Sponsoring Committee: Dr. Aaron M. Johnson, Chairperson
Dr. Agnieszka Roginska
Dr. Ronald C. Scherer
Dr. Ana Flavia Zuim

THE DEVELOPMENT OF THE OPERATIC VOICE DURING THE 20th CENTURY: AN ANALYSIS OF THE EFFECT OF EARLY RECORDING TECHNOLOGY

Joshua D. Glasner

Program in Voice Performance
Department of Music and Performing Arts Professions

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
Steinhardt School of Culture, Education, and Human Development
New York University
2019

ACKNOWLEDGEMENTS

Neither this dissertation, nor this degree would have been possible without the abundant support, mentorship, and encouragement that I have received from so many people throughout my professional and personal life. While I know that the completion of this project only marks the beginning of a career as a singing voice researcher, I want to take this moment to thank a few of the many individuals who have helped me to see this dream through to the end.

My dissertation chair, Dr. Aaron Johnson, for coming to the rescue when there was no one else at NYU to push this dissertation forward. Thank you for taking a chance on me, for pushing me to reach beyond the confines of what I thought I "knew," and for countless hours helping to turn this document into something about which I can be proud (especially through the flu in the last couple of weeks before submitting to the committee!).

Gerald Fabris, the Thomas Edison National Historical Park (West Orange, NJ), and Michael Devecka for donating their time and expertise to this project. Thank you, as well, to all of the wonderful singers who participated by lending their time and their voices. This project would not have been possible without your help.

The rest of my committee members: Dr. Agnieszka Roginska, Dr. Ronald Scherer, and Dr. Ana Zuim, for being there when I had questions that reached into areas with which I wasn't familiar. Thank you for your expertise, for the generous gift of your time, and for your encouragement throughout this process.

All of my past vocal pedagogy teachers, mentors, and colleagues who encouraged me to strive for excellence and to keep asking questions: Dr. Brian Gill, Dr. Kathy Price, Dr. Christopher Arneson, Prof. Margaret Cusack, Prof. Thomas Faracco, Dr. Noël Archambeault, Dr. Blake Smith, Dr. Nicholas Perna, and Dr. Ian Howell.   Thank you for helping me to grow and to mature as a teacher, as a student, as a colleague, and as a scholar.

Francine, Richard, and Tammy Black: there is no possible way any of this could have happened without you. Thank you for your kind, gracious hearts, and for doing more than what can be expected of any human beings. You have all of my respect and love.

Thank you to Kristen Ruiz and William Wesbrooks for fighting for me. I will forever be in your debt.

Thank you to my friends and family—too many to mention by name—for your support through the ups and downs of academic life throughout this entire process. I could not have done this without you.

Finally, thank you to Dr. Daniel and Elaine Glasner who have never ceased to find ways to convince me that I could accomplish this feat, and anything else I put my mind to. Thank you for always being there for me, even in spirit.

TABLE OF CONTENTS

continued

continued

LIST OF TABLES

continued

LIST OF FIGURES

continued

continued

continued

continued

continued

continued

continued

continued

continued

continued

continued

CHAPTER I

INTRODUCTION AND BACKGROUND

> "How interesting it will be to future generations to learn from the phonograph exactly how Rubinstein played a composition on the piano; and what a priceless possession it would have been to us, could we have Gen. Grant's memorable words, 'Let us have peace,' inscribed on the phonograph for perpetual reproduction in his own intonations!"
>
> --Thomas Alva Edison, "The Perfected Phonograph," 1888

1.1 Introduction

Those in vocal pedagogy and voice performance frequently discuss the decline of the art of singing. The debate often centers on topics such as interpretation (Crutchfield, 1983; Elliott, 2008; Kauffman, 1992) and vocal technique (Coffin, 2002; Crutchfield, 2012; Stark, 1999), but objective studies related to historical singing and the development of the operatic voice are scarce. As such, the fields of vocal pedagogy and vocal performance would benefit from an objective analytical method that clarifies and directs the ongoing conversation about the state and evolution of operatic singing. This dissertation is guided by a desire to understand how operatic singing has changed throughout the twentieth century. It is also fueled by a hope that a more objective method of analyzing historical voice recordings might improve the understanding of vocal pedagogy, modern evidence-based voice instruction, and historical performance practice.

This study attempts to connect the past to the present. In many ways, opera in the 21st century serves as a sort of museum piece; a way to view historical topics and to experience the emotions they evoke through an art form that developed over the course of four hundred years. Furthermore, performance practice and timbral aesthetic often are thought of as being static in the twenty-first century. Historical recordings, therefore, have the potential to give us many insights regarding historical performance practice, historical singing techniques, and even the way songs and operas were originally performed. In an essay concerning the nineteenth century singing voice, Crutchfield states:

> The purpose of studying performance's history, after all, is not to turn back the clock, but simply to see whether we might find something inspiring that could augment our work today, or could expand our understanding of the music we interpret. (Crutchfield, 2012)

Historical audio recordings give the modern voice teacher, performer, and historian the opportunity to hear and evaluate some of the original singers of oft-performed operas and art songs; however, early recording technology itself was limited in its potential to capture the entire spectrum of the voice of the operatic singer at the turn of the 20th century. Therefore, it is of great interest, and the objective of this study, to ascertain the effect that the earliest acoustic recording technology, namely, wax cylinders, had on the voice signal and to determine what aspects of the singer's voice was actually preserved on those historical recordings. It is important to note that it was not the aim nor the goal of this dissertation to recreate or in any way enhance historical recordings, but, instead, this study sought to understand how the wax cylinder phonograph system altered the sounds it recorded in the context of operatic singing.

1.2 Problem Statement

Most scholarly inquiries into the evolution of operatic singing rely on interpretations of historical texts and opinions about historical audio recordings. Inherently, these studies result in perspectives that are based on subjectivity. This dissertation investigates the spectral effect(s) of early audio recording technology on voice signals and uses those findings to provide an objective lens through which more nuanced analyses of historical singers, historical singing techniques, and performance practice can occur.

1.2.1 Statement of Sub-Problems

This dissertation aims to answer two research questions:

- To what extent and how does wax cylinder technology influence recorded voice output signal?[1]
- Can regression models be designed that estimate objective measures of historical voice quality? Such models would approximate the corrections of the distortion effects caused by historical audio recording technology.

This dissertation records and analyzes professional opera singers on two distinct recording technologies: a wax cylinder (digitized by an *Archéophone*[2]) and a flat-response microphone system. In order to understand how singing has evolved throughout the past century, it is important to determine how one of the earliest recording technologies—wax cylinders—altered or otherwise filtered the vocal signal. Historical recording technologies limited the recorded frequency range and sometimes attenuated the magnitude of certain frequencies—a topic that has been studied in more detail during the past few years (Feynberg, 2014). Furthermore, global and local degradations[3] attributed to historical audio recording technology (Välimäki et al., 2008) influence individual perception (Morange, Dubois, & Fontaine, 2010) and distort the results of modern voice

---

[1] Definition: A voice output signal is the digital or analog audio signal produced by a singer and captured by any recording technology.

[2] A transfer device that digitizes wax-cylinders (Chamoux, 2004).

[3] *Global degradations* refer to changes that affect the entire audio recordings (such as frequency modulations or filtering effects) and *local degradations* refer to changes that affect discrete points in the audio signal (such as clicks and low-frequency pulses) (Välimäki, González, Kimmelma, & Parviainen, 2008).

analysis methods. By studying the effects of wax cylinder technology on voice signals, it may be possible to analyze historical voices more objectively. It is a primary expectation of this study that this research outcome will allow for a more complete and accurate study of historical voices, vocal techniques, and performance practice.

The second research question uses regression analysis to determine how wax cylinder technology alters specific spectral and temporal measurements of the voice output signal. The values of the voice metrics used in this study have been shown to be correlated with phonation type (i.e. breathy, pressed, flow, or neutral phonation) and even glottal closure to some degree (Awan, 2015). Each voice metric value was analyzed as a one-to-one comparison, where the sole changing variable was the audio recording technology used to record an individual voice.

The acoustic measurements used in this study, such as the level difference between the first and second harmonics (L_1-L_2), harmonic-to-noise ratio (HNR), spectral moments (spectral mean, spectral standard deviation, spectral skewness, and spectral kurtosis), vary in their measurement values across the different phonation types, voice qualities, and perceived timbres. Therefore, through the use of regression analysis, it was a goal of this study to design linear models to use with historical audio recordings that estimate how wax cylinder technology alters the values of specific voice metrics that have been shown to be correlated with voice quality. By doing so, this study aimed to bypass perceptual analysis of singers—at least temporarily—by using objective measures to determine the way wax cylinder phonograph technology altered voice output signals. It is important to note that this study did not attempt to re-create or modify historical audio recordings of singers in order to hear early singing with the same fidelity of the original singing.

1.3 Hypothesis

This study tests the hypotheses that (1) past findings that wax cylinder technology produced inaccurate representation of sound production are valid, and that (2) the effects were primarily a limited frequency range and variable filtering of those sound productions, and, more specifically, singing. Furthermore, the study tests (3) the hypothesis that distortion of certain spectral characteristics (i.e. spectral moments and level differences between spectral peaks) of voice recordings is a main effect of the wax cylinder phonograph system. Additionally, this dissertation posits that (4) regression analysis of the acoustic characteristics of simultaneous digital and wax cylinder recordings of modern operatic singers can improve the accuracy of estimates of historical voice quality by indicating the extent to which those acoustic measurements are altered by historical recording technology in a controlled setting.

1.4 Background

1.4.1 Wax Cylinders

Different types of cylinder records were produced in the United States from 1877 to 1929 (Klinger, 2007). The original configuration was Thomas Edison's (1847-1931) tinfoil phonograph "which employed a thin sheet of metal foil, formed into a cylinder" (Klinger, 2007). Unfortunately, this early form of recording technology was prone to damage and few of these cylinders have been preserved.

In 1889, Edison began commercializing brown-wax cylinders after Charles Sumner Tainter (1854-1940), Chichester Bell (1848-1924), and Alexander Graham Bell (1847-1922) designed a new method and material for capturing sound (Osborne, 2012). These new records were produced by cutting a groove into a cylinder made of metallic-soap compounds (see Figure 1) which varied in color from light tan to dark brown (Klinger, 2007). In later years, brown-wax cylinders became used as the "master" records for mass-produced recordings. Early cylinders generally held no more than three minutes of audio and the first cylinders recorded two minutes. The literature shows that early wax-cylinder recordings were affected by "distorted spectral representation of the original sounds, low signal-to-noise ratio, large variations in record speeds, [which] all cause impairment of the interpretation of the recordings"

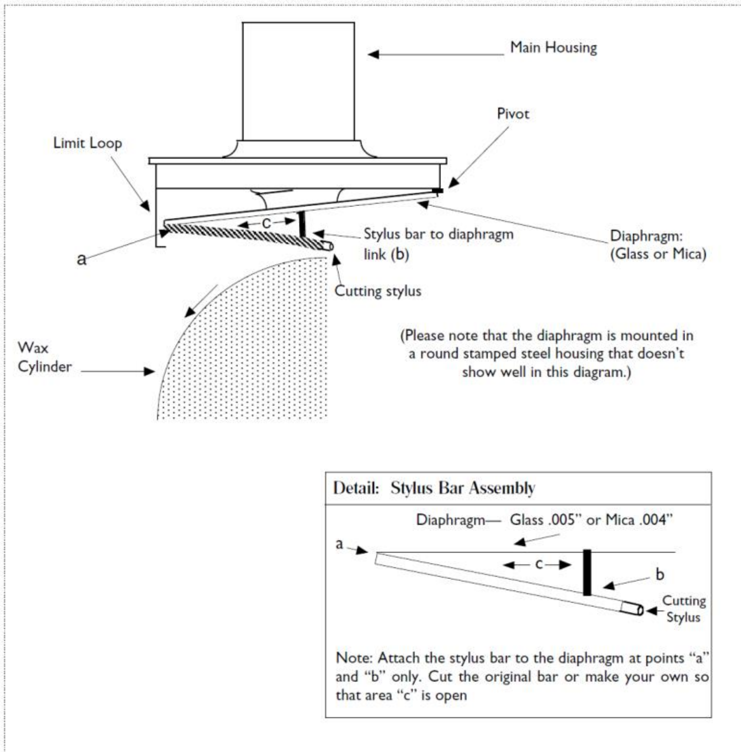


Figure 1. Side-view of an Edison Recorder (Diagram by Michael Devecka). Reprinted from "Modern Contact Transfer: Assessing Its Effect On Wax Cylinder Audio)." (p. 19), by R.Y. Feynberg, 2014, New York University, New York, NY: Unpublished Master's Thesis.

(Meulengracht-Madsen, 1976). Other studies (Owen & Fesler, 1983) discuss cylinder transfers, but seek to filter out local and global degradations—such as clicks and high frequency noise, respectively—from the original signal.

There is also evidence that sound recording techniques were proprietary in nature (Burkowitz, 1977; Feynberg, 2014). As such, it is not possible to know exactly how extant wax cylinder phonograph recordings were produced in the studio. There is some literature, for example, that suggests that professional singers were required to stand 3-4 inches (~7-10 cm) away from the phonograph

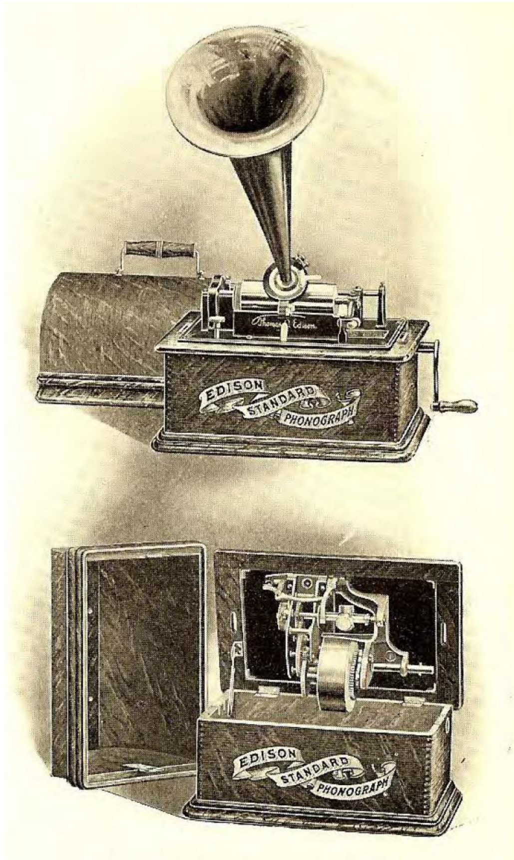


Figure 2. An Edison Standard Phonograph (identical to the one used for this study) from a 1903 National Phonograph Company catalogue (National Phonograph Company, 1903).

horn during recording sessions and, if their voice was "not strong enough," to step closer toward the horn during a high note (Torick, 1977, p. 881).

This historical note is interesting when sound pressure level (SPL) is taken into account. For example, Feynberg (2014) found that a 1 kHz sinewave at 90 dB visibly distorted the wax cylinder replica on an Edison Standard Phonograph. There is also evidence that historical singers were required to move further away from the horn if their singing was too loud—although the recording manual

Figure 3. Self-portrait of Enrico Caruso during a recording session (Public Domain attributed to RCA Records) (Torick, 1977).

mentions that it was more likely that the singer would be too quiet for the machine (National Phonograph Company, 1909, p. 10). All of these variables make it difficult to interpret historical recordings from the past, and necessitate a controlled study of this technology in the context of operatic singing.

Recently, the preservation of these cylinders has gained importance. In 2000, the United States Congress passed *The National Preservation Act of 2000 (Public Law 106-474).* This law led to the creation of *The National Recording Preservation Plan* (NRPP) in 2012 which, among other topics, established methods for transferring wax cylinders (Nelson-Strauss, Gevinson, Brylawski, & Loughney, 2012). Several studies have emerged in the last five years that address issues surrounding the transfer and playback of wax cylinders (Feynberg, 2014; Zakaria, 2016). These studies used a wax-cylinder transfer device—called the *Archéophone* (Chamoux, 2004) that uses a light-weight stylus to digitize both period and modern wax-cylinders. The aforementioned literature also suggests that society has a responsibility to transfer historical cylinders quickly and accurately so that the oldest audio recordings are not lost to time, and indicates that the *Archéophone* is a proper tool for such transfers.

The Department of Special Collections at University of California, Santa Barbara (over 10,000 historical cylinder transfers) has been at the forefront of historical recording transfer using this method; however, it is important to note that not all historical recordings that have been transferred to date have been transferred with the use of an *Archéophone*. For example, a relatively new transfer method has been developed by the Lawrence Berkeley National Laboratory and the Library of Congress, and has been put into use at the Northeast Document Conservation Center. IRENE uses optical-scanning technology which transfers audio signal without manually contacting the historical material. It was not the purpose of this study to compare historical recording transfers created by using these two transfer methods, but instead to analyze the way(s) in which historical recording technology alters voice output signals. In the future, the results of this study could be used to study historical recording transfers from both the *Archéophone* and IRENE, as both purport to deliver accurate representations of historical recording audio signal.

1.4.2 Spectral Analysis of Voices

The field of voice science has used spectrographic analysis for over seventy years (Fant, 1960) and, more recently, vocal pedagogy has sought to graft modern voice analysis techniques into its practice, both as biofeedback and as a way to study the singing voice objectively (D. G. Miller, 2008). This evolution in how the field looks at the singing voice has been brought about in no small part by the development of real-time spectral analysis applications, such as *Voce Vista* which was designed by D. Miller (2008).

While this study uses more robust applications and scripting languages such as *Praat* (Boersma & Weenink, 2018) and *Matlab* for quantitative spectral analysis and the creation of test signals, it relies on visual representations of the voices it studies by using both *Praat* and *Voce Vista Video Pro*. As such, it is necessary to introduce the type of information one can garner from such applications.

Formant tuning is a way to describe how voice resonances interact with harmonics to create peaks (local maxima) in a spectrum (D. G. Miller, 2008). In more conventional vernacular, formant tuning can be used in part to describe the style and/or vocal register a singer uses. A singer's method of formant tuning is notated by stating the vocal tract resonances and the harmonics they boost (Titze et al., 2015).

Vocal tract resonances are specific frequency regions that boost voice source harmonics when the source harmonics are within the passband of the resonances. In an acoustic spectrum, these will appear as local acoustic maxima. These local maxima have been defined as "formants," a word often used in place of the word "resonances." Many authors describe formant tuning methods that professional singers use to produce specific vocal styles (Bozeman, 2013;

Hallqvist, Lã, & Sundberg, 2017; McCoy, 2004; D. G. Miller, 2008; D. G. Miller & Schutte, 1990; Sundberg, 2015). Additionally, the literature describes a phenomenon called the *singer's formant cluster* (F_S)—a range of frequencies in the vicinity of 2.7-3.2 kHz for which the sound intensity level is boosted—that is produced by many operatic singers for stylistic quality requirements as well as for the benefit of those singers being more easily heard unamplified over an orchestra (Sundberg, 1987).

By observing the formant tuning method of a singer, it is possible to make inferences about a singer's technique—such as if he/she sings in head voice, chest voice, mixed voice, or belt (Bozeman, 2013; D. G. Miller, 2008; Sundberg, 2015). Of course, there are other techniques that can be used to augment the analysis of singing voice quality and function; however, these methods were beyond the scope of this dissertation and would have overly complicated the data collection process. As such, this dissertation takes into account previous research literature that has already studied adduction, phonation type, and aerodynamic measurements with spectral characteristics of the voice output signal.

Returning to the topic of the singer's formant cluster, it is possible that the source harmonics contained within the singer's formant cluster could be outside the captured frequency range of a wax cylinder recording depending on various factors (type of phonograph, number of playbacks, etc.). This characteristic would alter the listener's perception of the voice recorded onto a wax cylinder and subsequently played back. It is also possible that other filtering effects of wax cylinder phonograph technology might make traditional formant tuning analysis of historical recordings difficult or impossible.

VoceVista Video Pro and *Praat* are appropriate applications to analyze formant structure and to identify the presence of the singer's formant cluster. *VoceVista VP* can simultaneously display an audio waveform, a spectrogram, a long-term average spectrum, and a power spectrum ("spectral slice"). An older version of the application also allows the user to extract vibrato rate and extent. For the purposes of traditional formant tuning analysis, it is common to observe patterns in the power spectrum. Figure 4, for example, illustrates two singers who

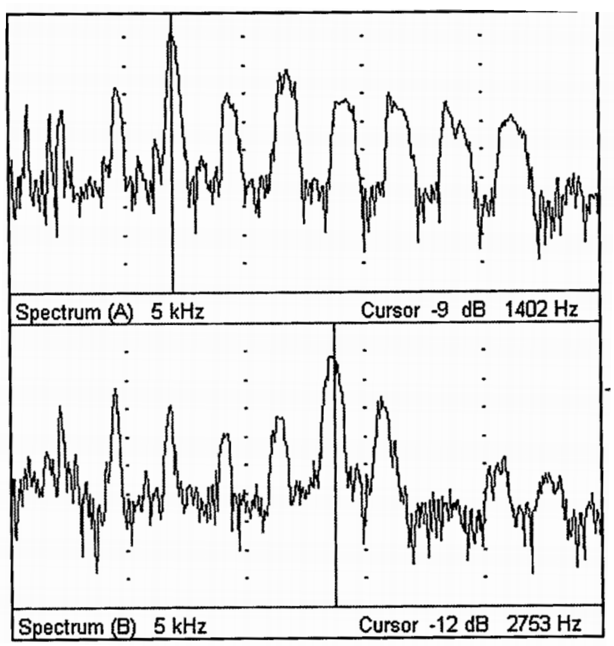


Figure 4. A power spectrum from the singing of Luciano Pavarotti (top trace) and Placido Domingo (bottom trace) singing a Bb4 (~466.16 Hz) from the aria "Celeste Aida." Reprinted with permission from *Resonance in Singing: Voice Building through Acoustic Feedback* (p. 2), by D.G. Miller, 2008, Princeton, NJ: Inside View Press. Copyright [2008] by Donald Gray Miller, and Inside View Press.

use distinct formant tuning methods. This graph represents frequency on the x-axis and sound intensity level (SIL) represented on the y-axis. By studying this power spectrum, it is possible to identify differences between singers.

The following examples are from the final B4-flat (~466.16 Hz) in "Celeste Aida" from Verdi's *Aida* and are sung by Luciano Pavarotti (1935-2007) and Placido Domingo (b. 1941), respectively. The power spectra represent a moment in time (window length = .250 s, bandwidth = 10 Hz) and demonstrate that the two tenors use markedly different resonance strategies when producing the same pitch in the same aria.

In order to analyze the formant tuning strategies used in Figure 4, it is necessary to observe the harmonic with the highest relative amplitude. Looking at the top spectrum, the harmonic with the highest relative amplitude is the third harmonic ($3f_o$). In comparison, looking at the bottom spectrum, the harmonic with the highest relative amplitude is the sixth harmonic ($6f_o$). Sometimes this type of analysis is obfuscated by inharmonic frequencies that are created by the voice source, by the accompaniment, or by background noise. In the case of the top spectrum, the audio recording sample includes inharmonic frequencies below 467 Hz (the approximate value of the fundamental frequency). Harmonic frequencies that are produced by the laryngeal source are integer multiples of the fundamental frequency (f_0) which are always equidistant on a power spectrum. Therefore, it is possible to determine that the harmonic of highest relative amplitude in the top spectrum is $3f_o$ and $6f_o$ in the bottom spectrum. It is also important to note that the source spectra may have been different for the two singers. That is to say, the two singers may have also been singing with a different degree of vocal fold closure and may have been using different aerodynamic and phonatory conditions.

The next step in the process of traditional formant tuning strategy analysis is to determine the vocal tract resonances that are used by the singer. As will be discussed further in Section 1.4.3, the vocal tract shape selectively filters specific frequencies. When an individual changes the configuration of his/her tongue, embouchure, or vertical laryngeal position, different laryngeal source harmonics are boosted. This action results in the perception of different vowels and timbres. Each vowel, therefore, is produced by differing ratios between the first and second vocal tract resonances.

When analyzing the formant tuning strategy of a singer, it is necessary to consider the vowel that he is producing while singing. The spectra represented above are both sung on a variation of the vowel [ɑ] at approximately 467 Hz. For male speakers, this vowel has an f_{R2} (second resonance) value of approximately 1090 Hz and an f_{R1} (first resonance) value of approximately 730 Hz (Ladofoged, 2005; Peterson & Barney, 1952; Titze, 2000). Therefore, it is unlikely that the first vocal tract resonance would be boosting the third harmonic and, instead, more likely that the second vocal tract resonance is boosting the third harmonic (notated as $f_{R2} \mid 3f_o$ tuning). In fact, it is probable that Pavarotti was singing a vowel closer to [æ] or [Λ] due to their approximate f_{R2} values (1720 Hz or 1190

Hz, respectively) for male speakers (Ladofoged, 2005; Peterson & Barney, 1952), which would be a reasonable technical expectation of an operatic tenor singing in the upper extreme of his range. A spectrum that shows the frequencies around 2700-3200 Hz having the highest relative amplitude can be described as using a resonance strategy that significantly boosts the relative amplitude of the *singer's formant cluster*. More traditionally in the field of vocal pedagogy, this resonance strategy is described as tuning to the singer's formant cluster (F_S tuning) (D. G. Miller, 2008).

When describing the two methods, one would say that Pavarotti uses $f_{R2}/3f_o$ tuning and Domingo uses F_S tuning. These two strategies result in perceptually distinct vocal outputs. The literature indicates that different vocal tract configurations and laryngeal sound sources are used when we see such different resonance strategies.

Sometimes conventional formant tuning analysis is not sufficient to describe the voice output signal. Recent study in music cognition and perception has suggested the possibility that a micro-analysis of the relationship between vocal tract resonances and source harmonics is not sufficient to describe the timbral complexity of the singing voice (Howell, 2016, p. 52). In these cases, more robust programs are needed in order to determine the values of specific quantitative measurements, which will be discussed in more detail in Section 1.4.3. These measurements compute the relative level of different frequencies and help to describe the shape of the spectrum. Applications and scripting languages such as *Praat* and *Matlab* allow for more precise spectral analysis.

1.4.3 Acoustic Correlates to Voice Quality

Voice and speech science has used source-filter theory to explain voice and speech production for nearly sixty years (Fant, 1960). While current research indicates that a simple linear model does not sufficiently explain all voice and speech phenomena (Titze, 2008), the basic concept that human vocal production can be described in two segments—a source and a filter—is still relevant.

The source-filter theory of voice and speech production states that the pulsatile air passing through the oscillating glottis creates a series of periodic pressure disturbances. The frequency of vocal fold vibration corresponds to the pitch that the listener perceives. Spectrum A of Figure 5 is a relatively typical spectrum of the laryngeal sound source. The glottal source signal excites the resonances of the vocal tract—which includes the laryngopharynx, oropharynx, and nasopharynx—and is filtered by those resonances that depend on the shape of those cavities. The filter can take many forms based on the configuration of the vocal tract. The shape and position of the tongue, the vertical position of the larynx, the mouth opening—or *embouchure*—and, in general, the length and shape of the vocal tract, all contribute to establishing the resonances of the vocal tract which characterize the filter (see spectrum B of Figure 5) affecting the source signal. These two components, in addition to radiation of the sound as it

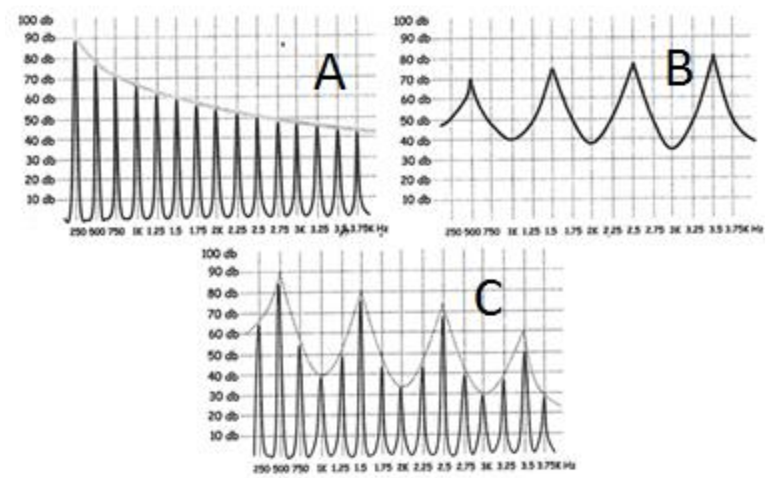


Figure 5. Glottal source spectrum (A), vocal tract filter representation (B), voice signal spectrum (C). Note that the radiation characteristic is not included in this graphical representation. Adapted with permission from *Your Voice: An Inside View* (p. 39), by S. McCoy, 2004, Princeton, NJ: Inside View Press. Copyright [2004] by Scott McCoy and Inside View Press.

moves away from the lips, combine to create the output sound produced by the singer. Spectrum C in Figure 5 shows the combination of the source and vocal tract filter as power spectra.

This discussion leads to a brief overview of the spectral measurements used for this study, which include: the level difference between the first and second spectral (harmonic) peaks (L_1-L_2), harmonic-to-noise ratio (HNR), and all four spectral moments (spectral mean, spectral standard deviation, spectral skewness, and spectral kurtosis). L_1-L_2 is the difference between the intensity level of the fundamental frequency (first harmonic) and the intensity level of the second harmonic, and has been used in the past as an acoustic correlate to breathy phonation (Hanson & Chuang, 1999; Klatt & Klatt, 1990), to style or genre of singing (Björkner, 2006), and to the closed quotient (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995; Sundberg, Andersson, & Hultqvist, 1999). That is to say, higher L_1-L_2 values are associated with breathy phonation, lower closed quotient values, the *falsetto* register, decreased sub-glottal pressure, and, overall, a more dominant fundamental frequency. L_1-L_2 has also been found to be lower in musical theatre singing compared to operatic singing (Björkner, 2006). Due to the relationship between vocal tract resonances and source harmonics, it is also expected that the measurement would increase as pitch ascends—particularly for female singers, as the second harmonic increases to come closer to the first resonance frequency from below the first resonance frequency. This dissertation will refer to the measurement as L_1-L_2 in keeping with more recent consensus in voice research nomenclature (Titze et al., 2015)

Harmonic-to-noise ratio (HNR) is commonly used as an indicator of the periodicity of a voice output signal. It has also been described as the amount of additive noise in a voice signal (Awan & Frenkel, 1994; Ferrand, 2002). More recently, HNR has been used as part of a regression model for and has been shown to be highly correlated with perceived voice quality (Maryn, Roy, Corthals, Van Cauwenberge, & De Bodt, 2009). For the purposes of comparison between voiced and non-voiced signals, it has also been shown to be reasonably correlated to the signal-to-noise ratio (SNR) of an audio sample (Boersma, 1993).

While the field of singing voice research has yet to fully explore the use of spectral moments of the Long-term Average Spectrum (LTAS) in the analysis of singers, recent research has suggested that there may be interesting and relevant relationships between the shape of the voice output spectrum and the voice source (Awan, 2015). While further study of the use of spectral moments of the LTAS to study the singing voice is necessary, there is an increasing foundation within the fields of linguistics and voice and speech science that points to spectral mean, spectral SD, spectral skewness, and spectral kurtosis as adequate descriptors of the shape of spectra (see Figure 6) (Anand, Kopf, Shrivastav, & Eddins, 2018; Forrest, Weismer, Milenkovic, & Dougall, 1988; Harwardt, 2011; Kardach et al., 2002; Lowell, Colton, Kelley, & Hahn, 2011; Tanner, Roy, Ash, & Buder, 2005) and, as such, of characteristics of the sound output related to timbre and voice

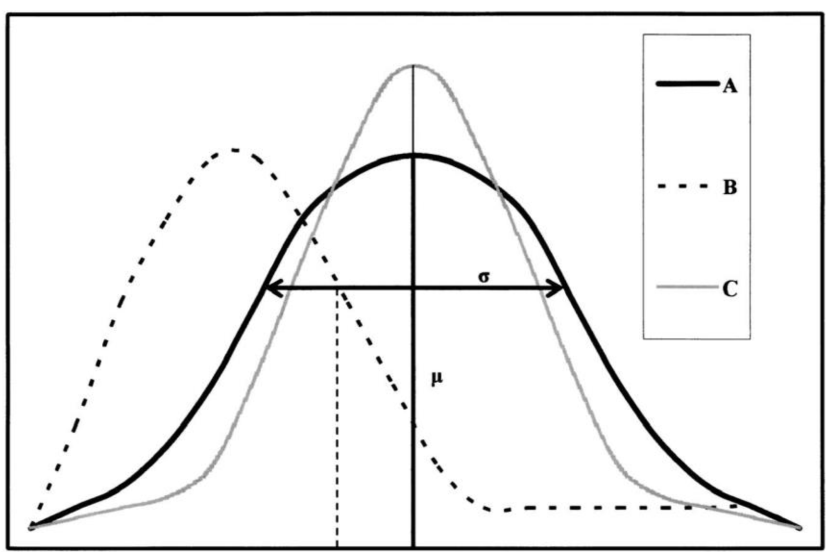


Figure 6. From Tanner, et al. (2005), "Mean (μ), standard deviation (σ), skewness, and kurtosis of a probability distribution. 'A' illustrates a normal distribution with corresponding mean (solid vertical line) and standard deviation (horizontal arrowed lines). 'B' illustrates a positively skewed distribution and corresponding shift in mean (dotted vertical line). 'C' illustrates a peaked distribution with a corresponding positive kurtosis (note that distributions A and C have the same mean)." Reprinted with permission from "Spectral Moments of the Long-term Average Spectrum: Sensitive Indices of Voice Change After Therapy?" by K. Tanner, N. Roy, A. Ash, & E.H. Buder, 2005, *Journal of Voice*, 19(2), p.214. Copyright [2005] by The Voice Foundation with permission from Elsevier.

quality.

Spectral mean (M1) indicates the central energy concentration of a spectrum (Kardach et al., 2002). Its value is found in *Praat* using the following equation[4]:

$$\frac{\int_0^{\infty} f|S(f)|^p df}{\int_0^{\infty} |S(f)|^p df} \tag{1}$$

where the numerator represents the centre of gravity and the denominator represents the weighted spectral energy. The resultant value is a weighted centre of gravity over the entire frequency domain (Boersma & Weenink, 2018). This measurement has been used in past literature to characterize phonemes (Forrest et al., 1988; Kardach et al., 2002) and effort level while speaking (Harwardt, 2011). One study using singers did not find a significant difference based on closed quotient (CQ) or electroglottographic (EGG) output measures for spectral mean, but the authors noted that a "concave down/knee" EGG output waveform shaping—commonly associated with more contact of the vocal folds (D. G. Miller, 2008)—seemed to correspond with lower spectral mean values (Awan, 2015). There is also recent literature that suggests that spectral centroid—a

[4] See *Praat* documentation ("Spectrum: Get centre of gravity…") for a succinct description of the variables in this equation (Boersma & Weenink, 2018).

similar measure to spectral mean—can be used in part to describe the complexities of timbre in the singing voice (Erickson, 2004; Howell, 2016).

Spectral standard deviation (M2) is a measure that describes the dispersion of spectral energy (Kardach et al., 2002). More simply put, a high spectral SD indicates a more spread out spectrum around the spectral mean, and a low spectral SD indicates that spectral energy is more concentrated around the spectral mean. Past research related to the singing voice has suggested that spectral SD could be related to the EGG output waveform shaping and, more specifically, that a "compact" spectrum in the lower frequency region occurs when CQ increases (Awan, 2015). In other words, spectral SD has been found to be inversely related with CQ. Spectral SD has also been shown to increase with loudness in speakers (Anand et al., 2018; Harwardt, 2011) due to an increased amount of energy in higher frequencies. While these two findings may seem to be contradictory, Awan, et al. (2015) noted that the relationship between spectral moment measurements and CQ accounted for a low proportion of the variance; referencing Alipour, et al. (2012) who reported that there is a complex relationship between increases in glottal adduction and the spectral energy of the first few harmonics. As such, spectral SD measurements should be interpreted with caution. The equation for the spectral SD used in *Praat* is found below[5]:

[5] See *Praat* documentation ("Spectrum: Get central moment…" and "Spectrum: Get standard deviation) for a succinct description of the variables in this equation (Boersma & Weenink, 2018).

$$\sqrt{\frac{\int_0^{\infty}(f-f_c)^2|S(f)|^p df}{\int_0^{\infty}|S(f)|^p df}} \quad (2)$$

where the numerator represents the second central moment of the spectrum and the denominator represents the spectral energy in the spectrum over the entire frequency domain.

Spectral skewness (M3) is a measure that describes the average symmetry of the spectral distribution (Kardach et al., 2002). More specifically, a positive spectral skew value indicates a higher concentration of spectral energy in lower frequency regions, whereas a negative skew value indicates that a high concentration of spectral energy resides in higher frequency regions. It has been shown to decrease with loudness increase in speakers (Anand et al., 2018; Harwardt, 2011) and, generally, to be lower—that is, have a flatter spectral slope—for obstruent consonants than for sonorant consonants (Harwardt, 2011). No relationship between EGG profile and spectral skewness has been found, but spectral skewness has been found to be generally higher for female speakers and singers (Awan, 2015). The equation[6] for the (normalized) spectral skewness as found in the *Praat* documentation can be seen in Equation 3:

[6] See *Praat* documentation ("Spectrum: Get skewness…") for a succinct description of the variables in this equation (Boersma & Weenink, 2018).

$$\frac{\int_0^\infty (f-f_c)^3 |S(f)|^p df}{\int_0^\infty |S(f)|^p df} \Bigg/ \left(\frac{\int_0^\infty (f-f_c)^2 |S(f)|^p df}{\int_0^\infty |S(f)|^p df}\right)^{3/2} \tag{3}$$

where the numerator represents the third central moment of the spectrum, and the denominator represents the 1.5 power of the second central moment of the spectrum.

The spectral slope (also called spectral tilt) of the glottal source signal has been shown to be indicative of voice quality and phonation type (see Figure 7)

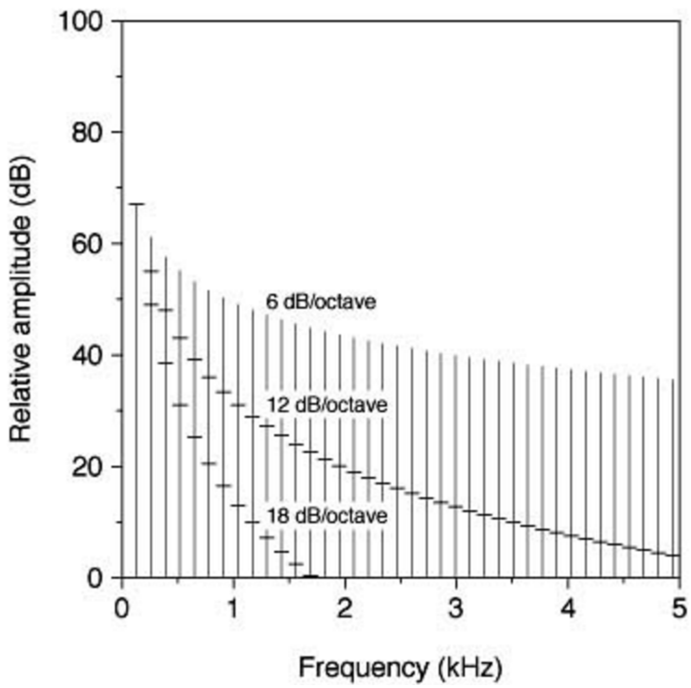


Figure 7. The horizontal bars indicate differing relative amplitude. This spectrum of three different glottal source signals represents “brassy,” “normal,” and “flute-like” vocal timbres (from top to bottom). Reprinted with permission from *Principles of Voice Production* (p. 130), by I. Titze, 2000, Iowa City, IA: National Center for Voice and Speech. Copyright [2000].

(Titze, 2000). This measure is similar, but not identical to spectral skewness. Different measures that compare the proportion of high frequency components of a spectrum with its low frequency components have been used to study the voice in the past, both with the glottal source spectrum and the radiated voice output spectrum (Cesari, Iengo, & Apisa, 2013; Duvvuru & Erickson, 2013; Hillenbrand, Cleveland, & Erickson, 1994; Sundberg & Nordenberg, 2006). This concept is perhaps the simplest example of a primary premise of the analytical method used in this dissertation: spectral measurements can be and have been used to estimate voice quality—a multifaceted description of resonance strategy, timbre, and glottal adduction—and can be used to make inferences about the singing voice as recorded with different technologies.

Spectral kurtosis (M4) is a measure of the definition of spectral peaks (Kardach et al., 2002, p. 535). The measure is used to indicate how much the shape of the spectrum around the *centre of gravity* is different from a Gaussian shape (i.e. a bell-curve) (Boersma & Weenink, 2018). More practically, it has been shown to be positively correlated with CQ (Awan, 2015, p. 525). That is to say, a higher CQ may lead to more concentrated peaks in the spectrum and, as such, may lead to a higher spectral kurtosis measurement. Conversely, it has been shown to negatively correlate with loudness; with spectral kurtosis being lower during loud speech samples—an indicator that the spectrum is flatter (Harwardt, 2011). The equation[7] used in *Praat* to determine the (normalized) kurtosis of a spectrum is found below:

$$\left(\frac{\int_0^\infty (f-f_c)^4 |S(f)|^p df}{\int_0^\infty |S(f)|^p df} \Bigg/ \left(\frac{\int_0^\infty (f-f_c)^2 |S(f)|^p df}{\int_0^\infty |S(f)|^p df} \right)^2 \right) - 3 \qquad (4)$$

where the numerator represents the fourth central moment of the spectrum, and the denominator represents the square of the second central moment of the spectrum.

1.4.4 Impulse Responses

An impulse response allows one to measure the filtering effect of a system on a source signal. It uses an impulse signal to determine the relative sound intensity of a range of frequencies and can be used to describe room acoustics, the frequency response of a filter, and the characteristics of the vocal tract. This method is only applicable to linear and time-invariant systems and can be measured with a variety of excitation signals such as Maximum Length Sequence (similar to pure white noise), Inverse Repeated Sequence, Time-Stretched Pulses,

[7] See *Praat* documentation ("Spectrum: Get kurtosis…") for a succinct description of the variables in this equation (Boersma & Weenink, 2018).

Golay codes, and Sinesweeps (Boren & Roginska, 2011; Farina, 2007; Stan, Embrechts, & Archambeau, 2002).

The result of these different impulse responses is a frequency response not unlike Figure 5B. Feynberg (2014) developed a method that used an airborne sinesweep (among other IR's) to describe the frequency response of wax cylinders. The series of test signals used for this study were adapted from Feynberg's work and is discussed in more detail in Section 3.4.1.

1.5 Need for Study

Historical operatic recordings are used for various purposes in academia, voice studios, and elsewhere. In academia, one of the uses of these recordings is to describe performance practice. For example, musicologist and conductor Crutchfield (2012) used historical recordings to transcribe recorded fragments of arias and pointed to those transcriptions as evidence that nineteenth century opera singers modified arias to fit their own individual voice.

Others such as Kauffman (1992) use recordings to point to changes in performance practice or interpretation. Many of these authors rely on their own perceptual analysis of recordings to make statements about the style used by singers recorded at the advent of recording technology:

> Listening to the recordings of our sample arias, it is immediately apparent that singers not only performed *portamenti* both with and without anticipations, they also sang any number of subtle gradations between the two extremes (Kauffman, 1992).

While this kind of analysis is a helpful addition to understanding historical performance practice, the perception of the singing voice is limited by audio recording technology (Morange et al., 2010). Therefore, statements that deviate from this model—such as commenting on timbral aspects of historical singing—may risk inaccuracy due to spectral distortions caused by antiquated technology. Once the fields of singing voice research and vocal pedagogy understand the effects of early audio recording technology on voice signals, it should possible to conduct perceptual studies using antiqued recordings and then make more definitive headway toward understanding historical vocal style and performance practice.

