Figure D.17. Plot of total population mean CPPs against T20. The shaded area designates the optimal T20 range for speech intelligibility. Higher T20 corresponds to lower CPPs.

D.3.2 Vocal Parameters Influenced by Gender against T20

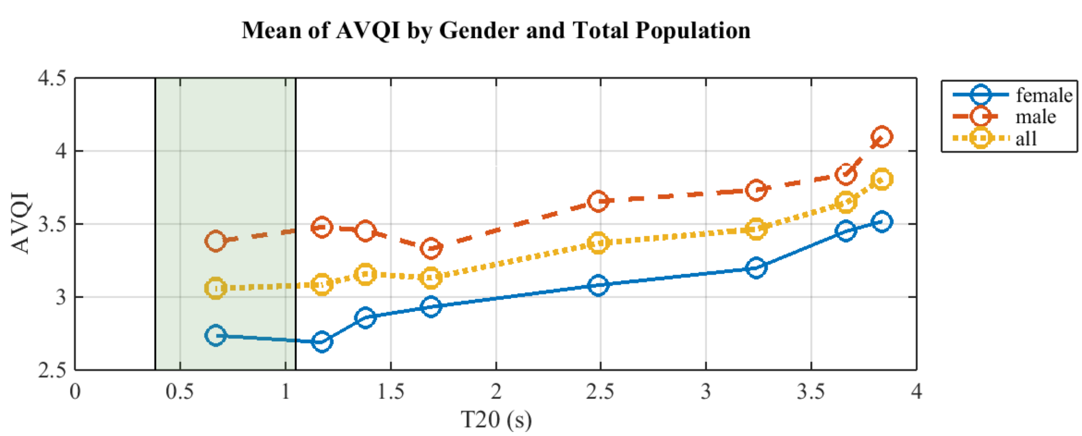

Figure D.18. Plot of AVQI means by gender against T20. The shaded area designates the optimal T20 range for speech intelligibility. Higher T20 corresponds to higher AVQI, and females have a lower AVQI than males.

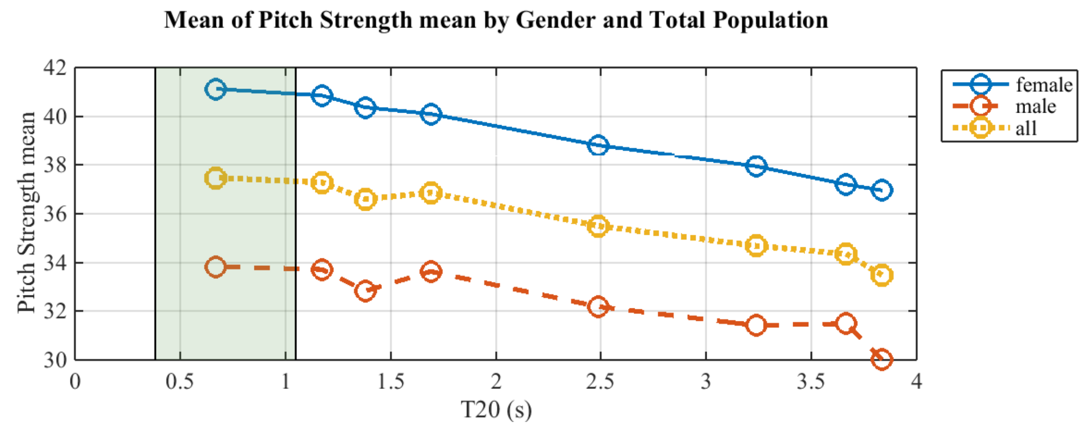

Figure D.19. Plot of pitch strength means by gender against T20. The shaded area designates the optimal T20 range for speech intelligibility. Higher T20 corresponds to lower pitch strength, and females have a higher pitch strength than males.

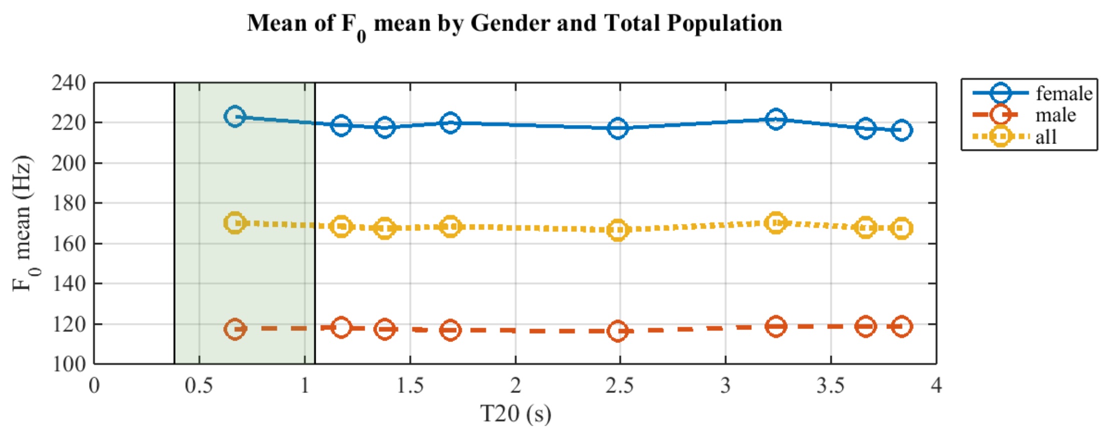

Figure D.20. Plot of F_0 means by gender against T20. The shaded area designates the optimal T20 range for speech intelligibility. Females have a higher F_0 than males.

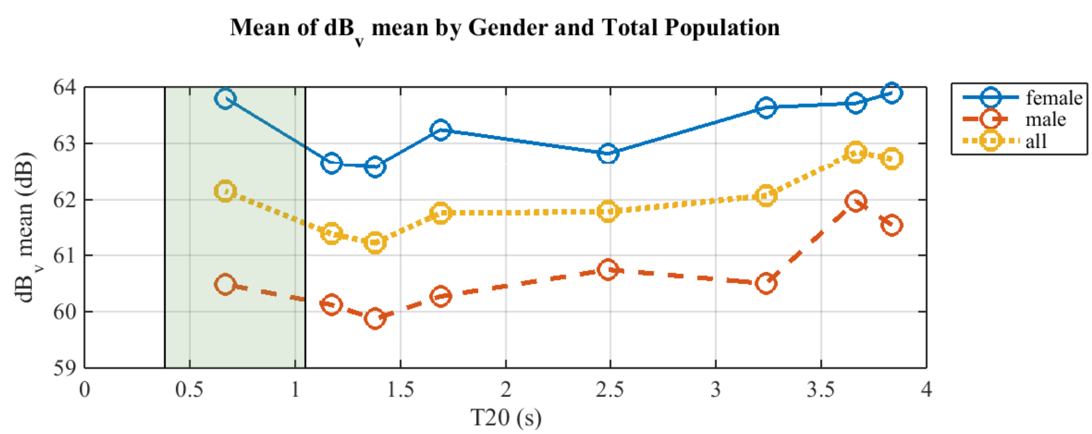

Figure D.21. Plot of voiced dB means by gender against T20. The shaded area designates the optimal T20 range for speech intelligibility. Females have a higher voiced dB than males.