Voice teachers also use recordings of historical singers to supplement their students' vocal education and development. Anecdotally, students are told to listen to historical recordings in order to hear specific singers' technical facility or their interpretive choices. What these students hear, however, may be inherently limited—whether they listen to historical recordings or modern recordings—because, as has been mentioned in this document, historical recording technology limits the amount of signal captured and modern sound engineering practices modify voice signals to fit commercial standards. These aspiring professional singers listen to recordings of professional singers and then attempt to mimic technical and stylistic practices. However, an inability to hear a true representation of the professional singer's voice may prevent the listener-student from attaining this goal. Studying the effect of wax cylinder technology on voiced signals serves as a way to inform voice teachers about how to use historical recordings in their pedagogy.

Although voice teachers and scholars such as Coffin (2002) and Kauffman (1992) posit that singing has declined since the nineteenth century, there are some who suggest that the "Decline of the Art of Singing" is merely a societal tendency to look back towards "an imaginary golden age" (Stark, 1999). Stark notes a number of pedagogues and critics who view the "golden age of singing" as having occurred in different time periods—for example, "the era of Caruso" or even "the era of the *castrati*."

Throughout the past three hundred years, singers, listeners, and voice teachers have lamented developments and stylistic changes as potentially damaging to the human voice as well as to the art of singing. Tosi complained that there were no longer great singers—despite being surrounded by the great castrati of the eighteenth-century (Stark, 1999). Similarly, Rossini remarked that Gilbert Duprez sounded "like the squalk of a capon whose throat is being cut" upon hearing the first "fully-chested high-C" (Stark, 1999). Even the use of vibrato was frequently looked down upon, despite evidence that vibrato is created organically in modern operatic singing due to "an undulation of the subglottal pressure…[and] of the adductive force in the larynx," (Sundberg, 1994) or periodic neurologic oscillations (Titze, 2000).

Past studies also support the notion that historical audio recording technology significantly altered recorded voices and, therefore, influences our perception of singers. Schutte, Miller, & Duijnstee (2005) studied operatic tenor formant tuning methods in the final sustained B4-flat of the aria “Celeste Aida” from Verdi’s *Aida*. The relevance of that project from the perspective of this dissertation proposal is not the authors’ conclusion about tenor resonance strategies, but, rather, the fact that they deemed it necessary to qualify the results of their study due to their use of commercial recordings. They stated:

> The validity of what is asserted here depends to a large extent on the degree to which these commercial releases give true acoustic data of the performances recorded. This is a complex issue with at least two major parts: **the characteristics of the various historical recordings systems**[8], from the early acoustic methods up through current digital techniques, and the matter of enhancements contributed in the processing of the original recordings. (Schutte et al., 2005)

The authors later indicate that their oldest samples have noticeable differences in spectral tilt (Schutte et al., 2005). This observation is important, because the two samples that the authors cite (Figure 8) are sung by the same singer (Enrico Caruso, 1873-1921). Due to the relatively short time period between the recordings, it is unlikely that the singer’s voice would have changed so drastically. Unfortunately, it is not yet possible to infer the reason behind these spectral differences, again suggesting the (technical) need for the current project.

Inquiries such as the one presented in this dissertation continue a pattern of bridging the gap between science and art that is ubiquitous throughout society in seemingly unrelated fields, such as sports, business, and the visual arts.

---

[8] This dissertation author’s emphasis added.

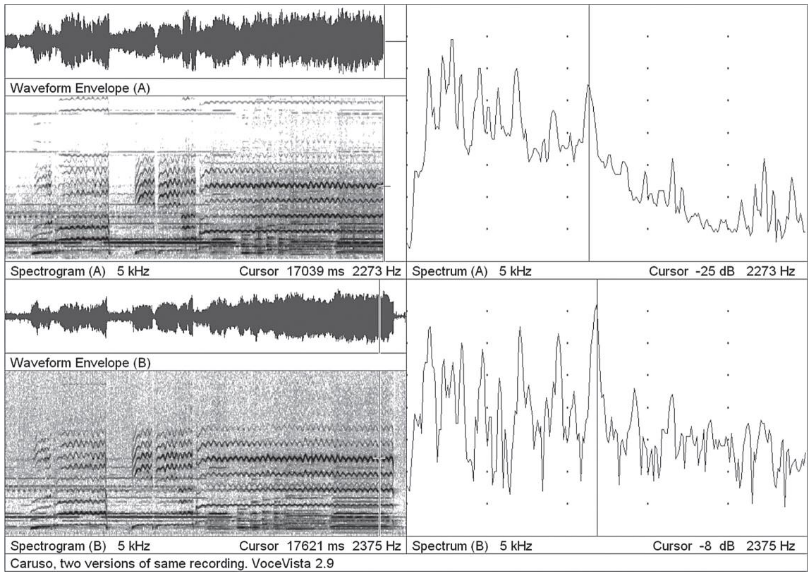


Figure 8. Two recordings of Enrico Caruso singing the final B4-Flat of "Celeste Aida." The sixth harmonic (seen in the power spectrum on the right) is approximately 10 dB higher in the second recording and has the highest measurement of magnitude when compared with all other harmonics in the spectrum. The different relative magnitude results in a different spectral slope. Reprinted with permission from "Resonance Strategies Revealed in Recorded Tenor High Notes." by H.K. Schutte, D.G. Miller, & M. Duijnstee, 2005, *Folia Phoniatrica et Logopaedica*, 57(5-6), p.305. Copyright [2005] by S. Karger AG, Basel.

Society as a whole continually attempts to connect the present with the past. As such, this study opens the door to the further contextualization of historical singing.

It is the intent of this study for the various voice professions to be able to use the resultant data to understand historical operatic singing more objectively. By doing so, a better understanding of historical vocal function and historical audio recording technology can be used to observe trends in technique and performance practice, as well as to make inferences about pedagogical trends (singing training).

1.6 Significance

This study has broad-reaching impact, but it is important to note that it is *foundational*. That is to say, its significance will be measured over time relative to how historical audio recordings will be analyzed and listened to in the lab and in the voice studio. Voice and speech science offers ways to quantitatively analyze voice and speech sounds and, furthermore, grants researchers tools to look back at the past from a more objective perspective. There is limited quantitative analysis of historical recordings of operatic singers and, due to the limitations of early recording technology, it is not possible to recreate the sound of great singers such as Enrico Caruso or Lilli Lehmann (1848-1929). However, this study uses advances in technical science that allow the estimation of the voice quality and other timbral characteristics of a historical singer with modern recording technology. The third chapter of this dissertation outlines a method to combine this knowledge from voice and speech research with common digital signal processing techniques in a novel way. It is the hope of this study that doing so will benefit a variety of fields such as vocal pedagogy, voice and speech science, musicology, music technology, and audio preservation.

CHAPTER II

RELATED LITERATURE

2.1 Introduction

Past research about the development of operatic singing is limited in its scope. The most rigorous of this literature, to be discussed below, investigates historical treatises, critiques of historical singers, and even looks to historical recordings as evidence of performance practice (Crutchfield, 2012; Crutchfield, 1983; Steane, 1993; Steane, 1996; Steane, 2003a; Steane, 2003b). Other literature relies on subjective analysis of historical recordings to determine how singing techniques and singers have changed over time.

The study of the development of the operatic voice can be categorized into four groups: performance practice as documented in audio recordings and treatises; the vocal demands of roles and changes in compositional style; pedagogical and technical practices as determined through the study of historical treatises; and the analysis of singers' voices. Furthermore, the latter group can be aptly described with three sub-categories: temporal measurements (such as vibrato), spectral measurements (such as spectral tilt), and perceptual analyses (that use listeners to describe or characterize the singing of performers). This chapter discusses literature that focuses on each of the four categories for studying the development of the operatic voice. It aims to situate this dissertation topic in a historical context and to demonstrate the state of the literature that studies historical voice recordings.

2.2 The Development of the Operatic Voice

Related studies can be grouped into three different categories: those that aim to track the development of specific voice types; those that utilize musicological methods to study changes in performance practice; and those that use acoustic analysis methods. Morrison's dissertation (2013) belongs in the first category. It compares historical vocal pedagogy treatises and analyzes historical recordings of soprano voices. The research design follows three steps:

- List information about the year of recording, recording label, recording method, and sound quality.
- List the *Fach* of the singer and the [perceived] treatment of the lower part of the singer's range.
- Detail biographical information about the singer (voice teacher and "synopsis of her career, popularity, and success") (Morrison, 2013).

While Morrison highlights an important topic, the dissertation's discussion of historical recordings does not contain any objective analysis and its technical discussion focuses on a concept that is still debated and studied in the singing voice research community: vocal registers. It concludes that sopranos who predate 1940 use a different type of chest-voice compared to modern sopranos who sing with mixed voice below F^4 (Morrison, 2013), but this assertion was determined solely by the author's perception. Future study could test this conclusion more objectively by comparing the values of acoustic measurements between sopranos who predate 1940 to their more modern counterparts; however, it is necessary to understand how early audio recording technology alters the voice output signal in order to progress toward more nuanced singing voice analysis such as vocal registration.

Knight (1988) specifically discusses the operatic baritone voice and its development from 1750 to 1830—an important period, because it directly precedes the advent of the modern tenor voice in 1831 (Stark, 1999). This dissertation observes changes in compositional style, vocal demands of each role, and the singers who were cast in operas during the aforementioned time period. As such, Knight's study focuses on one of the four criteria for understanding the development of the operatic singing voice mentioned at the beginning of this section: understanding the vocal demands placed upon operatic singers over time and how compositional style may have influenced those singers. He suggests that the baritone voice developed out of necessity due to the decline of the *castrato* voice in opera. This type of scholarly investigation aids in the understanding of the demands placed upon the baritone voice during the Classical era and allows researchers to make inferences about how those demands may have caused the operatic baritone voice to evolve technically.

Yet another member of the group of studies that relies on literature analysis is Reed (1983) which attempts to trace the development of the tenor voice from 1600 to 1983. The dissertation focuses on critical accounts of singers and historical treatises about singing in an effort to demonstrate that "lighter tenor voices" have a place historically in the "world of singing" (Reed, 1983). It presents and lists singers from throughout operatic history and draws connections between those singers and the composers they influenced. The author interprets events and uses those views to put forth a specific pedagogical perspective. In his study of both pedagogical practices and the vocal demands placed on singers, Reed concludes that voice teachers must strengthen and coordinate the "lighter mechanism" when training developing tenor voices.

These three dissertations attempt to answer a question that intrigues many voice teachers and performers: how has operatic singing evolved over time? Indeed, each contribute to the fields of vocal pedagogy, vocal performance, and musicology; and demonstrates current perspectives about the development of the operatic voice. Future study, however, might be aided by an improved understanding of singing voices that have been preserved by early audio recording technology. Unfortunately, it is still not completely understood how historical audio recording technology affects extant recordings from the past, or if those effects are constant (i.e. linear). This fact presents a barrier to investigating the development of the operatic voice; a barrier that this dissertation aims to address by determining how wax-cylinder technology alters the voice output signal.

2.3 Musicological Methods

Some literature focuses on musicological methods to interpret changes in operatic performance practice. Such writings use historical texts and recordings to provide insight about how singers at the turn of the century executed articulations such as *portamento* and trills. Some literature even dictates *cadenze* of historical singers from old recordings. These authors posit conclusions about topics such as the development of vibrato and some have measured vibrato rate and extent for different singers.

One of the most prolific vocal pedagogues of the twentieth-century, R. Miller, presents information about national schools of singing, or national styles of singing. In the text, the well-respected voice teacher dedicates an entire chapter to a discussion of vibrato and stylistic trends (R. Miller, 1997). Miller estimates the vibrato rate of singers trained from different national schools of singing, although not all of these observations include quantitative measurements and most include qualitative observations. The aforementioned chapter asserts the following claims:

- Singers trained in an Italianate style demonstrate a balanced (*chiaroscuro*) sound and a vibrato rate of 6-7 Hz (R. Miller, 1997).

- Singers trained in a German style demonstrate a vibrato rate of approximately 5 Hz which occurs due to a breathing technique that emphasizes abdominal contraction (R. Miller, 1997).
- Singers trained in a French style demonstrate a "rapid vibrato rate" that results from a "narrowed pharynx, elevated tongue, raised larynx, and the emphasis upon placing the tone in the masque" (R. Miller, 1997).
- Singers trained in an English style demonstrate a more rapid vibrato due to "exaggerated high-breath techniques" (R. Miller, 1997).

Miller's writings present a knowledgeable perspective regarding both the practice of vibrato and its cause. It is crucial to note that R. Miller—arguably *the* eminent scholar in vocal pedagogy during the twentieth-century—uses inferences based on historical vocal pedagogy to support perspectives about different national singing styles (R. Miller, 1977). As such, these writings are interesting from a historical perspective, illustrating how one prominent voice teacher and vocal pedagogue viewed national styles during the 1970's; however, they cannot be seen as conclusive.

Anecdotally, those in the voice performance community discuss the standardization of operatic singing at an international level. Miller's observations contradict that notion and might indicate that national styles of singing have become more uniform since the 1970's. Katz (2010) and Philip (1992) posit a similar concept: it is possible that the development of recording technology throughout the twentieth-century has standardized the modern aesthetic of our musical performances—both technically and stylistically. This idea seems to suggest the need to study the effect(s) of historical audio recording technology on voice output signals. In order to understand how recording technology might have changed or might have standardized performance practice and technical instruction, it is necessary first to seek to understand what it is that is heard when listening to historical audio recordings.

Other authors, such as Crutchfield (2012; 1983), firmly state that vibrato rate and extent have changed drastically throughout the past century. Crutchfield's writing is unique, because it strives to take into account current information from modern voice research before making conclusions about stylistic, technical, or aesthetic developments. For example, he cites a 1994 article (Sundberg) which confirms the landmark Seashore (1937a; 1937b; 1937c) study by presenting normative data on vibrato rate and extent (average vibrato rate range of 5-8 Hz). Crutchfield uses this information as a framework within which to construct his argument (Crutchfield, 2012), suggesting that singers from the turn-of-the-century demonstrate vibrato rates towards the higher end of Sundberg's published normative range and that modern singers demonstrate vibrato rates towards the lower end.

Crutchfield has also suggested a meticulous method of determining vibrato rate and extent using voice analysis software. The noted conductor and musicologist goes on to state that early recorded singers sang with a notably brighter timbre (Crutchfield, 2012). This notion, however, is difficult to justify due to the limitations of historical recording technology. It is the premise of this dissertation that a more objective analytical method would greatly benefit this type of musicological research.

Elliott (2008) similarly combines understanding garnered from voice research and historical treaties. This work is intriguing, because it details stylistic practices based on era and language. Each section mentions vibrato in some way as it relates to the time-period, style, or song-type (e.g. *Lied*, *mélodie*, or Italian art song) in question. This text is novel in its structure and its perspective that early music does not require robust singing and, therefore, does not induce vibrato. However, once again, it relies primarily on the interpretation of historical treatises, which is inherently filtered through a researcher's own scholarly focus and the opinions of the original authors.

Kauffman (1992)—previously discussed in this document—describes historical audio recordings and observes the use of *portamento* in recordings at the turn of the twentieth-century. The article claims that *portamento* has all but disappeared from modern performance practice and cites William James Henderson (1938) as having said, "[*portamento*] is capable of much expression when judiciously employed, but when it becomes a habit it is deplorable, because then it leads to scooping" (Kauffman, 1992). Kauffman argues that *portamento* is clearly used by early recorded singers and that it can be used to "heighten expression" (Kauffman, 1992). This paper demonstrates a potential use for historical audio recordings of singers from a musicological perspective and such observations about *portamento* are corroborated by authors such as Crutchfield (2012).

There is no way by which the modern voice teacher or voice researcher can travel back in time to hear singers who pre-date the advent of recording technology first-hand, but historical audio recordings from the turn-of-the-century can be used to gain insight about the performers who originally sang music from the nineteenth century and beyond. Scholars such as Crutchfield and Elliott begin this process and demonstrate ways by which such analyses could aid in our interpretation of the past and, by doing so, enhance our art form. These types of discussions about historical voice recordings provide important insights into historical performance practice and are feasible due to their reliance on temporal components of the audio recordings. This dissertation attempts to augment the study of such recordings by more accurately describing objective measurements of the voice output signal. By doing so, it may be possible to combine observations about musical gestures, notes, and rhythms performed on historical audio recordings with those that focus on specific topics concerning singing voice quality, timbre, and singing technique. This potential synergy between past scholarly inquiry and objective analytical methods could result in an enhanced understanding about historical voice recordings, historical singing techniques, and historical performance practice.

2.4 Quantitative Methods

As was previously mentioned, Seashore (1937a; 1937b; 1937c) and Sundberg (1994) present information regarding vibrato rate and extent, yet there is still a lack of consensus about how vibrato characteristics evolved during the twentieth century. Recent studies have sought to elucidate the cause of the common belief that vibrato rate has slowed over time.

In an unpublished article, Howell (2015) analyzes the vibrato rate and extent of Nellie Melba (1904 recording) singing "Porgi amor" from Mozart's *Le Nozze di Figaro* and compared the results to those of Leontyne Price, Renee Fleming, and Maria Callas. He notes that these four singers demonstrate a similar vibrato rate (approximately 6 Hz) and, with the exception of Callas, a similar vibrato extent. Figure 9 shows a spectrogram that was used to indicate that the recording of Melba is differentiated from the others by an increased signal-to-noise ratio. Howell suggests that our perception of Melba's vibrato and, thus, the difference between our perception of the four singers is caused by the higher signal-to-noise ratio in the older recording. He tests his hypothesis by removing all non-harmonic spectral information and provides the audio files for the results (See Figure 10).

Howell then continues by introducing additive noise from Melba's sample into the other three samples and, subsequently, amplifying the noise by 15 dB. He concludes by suggesting that Melba may have sounded similar to our modern aesthetic, but that the effects of the historical audio recording may distort our

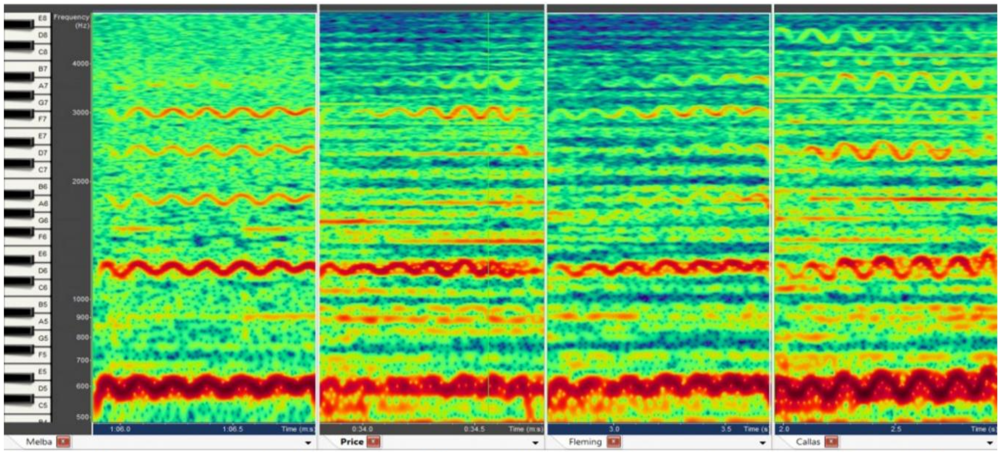


Figure 2: D5 on written vowel "ah" from m. 35 of Mozart's "Porgi amor" from Le Nozze di Figaro. From left to right, Melba, Price, Fleming, and Callas. In these one-second clips, note the similar rates of vibrato (around six cycles per second) of all four singers. Melba, Price, and Fleming all share a similar extent (how far up and down they move) of around a half step total; Callas' vibrato is differentiated by a wider total extent of almost a whole step, which may create the impression of a slower rate.

Figure 9. Refer to Howell's notes. Reprinted with permission from *NEC Vocal Pedagogy*, by I. Howell, 2015, Retrieved from http://vocped.ianhowell.net/melba/. Copyright [2014] by Ian Howell.

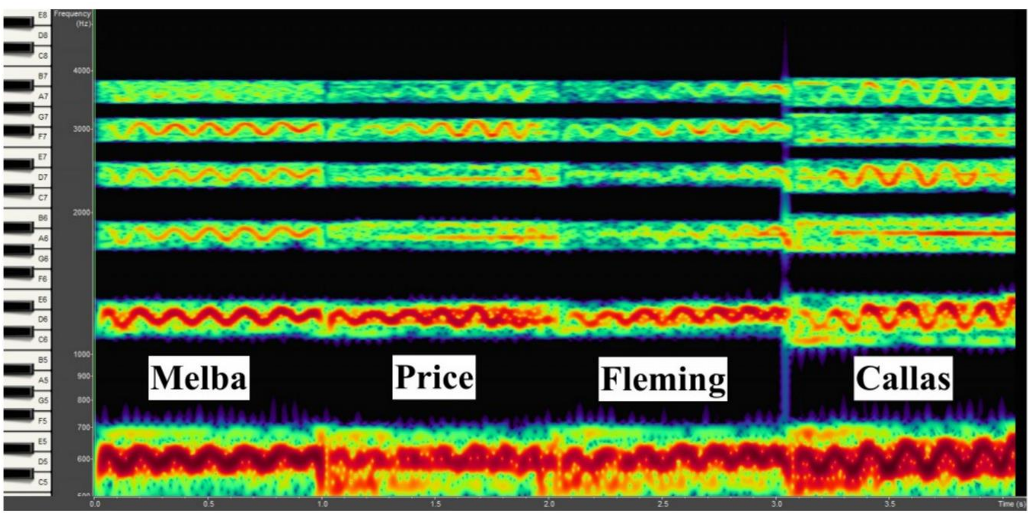


Figure 3: The audio samples from figure 2, with all sonic information above 3,680Hz and between the vocal harmonics deleted.

Figure 10. Note that the only spectral information in this spectrogram is harmonic (arguably only signal produced by the singer). Reprinted with permission from *NEC Vocal Pedagogy*, by I. Howell, 2015, Retrieved from http://vocped.ianhowell.net/melba/. Copyright [2014] by Ian Howell.

perception of her voice. This study is novel in its approach to analyzing vibrato from a historical perspective by using historical audio recordings and signal processing methods. In order to more accurately determine perceptual differences between voices, future studies should include the perceptual evaluations of multiple trained listeners. This dissertation may aid such inquiry by describing the way by which wax-cylinder technology alters the voice output signal (i.e. vibrato analysis) and by indicating acoustic measurements for which values are altered by historical audio recording technology.

Ferrante (2011) also studied historical differences between vibrato rate and extent produced by operatic soprano singers. He found a constant decrease in vibrato rate of 1.8 Hz and an increase in vibrato extent by approximately 56.3 cents since the advent of audio recording technology. The study analyzed 105 recordings by 75 different singers on the highest note (~923.33 Hz) of “Vissi d’arte” from *Tosca* (G. Puccini). A positive correlation between vibrato rate and extent within-participant was found.

Rothman et al. (Rothman, Diaz, & Vincent, 2000) investigated vibrato rate and extent by comparing the singing of historical and contemporary Jewish cantors to historical and contemporary operatic singers. The authors used samples from four singers in each category (n=16) and found that historical singers shared characteristics regardless of musical genre, as did contemporary singers. Historical singers demonstrated a higher vibrato rate (Historical Opera: 6.83 Hz; Historical Cantor: 6.73 Hz vs. Contemporary Opera: 6.0 Hz; Contemporary

Cantor: 5.6 Hz) and a narrower vibrato extent (Historical Opera: 2.11%; Historical Cantor: 2.75% vs. Contemporary Opera: 3.03%; Contemporary Cantor: 3.71%).

Their study also indicated that both historical and modern singers demonstrate the existence of the singer's formant (F_S) in their spectral data. The authors concluded that vibrato rate and extent form the primary difference between historical and contemporary singers and that intra-generational singers demonstrate more similarities than inter-generational singers, regardless of genre.

This study is perhaps the closest parallel to the one proposed in this document; however, it does not delve into spectral analysis enough to make the conclusion that vibrato is the primary difference between generations of singers. Furthermore, it does not note the possible effect of recording technology on spectral data. This dissertation intends to remove some of the limitations surrounding the study of historical voice recordings and, by doing so, will allow researchers to design more rigorous studies about the differences between modern and historical singers.

Välimaki et al. (2008) designed an algorithm to digitally "antique" audio files. The study classified audio degradations as being *global* (in the frequency domain) or *local* (in the time domain) (see Table 1). The goal of their study was to determine if it is possible to make a digital audio file with no defects to sound like it was originally recorded on a vinyl disk.

This research is unique, because it attempts to “backwards engineer” the effects of a historical recording. The same approach would be difficult to implement on voice recordings without further study due to missing frequency information in historical voice recordings and the current lack of understanding about the effects of early audio recording technology on the voice output signal.

Table 1

Common Degradations found in Historical Recordings

Global Degradations	Local Degradations
• Frequency response distortions (filtering effects)	• Tracking errors
• Limited signal bandwidth	• Clicks
• Dynamic range limitations	• Pulses (low-frequency)
• Constant distortion	• Speed inconsistencies (RPM)
• Pitch variation	
• Hiss	

(Table adapted from “Digital Audio Antiquing—Signal Processing Methods for Imitating the Sound Quality of Historical Recordings,” by V. Välimaki and S. Gonzalez, 2008, Journal of the Audio Engineering Society, 56(3), p. 119. Copyright [2008] by the Audio Engineering Society)

Morange et al. (2010) sought to determine how listeners perceive historical recordings of singers and the effects of historical audio recording technology. The authors used eleven commercial recordings of Enrico Caruso which were either unprocessed from the original recordings or digitally remastered for modern listeners. The latter process is usually accomplished with a low-pass filter or a series of filters that remove noise and local degradations created by the historical recording technology. After listening to each recording, both amateur listeners and musical experts (musicians or music acousticians) participated in a free commentary about the stimuli. The results of the study suggest that experts are more likely to rely on background noise and reverberation to determine their preference of sample (i.e. the quality of the recording), whereas amateurs are more likely to rely on how the excerpt made them feel (Morange et al., 2010). This study provides insight into how experts and non-experts perceive older audio recordings and voices of the past.

This chapter attempts to demonstrate that literature related to the development of operatic singing and historical voice recordings is limited. Such literature contributes to how the fields of singing voice performance and musicology understand operatic performance practice at the advent of early audio recording technology, but it does not fully elucidate how singing has changed. A combination of subjective and objective analytical methods may inform the discussion concerning the development of the operatic singing voice.

Anecdotally, the fields of vocal pedagogy and vocal performance are interested in the development of the operatic singing voice, but there is a hole in the literature where it concerns the analysis of historical voice recordings. Specifically, there is currently little understanding in the broader fields of voice performance and vocal pedagogy about how recording technology changes the voice output signal or our perception of historical voices. It is a fundamental hope of this study that an analysis of the effects of wax cylinder phonograph technology on recorded voice signals will help our field better understand its rich past.

CHAPTER III

METHODS

3.1 Overview

This study was observational and used a quantitative method. Statistical analyses were performed to test variables for normal distribution and heteroskedasticity, as well as to compare spectral and temporal measurements between recording conditions. Spectral measurements for which a significant main effect of recording condition was found were subsequently studied using regression analysis. The goal of this project was to examine the effect of wax cylinder audio recording technology on voice output signals and, ultimately, to be able to use that information to study singers from the turn of the 20th century more accurately.

The protocol used for data collection was approved by the University Committee on Activities Involving Human Subjects (UCAIHS) at New York University (IRB-FY2017-747).

3.2 Participant Selection

Participants were recruited by being referred to the investigator by their voice teachers and by their colleagues. Informed consent was obtained from 20 professional opera singers who took part in the study on a voluntary basis. They were selected using the following criteria:

- The participant must be an opera singer (the singer performs primarily Western operatic or Western classical repertoire).
- The participant must have no history of vocal injury or pathology within the last two years (self-reported).
- The participant must be between the age of 25-65.

AND

- The participant must earn a majority of his/her income—greater than or equal to 50%—as a singer (self-reported).

OR

- The participant must have sung a principal role in at least one AGMA production in the last two years.

A total of 20 professional opera singers were recruited (see recruitment letter in Appendix A). All participants signed a consent form and received identical directions during the recording session (see Appendix B). The voice types, average age, and experience-level of the participants are reported in Table 2. The sample size used for this project is consistent with previous studies that have analyzed various spectral moments of the LTAS and other spectral phenomena related to the singing voice (Cesari et al., 2013; M. Guzman, 2013; Holmberg et al., 1995). Various voice types were studied due to the inclusion of the test signal recording phase. It was determined that the test signals—and specifically various

Table 2

Selected Descriptive Statistics of Participants Arranged by Voice Type

		Age			Years Performing Full-Time		
Voice Type	**Number**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
Soprano	5	48	38	62	19.6	9	30
Mezzo Soprano	5	35	28	49	10.5	2	23
Tenor	5	35	32	41	12.0	4	17
Low Male Voice	5	39	29	50	13.0	5	22
Baritone	*2*	*45*	*40*	*50*	*19.0*	*16*	*22*
Bass-Baritone	*1*	*29*	-	-	*5.0*	-	-
Bass	*2*	*38*	*33*	*43*	*11.0*	*9*	*13*

impulse responses—would adequately characterize the frequency response of the wax cylinder system and that using multiple voice types would allow for a more in-depth study of differences between historical singers and modern singers.

Professional opera singers were recruited in order to accurately estimate the limitations of the wax cylinder recording equipment. For example, there is anecdotal evidence suggesting that operatic singing has changed due to stylistic constraints and orchestral textures of operas written by composers such as Richard Wagner (1813-1883) and Giacomo Puccini (1858-1924). More specifically, it is possible that operatic singers may have adopted more robust singing techniques since the end of the 19th century in order to be heard over a large orchestra (Stark, 1999, pp. 108-109)—or at least that lighter voices in the 19th century were more likely to be heard in roles that are now reserved for larger voices (Crutchfield, 2012).

Past research has shown that wax cylinder recordings become visibly distorted for a 1 kHz sine tone at approximately 90 dB (Feynberg, 2014). Furthermore, literature related to sound engineering techniques used for acoustic phonograph recordings indicates that singers at the turn-of-the-century would have been required to move in relation to the phonograph horn while being recorded in order to create dynamic contrast (Fesler, 1983, pp. 690-691; Torick, 1977, p. 881). Studying professional opera singers was likely to result in sound pressure levels in excess of 97 dB (at 50 cm mouth-to-microphone distance) (Björkner, 2008)—and higher sound pressure levels at closer distances[9] (Švec & Granqvist, 2018, pp. 442-444). Therefore, including professional opera singers provided comparable data points with which to contextualize historical voice recordings.

3.3 Equipment

3.3.1 Historical Recording Equipment

The wax-cylinder recording equipment that was used for this study matched that which was used by Zakaria (2016) and was similar to the equipment

[9] Note: According the Distance Law, the SPL of a voice decreases/increases by approximately 6 dB when the mouth-to-microphone distance is doubled/halved, respectively (e.g. a SPL of 97 dB measured at a mouth-to-microphone distance of 50 cm would be recorded as approximately 117 dB if measured at a mouth-to-microphone distance of 5 cm).

used by Feynberg (2014). Doing so allowed for cross-comparison between frequency response data and the voice recordings.

Data collection was completed in collaboration with the Thomas Edison National Historical Park (TENHP) (West Orange, New Jersey) using facilities at New York University over two, seven-hour recording sessions. The NYU Music Technology Program's Research Lab—a sound-treated room on a sound-isolated floor with a reconfigurable overhead grid and a 0.2 s reverb time—was used for all recording sessions. Wax cylinders were later digitized at the Thomas Edison NHP by museum curator, Gerald Fabris, with the use of an *Archéophone* owned by the museum. The historical recording equipment included:

- Edison Home Phonograph (serial # 378617D) (see Figures 11 and 12)
- Mica diaphragm recorder (serial # 262489: "New Edison Recorder" Dimensions: Outside diameter- 4.128 cm, Overall height- 3.334 cm, Outside diameter of top (connects to horn)- 1.588 cm)
- Black metal horn (Replica horn with dimensions: length- 68.58 cm, entrance diameter- 20.32 cm)

It is important to note that original horns are rare and this replica was created and designed by a horn repairer. Additionally, the museum curator offered the use of either a large or a small horn. According to a manual published by the National Phonograph Company in 1909, Edison-made cylinder phonographs at that time achieved best results with a horn between 26 to 30 inches (68.58-76.2 cm) long and a 6 inch (15.24 cm) wide opening (National Phonograph Company, 1909, p.

9). For this reason, this study used a horn approximating these measurements that was manufactured by Tom Govelitz (The Music Shop, Boonton, NJ), which was suspended from a horn crane during recording sessions.

Figure 11. The Edison Home Phonograph and the black metal phonograph horn replica used during data collection.

Wax-cylinder replicas were manufactured by and purchased from Paul Morris Music (Exeter, England) upon the recommendation by the museum curator. Each replica was designed to record approximately two minutes of audio signal. All wax cylinders were inspected at the Thomas Edison NHP by the museum curator and shaved if they were determined to be uneven. Prior to being used for recording, the cylinders were heated under a heating lamp to approximately 80.1°F (Zakaria, 2016) to soften the material—a process required in order to inscribe acoustic signal onto a cylinder.

During the first recording session (10/28/2018), it was discovered that the wax cylinder replicas were too thin for the Edison Home Phonograph. In order to compensate so that the stylus did not damage the cylinders, a 24-gauge brass wire was coiled and inserted inside the main housing, above the diaphragm. The decision to adjust the recording equipment was made by the researcher in consultation with Mr. Fabris and phonograph restorer and audio historian, Michael Devecka. Mr. Devecka also performed "swarf duty"[10] during both recording sessions.

Figure 12. A wax cylinder replica and the Edison Home Phonograph used for data collection.

3.3.2 Modern Recording Equipment

The modern recording equipment for the singing protocol consisted of three Earthworks M30 omnidirectional microphones (Earthworks, Inc., Milford, NH)[11]. The microphones were selected due to the flat frequency response

[10] "Swarf duty" is the act of blowing off the wax dust shavings (swarf) with a handheld device while the cylinder is being cut.
[11] See APPENDIX C for the polar response and the frequency response for the Earthworks M30.

reported by their manufacturer. By using a flat-response microphone, it is theoretically possible to record a human voice with high fidelity.

Two microphones were fixed to microphone stands and were placed on either side of the phonograph horn, exactly 20 cm from either side of the horn (see Figure 13). The third microphone was fixed to a microphone stand and placed

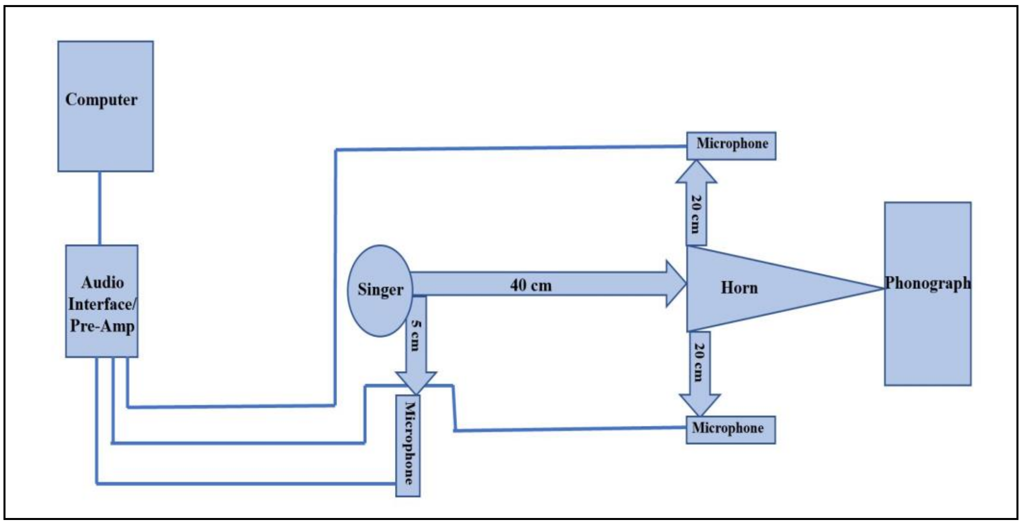


Figure 13. Diagram of recording device array for singing protocol.

perpendicular to the horn and other microphones, 40 cm in front of the mouth of the phonograph horn. They were connected via XLR cables to an Allen & Heath MixWizard WZ[3] 16:2 mixing board (Allen & Heath, Penryn, UK) which was fed through a wall panel using a masque connector to the CMR room SSL Alpha Link Madi AX a-d/d-a converter (Solid State Logic, Oxford, UK). The digital signal was then sent to an iMac desktop computer running MacOS Sierra 10.13.6 using an RME Madiface USB audio interface (RME, Haimhausen, Germany) where it was processed using *ProTools Ultimate 2018.10.0* (Avid Technology, Burlington, MA, USA) at a sampling rate of 44.1 kHz and a 32-bit float quantization. The recordings were saved as uncompressed *.wav* files. The sound pressure level was calibrated with a Reed Instruments Class 2 Sound Level Calibrator (Reed Instruments, Wilmington, NC) by introducing a 1 kHz sine-tone at 114 dB at various points during the recording sessions.

The original research design for this study intended to use a singer-to-horn and singer-to-microphone distance of 30 cm—which abides by standards in the field of voice and speech research (Hunter, Spielman, Starr, & Popolo, 2007; Svec, Sramkova, & Granqvist, 2009; Švec & Granqvist, 2018; Titze & Winholtz, 1993). Additionally, the microphone array was designed to be able to account for phase differences by placing a microphone to either side of the horn and by placing a third microphone directly below the horn. During the first recording session, it became apparent to the researcher that the horn moved approximately 40 cm horizontally—left to right—while the recording was occurring due to the configuration of the phonograph machine itself (i.e. the horn moved as the stylus tracked the cylinder during recording). As such, it was necessary to adjust the microphone array in order to document each singer's voice more accurately. Furthermore, the museum curator and his assistant noticed visible distortion on a test cylinder with the first singer 30 cm from the mouth of the horn. It was determined that a singer-to-horn distance of 40 cm would mitigate distortion of cylinder recordings and still limit the filtering effects of the room (Švec &

Granqvist, 2018). Participants were monitored in order to ensure that they maintained a 5 cm distance from the side microphone and a 40 cm distance from the horn.

3.3.3 Test Signal Recording Equipment

The recording device array was modified slightly in order to measure the frequency responses of the systems (see Figure 14). The test signals were

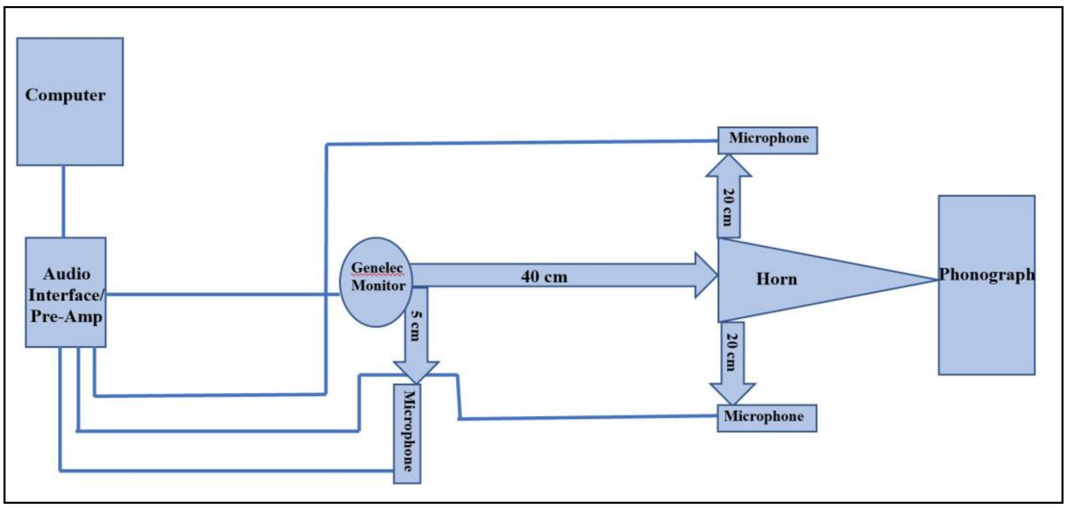


Figure 14. Diagram of recording device array for impulse response.

executed using *Protools Ultimate* 2018.10.0 on an iMac desktop computer running MacOS Sierra 10.13.6. The computer was connected to an RME Madiface USB audio interface, which was in-turn connected to the CMR room SSL Alpha Link Madi AX a-d/d-a converter. From there, the analog signal was sent through a wall panel to a Genelec 8030A nearfield monitor[12] (Genelec, Inc., Netick, MA) using an XLR cable. All other equipment was identical to the array described in Section 3.3.2.

3.3.4 Transfer Equipment

The wax-cylinder transfer equipment used for this study was identical to that used by Feynberg (2014) and Zakaria (2016). The cylinders were digitized by the museum curator at the Thomas Edison NHP using standards codified by IASA TC-04 (IASA Technical Committee, 2009).

Playback Equipment:

- Stanton 500-AL cartridge (wired stereo out)
- *Archéophone* cylinder playback machine (production number 11) with a tracking force of 4g
- KAB Souvenir EQS MK12 Preamplifier

WAV Digital Recording Equipment:

- KAB Souvenir EQS-MK12 preamplifier, balanced TRS stereo out
- Stereo TRS into return jacks of Metric Halo Mobile I/O ULN-2 A/D converter, firewire 6-pin out
- Firewire 9-pin into iMac computer
- SoundForge Pro 1 Software

---

[12] See APPENDIX C for specifications, including horizontal directivity and frequency response.

- iMac desktop computer

WAV Digital Recording Specifications:

- 96 kHz sampling rate
- 32 float bit quantization
- Stereo
- Stereo summed to Mono
- PCM (uncompressed) Broadcast Wave Format (*.wav*) file

3.4 Protocol

3.4.1 Phase One: Test Signal Recordings

A series of test signals was designed to analyze the two systems (see Table 3). It approximated those used in previous studies (Feynberg, 2014; Zakaria, 2016) (see Table 5 and Table 6), but included five seconds of silence and a modulated test signal to mimic vibrato rate and extent. Silence was included in the test signal recording in order to accurately measure the noise floor of the wax cylinder technology. A frequency-modulating sine-tone (Rate: 6.5 Hz, Extent: 54 cents) was created to study the effect of wax-cylinder technology on vibrato rate, vibrato extent, and fundamental frequency separate from human participants (see Figures 15 and 16). The 440 Hz sine tone was chosen given that it is in the middle of human phonation, and would be within the frequency range of the wax cylinder phonograph system.

Table 3

Contents of the Test Signal Recording

Start (min:sec:mil)	**End (min:sec:mil)**	**Test Signal Type**
0:00:00	0:05:00	Silence
0:05:00	0:10:00	440 Hz sine tone
0:10:00	0:15:00	Frequency modulating 440 Hz sine tone (6.5 Hz, 54 cents)
0:15:00	0:20:00	440 Hz sine tone
0:20:00	0:25:00	Frequency modulating 440 Hz sine tone (6.5 Hz, 54 cents)
0:25:00	0:30:00	440 Hz sine tone
0:30:00	0:35:00	Frequency modulating 440 Hz sine tone (6.5 Hz, 54 cents)
0:35:00	0:40:00	440 Hz sine tone
0:40:00	0:50:00	White noise (0 Hz - 22.05 kHz)
0:50:00	0:55:00	440 Hz sine tone
0:55:00	1:25:00	20 Hz - 20 kHz Linear sinesweep
1:25:00	1:30:00	440 Hz sine tone
1:30:00	2:00:00	20 Hz - 20 kHz Linear sinesweep

Table 4

Test Signal Suite from Feynberg (2014)

Start (min:sec:mil)	**End (min:sec:mil)**	**Track (test signal)**
00:00:00	00:00:10	1000 Hz alignment blip
00:02:00	00:12:92	20 Hz – 20 kHz Exponential Sine Sweep
00:15:00	00:15:10	1000 Hz alignment blip
00:17:00	00:27:00	500 Hz sine tone
00:29:00	00:29:10	1000 Hz alignment blip
00:31:00	00:41:00	1000 Hz sine tone
00:42:00	00:42:10	1000 Hz alignment blip
00:44:00	00:54:00	2000 Hz sine tone
00:56:00	00:56:10	1000 Hz alignment blip
00:58:00	01:08:00	1000 Hz intensifying sine tone beginning at -∞
01:10:00	01:10:10	1000 Hz alignment blip
01:12:00	01:22:00	White noise
01:24:00	01:24:10	1000 Hz alignment blip
01:26:00	01:25:99	Impulse
01:28:00	1:28:10	1000 Hz alignment blip

(Reprinted from “Modern contact transfer: assessing its effect on wax cylinder audio,” by R. Y. Feynberg, (p. 21), 2014, New York University, New York, NY: Unpublished Master’s Thesis)

Table 5

Test Signal Suite from Zakaria (2016)

Start (min:sec:mil)	End (min:sec:mil)	Test Signal
00:00:00	00:00:10	1kHz Alignment Blip
00:02:00	00:12:00	10s Sweep
00:14:00	00:14:10	1kHz Alignment Blip
00:16:00	00:46:00	30s Sweep
00:48:00	00:48:10	1kHz Alignment Blip
00:50:00	01:00:00	500Hz
01:02:00	01:02:10	1kHz Alignment Blip
01:04:00	01:14:00	1000Hz
01:16:00	01:16:10	1kHz Alignment Blip
01:18:00	01:28:00	2000Hz
01:30:00	01:30:10	1kHz Alignment Blip
01:32:00	01:42:00	White Noise
01:44:00	01:44:10	1kHz Alignment Blip
01:46:00	01:46:01	Impulse
01:48:00	01:48:10	1kHz Alignment Blip
01:50:00	02:00:00	10s Sweep

(Reprinted with permission from "The Edison Standard Phonograph: Digitally Recreating Three Phonograph Playback Horns for Tranferred Cylinders," by M.A. Zakaria, (p. 23), 2016, New York University, New York, NY: Unpublished Master's Thesis)

Both the exponential sine-sweep and the randomly generated white noise (see Figure 17) were included in order to adequately measure and compare the frequency responses of the wax cylinder technology and the flat-response

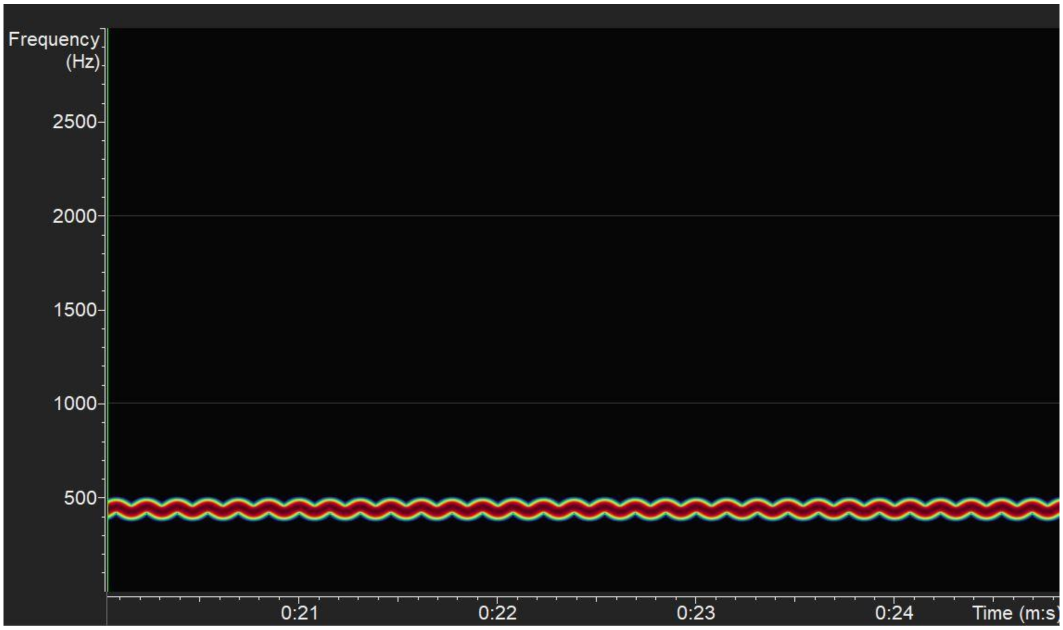


Figure 15. A spectrogram of the frequency-modulating 440 Hz sine-tone test signal (Rate: 6.5 Hz, Extent: 54 Cents). Time is shown on the x-axis, frequency is shown on the y-axis, and sound intensity level is shown as a color gradient (red is high intensity).

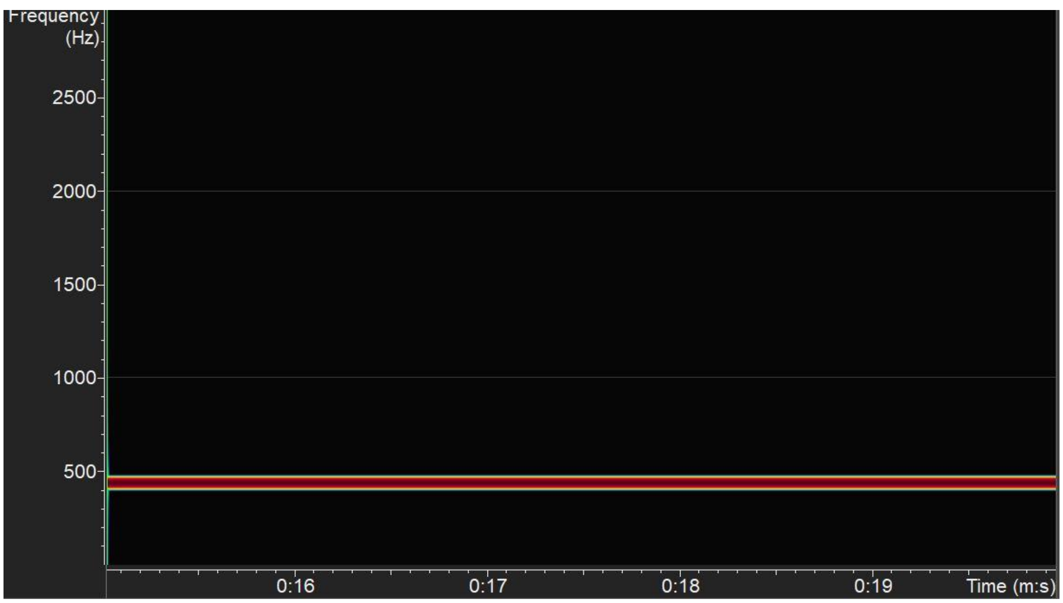


Figure 16. A spectrogram of the 440 Hz sine-tone test signal. Time is shown on the x-axis, frequency is shown on the y-axis, and sound intensity level is shown as a color gradient (red is high intensity).

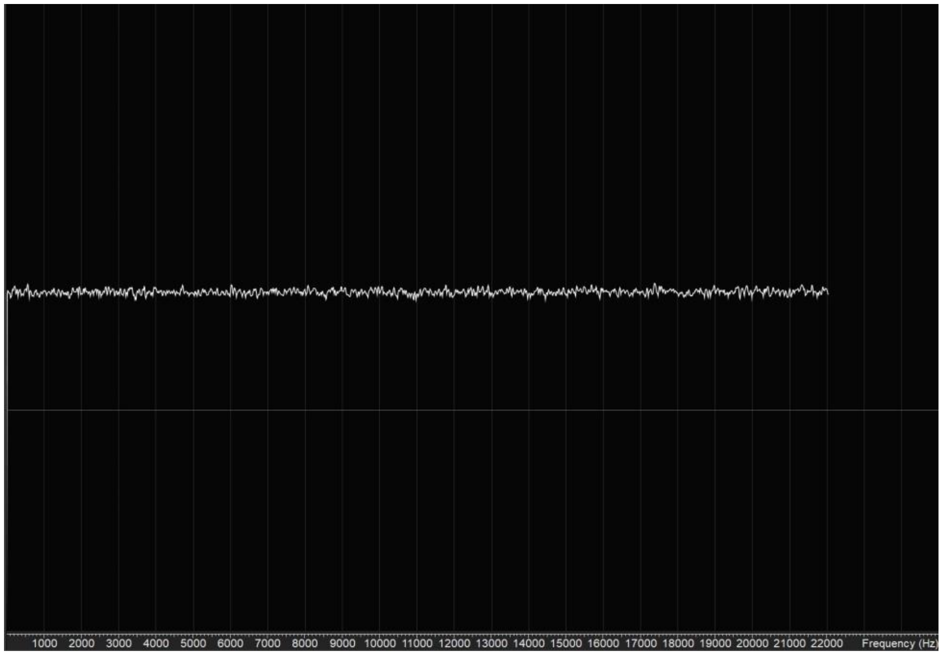


Figure 17. An LTAS of the 10-second randomly generated white noise test signal. Frequency is represented on the x-axis and sound intensity level (SIL) is represented on the y-axis. Notice that all frequencies from 0 Hz to 22.05 kHz have a relatively equal SIL (flat response).

omnidirectional microphones. The 30-second linear sine-sweep was identical to the signal that was used in Feynberg (2014) and Zakaria (2016) (see Figure 18). The test signals were created with *Matlab* scripts (see Appendix D) which resulted in five separate *.wav* files. These files were subsequently concatenated in *Audacity* at the original sampling rate of 44.1 kHz and a 32-bit float quantization.

During the second recording session (11/9/2018), the two-minute series of test signals was played with *Audacity* through a Genelec 8030A nearfield monitor. The Genelec monitor was suspended from the ceiling and placed exactly 40 cm in front of the opening to the phonograph horn to mimic the positioning of singing participants. The test signals were then recorded in *ProTools Ultimate*

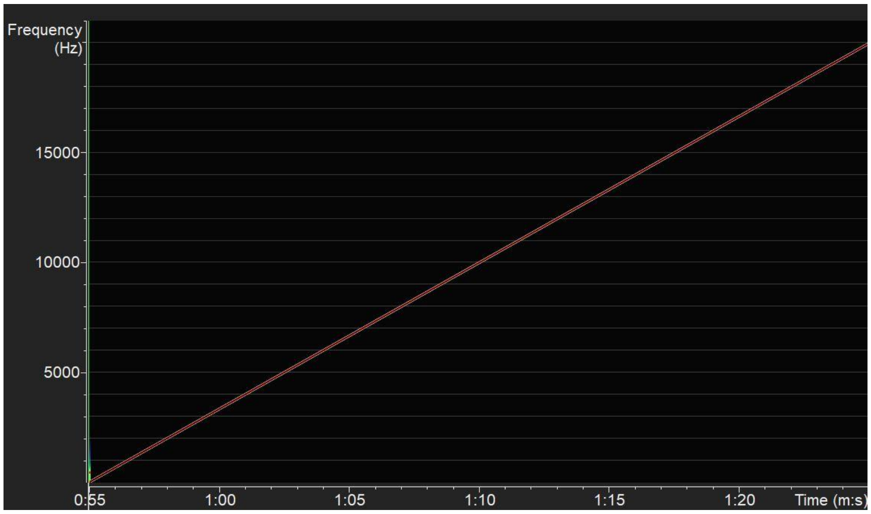


Figure 18. A spectrogram of the 30-second sinesweep test signal (20 Hz - 20 kHz). Time is shown on the x-axis, frequency is shown on the y-axis, and sound intensity level—which was constant throughout the range—is shown as a color gradient (red is high intensity).

v.2018.10.0 at a sampling rate of 44.1 kHz and a 32-bit float quantization, and onto a wax cylinder with the Edison Home Phonograph. Swarf[13] was brushed gently from the newly-cut wax cylinder and the cylinder was allowed to cool before it was placed back into its individual container.

3.4.2 Phase Two: Singer Recordings

As with Phase One, the singers were recorded using both an Edison Home Phonograph (see Section 3.3.1) and three Earthworks M30 omnidirectional microphones (see Section 3.3.2) simultaneously.

---

[13] Debris or shavings that result from cutting into a blank wax cylinder.

One week prior to each recording session, participants were emailed a list of vocal tasks in appropriate keys based on voice-type along with instructions. This was done in order to give the singers adequate time to feel prepared to be recorded onto a wax cylinder in one take. The document sent to participants included information about the tempi, vocal register, and vowels to be used for each task (see Appendix E).

Upon arriving for their recording session, each participant was asked to complete a questionnaire (see Appendix F) in order to collect demographic information and to better understand each participant's perception of recordings that pre-date 1920. While recruitment material explicitly stated the inclusion criteria, the questionnaire confirmed performance background, age, and history of vocal injury. Participants signed a consent form and were granted the opportunity to ask questions about the study.

Each participant was instructed to sing using their "best professional voice." Singers were given identical explanations about the vocal tasks and procedures and were given an opportunity to rehearse the vocal tasks before being recorded. The rehearsal was timed to ensure that the singer could complete the protocol in two minutes or less (the maximum duration of the wax cylinder replicas). The starting pitch was played with a pitch pipe application on a cell phone prior to each iteration of a vocal task. A printed vocal task sheet was provided to each singer based on voice type (Figures 19-24).

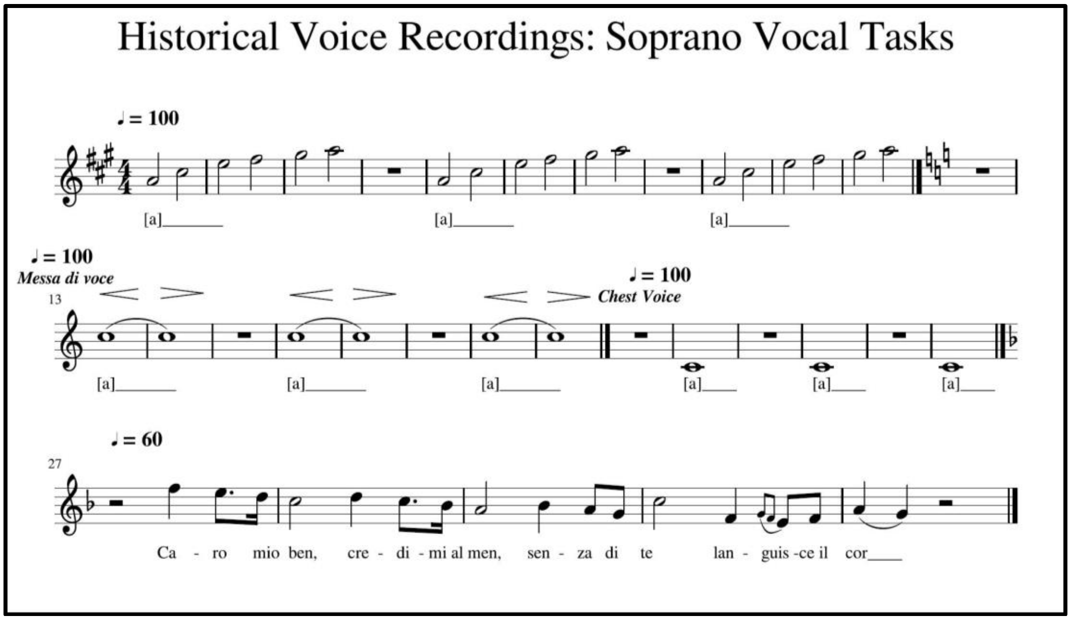


Figure 19. The vocal task list given to sopranos during recording sessions (Note: Soprano participants were given the option to sing the first task in C-Major).

Figure 20. The vocal task list given to mezzo sopranos during recording sessions.

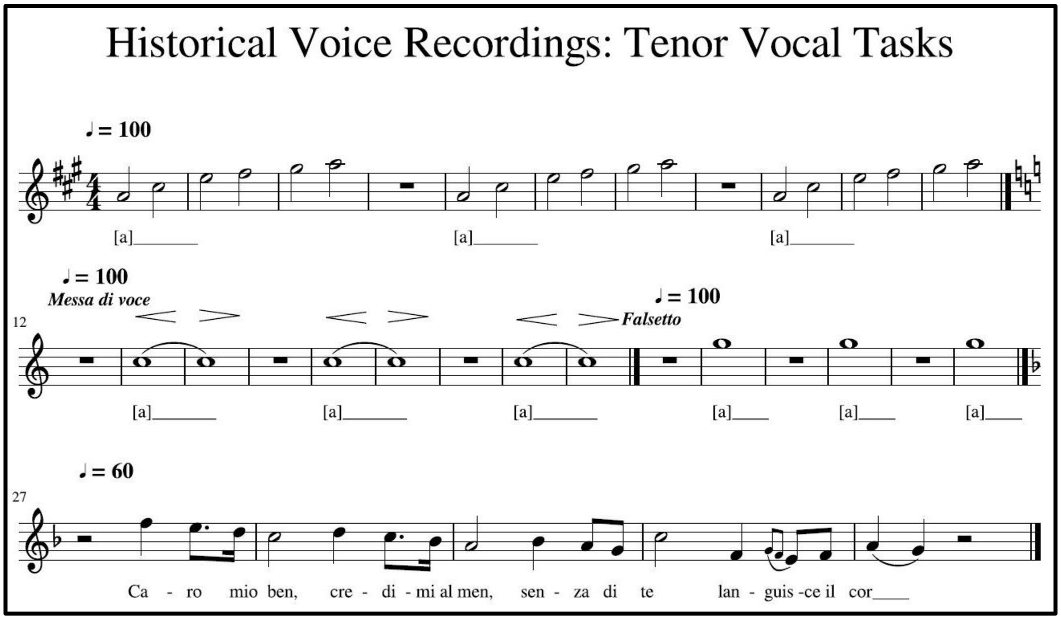


Figure 21.The vocal task list given to tenors during recording sessions (Note: Tenor participants were given the option to sing the first task in C-Major).

Figure 22. The vocal task list given to baritones during recording sessions.

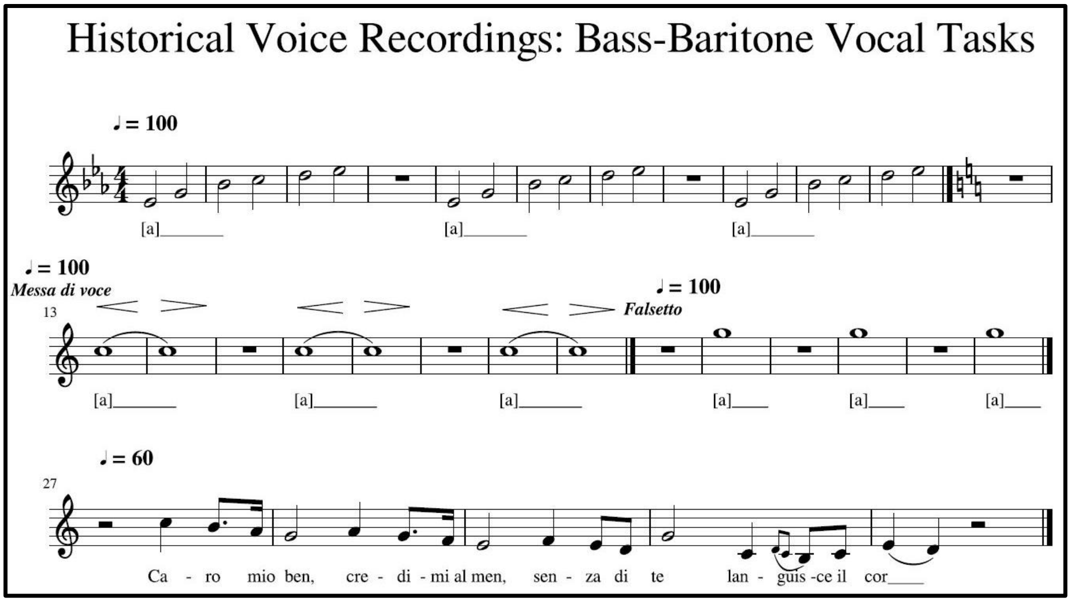


Figure 23. The vocal task list given to bass-baritones during recording sessions.

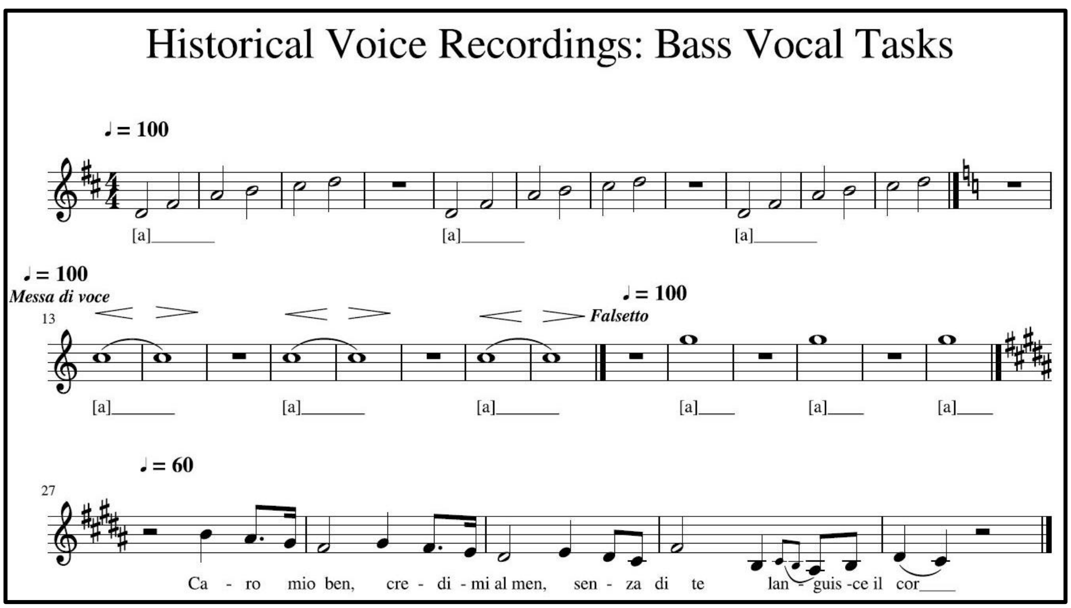


Figure 24. The vocal task list given to basses during recording sessions.

Studies that use spectral analysis techniques and observe formant tuning methods often use ascending octave scales in multiple keys (D. G. Miller & Schutte, 2005; Neumann, Schunda, Hoth, & Euler, 2005), scales that encompass all pitches in a singer's range (Henrich, Wolfe, & Smith, 2011; Henrich, Wolfe, & Smith, 2014; Scherer, Sundberg, Fantini, Trznadel, & Eyben, 2017), or ascending and descending octave scales in multiple keys (Sundberg, La, & Gill, 2011). Most studies also use tasks that incorporate different vowels, and involve different vocal fold lengths and variations of subglottal pressure (such as a *glissando*, *crescendo*, or *diminuendo*). This study, however, used wax-cylinders that were both analog recording devices and only able to record a limited duration of continuous signal (approximately two minutes). As such, it was not possible to adhere to common practice in singing voice research completely, because of the relatively short recording time. Therefore, these tasks were chosen to allow each participant to successfully complete the vocal battery within the allotted time and to allow them to complete the tasks without excessive vocal demands.

Due to the fact that this dissertation aimed to test the limitations of the wax cylinder system related to singing, it was necessary to record tasks that utilized different vocal registers, varying sound intensity levels (SIL), different parts of each singer's range, and in-context singing. The vocal tasks were designed to focus on these specific characteristics and to elicit certain responses from the participants. What follows is a brief discussion of the rationale behind the use of the vocal battery in this study.

It is important to note that this study used an ascending *arpeggio* followed by three diatonic pitches instead of full-octave scales. Doing so ensured that each subject would sing through register transitions—allowing for an estimation of formant tuning practices—and at the upper part of their vocal range. The vowel [a] was chosen in order to accommodate female singers due to formant tuning methods necessary for high-pitch female singing (i.e. transition between registers) (McCoy, 2004; D. G. Miller, 2008; Sundberg, 1987). While male operatic singers do not use the same vowels to accomplish formant tuning of their upper range (McCoy, 2004), it was determined by the author that it would be efficacious to use the same target vowel for all voice types in order to be able to compare spectral measurements across both sexes. The task was repeated three times so as to decrease the likelihood of aberrant results.

The *messa di voce* task was sung on C_4 (261.63 Hz) by male participants and C_5 (523.25 Hz) by female participants. This task was designed to study vibrato and sound intensity level (SIL) measurements on wax cylinder recordings. The [a] vowel was chosen to ensure that each participant was able to reach a maximal sound pressure level (SPL), thereby inducing vibrato. As with the *arpeggio*-task, a compromise was made by using the same vowel for all voice types in order to allow for accurate spectral comparisons. From a musicological perspective, it was important to study the *messa di voce* task as performed by modern professional opera singers, because there is evidence that the stylistic practice of singing a *messa di voce* has evolved over time to include two types: a *messa di voce* with vibrato throughout and a *messa di voce* that gradually adds vibrato as dynamic level (or sound intensity) increases (Gable, Frederick K. (tr.), 2009; F. K. Gable, 1992).

Additionally, chest-voice (female singers) and *falsetto* (male singers) tasks were used due to suggestions in the literature that female lower-voice and male upper-voice registration may have been different in the 19th century (Crutchfield, 2012). The chest-voice task was sung on a C_4 (261.63 Hz)—a pitch that could be sung in chest-voice regardless of female voice type (R. Miller, 1996)—and the *falsetto* task was sung on a G_4 (392.00 Hz)—a pitch that could be sung in *falsetto* regardless of male voice type.[14]

Lastly, many studies include in-context singing as a vocal task. While having subjects sing an excerpt from an aria would have resulted in an intriguing study, it was not feasible for this dissertation due to the time constraints of the recording media. Furthermore, it was deemed necessary to be able to compare in-context singing across voice types and participants. Requiring singers to perform an aria in their repertoire would have resulted in an unrelatable data set due to the inclusion of different pitches, ranges, and phonemic content. All of these factors would have resulted in spectral moment measurements that were vastly different between participants.

---

[14] *N.B.*-When writing about voice classification in Western classical singing, it is clear that our current gendered-terminology for voice types is not effective in the 21st-century. As such, this author recognizes that the language used in this dissertation to discuss voice type is not sufficient, but also notes that creating new terminology is beyond the scope of this work.

Past studies have used excerpts from "The Star Spangled Banner" (United States of America national anthem) (Johnson & Kempster, 2011), "Happy Birthday" in French ("*Joyeux anniversaire*") (Henrich, 2007), Gounod's "*Ave Maria*" (Henrich, 2007), or an excerpt from the participants' repertoire (Hallqvist et al., 2017; Sundberg, 2015). For this study, each subject sang the first four measures of "*Caro mio ben"* (G. Giordani, 1751-1798)—a widely known song in Italian—in a key to match their voice classification. This excerpt required each subject to sing a variety of vowels, multiple intervals (ascending and descending), and to sing in different parts of their range. The primary purpose of this task was twofold: to include an in-context singing task in the study and to have a longer task with a variety of pitch, consonant, and vowel combinations to analyze in comparison with historical recordings. The vocal task also allowed for an analysis of spectral moments of the LTAS over a longer sample duration.

Keys for each voice type were chosen in order to study registration events (*secondo* or "upper" *passaggio*[15])—in the case of the *arpeggio*-task—below or "lower" *passaggio* for female singers—in the case of the chest-voice task—and below *secondo* or "upper" *passaggio* for the in-context singing task. See below for an adapted version of R. Miller's (1996) descriptions of *approximate* register events for different voice types:

---

[15] As termed by R. Miller (1996). Discussions of registration terminology are beyond the scope of this dissertation.

Table 6

Approximate Registration Events for Female Voices, Adapted from R. Miller (1996)

Category of Voice	***Lower (primo passaggio)***	***Upper (secondo passaggio)***	***Flageolet***
Soprano	Eb4	F5-G5	D6
Mezzo-soprano	D4	D5-E5	C6
Contralto	C4	C5-D5	A5

Table 7

Approximate Registration events for Male Voices, Adapted from R. Miller (1996)

Category of Voice	***primo passaggio***	***secondo passaggio***
Tenorino	F4	Bb4
Tenore leggiero	E4 (Eb4)	A4 (Ab4)
Lyric Tenor	D4	G4
Spinto Tenor	D4 (C#4)	G4 (F#4)
Heldentenor	C4 (C#4)	F4 (F#4)
Lyric Baritone	B3	E4
Dramatic Baritone	Bb3	Eb4
Bass-Baritone	A3	D4
Bass	Ab3 (G3)	Db4 (C4)

3.4.3 Phase Three: Cylinder Digitization

Feynberg (2014) showed that wax cylinders record frequency information up to 8 kHz. Furthermore, findings from that project demonstrated that the frequency response was stable after the twelfth transfer despite losing the 8 kHz frequency band on two out of three cylinders (see Figure 25). During Phase Three, the test signal recording and one of the extra recordings were first digitized (-9 dB level adjustment) on an *Archéophone* (4.0 g tracking force) without

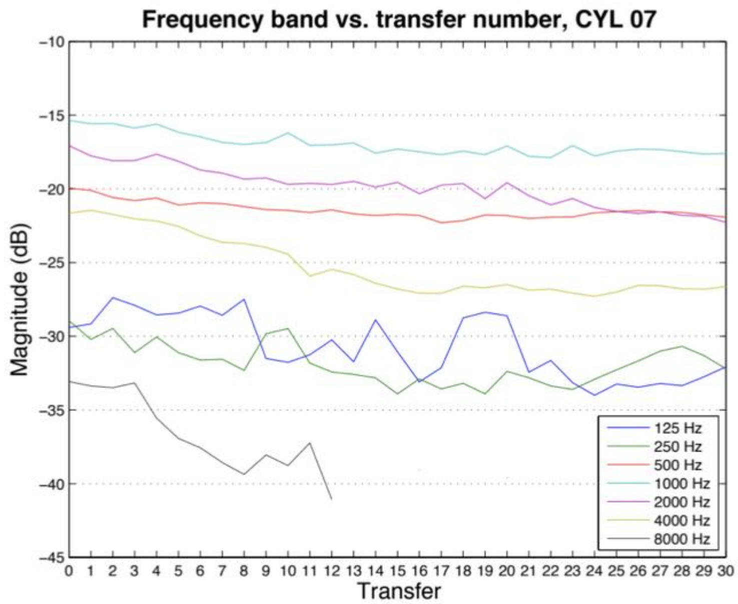


Figure 4.9 – Harmonic series magnitudes vs. transfer number of Cylinder 3

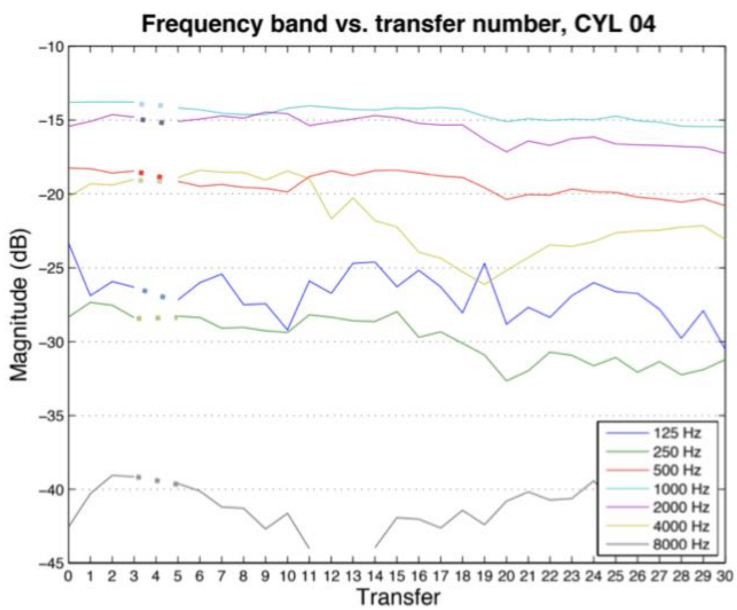


Figure 4.7 – Harmonic series magnitudes vs. transfer number. The dropout area (transfer 4) is interpolated with dots.

Figure 25. Note the dropoff in the 8 kHz band after transfer 13 and 12 as well as the comparatively lower magnitude of the 125 Hz and 250 Hz bands. Reprinted from "Modern contact transfer: assessing its effect on wax cylinder audio," by R. Y. Feynberg, (pp. 39-40), 2014, New York University, New York, NY: Unpublished Master's Thesis.

playing them on period phonograph equipment. The two recordings were subsequently played ten times on an Edison Fireside Phonograph (Edison Model C reproducer, 26.0 g tracking force) in order to simulate wear, and then digitized a second time.

Figure 26 and Figure 27 are long-term average spectra (LTAS) of the 10-second white noise test signal seen in Figure 17. Preliminary analysis during digitization indicated that the primary change after simulated wear was an increase in the noise floor above 3 kHz and a slight attenuation in the SIL at

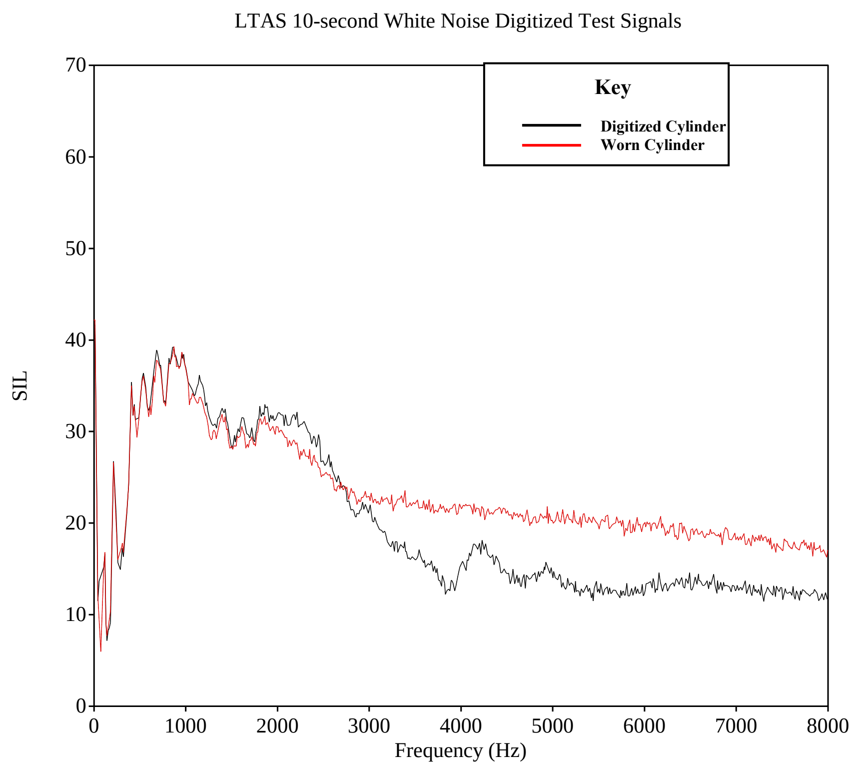


Figure 26. LTAS of a 10-second white noise test signal (from Figure 12) recorded by a wax cylinder and digitized with an Archéophone with no previous plays (black line) and after 10 plays on a period Edison Fireside Phonograph (red line). Notice the increase in frequency information (noise) above 3 kHz in the worn recording and the attenuation in SIL at approximately 2 kHz.

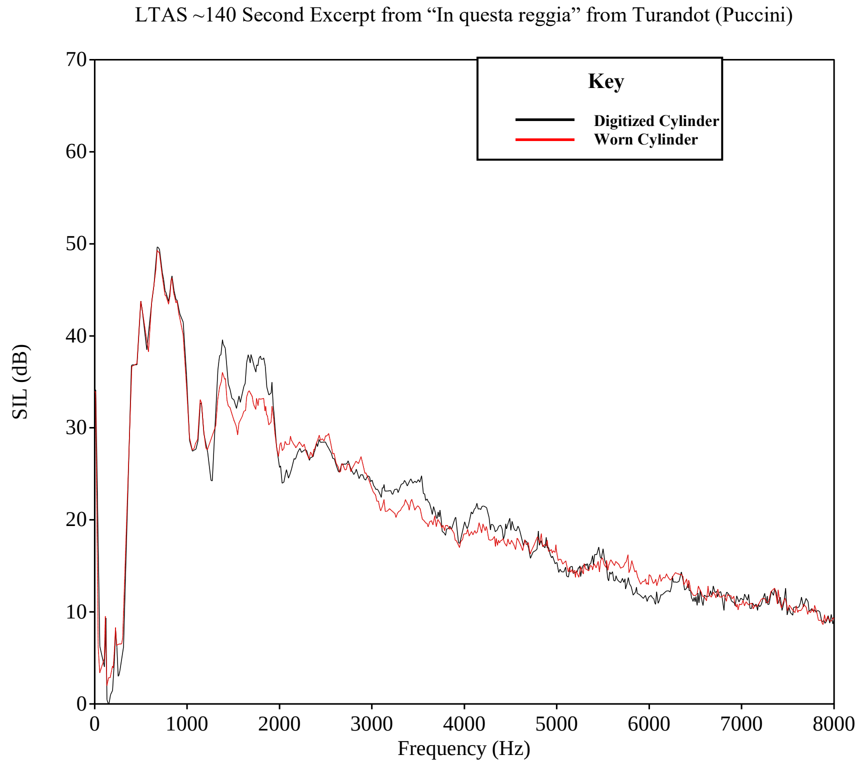


Figure 27. LTAS of an approximately 140 second excerpt from "In questa reggia" (Puccini) sung by a professional operatic soprano onto a wax cylinder. The cylinder was first digitized with no prior playback (black line) and then digitized after 10 plays on a period Edison Fireside Phonograph (red line) with a tracking force of 26.0 g.

approximately 2 kHz. Additionally, it was determined by the author that the resultant changes from simulated wear were not perceptually relevant enough to warrant digitizing cylinders before and after simulated wear.

Digitizing wax cylinders takes approximately 15-20 minutes per cylinder. Due to the number of cylinder recordings (20 participant recordings, 2 test signal recordings, and 2 extra recordings of in-context singing) and the spectral results seen in Figure 26 and Figure 27, it was determined that characterizing the phonograph system before and after wear with the test signal recording would be sufficient for this study. Additionally, there are not many extant "virgin" wax cylinders (cylinders that have never been played). As such, worn cylinders of singers provide a more similar comparison to digitized historical recordings.

Due to this preliminary analysis, "virgin" wax cylinders were not digitized. Instead, during Phase Three, wax cylinders were played back on a period Edison Fireside Phonograph 10 times with a tracking force of 26.0 g at 160 RPM in order to simulate wear. The cylinders were subsequently transferred with an *Archéophone* on the 11th playback with a tracking force of 4.0 grams at 160 RPM.

Following this method allowed this study to have like-comparisons to surviving wax cylinders. All digitized transfers were saved on an external hard drive in *.wav* format at a sampling rate of 96 kHz and 24-bit quantization.

3.5 Data Analysis

3.5.1 Pre-Processing Audio Files

Prior to analyzing the audio files, all digitized recordings were resampled from 96 kHz to 44.1 kHz in *Audacity* and saved in *.wav* format. It was discovered during the second recording session that some participants had been recorded on microphones at a sampling rate of 48 kHz due to settings in an adjacent computer room to the research lab. All files affected by this problem were resampled from

48 kHz to 44.1 kHz in *Audacity* and saved in *.wav* format. Resampling the audio signal for either recording condition resulted in all frequencies above the Nyquist frequency—in this case, 22.05 kHz—being omitted from the audio signal; however, this effect of resampling was deemed to be inconsequential since the test signal was indistinguishable from the noise floor above 3 kHz in the worn cylinder recording condition.

A lab technician assisted in recording participants during the first recording session. It is common practice in commercial sound engineering to adjust the gain for each performer/instrument during a recording session. Unfortunately, it is not common practice to do so for studies involving voice research. It was discovered after the first recording session that the gain had been adjusted for each singer. This error resulted in SPL measurements for the study being unsuitable for analysis. In order to adjust for this unforeseen issue, the noise floor for each of the microphone recordings during the first day of data collection was calculated and the level was adjusted in reference to the noise floor measurement from the second recording session (49.96 dB) by multiplying the audio signal in *Praat* using Equation 5 (Boersma & Weenink, 2018).

$$p = y * 10^{(\Delta L / 20)} \tag{5}$$

where ΔL is the level difference required in reference to the noise floor measurement. The RMS level difference (dB RMS) is shown in the Table 8. It is important to note that this step was specifically taken in order to visually represent the audio signals adequately. The errant SPL measurements did not affect the acoustic measurements used in this study, because each acoustic measurement was analyzed relative SIL's within a spectrum.

Table 8

Adjusted Levels Based on Noise Floor Estimates of Individual Recordings

Microphone	**Participant**	**Noise Floor (dB)**	**RMS Diff.**
C	S1	12.34	-37.63
C	S2	17.12	-32.84
C	B1	23.23	-26.74
C	T1	17.04	-32.92
C	B2	17.17	-32.79
C	B3	24.28	-25.68
C	M1	28.34	-21.62
C	B4	28.30	-21.66

3.5.2 Analysis of Vibrato

Vibrato measurements were collected from the digitized worn wax cylinder recording and a single microphone recording[16] using *Praat*. The *messa di voce* task was isolated for each recording and the middle 1-2 seconds of each sample were extracted from the audio file. The middle 1-2 seconds—the loudest

[16] The microphone placed 5 cm from the corner of the mouth of each participant was chosen so as to avoid analyzing artifacts resulting from the phonograph recording device.

segment of the task—were selected for analysis in order to ensure that participants were singing with vibrato. These samples resulted in voice output signals with at least 5 and at most 10 complete vibrato cycles (see Figure 28 & Figure 29) as per recommendations in past literature of at least 3-4 complete cycles or at least 500 ms of voice output signal for vibrato analysis (M. A. Guzman et al., 2012, p. 675; Herbst, Hertegard, Zangger-Borch, & Lindestad, 2017, p. 2; D. G. Miller & Horne, 2008, p. 33). The frequency modulated sinewave test signals were analyzed using an identical process.

Timestamps were documented in an *Excel* spreadsheet in order to ensure that the samples from the digitized wax cylinder audio files could be synchronized with the samples from the microphone recordings. The beginning timestamp of the *messa di voce* task was documented from the microphone recording in order to more accurately locate the same point in the digitized cylinder recording despite the raised noise floor that was found in digitized cylinder recordings. If the timestamp and sample duration resulted in a sample containing unequal segments of the vibrato cycle (i.e. the sample did not begin and end at an f_o peak in the vibrato cycle), the number of vibrato cycles following the beginning of voiced signal in the sample task was compared with the matching microphone recording and a sample was selected to include an identical number of vibrato cycles.

Frequency measurements were first extracted using *Praat*'s autocorrelation algorithm with the following parameters: Pitch range 75-1200 Hz,

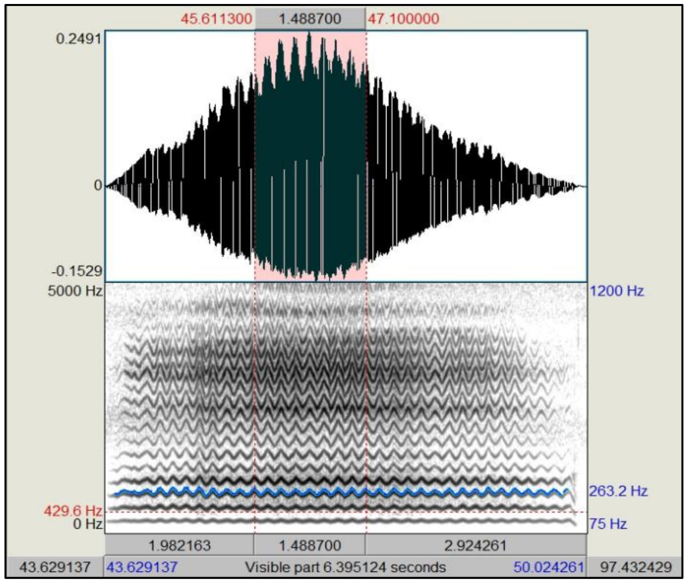


Figure 28. Microphone recording of a professional operatic tenor (T1) singing a *messa di voce* on C_4 (~261.63 Hz). This figure and the following figure demonstrate the position/duration of the vibrato samples extracted, matching that of the wax cylinder sample.

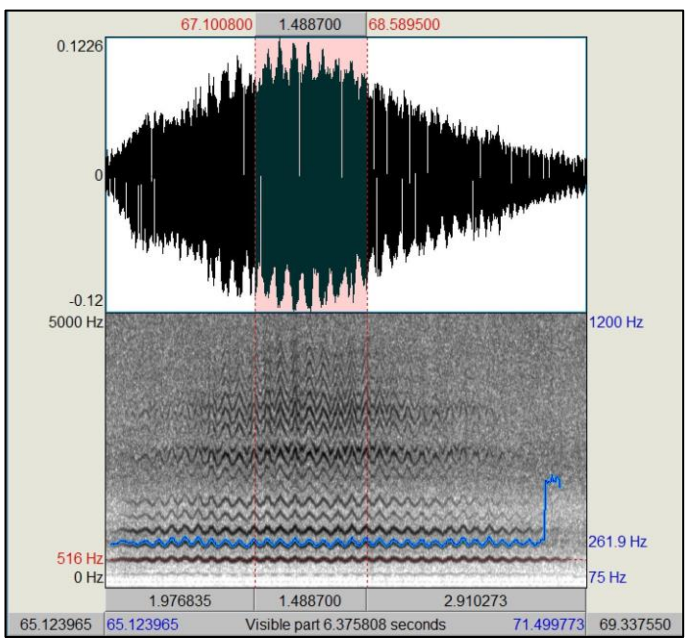


Figure 29. Digitized wax cylinder recording of a professional operatic tenor (T1) singing a *messa di voce* on C_4 (~261.63 Hz). This figure and the preceding figure demonstrate the position/duration of the vibrato samples extracted, matching that of the microphone sample.

Silence threshold 0.03, Voicing threshold 0.3, Octave cost 0.07, Octave-jump cost 0.35, Voiced/unvoiced cost 0.14. All of the parameters were chosen to follow those used by Herbst et al. (2017, p. 2) except for the pitch range. This alteration was chosen in order to use the same f_o analysis parameters throughout the study and future studies. More specifically, some soprano participants sang a C_6 for the first vocal task. Therefore, the maximum pitch range needed to exceed 1100 Hz.

Extracted frequency measurements were documented in a spreadsheet (*View & Edit-Pitch-Get pitch/Get minimum pitch/Get maximum pitch*), followed by the number of complete cycles and duration of the token (noted manually). The vibrato minimum and maximum frequencies over the duration of the sample were extracted by locating local f_o maxima and minima in the resulting data set (*Pitch-Pitch listing*). Vibrato rate was calculated in separate cell using Equation 6:

$$f_{vib} = \frac{x}{t} \tag{6}$$

where t is the duration of the token and x is the number of complete vibrato cycles within that token.

Average vibrato extent was calculated by first converting the each local f_o maximum and minimum to cents relation to middle C (C_4, ≈ 261.63 Hz)—a common unit with which to express vibrato excursion (Herbst et al., 2017, p. 2; D. G. Miller & Horne, 2008, p. 34)—with the following equation:

$$c[f_{o\_min}] = 1200\frac{\log(\frac{f_{o\_min}}{f_{o\_C_4}})}{\log(2)} \quad (7)$$

where f_{o_min} is the local fundamental frequency minimum (or, conversely, f_{o_max}) and $f_{o\_C_4}$ is the fundamental frequency of middle C (~261.63 Hz). The average fundamental frequency of each token's pitch contour as measured from *Praat*'s "Get pitch" command was also converted to cents using the same formula. Vibrato extent was calculated as the average absolute deviation $\overline{\Delta c}$ from the mean musical pitch of the pitch contour ($\bar{c}$) by using Equation 8:

$$\overline{\Delta c} = \frac{1}{n}\sum_{i=0}^{n-1}\frac{\left|\left(\bar{c} - c_i[f_{o\_min}]\right) - \left(\bar{c} - c_i[f_{o\_max}]\right)\right|}{2} \quad (8)$$

where $c_i[f_{o\_min}]$ is the indexed minimum fundamental frequency of the pitch contour of a phonatory cycle expressed in cents, and $c_i[f_{o\_max}]$ is the opposite peak of the same phonatory cycle. The vibrato extent calculations for each token were subsequently averaged to obtain a single vibrato extent measurement per participant for each recording condition.

3.5.3 Spectral Analysis

Spectral analysis was completed by first resampling all audio files in *Praat* to 22.05 kHz. The *Praat* resampling algorithm also applied a low pass filter (anti-aliasing) at half the new sampling rate (11 kHz). These steps were completed in order to approximate protocols used in past literature (Awan, 2015, p. 521; Kardach et al., 2002). Resampled audio files were subsequently segmented to isolate the sung phrase of "*Caro mio ben*" (approximately 15-23 seconds), the middle 1-2 seconds (6-8 vibrato cycles) of the *messa di voce* task (sustained [a]), and the middle 1-2 seconds of the chest voice task (female singers only).

The *Praat* script *ProsodyPro* (Xu, 2013) was edited to extract all spectral moments of the LTAS (window length: 0.005 s, Gaussian windowing function) in addition to Bio-Informational Dimensions (BIDs) already built into the algorithm (Xu, Kelly, & Smillie, 2013), and then executed (the edited script can be found in Appendix D). Acoustic measurements computed with this script included: *spectral mean (center of gravity)*, *spectral standard deviation*, *spectral skewness*, *spectral kurtosis*, L_1-L_2, *HNR (harmonic-to-noise ratio/harmonicity)*, and *cepstral peak prominence (CPP)*. The pitch range of the algorithm was set to 60 Hz – 1200 Hz and the maximum formant frequency was 5000 Hz for male participant samples and 5500 Hz for female participant samples as per the algorithm's manual.

This algorithm was chosen in order to further isolate audio samples efficiently and to confirm segmentation of audio files between microphone and digitized recordings (see Figure 30 for an example of segmentation in *ProsodyPro*). Doing so ensured that identical segments were compared between microphone and digitized recordings as all timestamps were saved in a separate

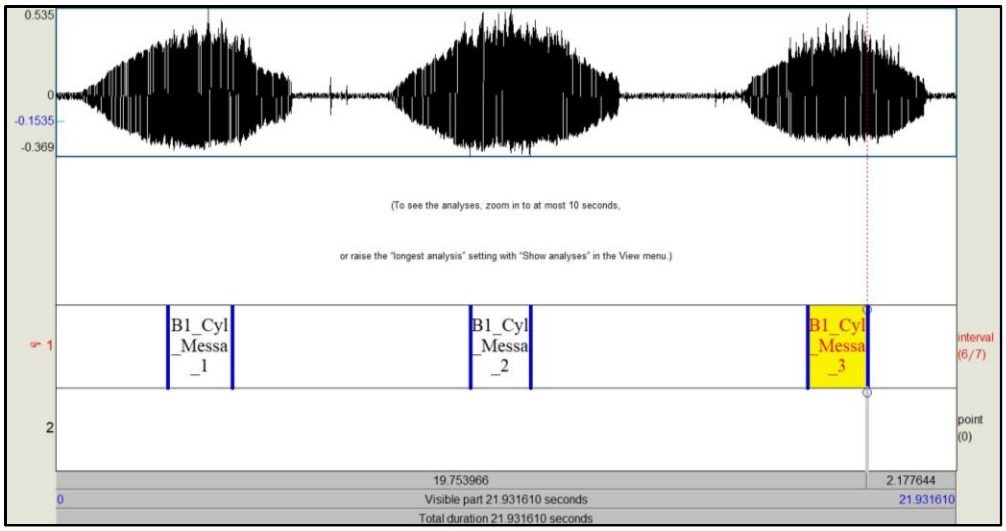


Figure 30. An example of segmentation in *Praat* using the *ProsodyPro* script (Xu, 2013).

text file. Spectral data were transferred to an *Excel* spreadsheet and acoustic measurements from the *messa di voce* task were averaged for each participant.

3.6 Statistical Analysis

The Shapiro-Wilk test and Levene's test for equality of error variances were first used to test for normality and heteroskedasticity, respectively. If homoskedasticity was found to be present in a dependent variable for both recording conditions, the spectral measurement was tested using a 2 x 2 mixed design Analysis of Variance (ANOVA) with recording condition as a two-level within-subjects factor (microphone, cylinder) and participant sex as a two-level between-subjects factor (male, female). Otherwise, in cases where Levene's test indicated the presence of heteroskedasticity, the non-parametric Wilcoxon Signed-ranks test was used (test value median = 0) to test for differences between recording conditions.

In order to make generalizations about historical recordings and, in the future, the singing by historical singers, variables that revealed significant main effects of recording condition were further studied using linear regression analysis. If the non-parametric test indicated significant differences based on recording condition, linear regression analysis with heteroskedasticity-consistent standard errors (Hayes & Cai, 2007) was performed. This comparison resulted in a series of models that could be used to study historical recordings.

All statistical analyses were calculated using SPSS Statistics, v. 25 software (IBM Armonk, New York). The confidence level for statistical analysis was set at 95% (α=0.05).

CHAPTER IV

RESULTS

4.1 System Analysis: Wax Cylinder Phonograph

4.1.1 Frequency-Modulated Sinewave Test Signal

Mean vibrato rate for the frequency-modulated sinewave test signal remained constant through all testing conditions (Table 9). Mean vibrato extent increased

Table 9

Mean Values for Frequency-Modulated Sinewave Test Signal Measurements

Measurement	Mean
Rate (Hz)	
Original	6.5
Digitized	6.5
Worn-Digitized	6.5
Extent (Cents)	
Original	54
Digitized	57
Worn-Digitized	59
Avg. f_0 (Hz)	
Original	440.00
Digitized	436.63
Worn-Digitized	436.76

once again, upon digitization after ten playbacks on period equipment. The average fundamental frequency was lower on wax cylinder recordings. No statistical analyses were performed, because the test signals were only recorded onto one cylinder.

Mean fundamental frequency (f_o) measurements between test conditions were calculated for the frequency-modulating test signals for a duration of approximately 4 seconds (Table 10). The f_o of the sinewave was observed to be lower on digitized wax cylinder recordings. Interestingly, while the maximum f_o was found to increase from the microphone recording condition (454 Hz) to the digitized wax cylinder condition (458 Hz), it decreased with simulated wear (455 Hz). Again, statistical analysis was not performed, because the test signals were only recorded onto one wax cylinder.

Table 10

Mean Frequency Values for the Frequency-Modulated Sinewave Test Signals

	Mean
Avg. f_0 (Hz)	
Original	440.00
Digitized	436.63
Worn-Digitized	436.76
Minimum f_0 Avg. (Hz)	
Original	426.11
Digitized	414.22
Worn-Digitized	414.53
Maximum f_0 Avg. (Hz)	
Original	453.89
Digitized	457.80
Worn-Digitized	455.10

4.1.2 Sinewave Test Signals

Pure sine tones (440 Hz) were segmented from downsampled test signal audio files using *ProsodyPro*. The *Praat* script extracted f_o measurements which were subsequently averaged over three tokens (22-25 complete cycles per token). in an *Excel* spreadsheet Table 11 lists average f_o measurements for three 440 Hz

Table 11.

Averaged Fundamental Frequency (f_o) Measurements for Three 440 Hz Sinewaves

Recording Condition	Mean f_o (Hz)	Min. f_o (Hz)	Max. f_o (Hz)
Test Signal	440.00	440.00	440.06
Microphone	439.99	439.93	440.01
Digitized Cylinder	436.73	427.32	446.27
Worn Cylinder	437.03	427.13	447.04

sine tones from the Test Signal Recording (Section 3.4.1). Observation indicates that f_o decreased from the microphone recording condition to the digitized wax cylinder recording condition, but increased slightly once the wax cylinder was worn. The f_o range increased with recording condition (see Figure 31). Statistical analysis was not performed, because the test signals were only recorded onto one cylinder.

Spectral analysis of the LTAS of each token indicated that the average difference between the levels of the first two harmonics in the spectrum (L_1-L_2), Harmonic-to-Noise Ratio (HNR), spectral skewness (M3), and spectral kurtosis (M4) were lower for the 440 Hz sine tone on digitized cylinder recordings (Table

12). The L_1-L_2 increased after simulated wear (i.e. the level of the second harmonic was lower

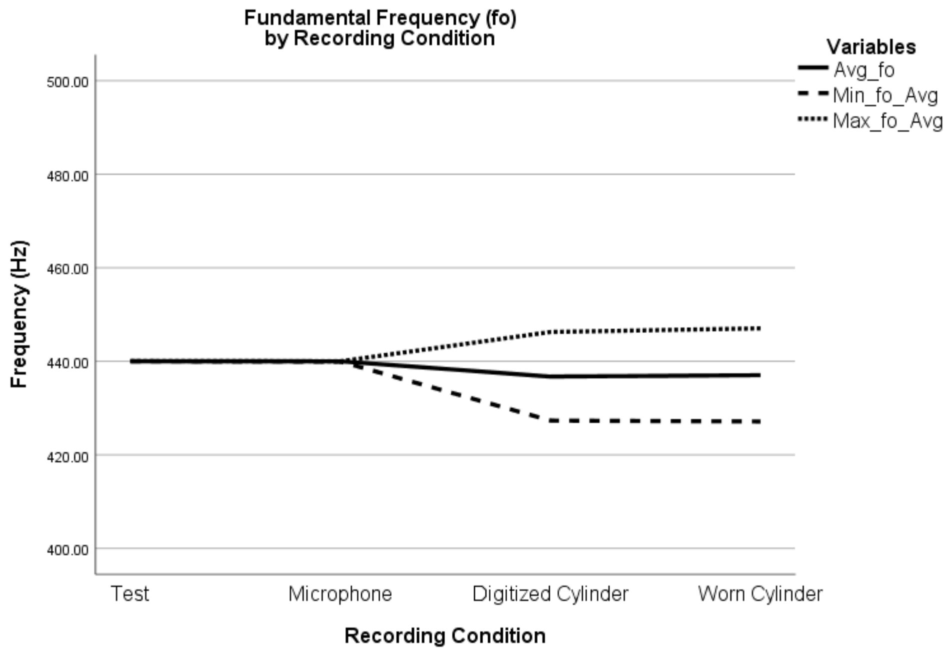


Figure 31. Line plot of fundamental frequency by recording condition/technology.

after 10 plays on period phonograph equipment). Theoretically, a pure sine tone should not include any overtones. This phenomenon will be discussed further in Chapter 5. HNR, spectral skewness, and spectral kurtosis all decreased further after simulated wear. Spectral mean and spectral standard deviation increased from the microphone recording to the digitized cylinder recording and, again, after simulated wear (Table 12).

Table 12

Averaged Spectral Measurements of the LTAS of Three Tokens of the 440 Hz Sinewave Test Signal

Recording Tech.	**L_1-L_2 (dB)**	**HNR (dB)**	**Spectral Mean (Hz)**	**Spectral SD (Hz)**	**Spectral Skewness**	**Spectral Kurtosis**
Test Signal	51.66	73.19	440.08	17.29	480.57	494230.14
Microphone	45.70	48.84	439.99	20.19	255.48	99170.46
Digitized Cylinder	23.61	22.33	462.06	372.48	19.71	431.64
Worn Cylinder	24.07	16.57	529.99	702.59	9.32	96.66

4.1.3 White Noise Test Signal

Spectral moment measurements and harmonic-to-noise ratio were calculated for the LTAS of the white noise test signal. The audio files were downsampled to 22.05 kHz prior to analysis in order to parallel post-processing that occurred on audio recordings of human subjects. As seen in Figure 32 and

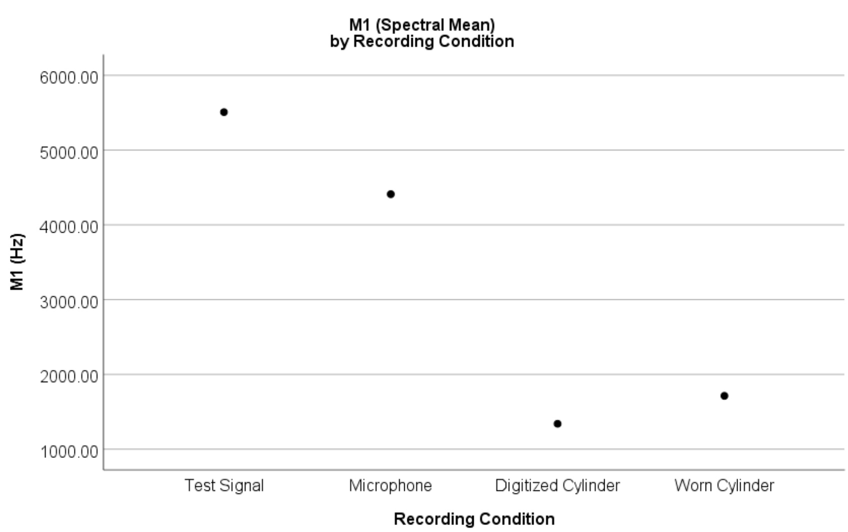


Figure 32. Changes in spectral mean (M1) for white noise (downsampled to 22.05 kHz) for the four recording conditions.

Table 13, spectral mean decreased from the microphone (40 cm distance) recording to the digitized wax cylinder recording, but increased after the cylinder was played ten times on period equipment and subsequently digitized. Spectral standard deviation followed the same pattern.

Table 13

Spectral Moments (of the LTAS) and HNR Measurements for the White Noise Test Signal

Recording Condition	Spectral Mean (Hz)	Spectral SD (Hz)	Spectral Skewness	Spectral Kurtosis	HNR (dB)
Test Signal	5507.55	3177.16	-0.003	-1.20	-6.40
Microphone	4409.98	3037.69	0.360	-0.98	-6.15
Digitized Cylinder	1339.32	1128.98	4.142	23.89	-2.54
Worn Cylinder	1712.69	1783.50	2.561	6.80	-2.95

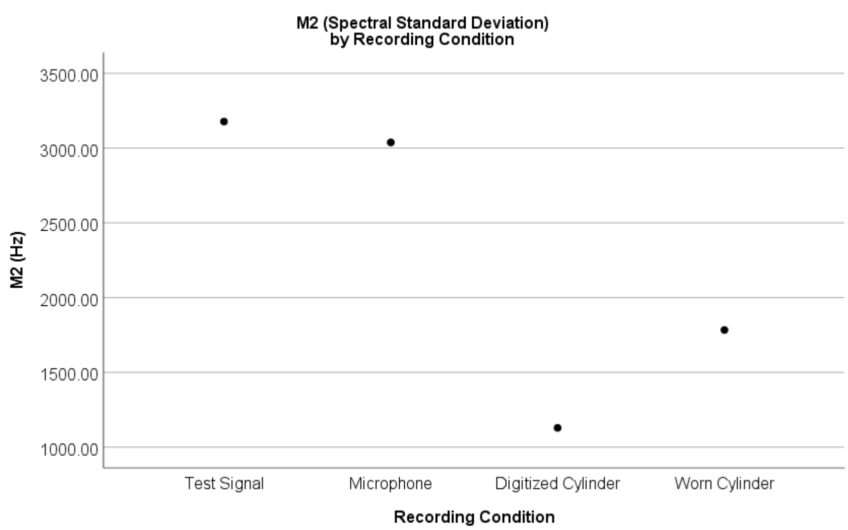


Figure 33. Changes in spectral standard deviation (M2) for white noise (downsampled to 22.05 kHz) for the four recording conditions.

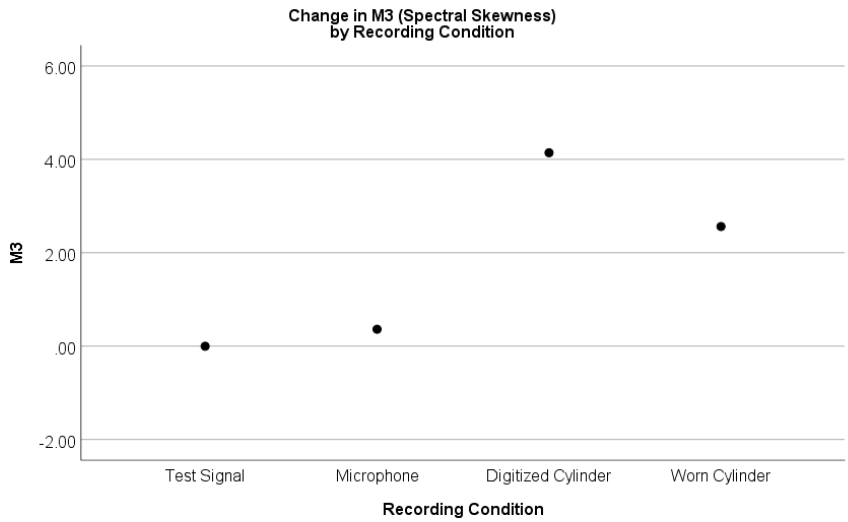


Figure 34. Changes in spectral skewness (M3) for white noise (downsampled to 22.05 kHz) for the four recording conditions.

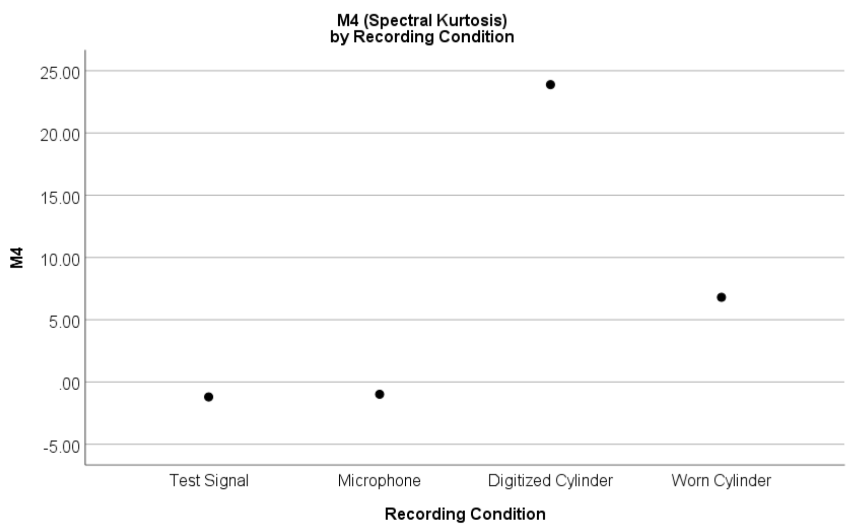


Figure 35. Changes in spectral kurtosis (M4) for white noise (downsampled to 22.05 kHz) for the four recording conditions.

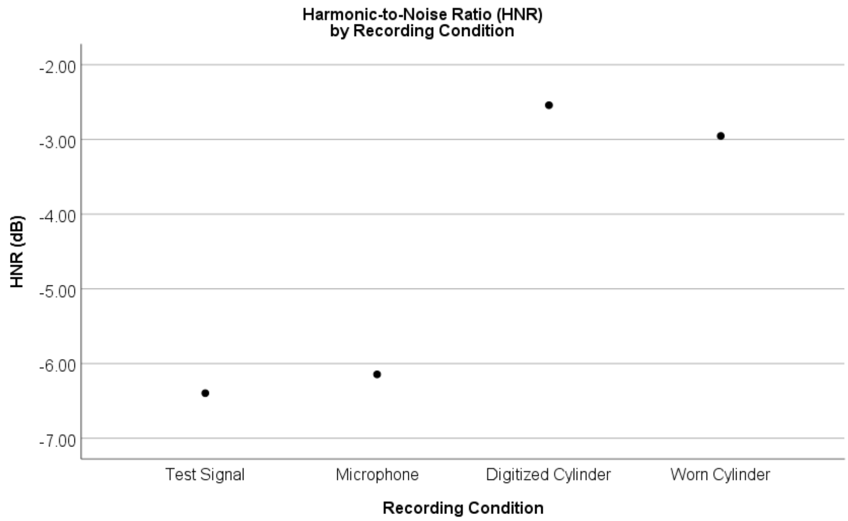


Figure 36. Changes in harmonic-to-noise ratio (HNR) for white noise (downsampled to 22.05 kHz) for the four recording conditions.

Spectral skewness was observed to increase from the microphone recording to the digitized wax cylinder recording, but dropped by nearly 50% after simulated wear. A similar pattern was found with spectral kurtosis and HNR measurements.

4.1.4 Sinesweep Test Signal

During data collection, it was observed that the Genelec speakers distorted the sinesweep test signals (see dotted lines in Figure 37). Anecdotally, the distortion was heard by this researcher as high pitches increasing and decreasing in frequency. While the LTAS of each recording condition was similar to the LTAS of the white noise test signal (Figure 38 & Figure 39), it was determined that the sinesweep data were unsuitable for spectral analysis.

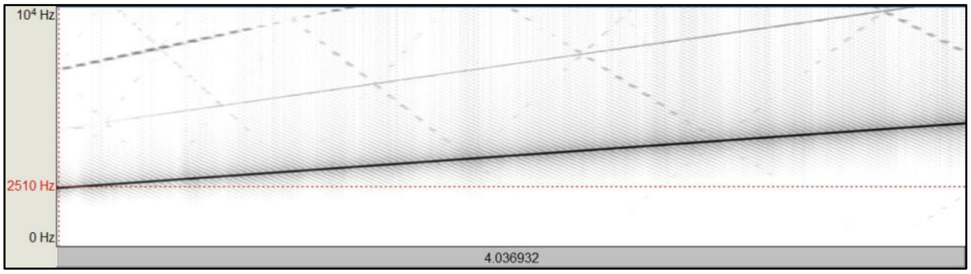


Figure 37. Distorted sinesweep from ~2.5 kHz - ~5 kHz as recorded by a flat-response omnidirectional microphone (Earthworks M30) placed exactly 40 cm in front of a Genelec 8030A nearfield monitor.

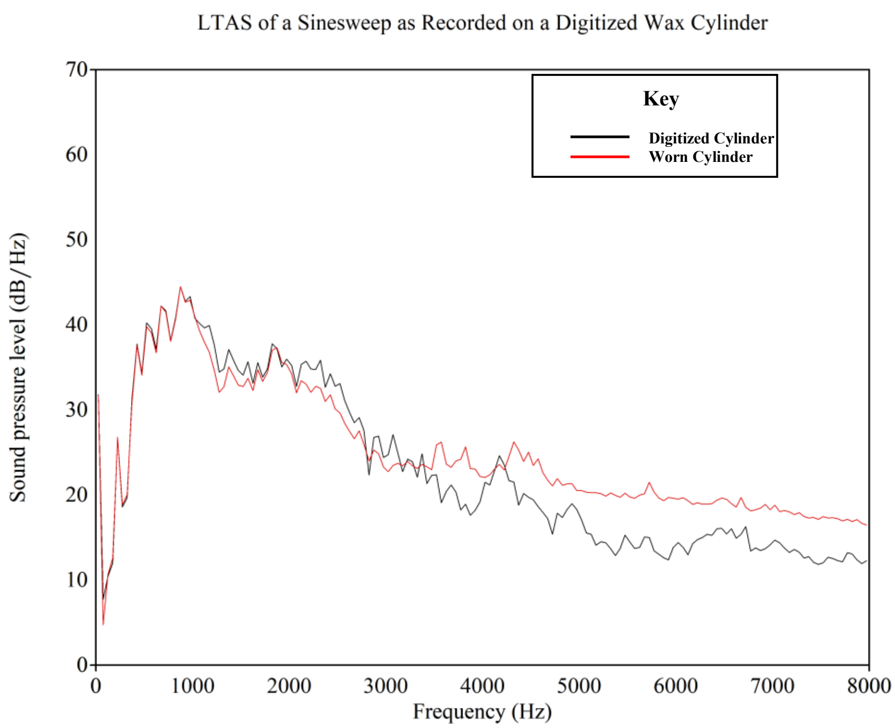


Figure 38. LTAS (50 Hz frequency resolution) of the first sinesweep test signal from 0-8000 Hz recorded on a wax cylinder. The black line represents the digitized wax cylinder (no prior playback) and the red line represents the wax cylinder digitized after simulated wear (10 playbacks).

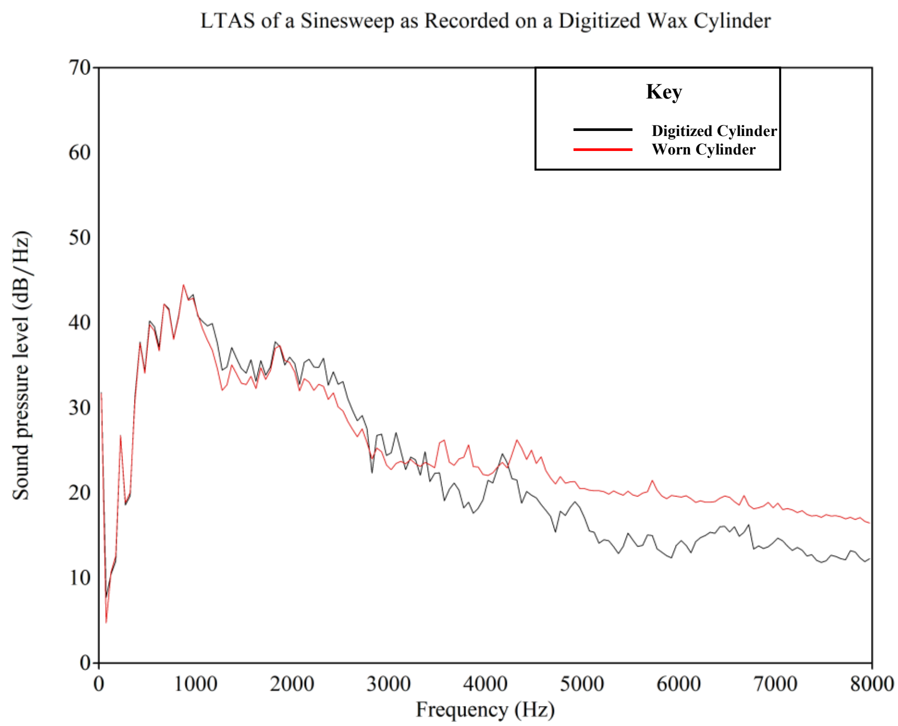


Figure 39. LTAS (50 Hz frequency resolution) of a 10-second white noise test signal (from Figure 12) recorded by a wax cylinder and digitized with an *Archéophone* with no previous plays (black line) and after 10 plays on a period Edison Fireside Phonograph (red line).

4.1.5 Noise Floor

The calibration signal was analyzed using *Praat's Query-Get Intensity* function. Due to noise from the wax cylinder equipment during recording, a silent portion of the calibration file was selected in order to calculate the noise floor at the microphone (5 cm mouth-to-microphone distance) and at the left microphone (40 cm mouth-to-microphone distance). The calibration recording for the 40 cm microphone indicated that there was a 35.86 dB difference between the recorded 1 kHz calibration signal and the 114 dB signal produced by the sound level calibrator. There was a 37.96 dB difference between the calibration signal produced and the calibration recording for the 5 cm microphone. These level differences were due to gain adjustments made prior to all recordings at the mixing board.

Digitized wax cylinder noise floor measurements were taken from a silent segment of the test signal recording. Digitized cylinder levels were adjusted to correct for the -9 dB gain adjustment during transfer. Refer to Table 14 for all noise floor measurements and level adjustments. Level adjusted noise floor measurements indicated that the SIL of the noise floor remained nearly constant between microphone recordings measured at different distances. The level adjusted noise floor measurements indicated that the noise floor increased from 52.13 dB (40 cm microphone) to 53.95 dB on digitized wax cylinders with no prior playback. Simulating wear on the wax cylinder by playing it on period phonograph equipment and subsequently digitizing the cylinder resulted in an increase in the SIL of the noise floor of 4.62 dB.

Table 14

Noise Floor Measurements Based on Measured Sound Intensity Level (SIL) with Level Adjustments

Recording Condition	**SIL (dB)**	**Adj. SIL (dB)**	**Adjustment**
Microphone (40 cm)	16.27	52.13	35.86
Microphone (5 cm)	15.15	53.10	37.96
Test Signal Recording	17.65	-	-
Digitized Cylinder	44.95	53.95	9.00
Worn Cylinder	49.57	58.57	9.00

4.1.6 Signal-to-Noise Ratio (SNR)

Due to the observed differences in the noise floor being relatively small, as well as the noise floor measurements for the research lab as measured by the microphones being higher than past reported measurements (30-40 dB), the signal-to-noise ratio was calculated across recording conditions. SNR measurements were determined by the use of the following equation:

$$SNR = L_S - L_N \tag{9}$$

where L_S is the SIL of the test signal and L_N is the SIL of the noise floor.

The microphone (5 cm) was level adjusted based on the Distance Law as suggested in Švec & Granqvist (2018) and as demonstrated by Braxton, Roginska, & Gill (2013) (see Equation 10). Both microphones were level adjusted based on the 114 dB 1 kHz calibration signal. Digitized wax cylinder signals were level adjusted to account for the -9 dB gain adjustment during the digitization procedure.

$$\Delta L = 20 \cdot \log_{10} \frac{L_{d_2}}{L_{d_1}} \tag{10}$$

where L_{d_1} and L_{d_2} are the sound intensity levels at two different distances (d_1 and d_2). The resultant calculations indicated that the SNR decreased from microphone (64.16-65.79 dB) to digitized wax cylinder recordings (37.49 dB), and decreased further upon simulated wear by period phonograph equipment (33.53 dB).

Table 15

Signal-to Noise Ratio (SNR) Measurements for each Recording Condition

Recording Condition	**SNR (dB)**
Microphone (40 cm)	64.16
Microphone (5 cm)	65.79
Digitized Cylinder	37.49
Digitized Worn Cylinder	33.53

Note: Microphone (5 cm) was level adjusted based on the Distance Law

4.2 Vibrato Analysis

The Shapiro-Wilks test indicated that vibrato rate and vibrato extent were normally distributed during the *messa di voce* task for both recording conditions; however, average fundamental frequency ($\bar{f_o}$) was non-normally distributed. Additionally, Levene's test of equality of error variances revealed the presence of heteroskedasticity for $\bar{f_o}$ (cylinder). See Table 16 for descriptive statistics of vibrato measurements.

A mixed-design ANOVA with recording condition (microphone, cylinder) as a within-subjects factor and participant sex (male, female) as a between-subjects factor indicated that there was no significant main effect of recording condition ($F_{(1,18)} = 0.010$, $p = 0.923$) or participant sex ($F_{(1,18)} = 2.162$, $p = 0.159$)

Table 16

Descriptive Statistics for Vibrato Measurements during the messa di voce Task

Measurement	All	Male	Female
Vibrato Rate (Hz)			
Microphone	5.3 (0.5)	5.1 (0.5)	5.4 (0.5)
Cylinder	5.3 (0.5)	5.1 (0.5)	5.4 (0.5)
Vibrato Extent (Cents)			
Microphone***	100 (31)	100 (36)	100 (27)
Cylinder	103 (30)	102 (35)	104 (27)
Avg. f_o (Hz)			
Microphone	390 (137)	256 (5)	523 (3)
Cylinder	389 (137)	256 (5)	522 (3)

Note: Data shown as mean (standard deviation)
*** Indicates significant group difference ($p < .001$) from cylinder.

on vibrato rate. No significant interaction between recording condition and participant sex ($F_{1,18}$ = 1.990, p = 0.175) was observed.

Analysis of vibrato extent indicated no significant interaction between recording condition and participant sex was found for vibrato extent ($F_{(1,18)}$ = 1.648, p = 0.216). Additionally, no significant main effect of participant sex ($F_{(1,18)}$ = 0.004, p = 0.950) was found; however, there was a significant main effect of recording condition on vibrato extent ($F_{(1,18)}$ = 37.860, $p < 0.001$, η_p^2 = 0.678). That is to say, group vibrato extent was significantly higher in the digitized wax cylinder recording condition than in the microphone recording condition.

There is a known nonlinear relationship between pitch and frequency—frequency and pitch do not increase at the same rate. As such, it was expected that $\bar{f_o}$ would be significantly different based on participant sex (see Figure 40)

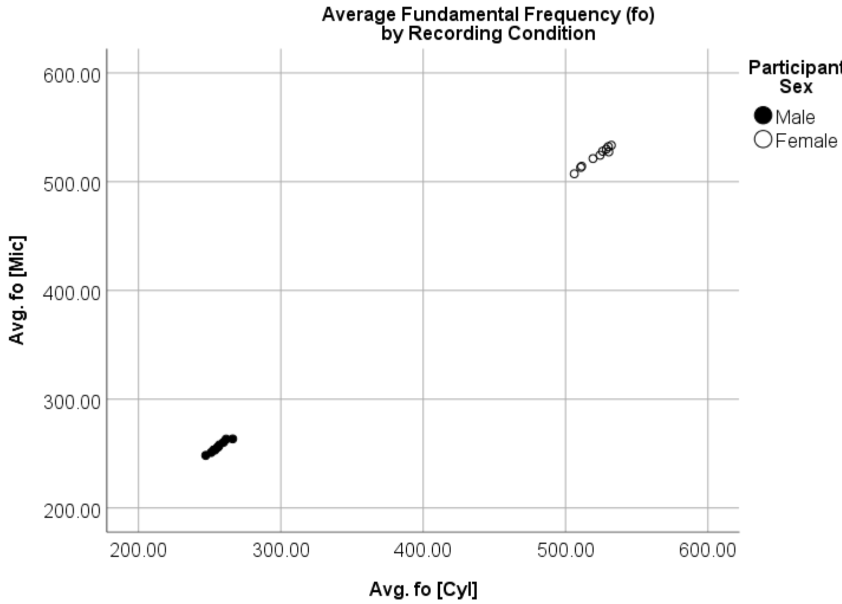


Figure 40. Scatter plot of fundamental frequency (f_o) by recording condition during the *messa di voce* task.

and participants were split by sub-group based on participant sex. Separate non-parametric Wilcoxon Signed-Ranks tests were used to study the differences between $\bar{f_o}$ measured by two different recording conditions. Results indicated that $\bar{f_o}$ between recording condition was not significantly different for male (Z = -1.326, p = 0.185) or female (Z = -1.784, p = 0.074) participants.

4.3 Spectral Analysis: In-Context Singing

Spectral measurements of the LTAS of the in-context singing task ("Caro mio ben") was first analyzed with the Shapiro-Wilk test of normality. L_1-L_2, harmonic-to-noise ratio (HNR), and spectral skewness (M3) were found to be normally distributed. The normality plot of the wax-cylinder spectral mean (M1) measurements was found to have an outlier (see Figure 41). Specifically, the in-context singing spectral mean in the digitized cylinder recording was roughly

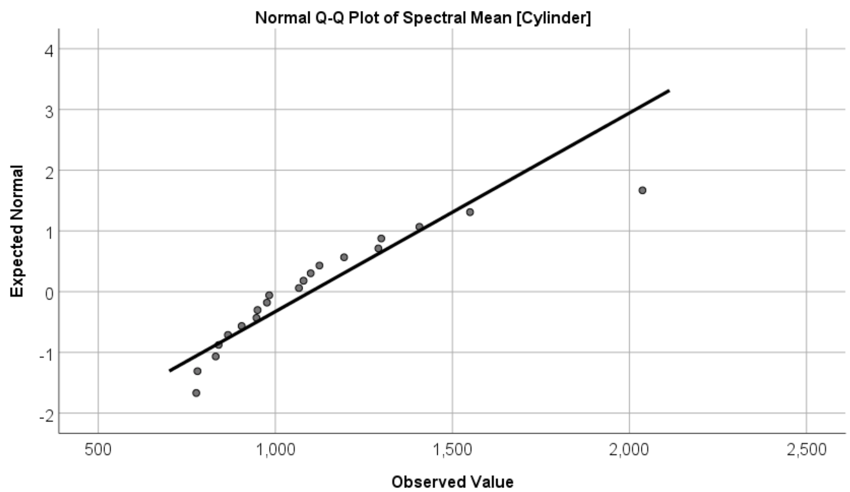


Figure 41. Normal Q-Q Plot of cylinder spectral mean measurements. Note the outlier beyond 2,000 Hz (B5).

double the value (2036.93 Hz) of the same sample in the microphone recording (1055.50 Hz). The B5 sample was subsequently found to have high frequency distortion during the in-context singing task—most likely caused by a tracking error during data collection. As such, it was excluded from the data set (see Figure 42).

Following the correction to the data set, the Shapiro-Wilk test was run again for all spectral measurements. It indicated that group means for spectral standard deviation (cylinder) and spectral kurtosis (both conditions) were not normally distributed, while the group mean for spectral mean was normally distributed for both recording conditions with the exclusion of the B5 sample.

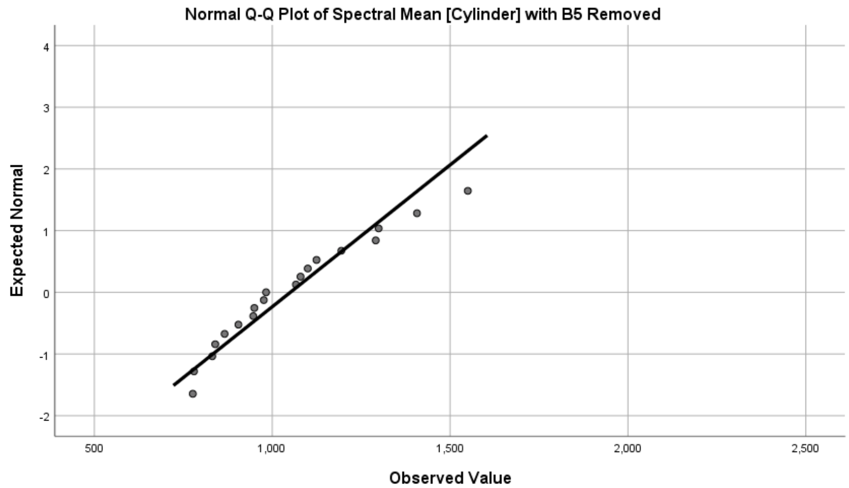


Figure 42. Normal Q-Q Plot of cylinder spectral mean measurements after the exclusion of the B5 sample.

Levene's test of equality of error variances indicated that HNR (cylinder) and spectral SD (cylinder) demonstrated heteroscedasticity. As such, HNR and spectral SD were analyzed later with the non-parametric Wilcoxon Signed-Ranks test. All other spectral measurements were tested using a 2 x 2 mixed design ANOVA with recording condition as a two-level within-subjects factor (microphone, cylinder) and participant sex as a two-level between-subjects factor (male, female).

Analysis of the L_1-L_2 during in-context singing indicated no significant interaction between participant sex and recording condition ($F_{(1, 17)} = 1.194$, $p = 0.290$). Significant main effects of both recording condition ($F_{1,17)} = 176.388$, $p < 0.001$, $\eta_p^2 = 0.912$) and participant sex ($F_{(1.17)} = 31.308$, $p < 0.001$, $\eta_p^2 = .648$)

Table 17

Descriptive Statistics of Spectral Measurements of the LTAS of In-Context Singing (Phrase from "Caro mio ben") by Recording Condition and Participant Sex

Measurement	All	Male	Female
L_1-L_2 (dB)			
Microphone	5.83 (6.03) *	0.78 (2.23) [s]	10.38 (4.44)
Cylinder	-5.26 (5.63)	-11.80 (2.93)	-1.57 (4.88)
Harmonic-to-Noise Ratio (HNR) [n]			
Microphone	21.83 (4.01) *	18.51 (2.49)	24.81 (2.41)
Cylinder	10.30 (4.34)	10.49 (2.56)	10.32 (5.64)
Spectral Mean (Hz)			
Microphone	902.56 (151.79) *	907.60 (163.40)	898.02 (149.32)
Cylinder	1050.84 (217.59)	1038.92 (138.30)	1086.67 (273.27)
Spectral Standard Deviation (Hz) [n]			
Microphone	607.11 (144.84) *	650.06 (156.09)	568.45 (129.59)
Cylinder	1010.72 (350.80)	915.89 (170.50)	1100.64 (273.27)
Spectral Skewness			
Microphone	2.42 (0.75) *	2.17 (0.81)	2.65 (0.64)
Cylinder	4.25 (1.46)	3.90 (1.53)	4.43 (1.45)
Spectral Kurtosis			
Microphone	9.65 (7.54) *	7.34 (7.58)	11.72 (7.24)
Cylinder	28.90 (20.38)	26.23 (19.82)	29.93 (21.88)

Note: Data shown as mean (standard deviation).
* Indicates significant difference ($p < .05$) from cylinder.
[s] Indicates significant difference ($p < .05$) between males and females regardless of recording condition.
[n] Indicates non-parametric testing due to heteroskedasticity.

were found, indicating that L_1-L_2 was significantly lower in digitized wax cylinder recording samples compared to equivalent microphone recording samples. Additionally, the analysis demonstrated that L_1-L_2 was significantly higher for female participants during in-context singing.

Analysis of spectral mean indicated no significant interaction between recording condition and participant sex ($F_{(1, 17)} = 0.730$, $p = 0.405$) and no significant main effect of participant sex ($F_{(1, 17)} = 0.209$, $p = 0.653$). A significant main effect of recording condition on spectral mean was observed ($F_{(1, 17)} = 8.580$, $p < 0.01$, $\eta_p^2 = 0.335$). That is to say, spectral mean was significantly higher in digitized wax cylinder recording samples during in-context singing.

No significant interaction between participant sex and recording condition for spectral skewness was observed during steady-state singing ($F_{(1.17)} = 0.031$, $p = 0.862$), nor was a significant main effect of participant sex found ($F_{(1.17)} = 0.907$, $p = 0.354$). A significant main effect of recording condition was found ($F_{(1.17)} = 41.985$, $p < 0.001$, $\eta_p^2 = 0.712$). This result indicated that spectral skewness was significantly greater in digitized wax cylinder recordings than in microphone recordings.

Analysis of spectral kurtosis during in-context singing revealed no significant interaction between participant sex and recording condition ($F_{(1.17)} = 0.065$, $p = 0.801$). No significant main effect of participant sex was found ($F_{(1.17)} = 0.321$, $p = 0.578$). A significant main effect of recording technology was observed ($F_{(1.17)} = 19.991$, $p < 0.001$, $\eta_p^2 = 0.540$), which suggested that M4 was significantly higher in digitized wax cylinder recordings than in microphone recordings.

Spectral measurements for which at least one recording condition demonstrated heteroscedasticity were analyzed using the Wilcoxon Signed-Ranks test. While previous analysis looked at the main effects and interaction of participant sex and recording condition on spectral measurements, this type of analysis is not available when using non-parametric analysis. As such, only the differences between spectral measurements and recording condition were analyzed when tests revealed heteroscedasticity. HNR was significantly lower in digitized wax cylinder recordings (Z = -3.823, p < 0.001). Spectral SD was found to be significantly higher in digitized wax cylinder recordings (Z =-3.823, p < 0.001). Figure 43 and Figure 44 demonstrate differences in dispersion pattern between male and female participants for HNR and spectral SD.

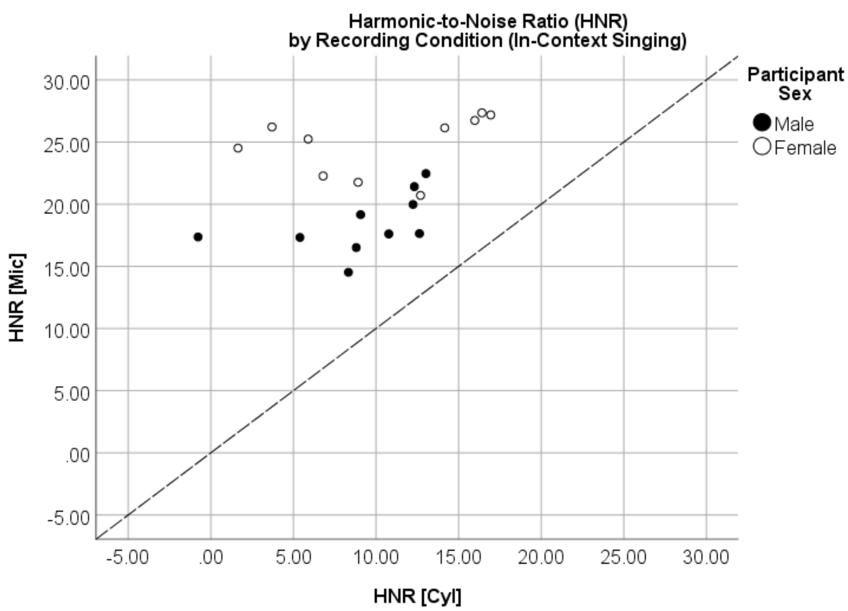


Figure 43. Scatter plot of the harmonic-to-noise ratio (HNR) during in-context singing on two different recording conditions. Participant sex highlighted to show differences between recording condition sub-group for HNR. Dashed line represents a one-to-one relationship between recording conditions.

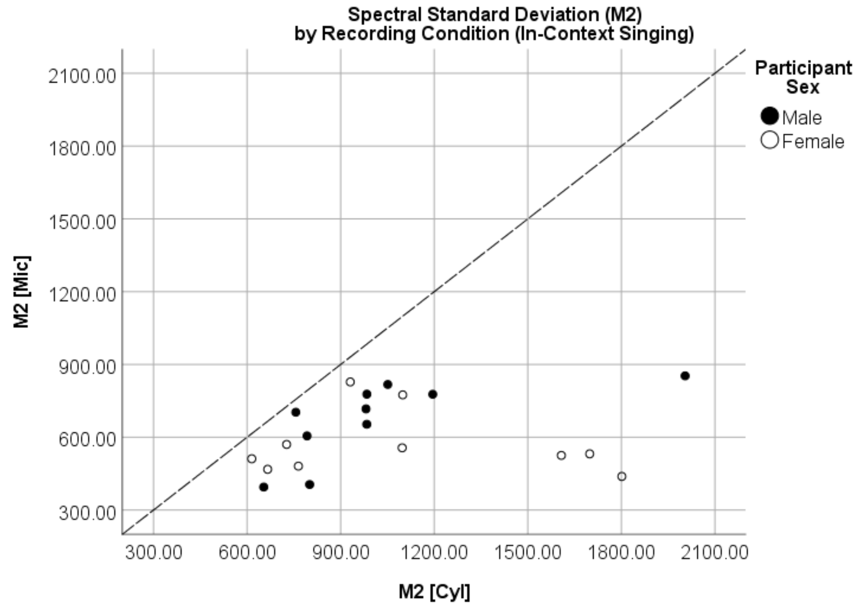


Figure 44. Scatter plot of the spectral standard deviation (M2) during in-context singing on two different recording conditions. Participant sex labeled to highlight differences in the dispersion of female participant M2 during the cylinder recording condition. Dashed line represents a one-to-one relationship between recording conditions.

4.4 Spectral Analysis: Steady-State Singing

Spectral measurements of the LTAS of the middle portion of the steady-state singing task (*messa di voce* task) were first tested for normality using the Shapiro-Wilk test. L_1-L_2 (cylinder) and spectral kurtosis under both recording conditions were found to be non-normally distributed. Levene's test for equality of error variances revealed the presence of heteroscedasticity for HNR (cylinder), spectral SD (cylinder), and spectral kurtosis (cylinder) measurements.

A 2 x 2 mixed design ANOVA was conducted on L_1-L_2 with recording condition (cylinder, microphone) as a within-subjects factor and participant sex (male, female) as a between-subjects factor. Analysis of L_1-L_2

Table 18

Descriptive Statistics of Spectral Measurements of the LTAS of Steady-State Singing (Middle 1-2 s of messa di voce Task) by Recording Condition and Participant Sex

Measurement	All	Male	Female
L1-L2 (dB) [i]			
Microphone	-4.39 (7.20) *	-9.29 (4.61) [s]	0.51 (5.91)
Cylinder	-11.58 (16.40)	-26.72 (4.80)	3.55 (5.96)
Harmonic-to-Noise Ratio (HNR) [n]			
Microphone	23.20 (4.37) *	19.93 (3.63)	26.47 (1.82)
Cylinder	14.57 (3.81)	14.35 (2.22)	14.78 (5.06)
Spectral Mean (Hz)			
Microphone	1064.62 (189.62)	1118.01 (213.39)	1011.22 (155.03)
Cylinder	1028.57 (220.87)	1070.93 (194.84)	1070.93 (247.18)
Spectral Standard Deviation (Hz) [n]			
Microphone	634.37 (219.73) *	716.99 (186.05)	551.75 (228.35)
Cylinder	852.21 (291.16)	870.21 (161.11)	834.22 (390.24)
Spectral Skewness			
Microphone	2.42 (1.29) *	1.65 (0.84) [s]	3.19 (1.22)
Cylinder	4.97 (2.31)	3.62 (1.79)	6.31 (2.01)
Spectral Kurtosis [n]			
Microphone	12.47 (16.76) *	4.17 (6.32)	20.76 (20.00)
Cylinder	44.54 (39.21)	24.46 (23.14)	64.63 (42.59)

Note: Data shown as mean (standard deviation)
* Indicates significant group difference ($p < .05$) from cylinder.
[s] Indicates significant difference ($p < .05$) between males and females regardless of recording condition.
[i] Indicates significant interaction between recording condition and participant sex.
[n] Indicates non-parametric testing due to heteroskedasticity.

indicated a statistically significant interaction between participant sex and recording condition ($F_{(1.18)} = 96.961$, $p < 0.001$, $\eta_p^2 = 0.843$). Significant main effects of both participant sex ($F_{(1.18)} = 86.161$, $p < 0.001$, $\eta_p^2 = 0.827$) and recording condition ($F_{(1.18)} = 47.846$, $p < 0.001$, $\eta_p^2 = 0.727$) were observed.

Post-hoc analysis completed using Fisher's Least Significant Difference t-test (LSD) revealed that the difference between L_1-L_2 for female participants during different recording conditions was not significant ($p = 0.212$). In other words, L_1-L_2 was significantly lower for male participants for the cylinder recording condition, but L_1-L_2 was not significantly different for female participants between recording conditions (Figure 45). Additionally, L_1-L_2 was significantly higher for female participants than for male participants in both recording conditions.

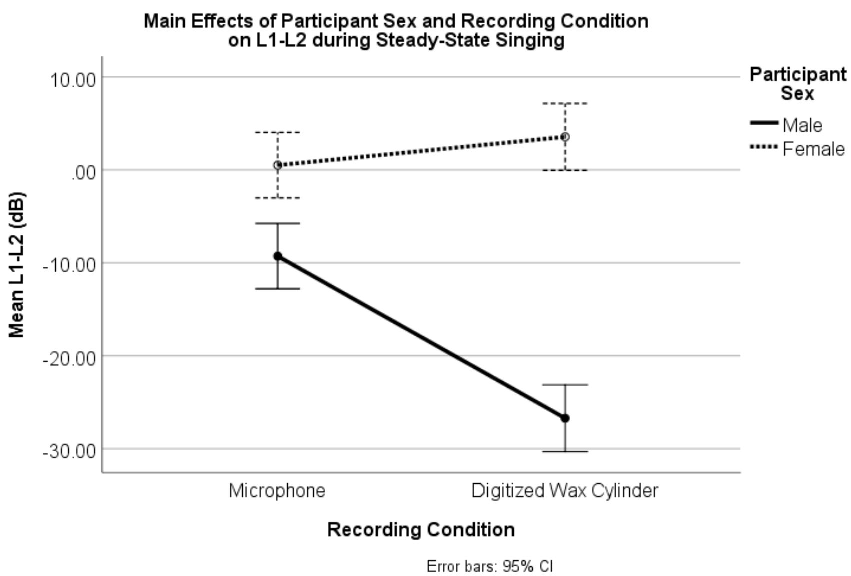


Figure 45. Main effects and interaction of participant sex and recording condition on the difference between the levels of the first two spectral peaks during steady-state singing (*messa di voce* task).

No statistically significant interaction was found between participant sex and recording condition for spectral mean during steady-state singing ($F_{(1.18)}$ =

0.043, $p = 0.837$). No main effects of participant sex ($F_{(1.18)} = 1.631$, $p = 0.218$) or recording condition ($F_{(1.18)} = 0.463$, $p =0.505$) were observed.

Spectral skewness was subjected to a 2 x 2 mixed design ANOVA with two levels of a within-subjects factor (recording condition) and two levels of a between-subjects factor (participant sex). Analysis of spectral skewness during steady-state singing revealed no statistically significant interaction between participant sex and recording condition ($F_{(1.18)} = 2.495$, $p = 0.132$). Significant main effects of participant sex ($F_{(1.18)} = 13.036$, $p = 0.01$, $\eta_p^2 = 0.420$) and recording condition were observed ($F_{(1.18)} = 50.344$, $p < 0.001$, $\eta_p^2 = 0.737$). This

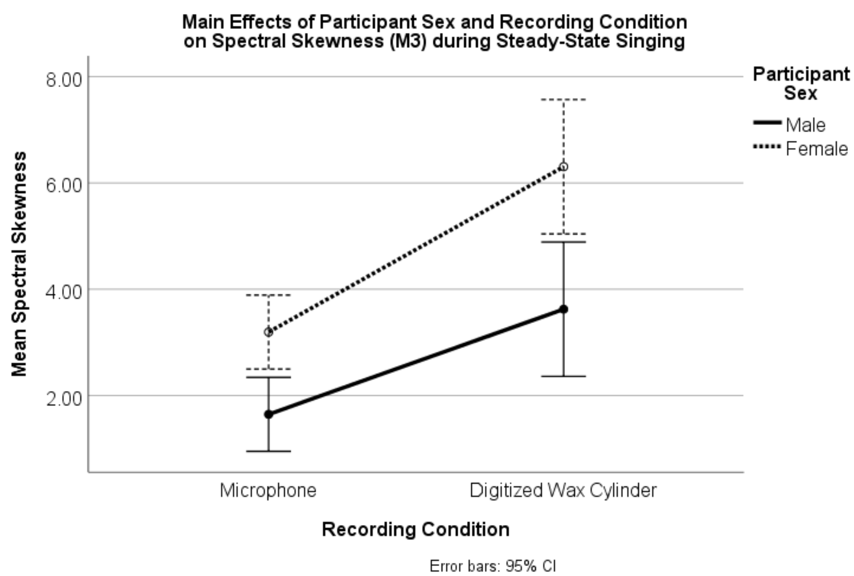


Figure 46. Main effects and interaction of participant sex and recording condition on the spectral skewness (M3) during steady-state singing (*messa di voce* task).

result indicated that M3 was significantly larger in digitized wax cylinder recordings than in microphone recordings, and that M3 was significantly larger for female participants than for male participants (Figure 46).

The three measurements for which the recording condition sub-group demonstrated heteroscedasticity were analyzed using the non-parametric Wilcoxon Signed-Ranks test. HNR was found to be significantly lower in digitized wax cylinder recordings than in microphone recordings ($Z = -3.920$, $p < 0.001$). Visual inspection of the data (see Figure 47) suggested that HNR_{Fem} decreased by as much as 17 dB in digitized wax cylinder recordings.

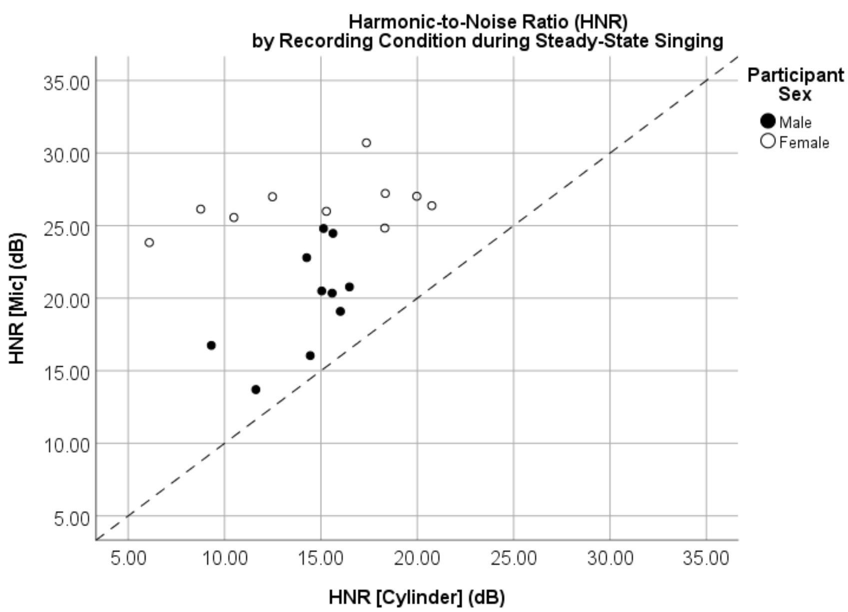


Figure 47. Scatter plot of the harmonic-to-noise ratio (HNR) during steady-state singing on two different recording conditions. Participant sex highlighted to show a potential interaction between recording condition and participant sex for HNR. Dashed line represents a one-to-one relationship between recording conditions.

Non-parametric analysis of spectral SD indicated significant differences due to recording condition ($Z = -2.875$, $p < 0.01$). Namely, spectral SD was significantly higher in digitized wax cylinder recordings than in microphone recordings. As seen in Figure 48, the range of female participant spectral SD in

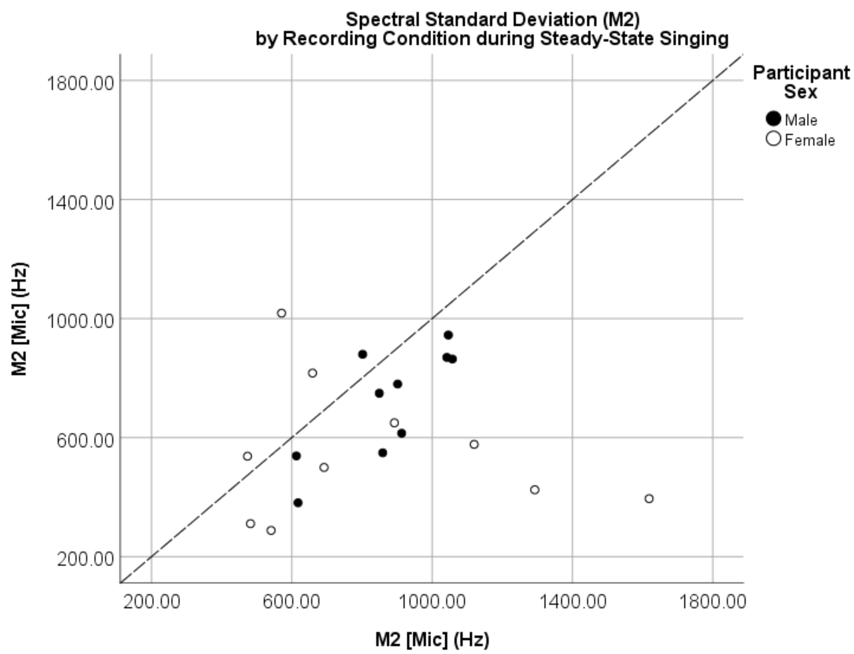


Figure 48. Scatter plot of the spectral standard deviation (M2) during steady-state singing on two different recording conditions. Participant sex labeled to highlight differences in the dispersion of female participant M2. Dashed line represents a one-to-one relationship between recording conditions.

digitized wax cylinder recordings was found to be larger than the range of male participant spectral SD from either recording condition.

Spectral kurtosis was observed to be significantly greater in the digitized wax cylinder recording condition than in the microphone recording condition ($Z = -3.771$, $p < 0.001$). While visual inspection of the relationship between spectral kurtosis measured from different recording conditions suggested the presence of an outlier (see Figure 49), it was decided to keep the datum point in the data set

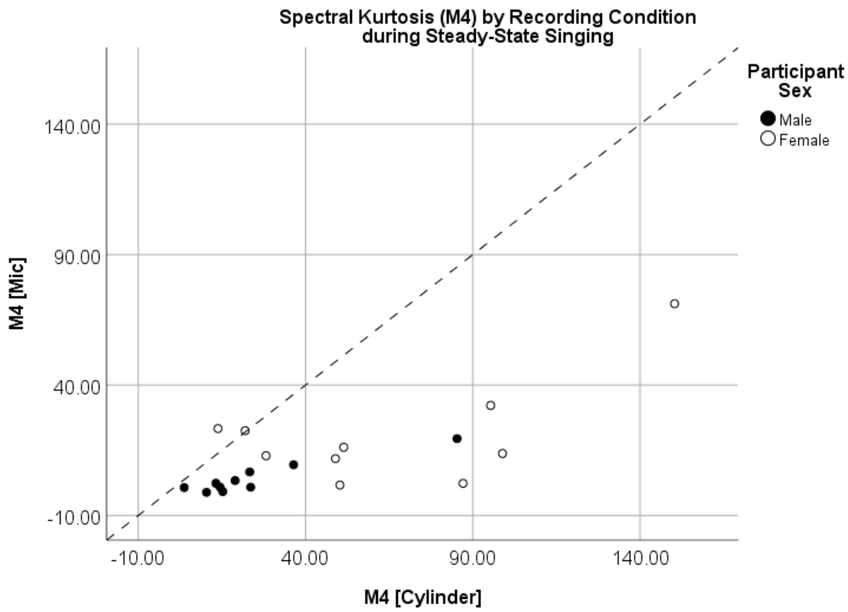


Figure 49. Scatter plot of the spectral kurtosis (M4) during steady-state singing on two different recording conditions. Dashed line represents a one-to-one relationship between recording conditions.

for steady-state singing, because the spectral kurtosis measurement was consistently high across both recording conditions. It should be noted that the outlier corresponded with participant S3 and was not related to the outlier found during in-context singing. Additionally, no audible distortions or tracking errors (including local degradations) were found upon inspection of the audio sample.

4.5 Spectral Analysis: Female Chest Voice Singing

Spectral measurements of the LTAS for the chest voice task (female participants only) were first tested for normality using the Shapiro-Wilk test. Spectral SD (microphone) and spectral kurtosis (cylinder) were found to be non-normally distributed. Levene's test for equality of error variances indicated that spectral mean (microphone) demonstrated heteroscedasticity. All spectral

Table 19

Descriptive Statistics of Spectral Measurements of the LTAS during the Chest Voice Task (Female Participants only) Organized by Recording Condition and Voice Type

Measurement	All	Soprano	Mezzo Soprano
L_1-L_2 (dB)			
Microphone	-1.74 (2.44) *	-0.51 (2.79)	-2.96 (1.35)
Cylinder	-17.74 (5.88)	-16.50 (7.92)	-18.99 (3.33)
Harmonic-to-Noise Ratio (HNR)			
Microphone	19.19 (2.16) *	19.15 (3.16)	19.23 (0.66)
Cylinder	6.97 (4.24)	6.30 (4.79)	7.63 (4.05)
Spectral Mean (Hz)^n			
Microphone	953.34 (139.61) *	900.53 (39.12)	1006.15 (188.01)
Cylinder	1359.41 (370.04)	1396.89 (434.24)	1321.92 (340.61)
Spectral Standard Deviation (Hz)			
Microphone	542.49 (171.64) *	494.04 (143.11)	590.95 (199.84)
Cylinder	1492.60 (522.65)	1568.99 (565.37)	1416.22 (518.64)
Spectral Skewness			
Microphone	3.36 (0.87)	3.57 (0.81)	3.15 (0.97)
Cylinder	3.78 (1.85)	3.72 (2.29)	3.84 (1.57)
Spectral Kurtosis			
Microphone	21.26 (13.87)	26.87 (15.56)	15.64 (10.59)
Cylinder	21.09 (22.99)	21.91 (29.99)	20.27 (16.96)

Note: Data shown as mean (standard deviation)
* Indicates significant group difference ($p < .05$) from cylinder.
^{n} Indicates non-parametric testing due to heteroskedasticity.

measurements were analyzed using a 2 x 2 mixed design ANOVA with recording technology as a two-level within-subjects factor (microphone, cylinder) and voice type as a two-level between-subjects factor (soprano, mezzo soprano).

Analysis of the L_1-L_2 during female chest voice singing demonstrated no significant interaction between voice type and recording condition ($F_{(1.8)} = 0.000$, $p = 0.986$). Likewise, no significant main effect of voice type was found ($F_{(1.8)} = 0.969$, $p = 0.354$). A statistically significant main effect of recording condition for L_1-L_2 was observed($F_{(1.8)} = 125.567$, $p = 0.001$, $\eta_p^2 = 0.940$). That is to say, L_1-L_2 was significantly lower when measured from the digitized wax cylinder recording condition compared to the microphone recording condition.

No significant interaction between voice type and recording condition for HNR during female chest voice singing was found ($F_{(1.8)} = 0.220$, $p = 0.651$). There were no significant main effect of voice type ($F_{(1.8)} = 0.154$, $p = 0.705$); but a significant main effect of recording condition was observed ($F_{(1.8)} = 83.660$, $p < 0.001$, $\eta_p^2 = 0.913$); indicating that the HNR was significantly lower on digitized wax cylinder audio samples (Figure 50).

No significant interaction between voice type and recording condition for spectral SD was observed ($F_{(1.8)} = 0.404$, $p = 0.543$). Additionally, no statistically significant main effect of voice type was observed ($F_{(1.8)} = 0.028$, $p = 0.870$); however, a significant main effect of recording condition was found ($F_{(1.8)} = 23.408$, $p = 0.001$, $\eta_p^2 = 0.745$). This result indicates that spectral SD is significantly greater in the digitized wax cylinder recording condition than in the

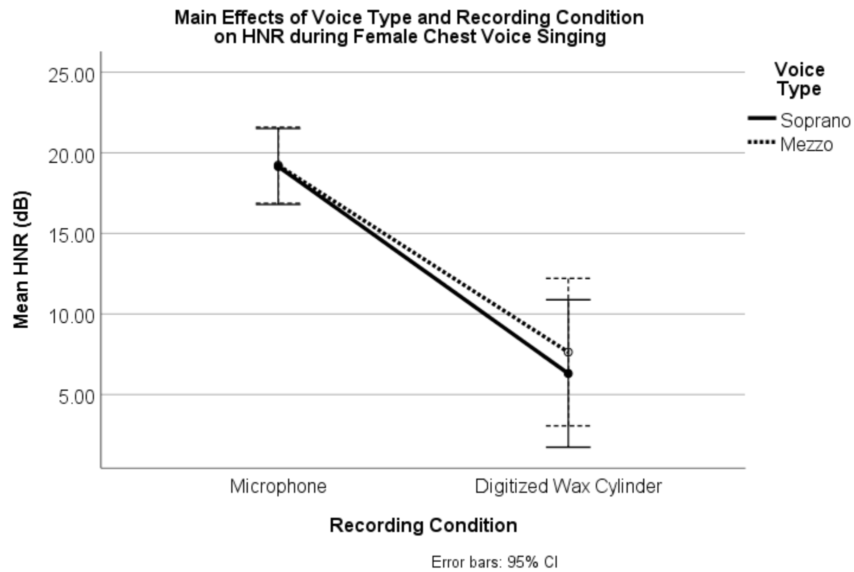


Figure 50. Main effects and interaction of voice type and recording condition on the harmonic-to-noise ratio (HNR) during the female chest voice singing task.

microphone recording condition (Figure 51).

Analysis of spectral skewness during female chest voice singing revealed no significant interaction between voice type and recording condition ($F_{(1.8)}$ = 0.142, p = 0.716). Likewise, no significant main effects of voice type ($F_{(1.8)}$ = 0.054, p = 0.822) or recording condition ($F_{(1.8)}$ = 0.343, p = 0.574) were observed.

Similarly, analysis of spectral kurtosis indicated that there was no statistically significant interaction between voice type and recording condition ($F_{(1.8)}$ = 0.268, p = 0.619). No significant main effects of recording condition ($F_{(1.8)}$ = 0.000, p = 0.986) or voice type ($F_{(1.8)}$ = 0.604, p = 0.459) were found. Non-parametric analysis of spectral mean during female chest voice singing revealed that spectral mean was significantly higher in digitized wax cylinder

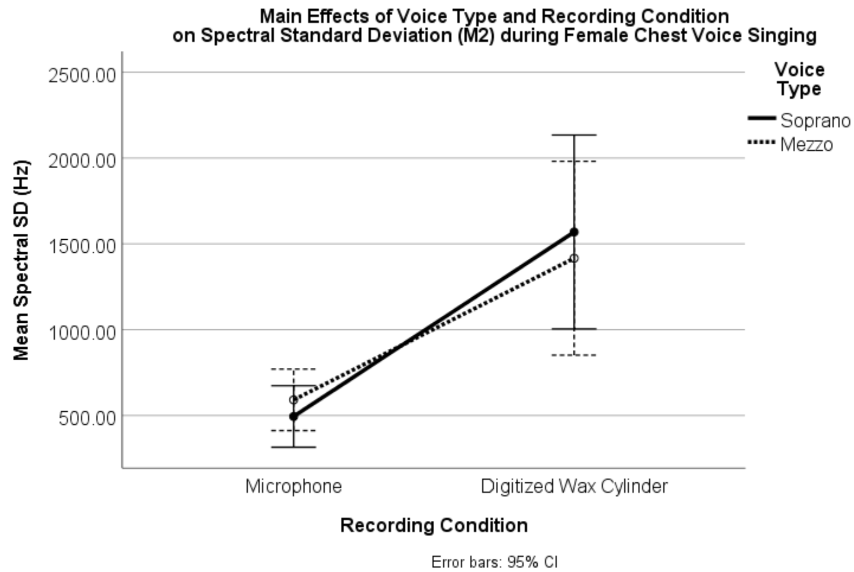


Figure 51. Main effects and interaction of voice type and recording condition on the spectral standard deviation (M2) during the female chest voice singing task.

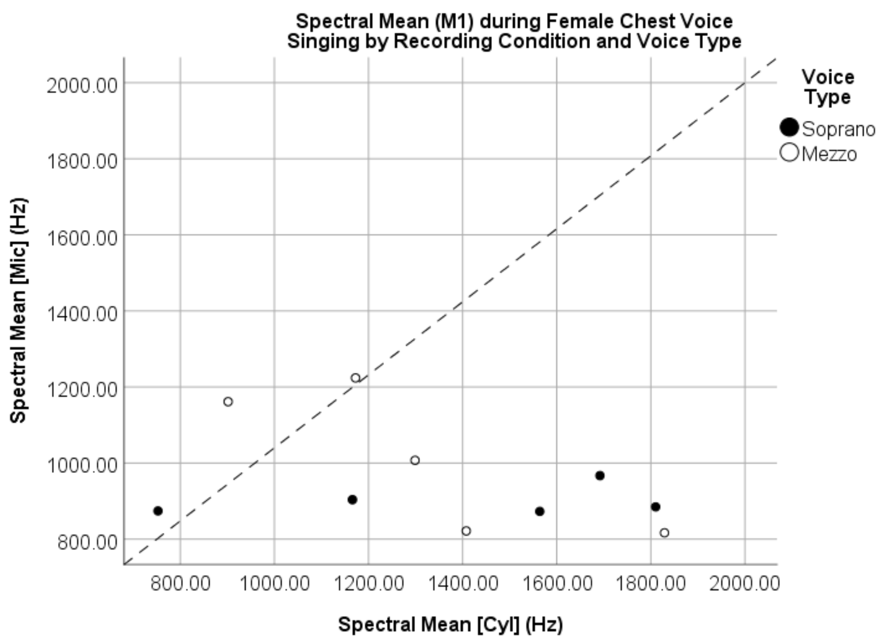


Figure 52: Scatterplot of Spectral Mean (microphone) and Spectral Mean (cylinder) for female chest voice singing. The dashed line represents a 1-1 relationship between the microphone and cylinder recording conditions. Labels for values based on voice type are included.

recordings than in microphone recordings ($Z = -2.191$, $p < 0.05$). Visual inspection of the scatter plot of spectral mean between recording conditions did not show any noticeable pattern based on voice type (see Figure 52).

4.6 Regression Analysis

As stated in Section 3.6, spectral measurements for which a significant main effect of recording condition was found were subsequently studied using linear regression analysis (see Table 20). This section will present the results of regression analysis for the following spectral measurements by task. Refer to Table 21 for all statistics related to the regression analyses.

Table 20

List of Significant Differences based on Recording Condition

Task	**L_1-L_2**	**HNR**	**M1**	**M2**	**M3**	**M4**
In-Context Singing	X	X	X	X	X	X
Steady-State Singing	X	X		X	X	
Chest Voice Singing (Female Only)	X	X	X	X		

Note: Regression analyses was performed on each of the measures noted with an "X" based on results of statistical tests reported in Sections 4.2-4.4.

4.6.1 L_1-L_2

Simple linear regression with L_1-L_2(Mic) as the dependent variable and L_1-L_2(Cyl) as the independent variable was performed for each of the three tasks. Each regression model significantly predicted L_1-L_2(Mic). That is, the regression equation manipulates the value of L_1-L_2(Cyl) to nearly equal the value of L_1-L_2(Mic). The results of each model can be found in Table 21 and the accompanying scatter plots can be found below.

Table 21

Results[17] of Regression Analysis of Spectral Measurements of the LTAS during Different Tasks

Variable	Model	*B*	*SE B*	β	R^2	*F*
L_1-L_2	In-Context ***	0.864	0.153	0.807	.651	31.76
	Steady-State ***	0.353	0.062	0.804	.646	32.92
	Female Chest Voice **	0.323	0.092	0.780	.608	12.41
HNR	In-Context ^{h}	0.223	0.285		.058	0.62
	Steady-State ^{h}	0.345	0.280		.090	1.52
	Female Chest Voice	0.176	0.169	0.346	.120	1.09
Spectral Mean (M1)	In-Context	0.253	0.158	0.363	.132	2.58
	Female Chest Voice Model ^{h}	-0.166	0.173		.194	0.93
Spectral SD (M2)	In-Context ^{h}	0.005	0.119		.000	0.002
	Steady-State ^{h}	0.049	0.284		.004	0.03
	Female Chest Voice	-0.105	0.110	-0.320	.103	0.92
Spectral Skewness (M3)	In-Context **	0.294	0.101	0.577	.333	8.49
	Steady-State ***	0.396	0.093	0.707	.500	17.99
Spectral Kurtosis (M4)	In-Context	0.165	0.080	0.445	.445	4.21
	Steady-State ^{h}	0.313	0.174		.538	3.24

* Indicates significant ($p \leq .05$) regression model.
** Indicates significant ($p \leq .01$) regression model.
*** Indicates significant ($p \leq .001$) regression model.
^{h} Indicates regression analysis performed with heteroskedasticity-consistent standard errors.

[17] *B* is the unstandardized Beta coefficient.
SE B is the standard error of the unstandardized Beta coefficient.
β is the standardized Beta coefficient (note: regression analyses performed with heteroskedasticity-consistent errors do not include standardized Beta coefficients).
R^2 or R-squared is the coefficient of determination.
F is the F-statistic

As reported in Section 4.4, a significant interaction between participant sex and recording condition was found. As such, a separate regression analysis was performed with L_1-L_2(Cyl) as a continuous independent variable, participant sex as an independent dummy variable, and participant sex*L_1-L_2(Cyl) as a moderator variable. Neither participant sex ($\beta = -0.447$, $t_{(16)} = -0.748$, $p = .465$), nor the moderator variable ($\beta = 0.102$, $t_{(16)} = 0.419$, $p = .681$) significantly contributed to the multiple regression model.

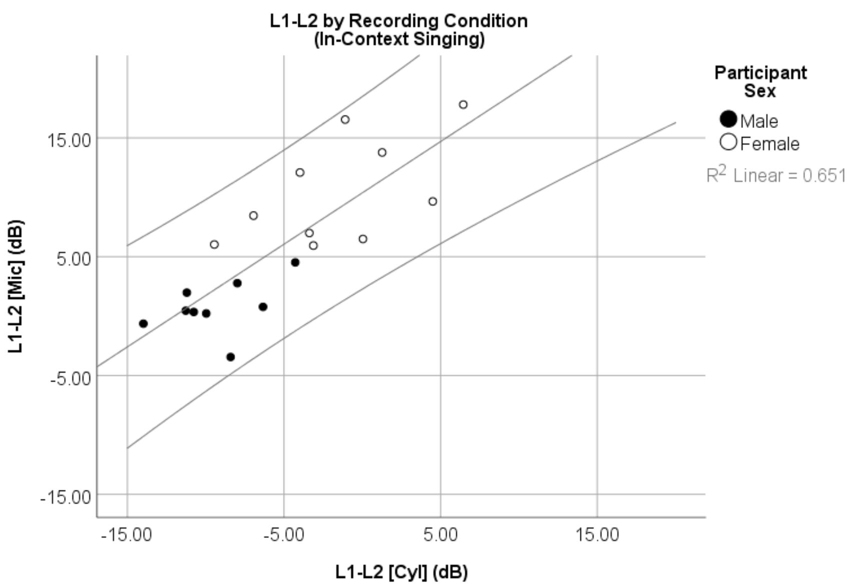


Figure 53. Scatterplot of L_1-L_2(Mic) and L_1-L_2(Cyl) for in-context singing. The line of best fit is equivalent to the regression equation computed from the regression analysis ($R^2 = .651$). Confidence Intervals (95% Confidence Level) are included for reference as well as labels for values based on participant sex. The regression equation is: L_1-L_2(Mic) = 10.376 + 0.864(L_1-L_2(Cyl)).

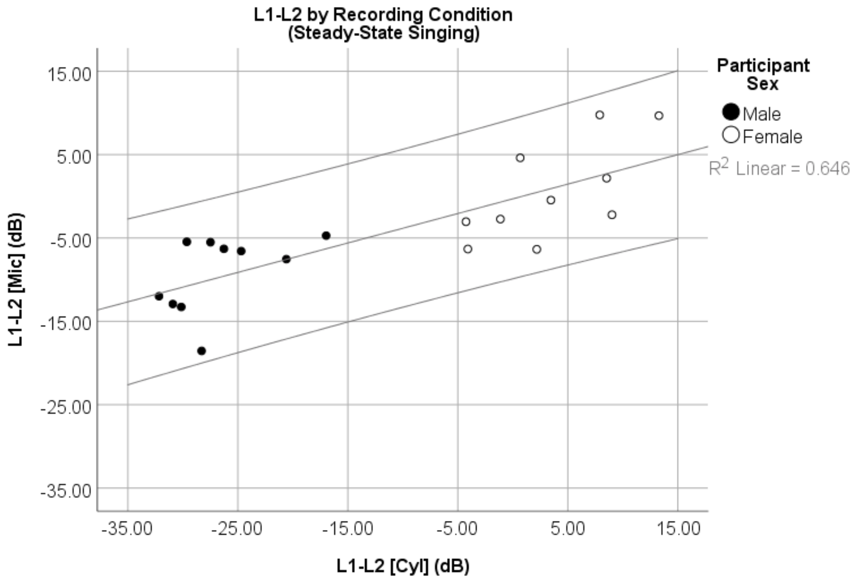


Figure 54. Scatterplot of L_1-L_2(Mic) and L_1-L_2(Cyl) during steady-state singing (*messa di voce* task). The line of best fit is equivalent to the regression equation computed from the regression analysis (R^2 = .646). The regression equation is: L_1-L_2(Mic) = -0.300 + 0.353(L_1-L_2(Cyl))

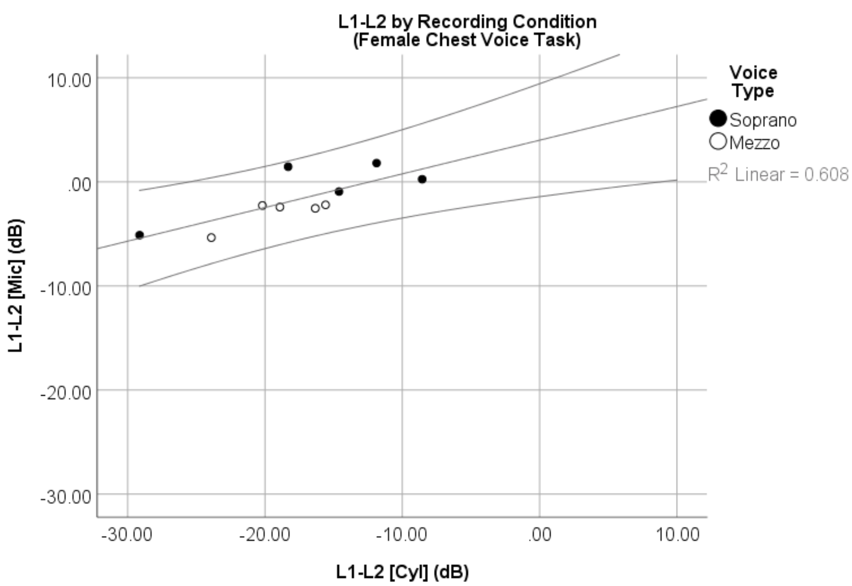


Figure 55. Scatterplot of L_1-L_2(Mic) and L_1-L_2(Cyl) for female chest voice during steady-state singing. The line of best fit is equivalent to the regression equation computed from the regression analysis (R^2 = .608). The regression equation is: L_1-L_2(Mic) = 4.004 + 0.323(L_1-L_2(Cyl))

4.6.2 Harmonic-to-Noise Ratio (HNR)

Simple linear regression analysis was performed with heteroscedasticity-consistent standard errors for HNR during in-context singing and steady-state singing using HNR_{Cyl} as the independent variable and HNR_{Mic} as the dependent variable. Simple linear regression analysis was performed with standard methods for HNR during female chest voice singing. No significant models resulted from the regression analyses.

4.6.3 Spectral Mean (M1)

Analysis of spectral mean during in-context singing and female chest voice singing indicated a significant main effect of recording condition. For in-context singing, simple linear regression analysis was performed on spectral mean measurements. Linear regression analysis with heteroskedasticity-consistent standard errors was conducted for female chest voice singing spectral mean. Both analyses used $M1_{Mic}$ as the dependent variable and $M1_{Cyl}$ as the independent variable. Neither regression model was significant.

4.6.4 Spectral Standard Deviation (M2)

Spectral SD during in-context singing and steady-state singing were analyzed using linear regression with heteroskedasticity-consistent standard errors. Standard simple linear regression was conducted on spectral SD during female chest voice singing. All analyses used $M2_{Mic}$ as the dependent variable and $M2_{Cyl}$ as the independent variable. As seen in Table 21, the models were not significant predictors of $M2_{Mic}$.

4.6.5 Spectral Skewness (M3)

Simple linear regression analysis was performed on spectral skewness measurements during in-context singing and steady-state singing; using $M3_{Mic}$ as the dependent variable and $M3_{Cyl}$ as the independent variable. Both regression models were found to be significant predictors of $M3_{Mic}$; however, the steady-state model explained a higher percentage of the variance than the in-context model. This result might be partially explained by the existence of an outlier in the spectral skewness data (the data point outside of the 95% confidence interval) during in-context singing (seen in Figure 56).

4.6.6 Spectral Kurtosis (M4)

Spectral kurtosis was found to demonstrate a significant main effect of recording condition for in-context singing and steady-state singing. As such, simple linear regression analysis with $M4_{Cyl}$ as the independent variable and $M4_{Mic}$ as the dependent variable was performed. Results indicated that neither model was a significant predictor of $M4_{Mic}$; however, the scatter plot of spectral kurtosis during steady-state singing revealed an outlier that warranted further investigation. The corresponding sample will be discussed in Chapter 5.

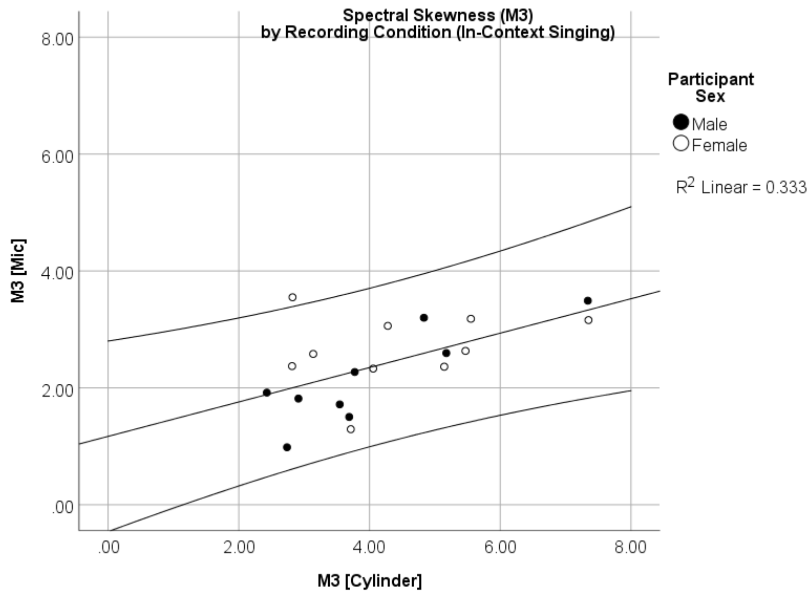


Figure 56. Scatterplot of $M3_{Mic}$ and $M3_{Cyl}$ during in-context singing (*"Caro mio ben"* phrase). Note the outlier (outside of the 95% confidence interval) that corresponds to participant-M2, which may influence the accuracy of the model. The regression equation is: $M3_{Mic} = 1.170 + 0.294(M3_{Cyl})$.

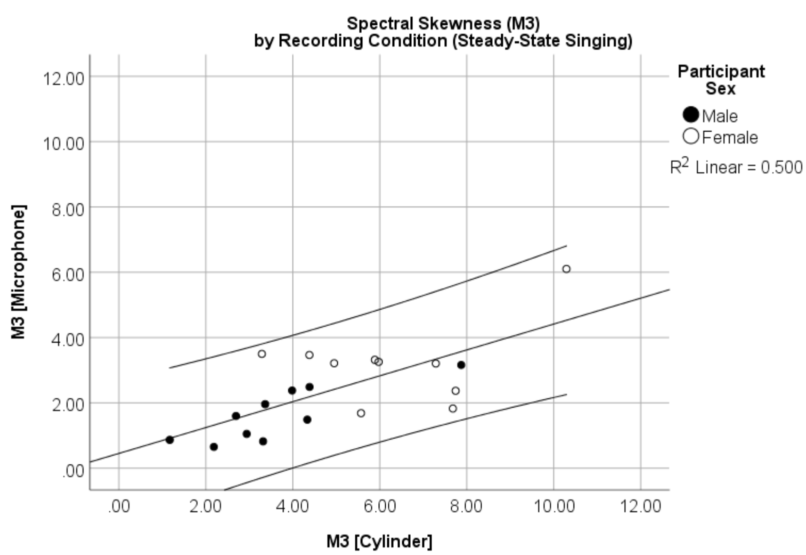


Figure 57. Scatterplot of spectral skewness (M3) during steady-state singing (*messa di voce* task). The line of best fit is equivalent to the regression equation computed from the regression analysis ($R^2 = .500$). Confidence Intervals (95% Confidence Level) and labels based on sub-group are included for reference. The regression equation is: $M3_{Mic} = 0.453 + 0.396(M3_{Cyl})$.

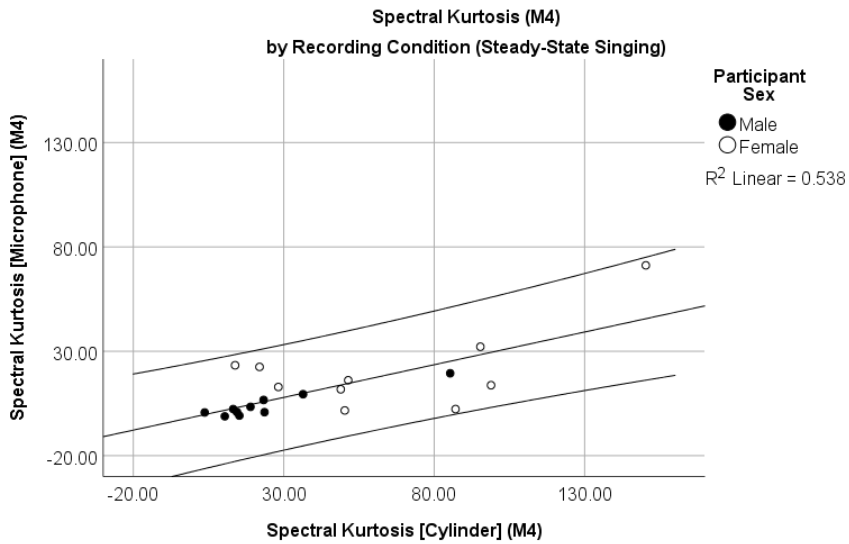


Figure 58. Scatterplot of $M4_{Mic}$ and $M4_{Cyl}$ during steady-state singing (*messa di voce* task). Note the outlier that corresponds to participant-S3.

4.7 Summary

4.7.1 System Analysis

Section 4.1 reported spectral measurement results for the following test signals from the Test Signal Recording:

- Frequency-modulating sine waves

  ($\overline{f_o}$ = ~440 Hz, Rate = 6.5 Hz, Extent = 54 Cents):

  - $\overline{f_o}$ was found to decrease by approximately 3 Hz between the microphone and the digitized worn cylinder recording condition.
  - Vibrato rate was found to remain constant between recording conditions.

    - Vibrato extent was found to increase approximately 2 cents from the microphone to the digitized worn cylinder recording condition.
- Pure tone sine waves ($\overline{f_o}$ = 440 Hz):
    - $\overline{f_o}$ was observed to decrease approximate 3 Hz from the microphone to the digitized worn cylinder recording condition.
    - Minimum $\overline{f_o}$ was found to decrease approximately 14 Hz from the microphone to the digitized worn cylinder recording condition.
    - Maximum $\overline{f_o}$ was found to increase approximately 8 Hz from the microphone to the digitized worn cylinder recording condition.
- White noise (0-22 kHz, downsampled to 0-11 kHz):
    - Spectral mean (M1) was found to decrease from the microphone to the digitized wax cylinder recording condition, but increased slightly after wear.
    - Spectral SD (M2) decreased from the microphone recording condition to the digitized wax cylinder recording condition, but increased slightly after wear.
    - Spectral skewness (M3) increased from the microphone to the digitized wax cylinder recording condition and decreased by nearly 40% after simulated wear.
    - Spectral kurtosis (M4) increased from the microphone to the digitized wax cylinder recording condition, and decreased by approximately 70% after simulated wear.

    - Harmonic-to-Noise Ratio (HNR) was found to decrease from the digitized wax cylinder recording condition to the digitized worn wax cylinder recording condition.
- Sinesweep (20 Hz – 20 kHz): The sinesweep test signal was visibly distorted by the nearfield monitor upon inspection of the spectrographic output. While it demonstrated similar characteristics to the spectral envelope of the white noise test signal, it was determined that the sinesweep signal was not suitable for further analysis.

The noise floor was analyzed and level adjustments based on calibration signals were reported. The signal-to-noise ratio (SNR) was found to decrease by nearly 50% from the microphone recording condition to the digitized worn wax cylinder recording condition.

4.7.2 Vibrato Analysis

Vibrato rate, vibrato extent, and fundamental frequency ($\bar{f_o}$) measurements during steady-state singing (*messa di voce* task) were reported. Analysis of vibrato indicated no statistically significant difference between the group mean of **vibrato rate** extracted from microphone and digitized wax cylinder recordings. $\boldsymbol{\bar{f_o}}$ for male and female sub-groups was not found to be significantly different between recording conditions. **Vibrato extent** group mean measurements from microphone and digitized wax cylinder recordings were significantly higher in the digitized wax cylinder recording condition.

4.7.3 Spectral Analysis

In Sections 4.3, 4.4, and 4.5, spectral measurements from microphone and digitized wax cylinder recordings of three tasks were reported: in-context singing (phrase from "*Caro mio ben*"), steady-state singing (middle portion of the *messa di voce* task), and female chest voice singing (female participant-only chest voice task on a constant pitch). Statistical analysis studied main effects and interactions of recording condition and participant sex with spectral measurements. Spectral measurements for which heteroskedasticity was present were analyzed with non-parametric tests to study differences between recording condition. Statistical analyses indicated the following:

- **L_1-L_2 difference**:
    - In-context: Significant main effects of participant sex and recording condition were found during in-context singing. Overall, the level of the second harmonic was significantly higher than the level of the first harmonic in the cylinder recording condition, and female L_1-L_2 measurements were significantly higher than for male participants.
    - Steady-state: A significant interaction was found between participant sex and recording condition during steady-state singing (main effects of participant sex and recording condition were also observed). L_1-L_2 measurements significantly decreased for male participants for the wax cylinder recording condition, but

significantly increased for female participants for the same condition.

  - Female chest voice: A significant main effect of recording condition was found during female chest voice singing. Overall, L_1-L_2 measurements significantly decreased for the wax cylinder recording condition.

- **HNR**:
  - In-context: HNR was found to have a significant main effect of recording condition during in-context singing. Overall, HNR significantly decreased after being recorded with the wax cylinder system.
  - Steady-state: A significant main effect of recording condition was observed during steady-state singing. Overall, HNR significantly decreased for the wax cylinder recording condition during steady-state singing.
  - Female chest voice: A significant main effect of recording condition was found during female chest voice singing. HNR significantly decreased for the wax cylinder recording condition.
- **Spectral Mean (M1):**
  - In-context: A significant main effect of recording condition was found during in-context singing. Overall, spectral mean increased from the microphone to the cylinder recording condition.

- Female chest voice: Spectral mean was found to be significantly higher in the digitized wax cylinder recording condition during the female chest voice singing task.

- **Spectral Standard Deviation (M2):**
  - In-context: Spectral SD was found to be significantly larger in the digitized wax cylinder recording condition during in-context singing.
  - Steady-state: A significant main effect of recording condition was found during steady-state singing. Spectral standard deviation significantly increased for the cylinder recording condition.
  - Female chest voice: A significant main effect of recording condition was observed during the female chest voice singing task. Spectral standard deviation was found to be significantly higher in the wax cylinder recording condition.
- **Spectral Skewness (M3):**
  - In-context: Spectral skewness was found to be significantly greater in the wax cylinder recording condition during in-context singing.
  - Steady-state: Spectral skewness was found to be significantly greater in the digitized wax cylinder recording condition during steady-state singing. Spectral skewness significantly increased for the wax cylinder recording condition.
- **Spectral Kurtosis (M4):**

- In-context: A significant main effect of recording technology was found during in-context singing. Overall, spectral kurtosis was found to be significantly higher in the digitized wax cylinder recording condition during in-context singing.
- Steady-state: Spectral kurtosis was found to be significantly higher in the digitized wax cylinder recording condition during steady-state singing.

4.7.4 Regression Analysis

Regression analyses were performed on spectral measurements for which a significant main effect of recording condition was found. Simple linear regression was performed with the digitized wax cylinder measurement as the independent variable and the microphone spectral measurement as the dependent variable

- **L1-L2 difference** regression analysis resulted in the creation of four statistically significant models (in-context singing, steady-state singing, and female chest voice singing).
- **HNR** regression analysis resulted in the creation of no statistically significant models.
- **Spectral mean (M1)** regression analysis resulted in the creation of no statistically significant models.

- **Spectral standard deviation (M2)** regression analysis resulted in the creation of no statistically significant models.
- **Spectral skewness (M3)** regression analysis resulted in the creation of two statistically significant models (in-context singing and steady-state singing).
- **Spectral kurtosis (M4)** regression analysis resulted in the creation of no statistically significant models.

CHAPTER V

DISCUSSION

5.1 Introduction

Chapter Four presented the results of statistical analyses of the differences between recordings of singers on two distinct and dissimilar audio recording technologies (i.e. recording conditions). It was the aim of this study to use flat-response microphone recordings as a control (i.e. a near-perfect representation of each participant's voice) in order to study modern opera singers recorded on period wax cylinder phonograph equipment. This chapter discusses those results and contextualizes them in relation to current singing voice research and vocal pedagogy. It also suggests ways in which these results might influence future analysis of digitized wax cylinder recordings of singers.

5.2 Limitations of Wax Cylinder Phonographs

This study aimed to investigate historical recording technology in the context of its effect on operatic voices. While there has been previous inquiry pertaining to phonograph recordings, including analysis of frequency responses (Fesler, 1983; Feynberg, 2014; Owen & Fesler, 1983; Zakaria, 2016) and spectrographic analysis (Feynberg, 2014; Schutte et al., 2005; Välimäki et al., 2008), none have paired a focused study of wax cylinder phonograph technology with an analysis of recordings of singers on that technology. In order to understand the effects of the wax cylinder phonograph on recordings of professional singers, it is worthwhile to discuss first the ways in which the equipment altered the test signals.

5.2.1 Noise Floor

Most broadly, analysis of the wax cylinder system indicated that the noise floor is higher on "virgin," or never-played, digitized wax cylinder recordings (53.95 dB) than on flat-response microphone recordings (52.13 dB), as seen in Figure 59. The noise floor was found to increase after simulated wear on digitized wax cylinder recordings by approximately 4.62 dB (58.57 dB). These

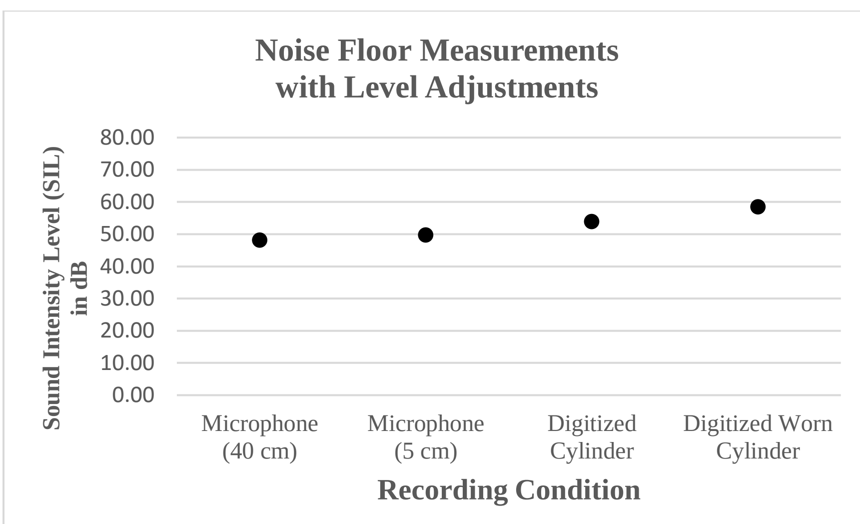


Figure 59. Adjusted SIL (in dB) of the noise floor based on recording condition.

noise floor measurements for digitized wax cylinders agree with those from past research (Feynberg, 2014, p. 43), which indicated an increase of approximately 1.5 dB to 4 dB in the noise floor with simulated wear.

The noise floor as measured from microphone recordings did not match with previous estimates of the noise floor in the sound-treated research lab (30-40 dB), which may have been caused by mechanical noise from the wax cylinder equipment. Furthermore, the horn was observed to project noise from the stylus/reproducer during recording sessions. Any of these factors could have resulted in a higher-than-expected noise floor. Subsequently, the SNR was studied in order to understand the impact of the noise floor on digitized wax cylinder recordings.

5.2.2 Signal-to-Noise Ratio

As expected, the SNR was found to be lower on digitized wax cylinder recordings (37.49 dB) compared to flat-response microphone recordings (64.16-65.79 dB) (Figure 60). Simulated wear on an Edison Fireside Phonograph with a 26.0 g tracking force resulted in a SNR measurement of 33.53 dB. These findings are relatively similar to previous research that found the SNR on digitized wax cylinder recordings with simulated wear to be between approximately 30.0-32.5 dB (Feynberg, 2014, p. 44). Additionally, previous work that focused on designing a vinyl record filtering algorithm also implemented SNR's of 30-37 dB, further situating the findings of the current study related to SNR within the literature.

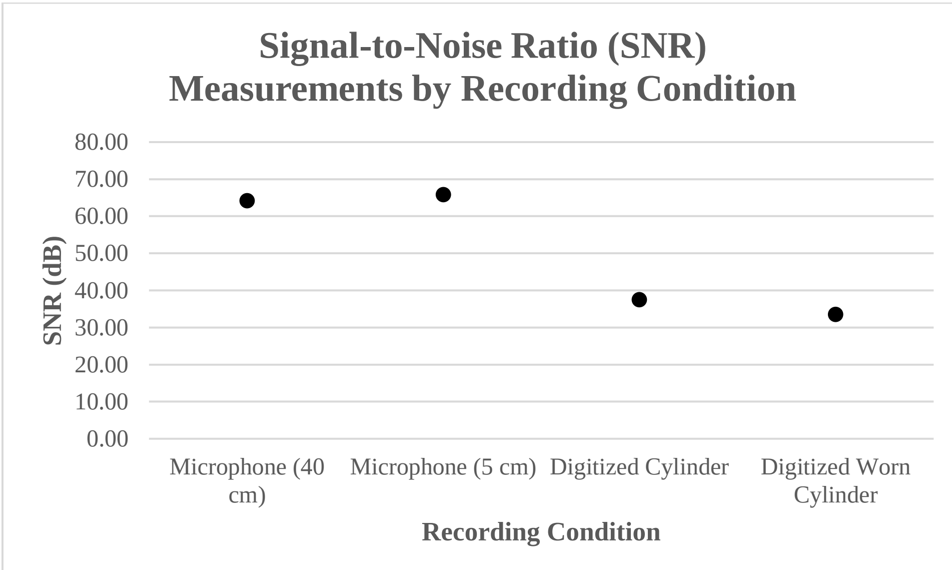


Figure 60. Signal-to-noise ratio (SNR) changes by recording condition.

5.2.3 System Frequency Response

It is important to consider the SNR and noise floor findings when reviewing the frequency response of the digitized wax cylinder recordings and related spectral measurements. For example, Figure 61 is an LTAS of the white noise test signal over 9 seconds. As discussed previously, the noise floor increased (Figure 59) and the SNR decreased (Figure 60) with each digitized wax cylinder recording condition. This effect is observed in the spectrum of Figure 61 above approximately 3 kHz for the worn cylinder condition. The “signal” portion is found below 3 kHz and the “noise” is seen from 3-10 kHz. The original

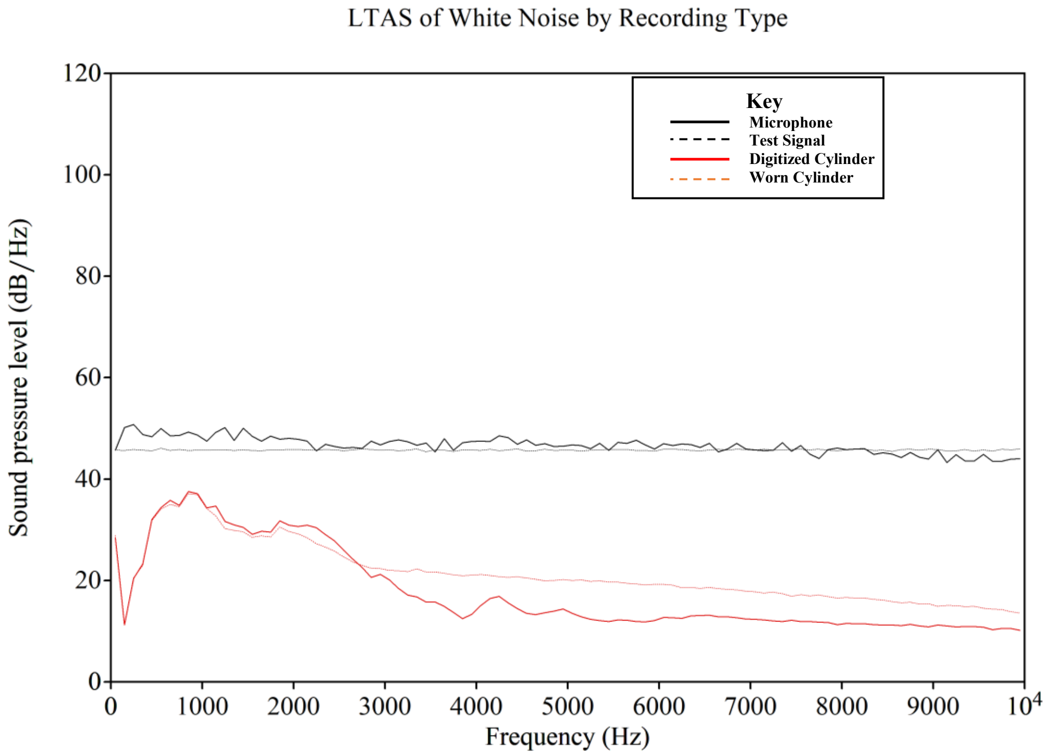


Figure 61. LTAS (200 Hz frequency resolution) of 9 seconds of the white noise test signal (dotted black line), as recorded by a microphone (solid black line), a digitized wax cylinder (solid red line), and a digitized wax cylinder after simulated wear (dotted red line).

digitized cylinder demonstrates similar filtering effects, but spectral components of the input audio signal are observed even up to 10 kHz—although frequencies above 5 kHz are significantly attenuated. While this study did not focus on SPL measurements, Figure 62 suggests a decrease in the SPL by at least 10 dB, agreeing with previous research (Feynberg, 2014; Zakaria, 2016, p. 34). It seems likely that the decreased SNR and increased noise floor was caused due to simulated wear on the wax cylinder and that these changes were reflected in the analysis of spectral moments of the LTAS. For example, the spectral mean was observed to decrease from the microphone recording condition (4409.98 Hz) to

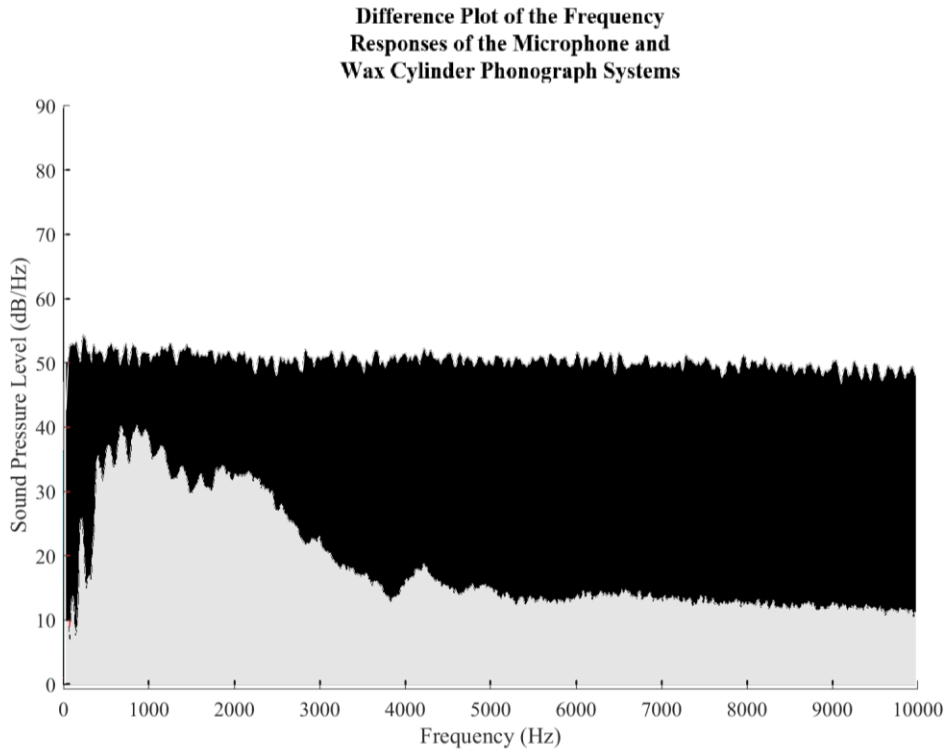


Figure 62. A difference plot of the frequency responses of the microphone and the digitized wax cylinder systems. The light gray area is the frequency response of the digitized wax cylinder system response and the black area is the difference between the two systems.

the digitized wax cylinder recording condition (1339.32 Hz), but increased after simulated wear (1712.69 Hz). The decrease from microphone to digitized wax cylinder recording condition was likely caused by the frequency band-limit around 4 kHz. Conversely, the increase in spectral mean seen in the simulated wear recording condition was likely due to the increase of noise above 3 kHz that was introduced after being played on period phonograph equipment 10 times (see Figure 63 and Figure 64). For reference, the spectral mean of white noise should be approximately half of the Nyquist frequency, which was 11.025 kHz for the test signal in question. It should be noted that the difference between the test

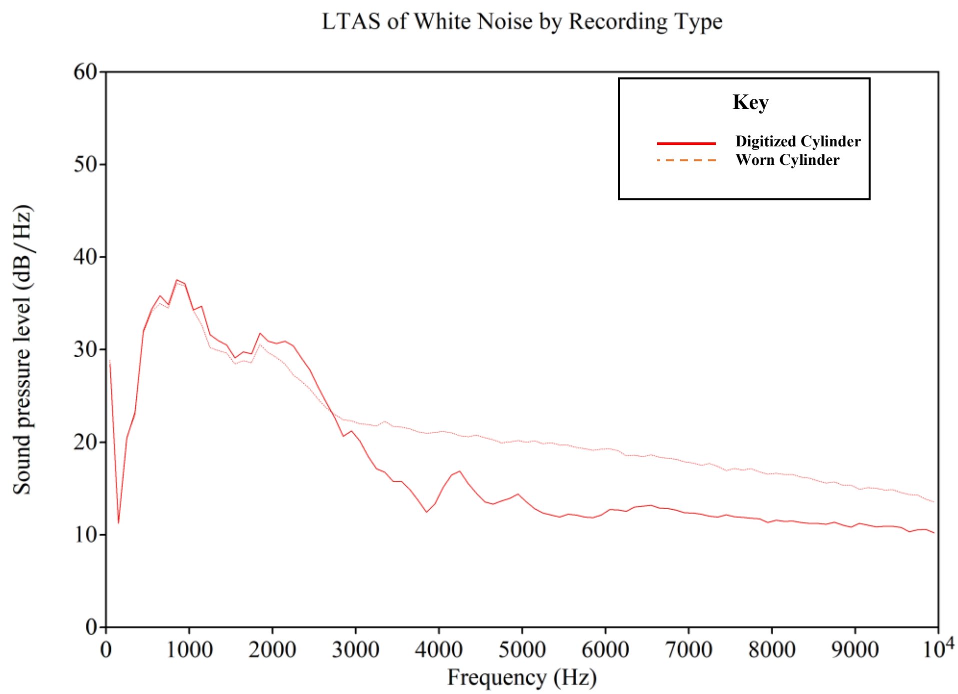


Figure 63. Comparison of the LTAS (200 Hz frequency resolution) of 9 seconds of the white noise test signal as recorded by a digitized wax cylinder (solid red line) and a digitized wax cylinder after simulated wear (dotted red line).

signal spectral mean (5508 Hz) and the microphone spectral mean (4410 Hz) may have been caused by the filtering characteristics of the room.

There is recent thought in vocal pedagogy literature that frequency components around 2 kHz – 4 kHz may contribute to auditory roughness and perceptual "brightness" in a sound (Howell, 2016; 2017). As such, it is possible that frequency range limitations as identified with a lower spectral mean measurement in the white noise test signal may result in a perceptually "darker" sounding recording compared to an identical signal recorded with a flat-response microphone. If there is a perceptual relationship between a lower spectral mean in digitized wax cylinder recordings and timbral qualities of singers inherent to

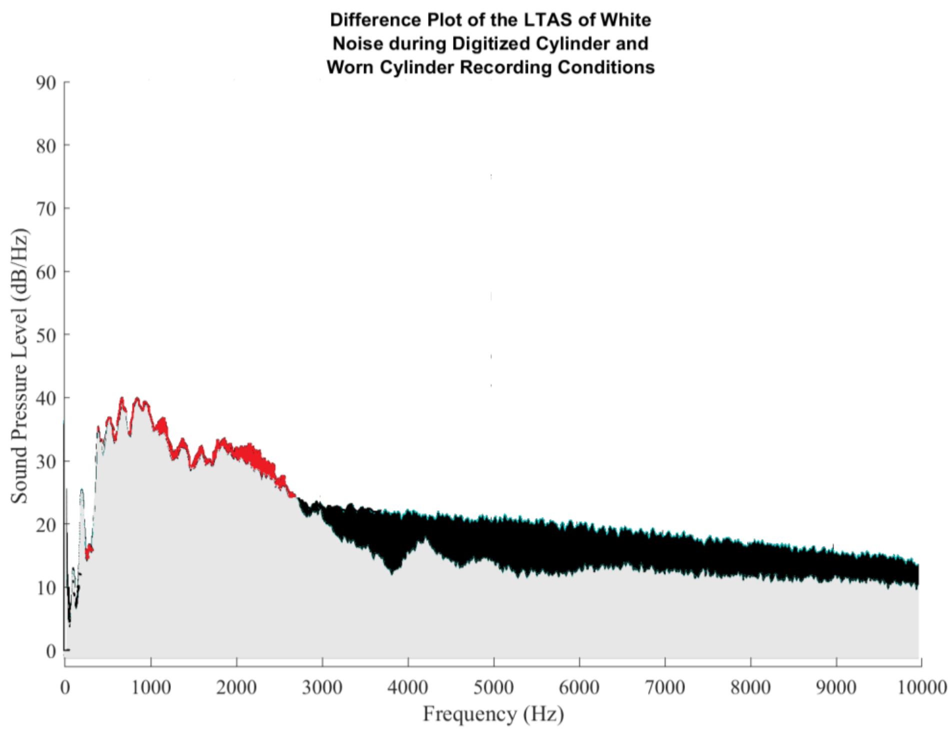


Figure 64. A difference plot of the frequency responses of the digitized wax cylinder and worn wax cylinder systems. The light gray area is the frequency response of the digitized wax cylinder system response and the black area is the difference between the two systems. The red area indicates where the frequency response of the digitized wax cylinder system has a higher SIL than the worn wax cylinder system.

frequency bands that are limited in those recordings, it is possible that broader generalizations can be made about timbral qualities of historical singers. This aspect related to listener perception of singers and historical recordings warrants further investigation.

Spectral skewness was found to increase from the microphone recording condition of the white noise (0.36) to the digitized wax cylinder recording condition (4.14). This effect was most likely caused by the decrease in spectral mean in combination with the decrease in frequency range of the wax cylinder phonograph system. In other words, the wax cylinder technology inaccurately represented frequencies below 440 Hz and above 2.5 kHz, but the signal was indistinguishable from the noise floor above 5 kHz.

In contrast, the worn cylinder recording condition demonstrated a decrease in spectral skewness (2.56). This effect was likely due to the increase in the SIL of the noise floor and its distortion of frequencies above 3 kHz in the worn digitized cylinder recording condition—approximately 2 kHz lower than was observed on “virgin” digitized wax cylinders. These changes can be more easily observed in spectrograms of the two recording conditions (Figure 65)

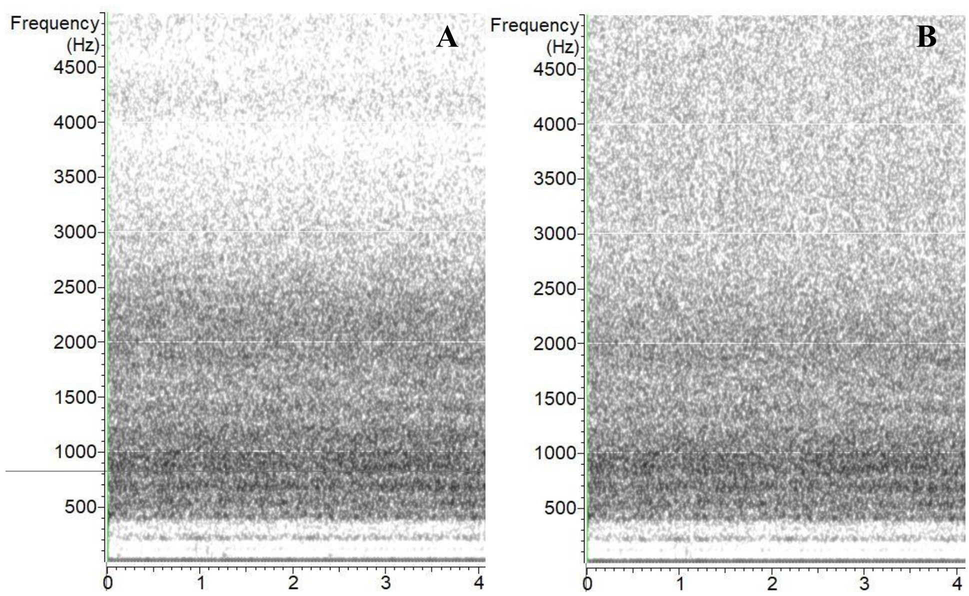


Figure 65. Spectrogram of white noise as recorded by a wax cylinder with no prior playback (A) and with 10 plays on a period wax cylinder phonograph (B).

In the first spectrogram (Figure 65A), the frequency region above 2.5 kHz is seen to have a relatively low SIL compared to lower frequency regions. When compared with the spectrogram of the digitized worn wax cylinder (Figure 65B), it can be seen that the SIL of the frequency region above 3 kHz increases. It is also possible to see that the frequency region from 2 kHz – 2.5 kHz is less prominent in the worn recording condition.

Inspection of the low frequency content of these spectrograms presents an interesting phenomenon (highlighted in Figure 66): frequencies below 440 Hz are severely attenuated or completely negated on wax cylinder recordings before and after simulated wear—confirming previous studies related to acoustic recording technology (Fesler, 1983; Feynberg, 2014; Välimäki et al., 2008).

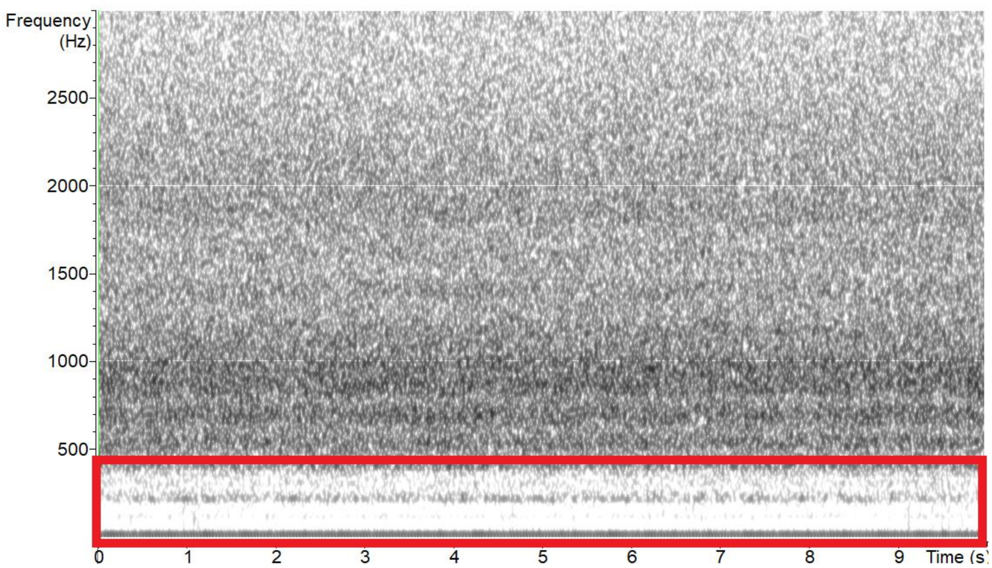


Figure 66. Spectrogram of white noise (0-22.05 kHz) recorded on a wax cylinder. This spectrogram is identical to Figure 52, but has a frequency range of 0-3 kHz in order to highlight the attenuation of frequencies below 440 Hz (marked with the red box).

This finding is important in the context of how this historical technology effects voice recordings. For example, singers for this study were required to sing the pitch C_4 (~261.63 Hz)—male participants while singing a *messa di voce* and female participants while demonstrating chest voice. These tasks tested the recording technology by potentially providing a frequency input outside of its frequency range and will be discussed further in Section 5.4.

Upon further inspection, a narrowband LTAS seemed to indicate the presence of harmonic frequencies (Figure 67). In order to understand why these frequencies might have been amplified, the resonance frequencies of the phonograph horn used for this study were estimated using the following classical equation:

$$f_{R_n} = \frac{nv}{2(L+0.8d)} \qquad (11)$$

where f_{Rn} is the frequency of resonance n, n is a whole number integer, v is the speed of sound, L is the length of the conical structure, and d is the diameter of the opening of the conical structure.

The estimated resonance frequencies of a cone with the dimensions of the phonograph horn used for this study[18] are listed in Table 22. Using the same equation, the mica recorder used[19] was estimated to have resonances at multiples of 2.5 kHz. The local spectral maxima as measured from the LTAS of the white

---

[18] Replica horn dimensions: length- 68.58 cm, diameter- 20.32 cm.
[19] New Edison Recorder" Dimensions: Outside diameter- 4.128 cm, Overall height- 3.334 cm, Outside diameter of top (connects to horn)- 1.588 cm.

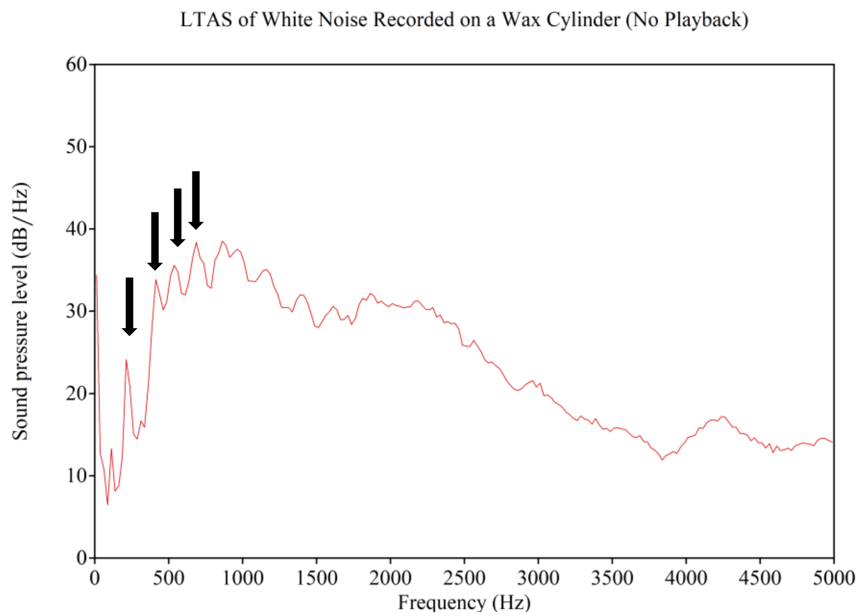


Figure 67. LTAS (100 Hz frequency resolution) of white noise recorded on a wax cylinder (50 Hz bandwidth). Black arrows point to possible resonance frequencies of the phonograph horn.

Table 22

Estimated Resonance Frequencies of the Phonograph Horn Compared with Measured Local Spectral Maxima

Integer/Local Spectral Maxima	Estimated Resonance Frequency (Hz)	Measured Local Maxima Frequency (Hz)
1	202.15	~218
2	404.31	~410
3	606.46	~687
4	808.62	~865
5	1010.77	~1159
6	1212.93	~1345

Note: From an LTAS (Blackman-Nuttal windowing function with a spectral resolution of 0.67 Hz) of white noise recorded by a wax cylinder phonograph.

noise signal do not perfectly align with the estimated resonance frequencies of a conical structure with the dimensions of the phonograph horn used for this study.

It is possible that the complex interaction between horn, reproducer, and other mechanical components of the phonograph resulted in slightly different resonant frequencies for the system. It is also possible that the exact frequencies of the local spectral maxima are not exact despite the high frequency resolution (0.67 Hz) of the LTAS due to artifacts caused by windowing.

The sinesweep test signal was not used to study the phonograph system's frequency response due to high frequency distortions that were apparent during data collection. Visual inspection of the spectrogram of the microphone recording confirmed those observations (Figure 68), but also revealed interesting aspects

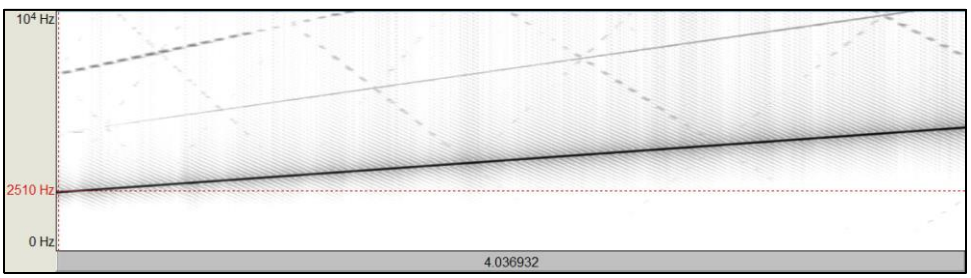


Figure 68. Distorted sinesweep from ~2.5 kHz - ~5 kHz as recorded by a flat-response omnidirectional microphone (Earthworks M30) placed exactly 40 cm in front of a Genelec 8030A nearfield monitor.

about the frequency range of the cylinders before and after simulated wear. The spectrographic representation indicated that frequency information was recorded on "virgin" digitized wax cylinder from approximately 160 Hz - 9 kHz. After simulated wear, the frequency range seemed to decrease to approximately 160 Hz - 7 kHz, with a gradual fade into the noise floor starting near 4.5 kHz.

Harmonic content was found (see Figure 69 and Figure 70) which may have been caused by phonograph horn resonances or by the SIL of the nearfield

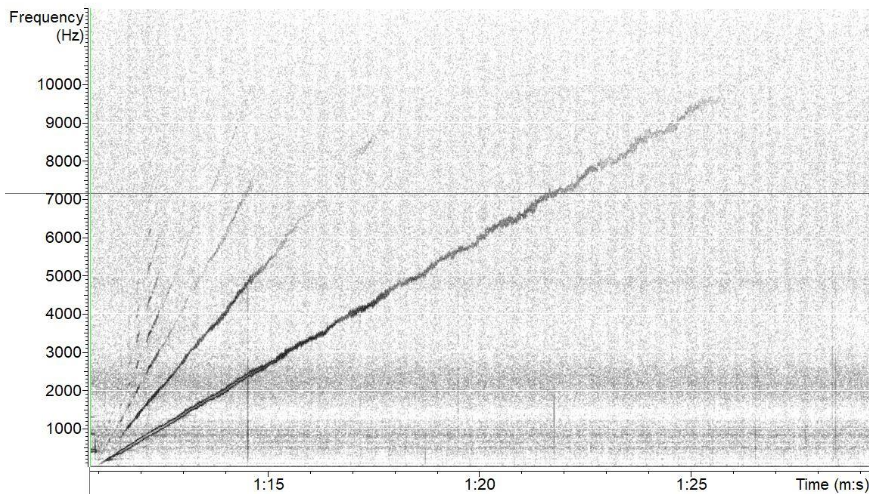


Figure 69. Spectrogram (0 Hz - 11 kHz) of a digitized wax cylinder recording with no prior playback.

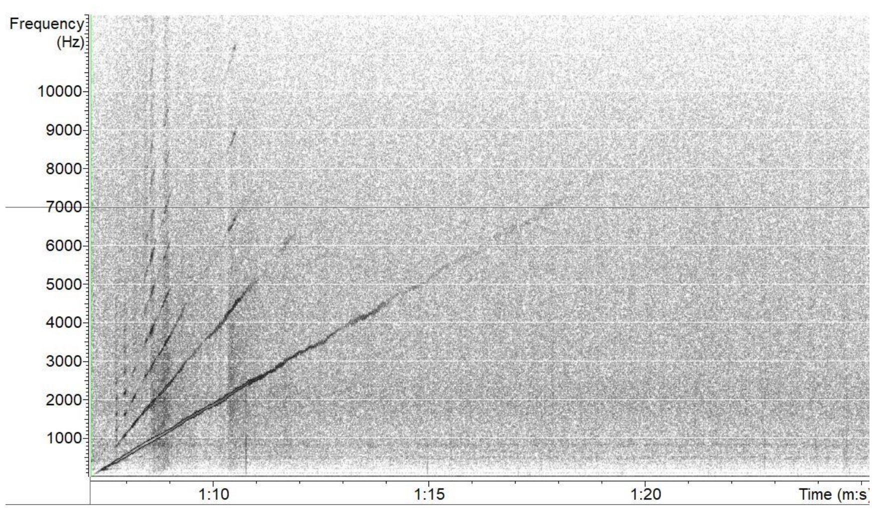


Figure 70: Spectrogram (0 Hz - 11 kHz) of a digitized wax cylinder recording with simulated wear (10 plays on period equipment).

monitor being too high during the sinesweep test signal. For example, if the SIL during the sinesweep was higher than the speaker's linear amplitude zone, harmonic distortions would have been created. It seems possible that harmonic distortions were caused by this effect. It also seems likely that the high frequency distortions were caused by an interaction in the *ProTools* algorithm due to simultaneous playback and recording through using application. While these distortions of the sinesweep test signals resulted in data that are not suitable for extensive analysis, it is interesting to note that the resultant frequency response is comparable to that of the white noise test signal, which was recorded at a lower SIL and did not produce similar distortions upon playback (Figure 71).

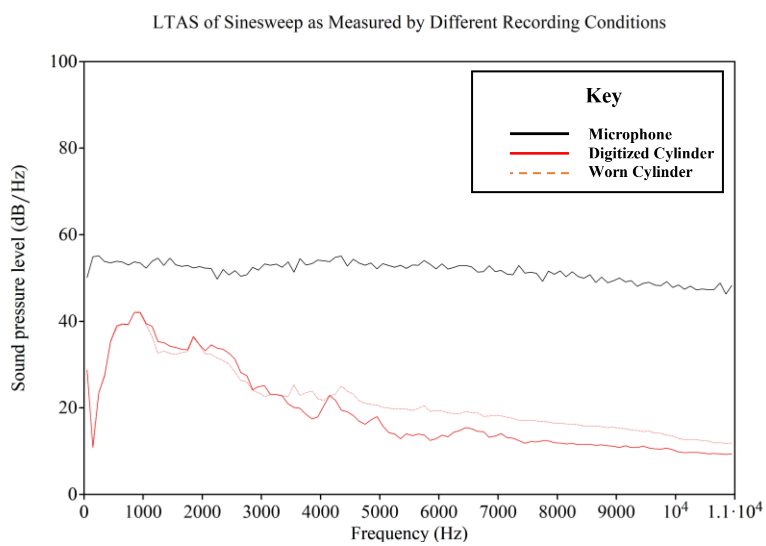


Figure 71. LTAS (0 Hz - 11 kHz, 200 Hz frequency resolution) of the sinesweep as recorded by a microphone (black line), a digitized wax cylinder (solid red line), and a digitized wax cylinder after simulated wear (dotted red line).

What is evident from this analysis of the wax cylinder phonograph system's frequency response is that the technology has an approximate frequency range from 160 Hz to 5,000 Hz on digitized wax cylinders prior to playback. The range is not flat (± 3 dB), however, but spans approximately 20 dB with the most prominent frequencies between 900 and 2200 Hz). Furthermore, the frequency range decreases to a range of 160 Hz to 3,000 Hz on digitized wax cylinders after being played 10 times on period phonograph equipment. The technology drastically attenuates frequencies below 440 Hz on both cylinder recording conditions, and the horn used for this study seems to have resonance frequencies at whole number multiples of approximately 218 Hz. It also seems likely from results of this analysis in combination with other recent studies using the same or similar wax cylinder phonograph equipment (Feynberg, 2014; Zakaria, 2016) that the noise floor rises in worn wax cylinder recordings to encompass frequencies higher than 3 kHz.

These observations related to the broad filtering effects of wax cylinder phonographs before and after simulated wear provide context for this study. While this project sought to distill the ways in which historical recording technology accurately or inaccurately represents voices from the past, it is helpful to characterize the system as a whole to frame the analysis of recordings of singers that follows in subsequent sections.

5.2.4 Sinewave Test Signal

A steady 440 Hz sinewave was produced by a nearfield monitor and recorded by a flat-response microphone array and a period wax cylinder phonograph in order to study the effect of the wax cylinder system on steady tones. Theoretically, the only reason that frequency should be inaccurately represented on a digitized wax cylinder recording is due to inconsistent speed during recording or cylinder transfer.

It was found that f_o measurements decreased slightly from the microphone recording condition (440 Hz) to the digitized cylinder conditions (~437 Hz). F_o range increased when recorded on digitized wax cylinders. It is possible that the increase in f_o range on digitized cylinder recordings was an artifact of the acoustic recording technology, but it is similarly possible that minute fluctuations by the phonograph's rotating mechanism during recording or by the *Archéophone* during digitization could have caused these deviations from the original test signal. Practically, it seems likely that a slight lowering of f_o should be expected in digitized wax cylinder phonograph recordings.

Visual inspection of the 440 Hz sinewave test signal revealed harmonics at whole number multiples of the f_o (Figure 72). While it is possible that these harmonics were created due to the test signal amplitude exceeding the speaker's maximum output levels as discussed in Section 5.2.3, inspection of the data sheet for the nearfield monitor (Genelec, 2007) indicated that the maximum short term sinewave acoustic output at 0.4 m should have been approximately 108 dB (maximum long term acoustic output = 109 dB). By adjusting the SIL as discussed in Section 4.1.5, it was determined that the SIL for the 440 Hz sinewave test signal was approximately 92.75 dB—below the threshold above which one would expect harmonic distortions from the nearfield monitor.

Given the information reported by the manufacturer of the nearfield monitor, it is possible that the harmonic frequencies in the microphone and cylinder recordings are actually resonant frequencies imprinted onto the wax

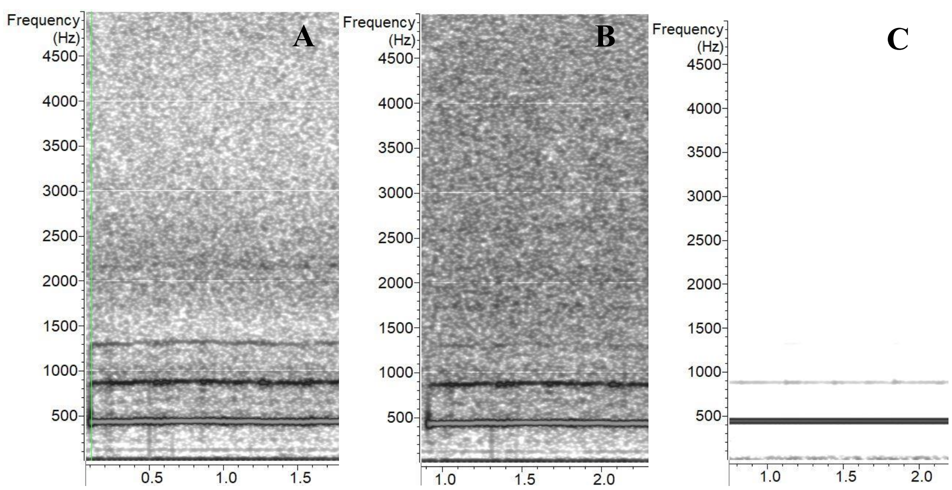


Figure 72. Spectrograms of a 440 Hz sinewave test signal (A: Digitized Wax Cylinder; B: Digitized Worn Wax Cylinder; C: Flat-Response Microphone at a distance of 40 cm from the nearfield monitor).

cylinder by the phonograph reproducer and *reflected back* toward the microphone array. This explanation would be in line with conclusions from past studies using the same or similar phonograph equipment (Feynberg, 2014; Zakaria, 2016) that hypothesized that the test signals excited resonance frequencies of the horn.

The phonograph horn can be thought of as a cylindrical tube that is open at both ends. As such, it is to be expected that resonant frequencies would transmit sound pressure waves past the larger opening of the horn. Therefore, it is plausible that the harmonic frequencies seen in all recording conditions could be a result of horn resonances and not due to mechanical artifacts from the nearfield monitor.

Regardless, harmonic content was present in all recording conditions and, thus, provided the possibility for a study of interactions between recording condition and the levels of the first two harmonics. It was found that for the 440 Hz sinewave, L_1-L_2 decreased from the microphone recording condition (45.70 dB) to the digitized wax cylinder condition (23.61 dB), but increased upon simulated wear (24.07 dB). This effect is interesting and continued as a pattern throughout most recording samples in this study. A positive L_1-L_2 indicates that the intensity level of the first harmonic (f_o) is higher than the intensity level of the second harmonic ($2f_o$). This analysis of the L_1-L_2 as related to a steady sinewave shows that there is, indeed, harmonic information created at some point in the recording process. It also suggests one of two possibilities: either f_o is attenuated, or $2f_o$ is boosted on cylinder recordings. This phenomenon will be discussed further in this chapter in relation to the recordings of the singers.

HNR was also studied between recording samples (Figure 73). It was discovered that HNR decreased with each successive recording condition. While this use of HNR is not a perfect parallel to a voiced signal, it is interesting to note

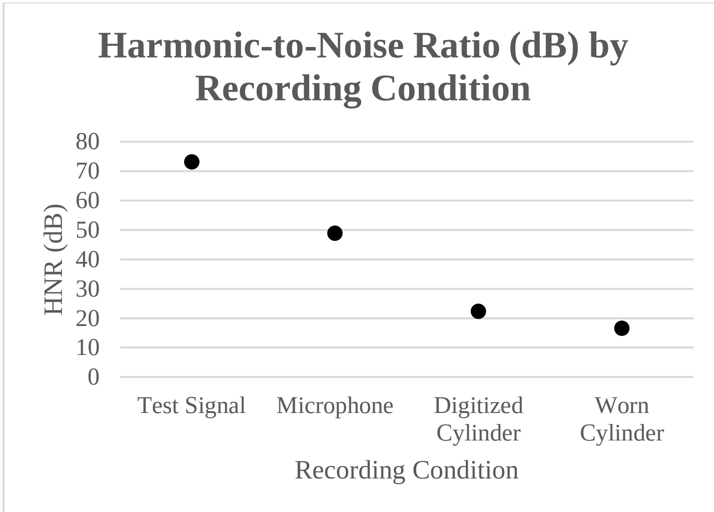


Figure 73. Harmonic-to-Noise Ratio (HNR) based on recording condition.

that only the worn cylinder condition fell below the 20 dB threshold. That is to say, the HNR of the digitized cylinder condition without simulated wear decreased by about 50% compared to the microphone condition. The HNR for the digitize wax cylinder condition was slightly above 20 dB—the threshold for healthy phonation on the vowels [a] or [i] according to *Praat* documentation (Boersma & Weenink, 2018)—but below that threshold for the worn cylinder condition. As such, it is expected to see similar changes in the HNR of recordings of singers between recording conditions.

Both the spectral mean and spectral standard deviation increased with each successive recording condition. Conversely, spectral skewness and spectral kurtosis decreased with each successive recording condition. These effects are likely directly related to the introduction of noise into the cylinder recordings and increased high frequency noise upon simulated wear. The addition of harmonic content in microphone and digitized cylinder recording conditions also explains these changes in relation to the original test signal, because the proportion of spectral information above the spectral mean increased. This type of change would theoretically—and practically—result in spectral skewness and spectral kurtosis values closer to zero (i.e. a decrease from the original value).

5.2.5 Frequency-Modulated Test Signal

As discussed in Section 3.4.1, a frequency-modulated 440 Hz sine tone with a vibrato rate of 6.5 Hz and a vibrato extent of 54 cents was introduced to the wax cylinder phonograph system to study the system's effect on vibrato. This test was particularly relevant due to past scholarly inquiry regarding historical changes in vibrato (see Sections 2.3 & 2.4). Spectrograms of one of the frequency-modulating test signals on the three different recording conditions can be seen in Figure 74.

Theoretically, modulations that can be analyzed in the time domain such as vibrato rate or vibrato extent should not deviate from the original test signal when recorded by cylinder phonographs unless, as mentioned in Section 5.2.4, there are variations in the rotation speed during recording or during digitization. As expected, vibrato rate did not change between recording conditions; however, vibrato extent increased with each successive recording condition. Avg. f_o decreased from the microphone recording condition to the digitized cylinder condition (~4 Hz), but did not show marked changes with simulated wear. In

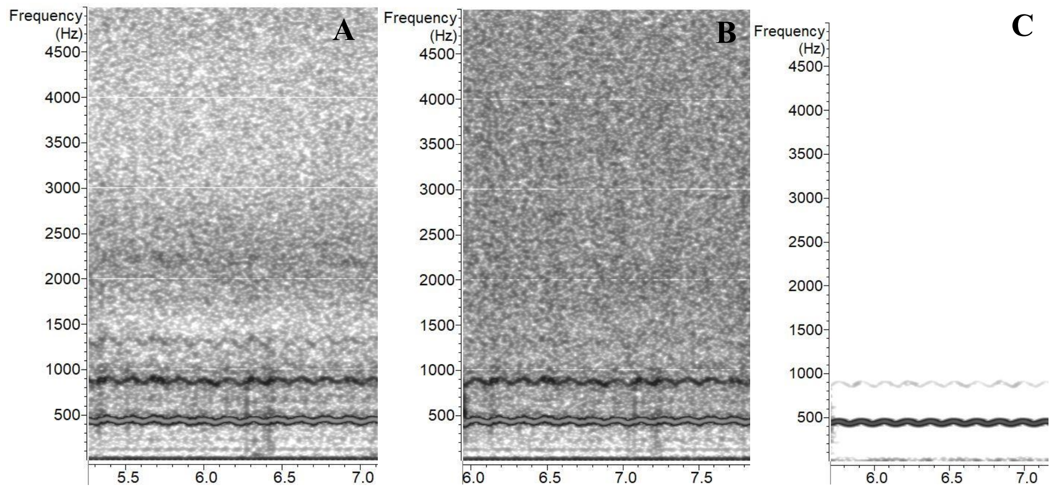


Figure 74. Spectrogram comparison of a 440 Hz sinewave with a vibrato rate of 6.5 Hz and an original vibrato extent of 54 Hz (Spectrogram A: digitized wax cylinder; Spectrogram B: digitized worn wax cylinder; Spectrogram C: microphone)

other words, the f_o range as measured by vibrato extent increased from the microphone recording condition (54 cents) to the digitized wax cylinder recording condition (56 cents) to the digitized worn wax cylinder recording condition (58 cents), but the average f_o—which is perceived as the mean between the two extremes of the f_o (Sundberg, 1987)—was not changed drastically upon simulated wear.

The preceding spectrographic images clearly show the presence of harmonics. As was previously discussed in Section 5.2.4 these harmonics were likely caused by resonances of the phonograph horn. The microphone recording condition (Figure 74C) also picked up a harmonic at $2f_o$, which was likely caused by a radiated resonance of the phonograph horn.

The results of the frequency-modulating sinewave test signal suggest that the vibrato rate of frequency-modulating sounds recorded on wax cylinder phonograph technology can be viewed as a direct and accurate representation of the original input signal. It seems likely that vibrato extent increases upon being recorded on wax cylinders and that it is distorted further upon playback. Lastly, the average f_o can be expected to decrease slightly upon being recorded on wax cylinder phonograph technology. These findings will be discussed further in the following section in the context of recordings of professional opera singers.

5.3 Vibrato

5.3.1 Vibrato Rate

Analysis of vibrato during the *messa di voce* task (Section 4.2) indicated that there were no significant differences between the vibrato rate across recording conditions (Figure 75). This result is in line with the analysis of the frequency-modulating test signal discussed in Section 5.2.5. Figure 75 further demonstrates similarities between vibrato rates measured from different recording conditions with participants categorized by sex.

These vibrato rate values seem to agree with past studies related to vibrato rate differences between male and female participants (M. A. Guzman et al., 2012; Nandamudi, 2017; Nix, Perna, James, & Allen, 2016; Sundberg, 1994).

More specifically, this study and previous literature all point to the fact that female singers generally have a faster vibrato rate on average than male singers.

Average vibrato rate (5.3 Hz) from participants in this study also fits within norms (5-7 Hz) reported in past literature (Nandamudi, 2017; Sundberg, 1994; Titze, 2000). In light of these results, it seems that wax cylinder phonograph technology does not significantly alter vibrato rate when the rotating mechanisms of the phonograph used during recording and the *Archéophone* used during digitization operate at a constant speed.

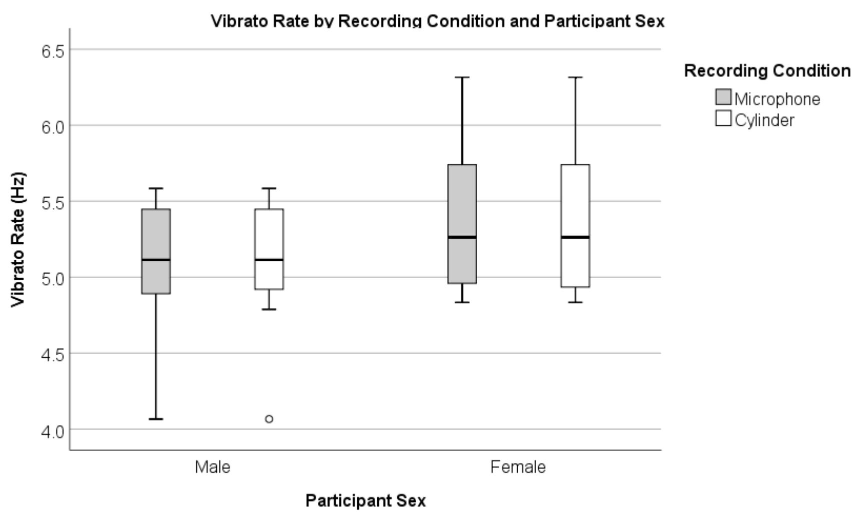


Figure 75. Vibrato Rate (Hz) by participant sex as measured from microphone and digitize wax cylinder recordings.

5.3.2 Vibrato Extent

Statistical analysis of vibrato extent revealed a significant main effect of recording condition. Overall, vibrato extent was found to be slightly higher when extracted from digitized wax cylinder recordings (103 cents) than when extracted from microphone recordings (100 cents) (see Figure 76). This finding agrees with

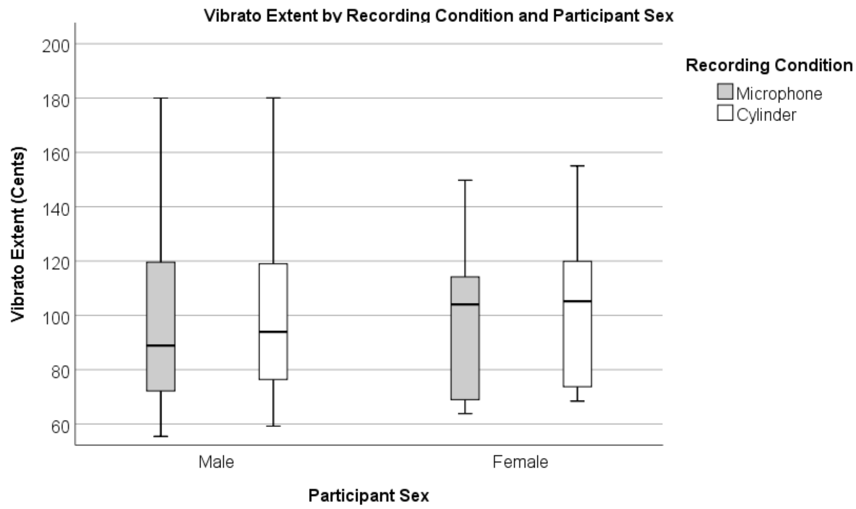


Figure 76. Vibrato Extent (Cents) by participant sex as measured from microphone and digitize wax cylinder recordings.

results from the frequency-modulating test signal (Section 5.2.5). Given that the just noticeable difference (JND) of pitch at 1000 Hz is approximately 8 cents (Howard & Angus, 2009, p. 137) and approximately 11 cents at 440 Hz (M. F. Davis, 2014, p. 784). Furthermore, the frequency difference limen can change based on the duration of the stimulus sound. Moore (1973, p. 612), for example, found the frequency difference limen is approximately 2.6 cents for a sine tone at 500 Hz with a duration of 200 ms and approximately 29.5 cents for a 500 Hz sine tone with a duration of 6.25 ms. Therefore, it is important to note that the aforementioned effect of wax cylinder technology is mathematically significant, but that it might not be perceptually relevant.

Average vibrato extent measurements from this study also agree with expected norms found in past literature (M. A. Guzman et al., 2012; Hakes, Shipp, & Doherty, 1988; Nandamudi, 2017; Nix et al., 2016; Sundberg, 1994; Vennard, 1968). It seems likely that the discrepancies found between vibrato extent measurements from the different recording conditions were the result of sound transmission inaccuracies or non-linear effects related to acoustic recording technology; however, the specific cause of these inaccuracies is unknown at this time. The results from this portion of this study suggest that vibrato extent measurements extracted from digitized wax cylinder recordings might be overestimated by an average of up to three cents in studies of historical singers.

5.3.3 Average Fundamental Frequency

Average f_o was not found to be significantly different between recording conditions (Figure 77 and Figure 78). While some inaccuracy for $\bar{f}_o$ measurements from digitized wax cylinder recordings should be expected due to the acoustic technology, differences should not significantly deviate from expected values. Otherwise, the audio recording technology at the turn of the 20th century would not have been practical or commercially marketable. That is to say, it is to be expected that even acoustic recording technology would be able to reproduce the perceived pitch of an instrument or a voice as long as the f_o was within the frequency range of that technology.

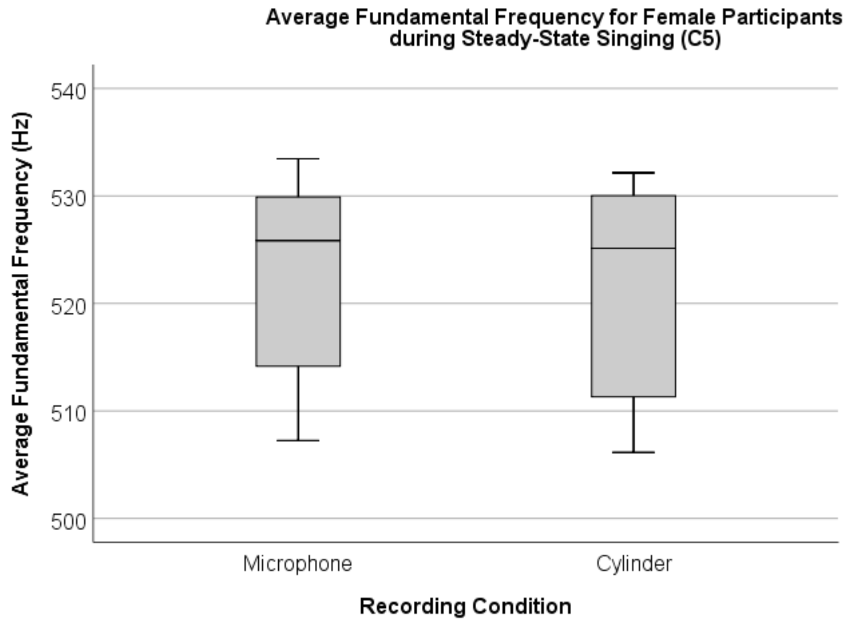


Figure 77. Comparison between average fundamental frequency measurements for female participants as measured from microphone and digitized wax cylinder recordings during steady-state singing (*messa di voce* task).

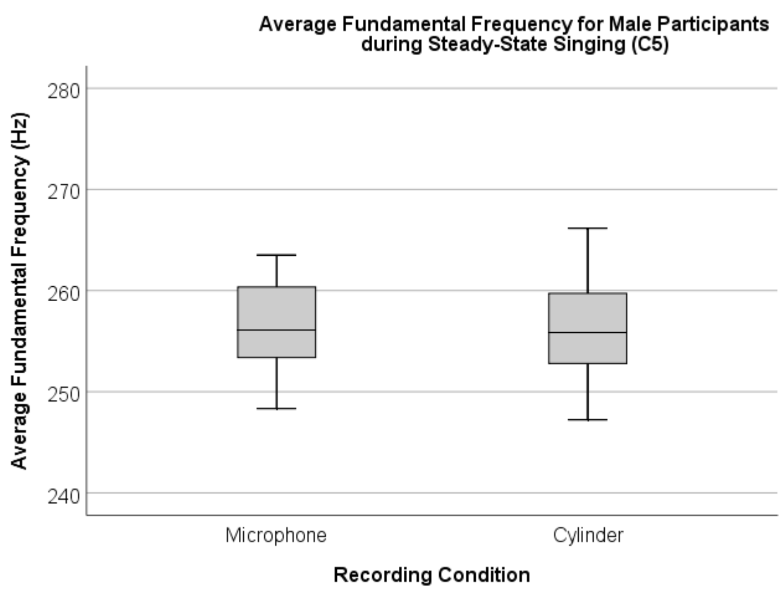


Figure 78. Comparison between average fundamental frequency measurements for male participants as measured from microphone and digitized wax cylinder recordings during steady-state singing (*messa di voce* task).

5.3.4 Impact on Analysis

Changes in the use of vibrato have been studied and discussed extensively in scholarly literature from related fields such as musicology (Crutchfield, 2012; Elliott, 2008; Gable, Frederick K. (tr.), 2009; F. K. Gable, 1992), vocal pedagogy (Doscher, 1994; Ferrante, 2011; Johnson-Read, Chmiel, Schubert, & Wolfe, 2015; R. Miller, 1997; Vennard, 1968), and singing voice research (Hakes et al., 1988; Howell, 2015; Howes, 2004; Prame, 1994; Rothman et al., 2000; Seashore, 1937a; Sundberg, 1994). The general consensus within the literature points toward a gradual decrease in vibrato rate and an increase in vibrato extent over time; however, there has not been any inquiry to date into the effects of historical recording technology on these measurements.

This study has shown that digitized wax cylinder phonograph recordings can be analyzed in order to better understand vibrato rate. As such, this study validates previous analyses of vibrato rate as extracted from digitized wax cylinder recordings, that suggest that vibrato rates were faster in the early 20th-century. The cause of this change warrants further investigation.

Results also indicated that vibrato extent was slightly higher on digitized wax cylinder samples—though the difference might not have been perceptually relevant. This finding is interesting in light of previous research that states that vibrato extent has increased over time (Ferrante, 2011). If past inquiry and conventional wisdom regarding the steady increase of vibrato extent in Western classical singing holds true, this effect of historical recording technology may indicate that vibrato extent as measured from digitized wax cylinder recordings might have been *lower* in reality.

The results from this project suggest that archived materials that have been digitized such as those found at the UCSB Cylinder Audio Archive, the Library of Congress's Recorded Sound Research Center, and the Thomas Edison National Historical Park can be studied in order to understand how vibrato may have been used technically and aesthetically at the turn of the twentieth century. Furthermore, it is important to emphasize that any differences in vibrato rate that listeners may hear in historical recordings are not due to significant differences in the singer's voice as recorded by a wax cylinder phonograph. On the other hand, vibrato extent measurements may be affected by wax cylinder technology and future research related to studying digitized cylinder recordings of singers should consider that possibility.

Lastly, previous study has suggested that a lower SNR and a limited frequency range result in discrepancies in a listener's ability to discern vibrato rate and extent (Howell, 2015). As was discussed previously, the results presented in this document suggest that vibrato rate is not significantly different based on recording technology. It also presents evidence that vibrato extent, while it is a significant main effect of recording condition, does not demonstrate perceptible differences as justified by the JND of pitch for human hearing. Therefore, if vibrato is perceived by the listener as different in historical recordings compared to modern recordings (e.g. the participants recorded for this study on both technologies), the results from this analysis point toward the need for future study related to the perception of vibrato in historical recordings. The results also suggest the need for future study related to specific technical or functional causes of changes in vibrato rate and/or extent in the singing voice.

5.4 Spectral Analysis

Spectral analysis focused on six different voice metrics: L_1-L_2 (level difference between the first two harmonics), HNR (harmonic-to-noise ratio), and spectral moments of the LTAS (spectral mean, spectral standard deviation, spectral skewness, and spectral kurtosis). These spectral measurements have been used in past literature to show differences between phonation type, to distinguish between healthy and dysphonic voices, to study singing style, and to analyze vocal effort (Anand et al., 2018; Awan, 2015; Björkner, 2006; Björkner, 2008; Bloothooft & Plomp, 1986; Flynn, Trudeau, & Johnson, 2018; Harwardt, 2011; Klatt & Klatt, 1990; Sjölander & Sundberg, 2004; Tanner et al., 2005). Past study also has shown that the LTAS of in-context singing or speaking—continuous tasks requiring both consonants and vowels—can be useful when analyzed with these measurements (Sundberg & Nordenberg, 2006; Tanner et al., 2005). As such, this study used both steady-state (sustained) and in-context (continuous) singing tasks to characterize short- and long-term spectral effects of the wax cylinder technology on recorded voice signals.

5.4.1 L_1-L_2

Results during all tasks indicated the presence of a significant main effect of recording technology on L_1-L_2, as well as a significant interaction between participant sex and recording condition during steady-state singing. That is to say, L_1-L_2 was significantly lower in digitized wax cylinder recordings compared to microphone recordings, except for female participant steady-state singing. These differences can be seen in Figure 79.

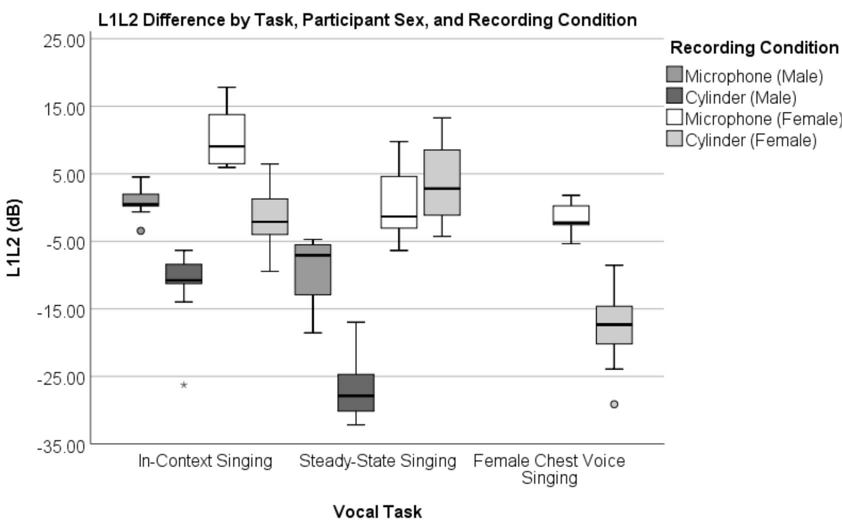


Figure 79. L1-L2 differences by task, recording condition, and participant sex. Notice that female participant L_1-L_2 during steady-state singing is the only sub-group and task to show an increase from microphone to cylinder.

It is likely that the discrepancy in L_1-L_2 during female steady-state singing compared to the other tasks was caused by differences in the frequency range of each task. Specifically, the in-context singing task included pitches from C_4-F_5 (~261.63 Hz – ~698.46 Hz) for female participants and A_2-F_4 (~110.00 Hz -

~349.23 Hz) for male participants depending on voice type (see Section 3.4.2 Phase Two: Singer Recordings). On the other hand, the steady-state singing task included only one pitch, which was based on participant sex: C_4 (~261.63 Hz) for male participants and C_5 (~523.25 Hz) for female participants.

As shown previously in Section 5.2.3, this study replicated findings regarding the frequency response of wax cylinder phonograph systems found in previous literature (Feynberg, 2014, pp. 39-41). Based on the frequency response of the wax cylinder system, it was expected to see a steep attenuation in the SIL of frequencies below 440 Hz (the second resonance of the phonograph horn as discussed in Section 5.2.3). As can be seen in Figure 80, digitized wax cylinder recordings of male participants during in-context singing drastically limited

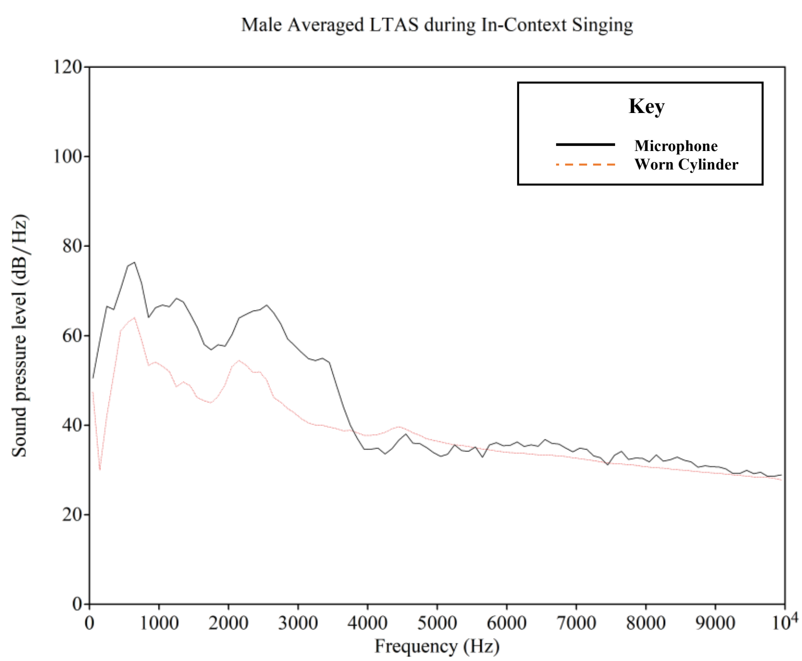


Figure 80. LTAS (200 Hz frequency resolution) of all male participants during in-context singing ("*Caro mio ben*" task).

frequency content below 500 Hz. Figure 81 demonstrates a similar effect in the LTAS of all female participants during in-context singing, but to a lesser degree

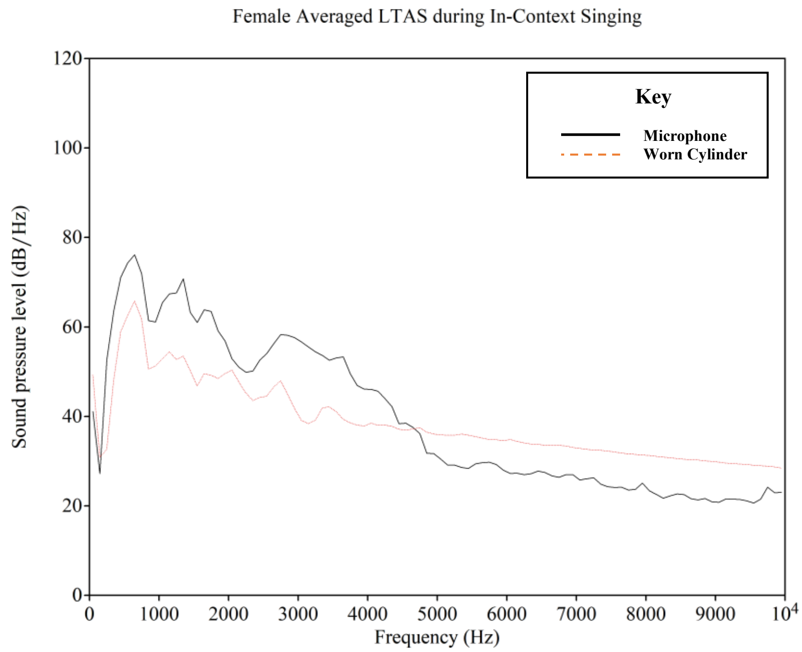


Figure 81. LTAS (200 Hz frequency resolution) of all female participants during in-context singing ("*Caro mio ben*" task).

being as female participants sang above 440 Hz for the majority of the in-context singing task. This phenomenon is further clarified in Figure 82, which demonstrates that the frequency region below 500 Hz—a part of the sung task—is severely attenuated on the digitized wax cylinder recording condition during steady-state singing.

When comparing Figure 82 and Figure 83, it is clear that the frequency range of the steady-state singing task for female participants completely avoids

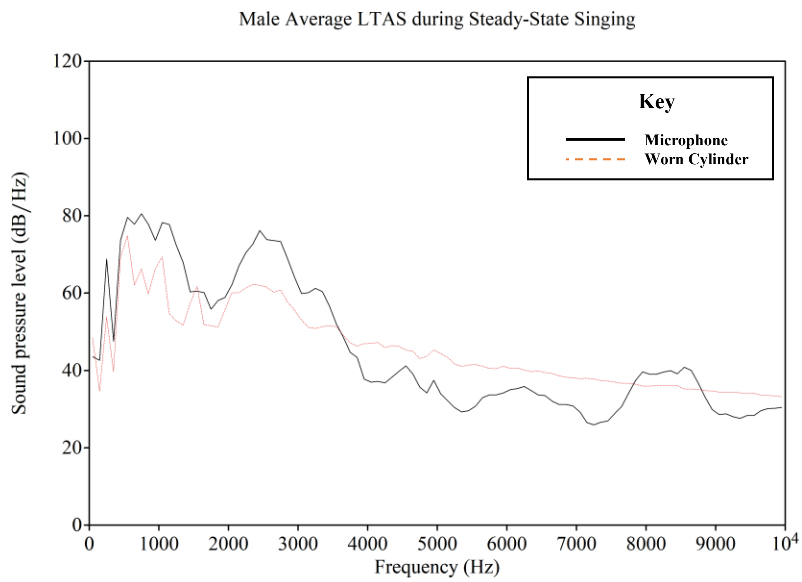


Figure 82. LTAS (200 Hz frequency resolution) of all male participants during steady-state singing (*messa di voce* task).

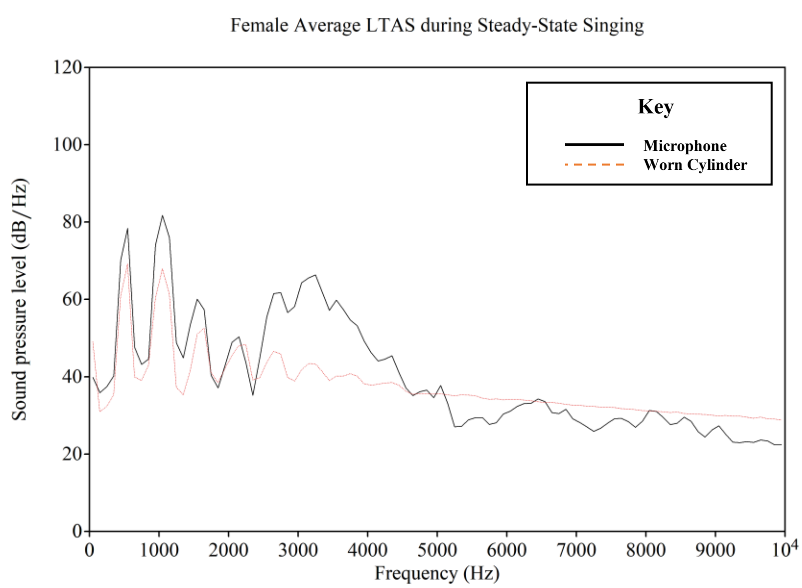


Figure 83. LTAS (200 Hz frequency resolution) of all female participants during steady-state singing (*messa di voce* task).

the lower boundary of the wax cylinder system's frequency response. Furthermore, as was discussed in Section 1.4.3, L_1-L_2 has been shown to correlate with phonation type (Björkner, 2006; Hanson & Chuang, 1999; Hillenbrand et al., 1994; Klatt & Klatt, 1990; Sundberg & Högset, 2001) and its value generally increases as pitch ascends (Hanson, 1997; Sundberg & Högset, 2001). As such, it was apparent that an analysis of the female chest voice task was necessary given that the pitch of that task (C_4) matched that which was sung by male participants during the steady-state singing task.

Referring back to Figure 79, it seems that the frequency range of the task played a pivotal role in the discrepancy seen in L_1-L_2 measurements. Furthermore, Figure 84 seems to show that the first spectral peak (L_1) is

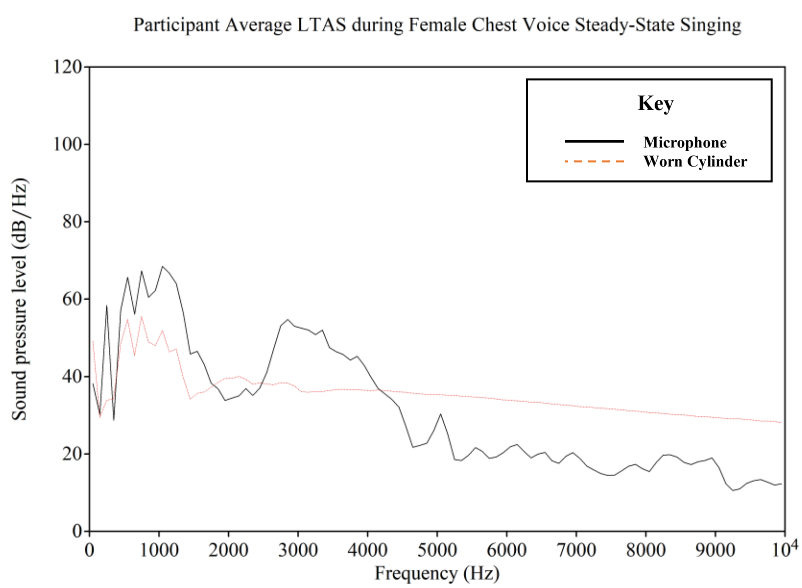


Figure 84. LTAS (200 Hz frequency resolution) of all female participants during the chest voice singing task.

completely filtered out of the digitized cylinder sample during the female chest voice task. In other words, L_1-L_2 exhibited the same pattern during the female chest voice task as during the in-context singing and the male steady-state singing tasks. The results from the female chest voice task in combination with the data from the in-context and steady-state singing tasks indicate that L_1-L_2 significantly lowers when a digitized wax cylinder voice recording sample includes frequency content below the frequency range of the wax cylinder system, but that the metric might be suitable for analysis if the f_o is at least 523 Hz (i.e. the pitch sung by female participants during the steady-state singing task).

More broadly, it seems that any analysis of digitized wax cylinder recordings of singers that involves measuring a f_o below the range of 220 Hz – 523 Hz must first confirm that the f_o is present in the signal, which, according to this study, will be highly attenuated. If it is not present or if its level is significantly attenuated, traditional formant tuning analysis and more advanced analytical methods using the first harmonic or the level of the first spectral peak will not be valid. Finally, it seems reasonable to conclude that a singer's L_1-L_2 difference being lower in wax cylinder recordings—if the f_o is below 523 Hz—adequately describes an effect of the wax cylinder phonograph system on recorded voices.

5.4.2 Harmonic-to-Noise Ratio (HNR)

The HNR was found to be significantly lower in digitized wax cylinder recordings than in microphone recordings during all three vocal tasks (Figure 85). HNR is a measurement that can be used to indicate voice quality or the SNR

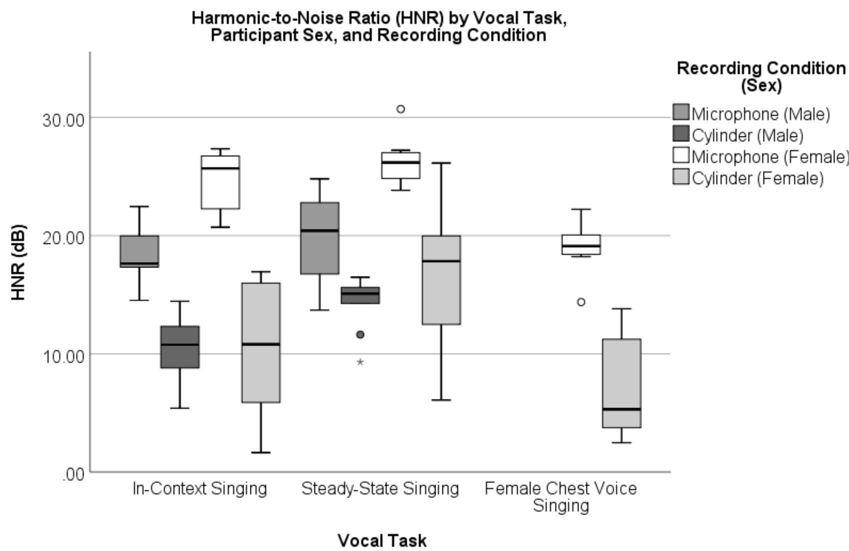


Figure 85. HNR differences between task, participant sex, and recording condition.

when a signal is periodic (Boersma, 1993). Like L_1-L_2, HNR is primarily used to analyze steady-state (sustained) tasks, but the results seem to suggest that HNR can adequately characterize the effects of the wax cylinder system on recorded voice signals. Additionally, it is possible to compare the SNR of test signals to the HNR of voice samples.

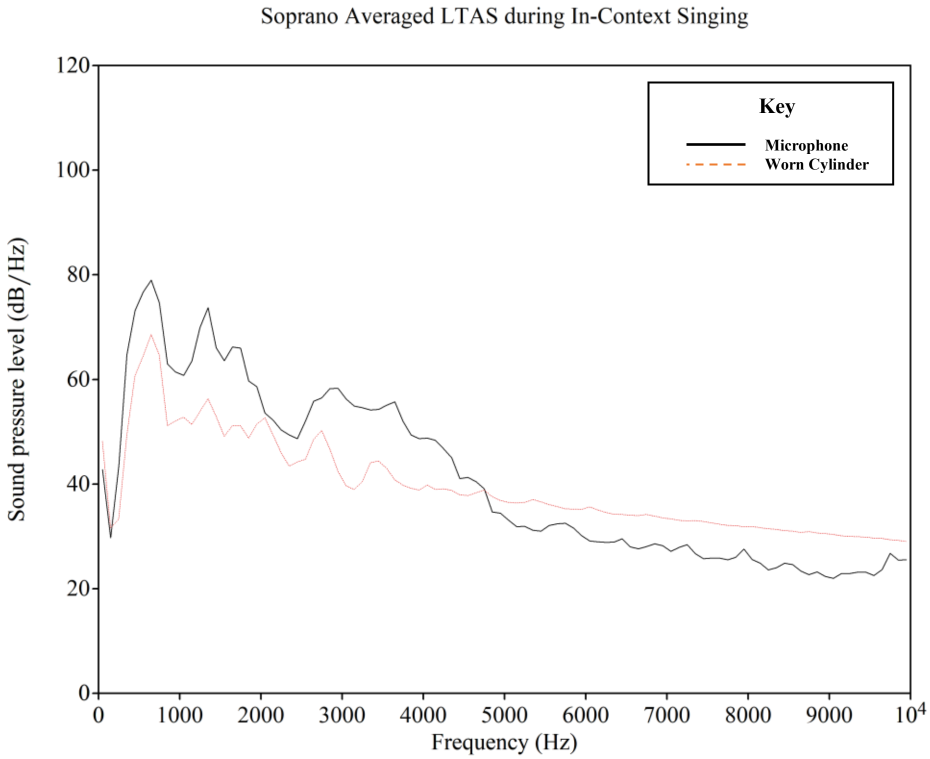
Given that the SNR was found to decrease based on recording condition (see Section 5.2.2), it was to be expected that the HNR would also decrease. The

HNR group mean was within reported norms for healthy phonation (Awan & Frenkel, 1994; Boersma & Weenink, 2018; Ferrand, 2002) and decreased by more than 50% when recorded on a wax cylinder during in-context singing and female chest voice singing tasks. However, that is not to say that voices recorded on wax cylinder phonographs sound dysphonic, but rather that HNR as a measure of the SNR for voice recordings is reduced on digitized wax cylinders. More simply put, the recordings will sound noisier.

In addition to significant differences found in HNR measurements between recording conditions, female participants had a higher absolute ΔHNR (Δ = -14.50 dB) than male participants (Δ = -8.23 dB) during in-context singing. This pattern continued during steady-state singing ($\Delta HNR_{Male} = -5.57$; $\Delta HNR_{Fem} = -11.68$) and might indicate that the wax cylinder phonograph technology records male voices with a higher fidelity than female voices.

Another explanation of this difference could relate to the frequency range of the wax cylinder technology: it is possible that singing in a higher range—as the female participants were asked to do for this study—could result in a lower proportion of the radiated sound spectrum being within the frequency range (~220 Hz–4,000Hz) of the wax cylinder system. As such, the HNR would be expected to be lower for female singers than for male singers.

A clear example of this issue related to the wax cylinder system's frequency range can be seen in the following spectra. The figures below illustrate differences between microphone and digitized wax cylinder recordings by voice



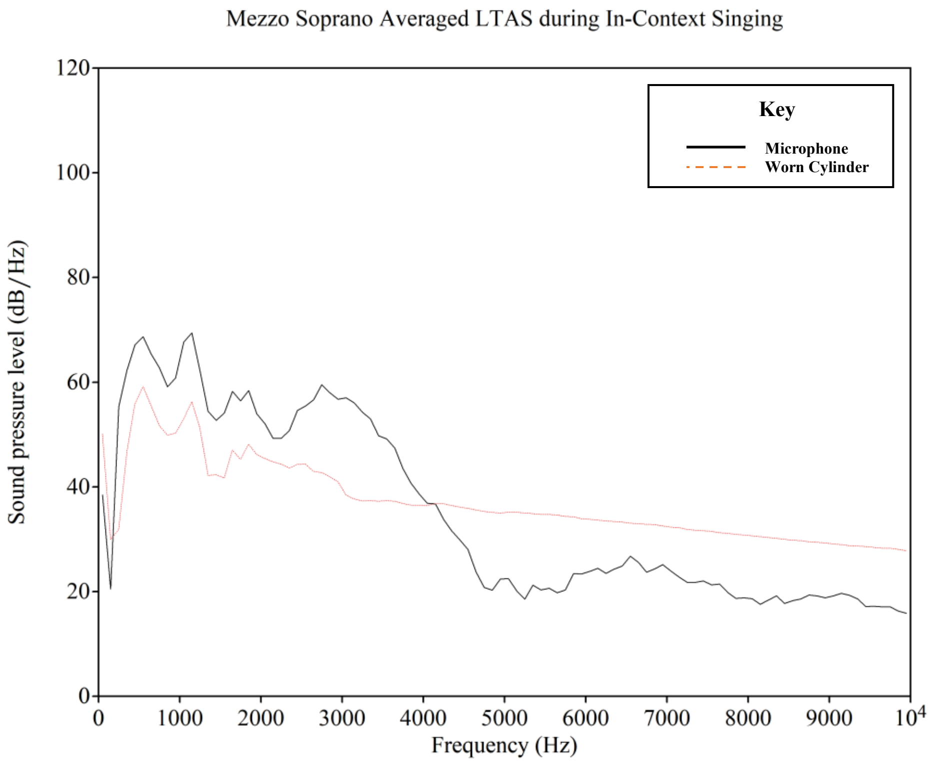
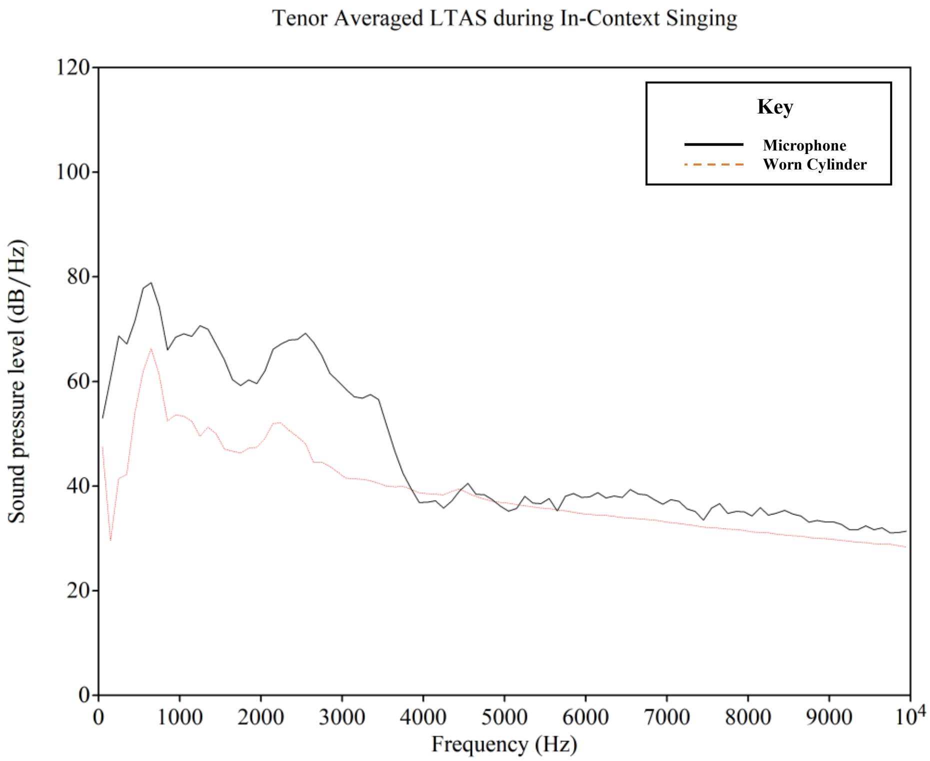


Figure 86. LTAS (200 Hz frequency resolution) of Soprano and Mezzo Soprano participants, respectively, during in-context singing (*"Caro mio ben*" task) by voice type.



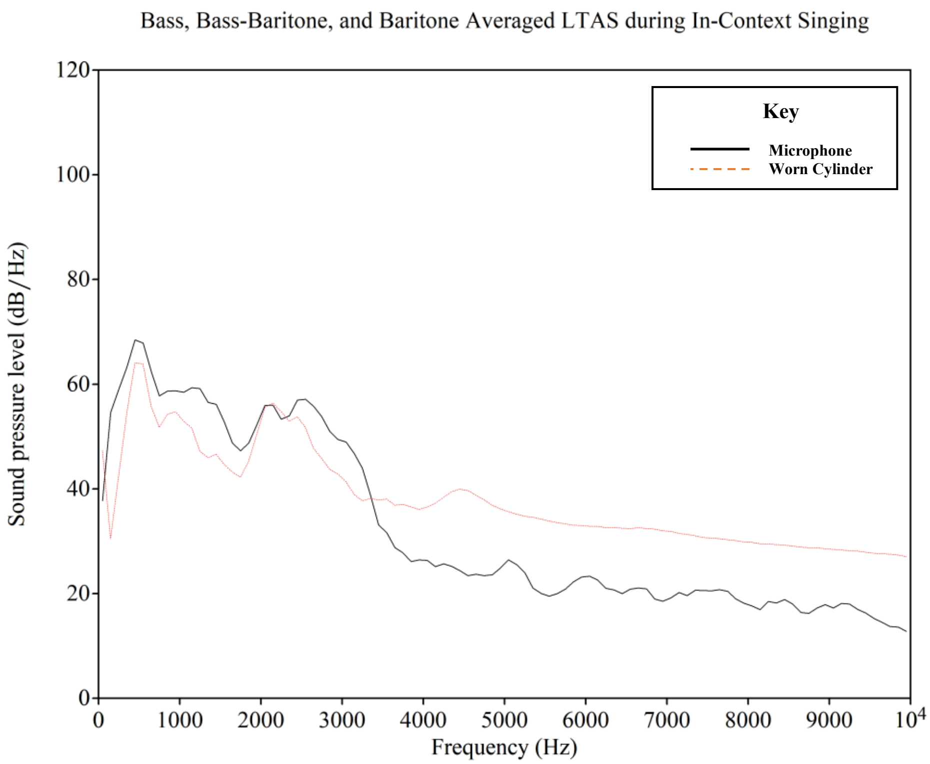


Figure 87. LTAS (200 Hz frequency resolution) of Tenor and Low Male Voice participants, respectively, during in-context singing ("*Caro mio ben*" task) by voice type.

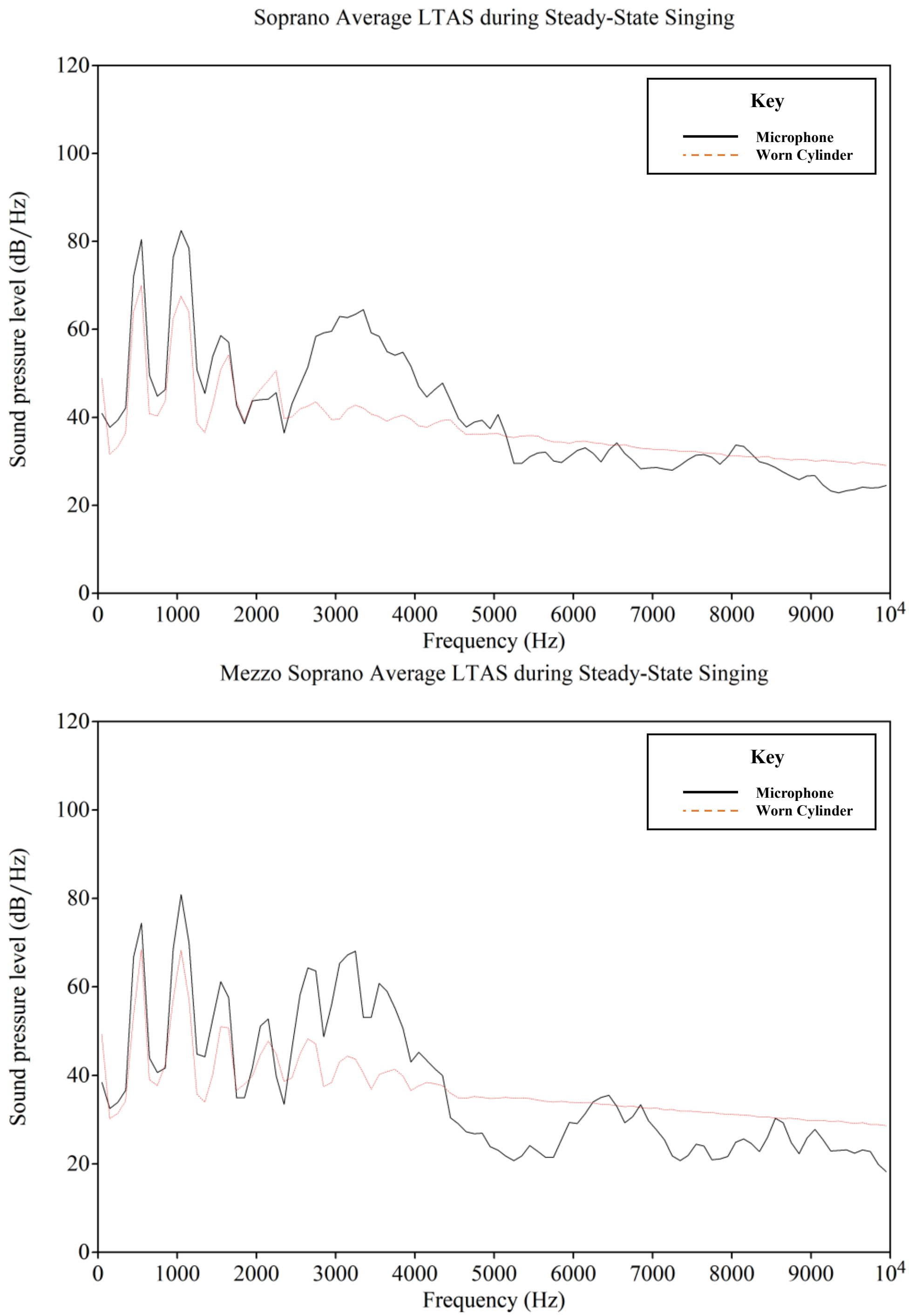
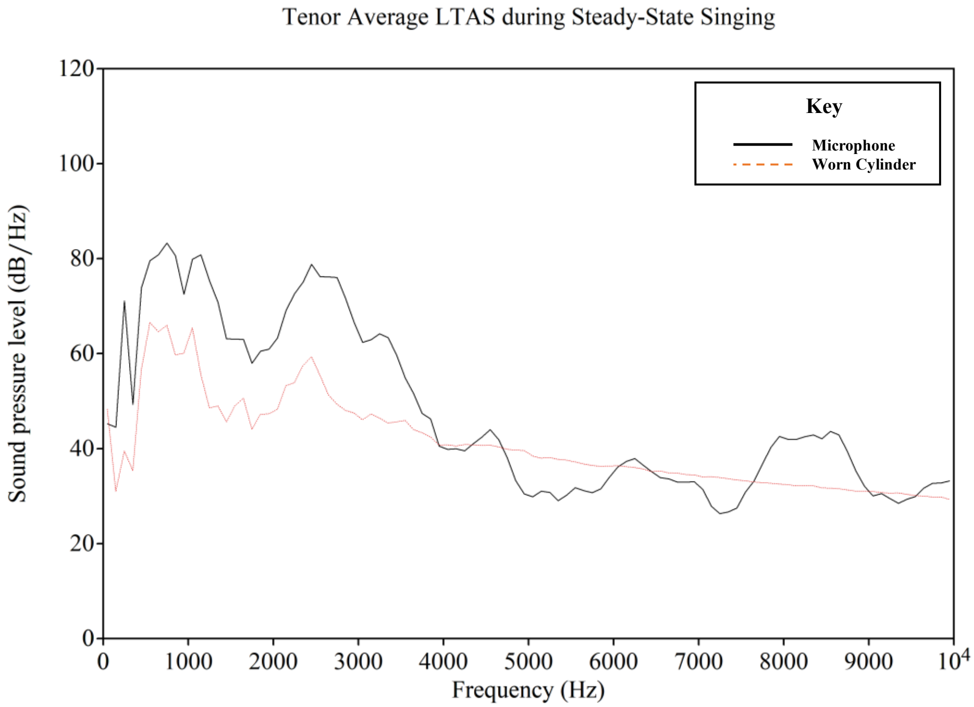


Figure 88. LTAS (200 Hz frequency resolution) of Soprano and Mezzo Soprano participants, respectively, during steady-state singing (*messa di voce* task) by voice type.



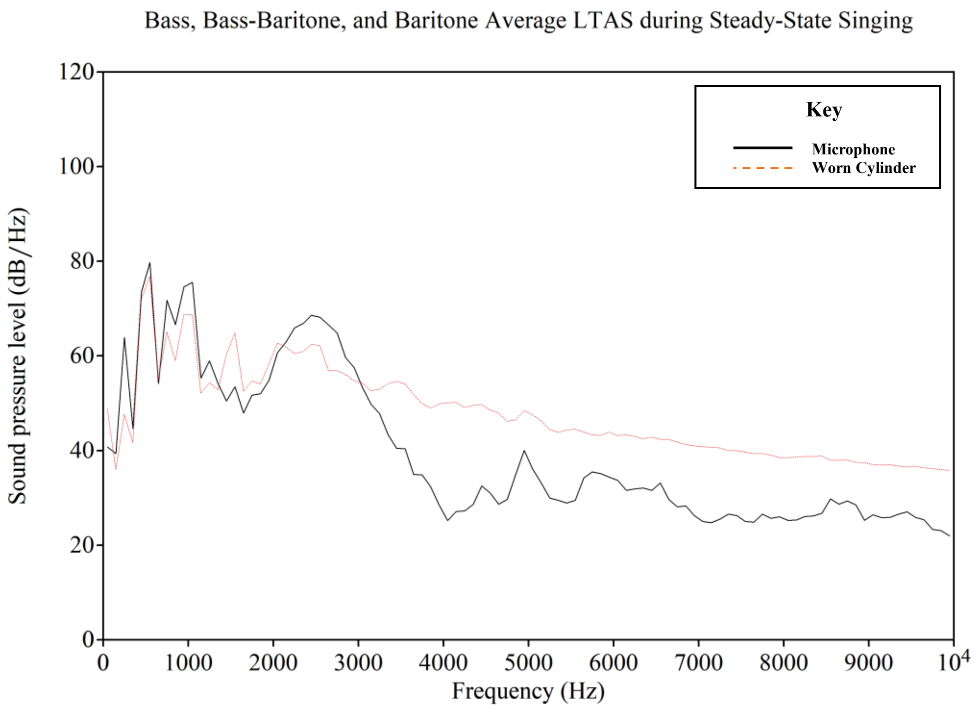


Figure 89. LTAS (200 Hz frequency resolution) of Tenor and Low Male Voice participants, respectively, during steady-state singing (*messa di voce* task) by voice type.

type. Both the soprano and mezzo soprano spectra clearly show a decrease in the level of the singer's formant cluster (SFC) or a nearly complete attenuation of the level of the 2.2 kHz – 4 kHz frequency band. Foregoing a discussion of the presence or absence of the SFC in female singers, a slightly attenuated version of the relative local maximum in the spectra of male participants seems to be present in digitized wax cylinder recordings. This phenomenon seems to be due to the SFC occupying a lower frequency area in male singing samples by approximately 500 Hz–1,000 Hz and results in the SFC of male participants being within the frequency range of the wax cylinder system.

If the SFC, indeed, contributes to the measured HNR in this way, the attenuation in the sound spectrum above 3 kHz as seen in digitized wax cylinder recordings could explain the differences in ΔHNR between male and female participants. More practically, these results seem to suggest one reason why wax cylinder phonograph recordings of female singers might sound more degraded than those of their male counterparts: HNR was found to be lower in digitized wax cylinder recordings of female singers. In summary, analysis of HNR in this study demonstrates that decreased HNR is an effect of wax cylinder phonograph technology on the voice output signal.

5.4.3 Spectral Mean (M1)

This discussion regarding relative spectral differences between recording conditions and voice types leads to this project's findings regarding spectral moments of the LTAS as measured from different recording conditions. Spectral mean was found to significantly increase during in-context singing and female chest voice singing; however, the measurement decreased and was not found to be significantly different between recording conditions when measured during 1-2 seconds of steady-state singing (see Figure 90).

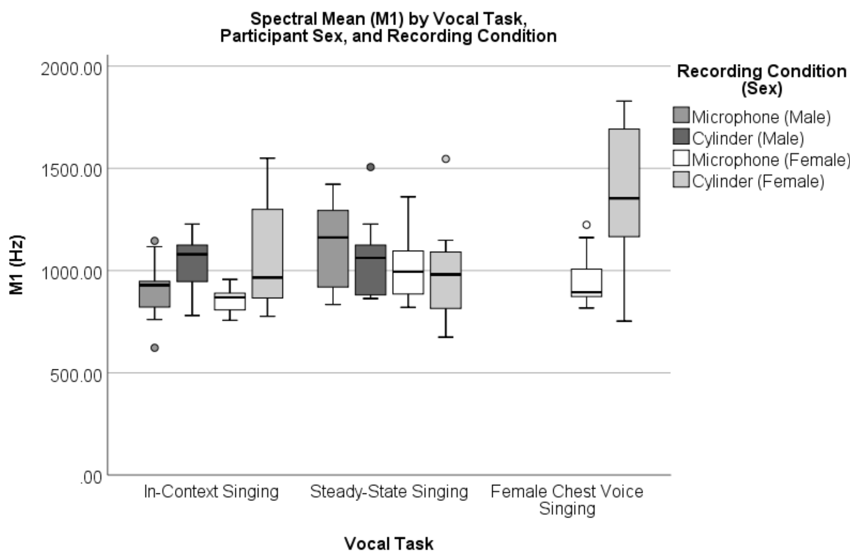


Figure 90. Spectral mean (M1) differences by task, participant sex, and recording condition.

While spectral moments of the LTAS and, specifically, spectral mean have been used in the past to suggest changes in the EGG profile (i.e. if the EGG signal was "concave up" or "concave down") (Awan, 2015, p. 524) and spectral differences for patients after voice therapy (Tanner et al., 2005, p. 216), only the former has used these types of measurements in the context of singing voice research. Furthermore, many studies that use spectral moments of the LTAS to study the voice (Awan, 2015; Harwardt, 2011; Tanner et al., 2005) or linguistics (Forrest et al., 1988) have used analysis windows as short as 20 ms or as long as 9 s. It is unclear what effect the length of the analysis window has on spectral mean and, as such, an audio sample of a female participant (S2) was selected to study why spectral mean may have increased during in-context singing and female chest voice singing, but not during steady-state singing.

The in-context sample of participant S2 was segmented into ten separate words and one segment of silence using the *ProsodyPro Praat* script. Spectral moment measurements, $\bar{f}_o$, and segment duration were calculated and subsequently transferred to *SPSS*. The different segment durations, vowel and consonant combinations, and $\bar{f}_o$'s allowed for an estimation of the change in spectral mean over the course of the phrase and under different conditions. It was hypothesized prior to this *post-hoc* analysis that the duration of the in-context singing task resulted in the different patterns in spectral mean values between tasks.

As can be seen in Figure 91, spectral mean increased at three points in the phrase for the cylinder recording condition: during the silent segment, during the word "languisce," and during the word "il." The change in spectral mean between recording conditions for each of the segments was higher than in other segments of the phrase. Upon further analysis, it seemed apparent that spectral mean increased when the $\bar{f}_o$ was lower than 400 Hz, or when there was silence in

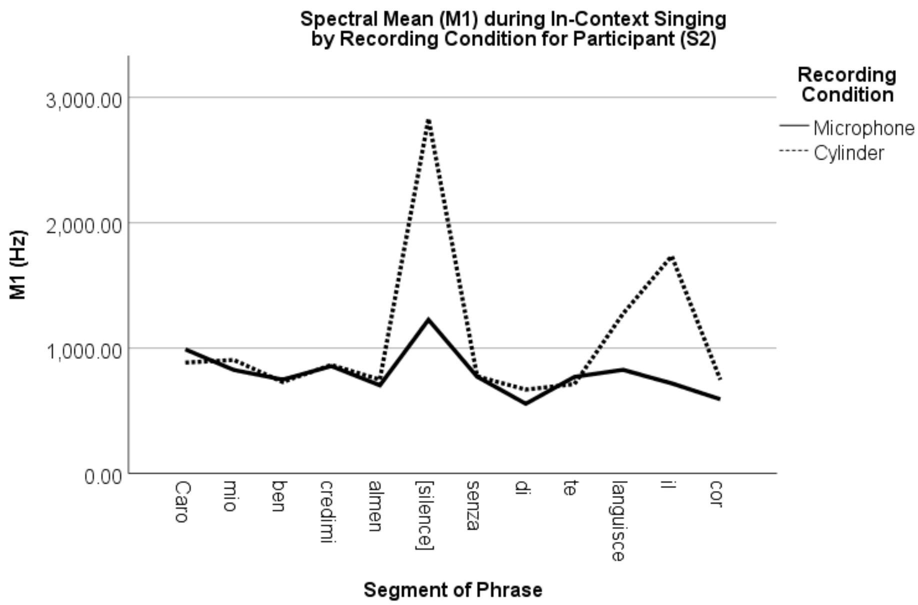


Figure 91. Spectral mean (M1) as measured for each word during the in-context singing task ("*Caro mio ben*" phrase) for one soprano participant.

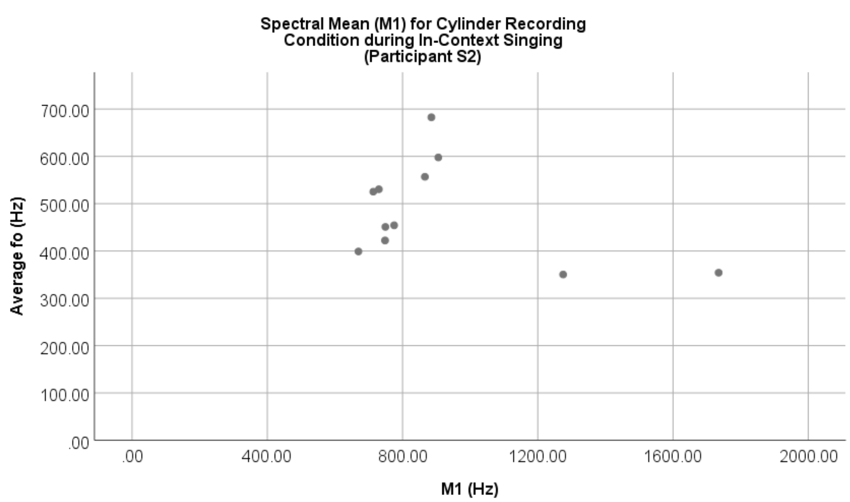


Figure 92. Relationship between average fundamental frequency (f_o) and spectral mean (M1) for participant S2 during in-context singing as extracted from a digitized wax cylinder recording.

the segment. Figure 92 demonstrates the relationship between $\bar{f_o}$ and spectral mean for the digitized wax cylinder recording condition.

Observing the relationship between $\bar{f_o}$ and spectral mean for a female participant indicated the possibility that a male participant spectral mean would be higher for every word segment below 400 Hz. As such, a Bass participant (B1) was selected to continue the case study, because all pitches sung by basses for the in-context singing task were below 400 Hz. Figure 93 demonstrates the relationship between spectral mean by segment between the two recording conditions for participant B1.

As can be seen in the figure below, spectral mean during in-context singing for participant B1 was consistently higher for the digitized wax cylinder

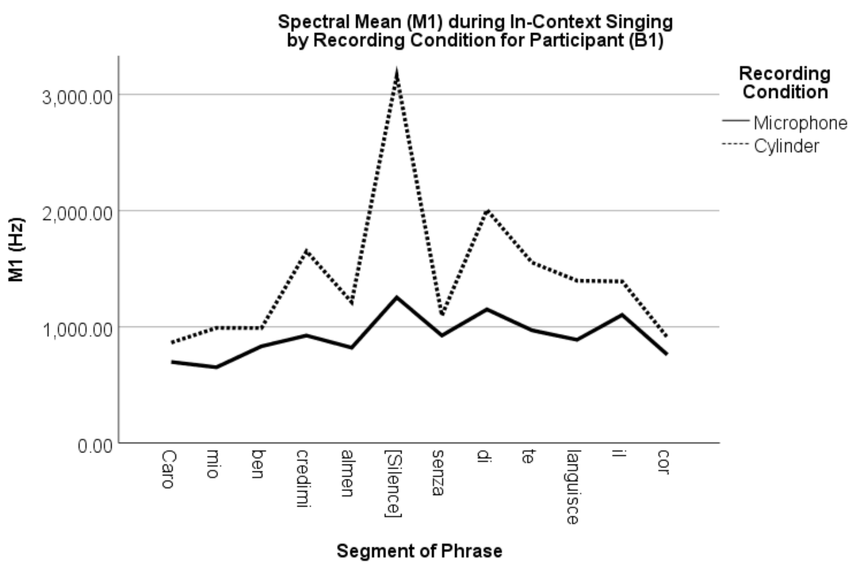


Figure 93. Spectral mean (M1) as measured for each word during the in-context singing task ("*Caro mio ben*" phrase) for one bass participant.

recording condition. Again, it should be noted that participant B1 did not sing any pitches higher than B_3 (~246.94 Hz) for the in-context singing task. That is to say, the majority of the in-context singing task for participant B1 was *below* the measured frequency range of the wax cylinder phonograph system.

The relationship between $\bar{f_o}$ and spectral mean for the digitized wax cylinder recording condition can be seen in Figure 94. Obviously, two case

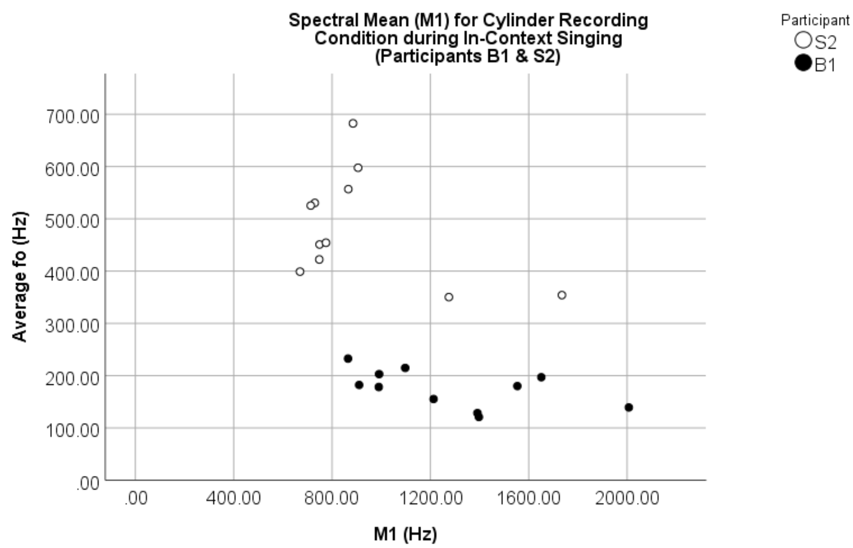


Figure 94. Relationship between average fundamental frequency (f_o) and spectral mean (M1) for participants S2 and B1 during in-context singing as extracted from digitized wax cylinder recordings.

studies cannot lead to a definitive conclusion about the effects of $\bar{f_o}$ or ambient noise on the spectral mean for either recording condition. That being said, it seems plausible that these two variables might explain the discrepancy found between spectral mean measurements for different types of tasks. More specifically, significant increases in spectral mean measurements from digitized wax cylinder recordings during in-context singing seem to be caused by the inclusion of pauses in singing (i.e. only high-frequency noise) and by pitches being sung that are outside of the frequency range of the phonograph system. In other words, it seems likely that differences in spectral mean values were not caused by variable sample duration. Instead, significant increases in spectral mean from digitized wax cylinder recordings during the female chest voice task seem to be due to the pitch being sung being near the lower boundary of the phonograph's frequency range.

While not statistically significant, it is interesting that spectral mean was *lower* during steady-state singing for both sub-groups (see Figure 95). This result corresponds with anecdotal perceptual evaluations of the digitized wax cylinder

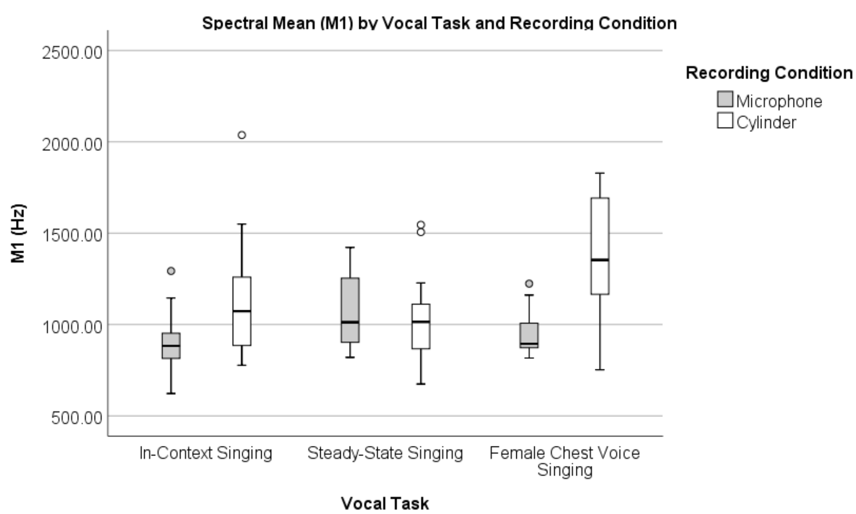


Figure 95. Group spectral mean (M1) based on vocal task and recording condition.

recordings that the recordings sound “darker” or “warmer” than their microphone counterparts. The possibility that spectral mean might be a reasonable metric for perceptual comparisons of timbral quality warrants further investigation. For example, one potential method could involve studying the perception of professional listeners and professional voice teachers of a series of controlled voice recordings for which the spectral mean value is known.

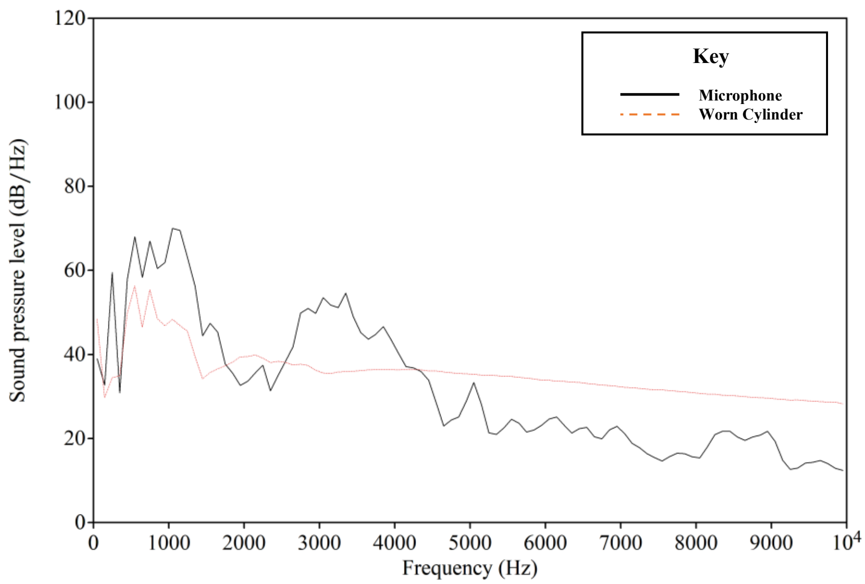
Past study has suggested that spectral mean is not significantly affected by closed quotient or EGG profile during steady-state singing tasks (Awan, 2015); however, the literature also indicates that spectral mean rises when speakers perform a “loud” vocal task (Anand et al., 2018; Harwardt, 2011). While further study about spectral mean and its relationship to the voice is needed, it seems that the measurement can explain differences between recording conditions when the LTAS is exposed to high frequency noise such as during in-context singing. If that is the case, it is likely that spectral mean during in-context singing explains the addition of noise above 3.5 kHz. It also seems likely that spectral mean can explain an effect of wax cylinder phonograph technology on voice recordings when the pitch is below the frequency range of the phonograph system as seen in the female chest voice task. However, based on the results from this study, spectral mean cannot be used to make inferences about digitized wax cylinder voice recordings during steady-state singing within the system’s frequency range. To summarize, spectral mean can describe an effect of the wax cylinder phonograph system under certain conditions, but the measurement primarily seems to indicate the presence of high frequency noise or that the fundamental frequency being sung is below the system's frequency range.

5.4.4 Spectral Standard Deviation (M2)

Results from this study indicating that spectral SD was significantly higher in digitized wax cylinder recordings than in microphone recordings (Figure 96) seem to be a result of a less "compact" spectrum. In other words, the spectral information within all groups was more "spread out" in digitized wax cylinder recording samples. This effect of the wax cylinder technology was likely due to the decreased SNR as discussed in Section 5.2.2. Spectral SD results from this study also seem to be larger than measurements reported in past literature.

An interesting aspect of the spectral SD results from this study is the large difference between recording condition found during the female chest voice task (see Figure 97). Furthermore, the change between recording conditions for male participants during the *messa di voce* task ($\overline{\Delta}$ = 153 Hz) was less than the same pitch sung by female participants in chest voice ($\overline{\Delta}$ = 950 Hz). As can be seen in Figure 96, spectral energy was concentrated below 5 kHz for both voice types in themicrophone recordings. Additionally, a prominent SFC can be seen between 3 kHz to 4 kHz in both spectra; however, the same spectra as measured from digitized wax cylinder recordings completely negate that spectral peak. The wax cylinder system also severely attenuates the fundamental frequency (~261.63 Hz) for all female participants.

Soprano Average LTAS during Chest Voice Steady-State Singing



Mezzo Soprano Average LTAS during Chest Voice Steady-State Singing

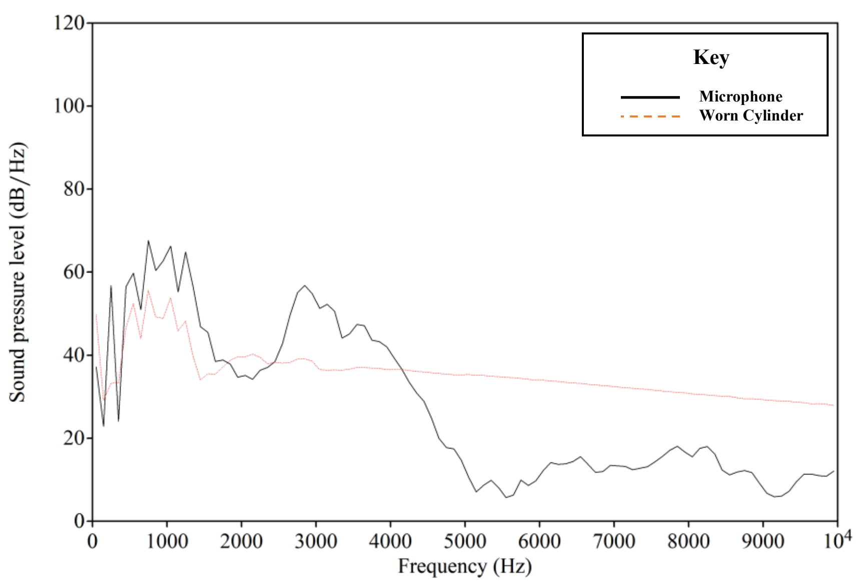


Figure 96. LTAS (200 Hz frequency resolution) of all participants during female chest voice singing by voice type.

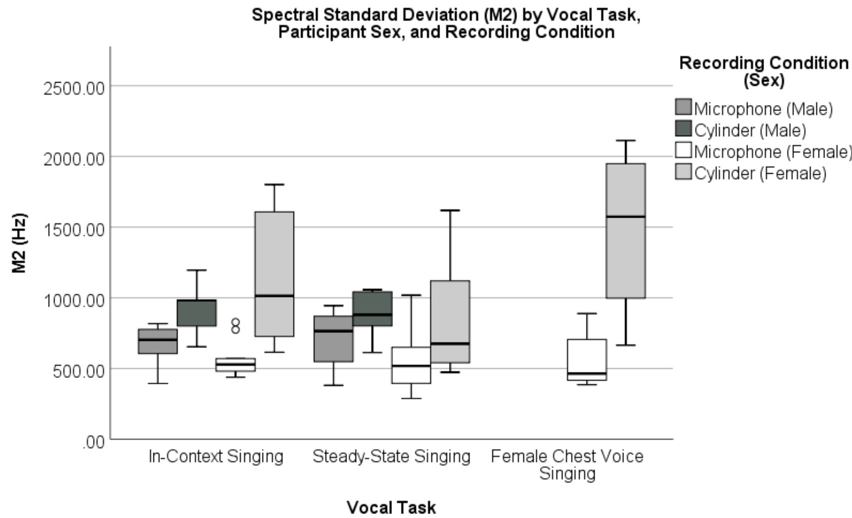


Figure 97. Spectral standard deviation (M2) differences by task, participant sex, and recording condition.

As was discussed in Section 2.2, the female chest voice register during the turn of the twentieth century is of particular interest to voice teachers, musicologists, and vocal pedagogues. If, as suggested by Awan et al. (2015), spectral SD can be used to describe CQ and EGG profile, it is possible that spectral SD could be an effective measure with which to characterize and study female chest voice in digitized wax cylinder recordings from the early 20^{th}-century. In other words, if future analysis of spectral SD in archived digitized wax cylinder recordings for female chest voice is found to have a higher value—more similar to male participants on the same pitch in this study—then it may be possible to infer that female operatic singers used to sing with a heavier mechanism when singing in chest voice in the early 20^{th}-century. However, it is also possible that increased spectral SD for digitized wax cylinder recordings is simply an artifact of an acoustic technology and is solely caused by increased high frequency noise. In summary, spectral SD—a measure of the dispersion of spectral information in a signal—was found to be higher in digitized wax cylinder recordings of singers. This finding represents a clear effect of the wax cylinder phonograph system.

5.4.5 Spectral Skewness (M3)

Results indicated that spectral skewness was significantly higher in digitized wax cylinder recordings during in-context and steady-state singing (Figure 98). Spectral skewness measurements reported in this study for male participants were generally lower than in past studies, but spectral skewness results for female participants seemed to agree with prior data.

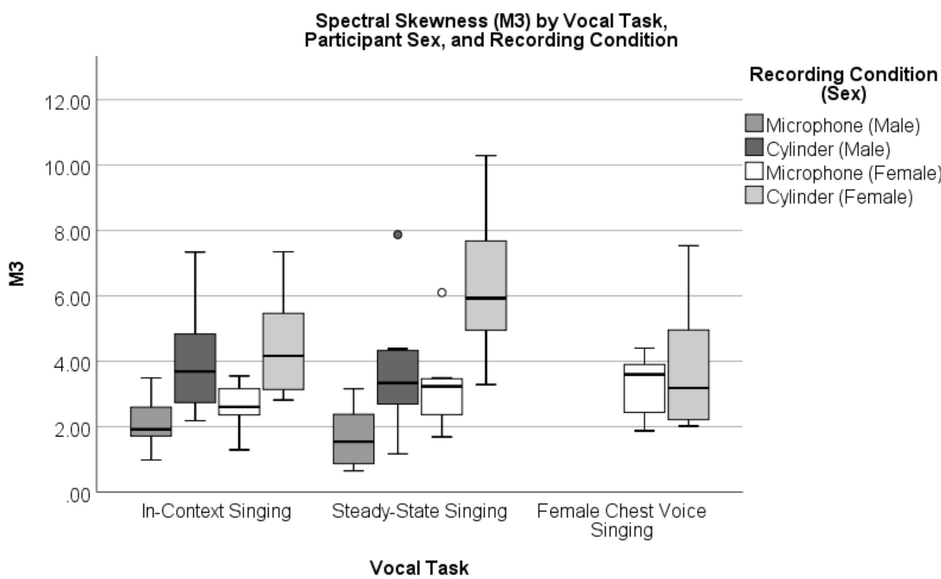


Figure 98. Spectral skewness (M3) differences by task, participant sex, and recording condition.

Many studies in the past have attempted to quantify voice quality or phonation type with various measures of spectral slope (Cesari et al., 2013; Duvvuru & Erickson, 2013; E Mendoza, N Valencia, J Muñoz, & H Trujillo, 1996; M. Guzman, 2013; Hillenbrand et al., 1994; Schutte et al., 2005); however, there is evidence that spectral slope as measured from the harmonic that exhibits the highest relative SIL can lead to more accurate results (Alipour, Scherer, & Finnegan, 2012). As such, spectral skewness may provide a compromise between analytical techniques and has been shown to be indicative of phonation type and CQ (Awan, 2015; Tanner et al., 2005), as well as loudness (Harwardt, 2011).

As stated at the beginning of this sub-section, spectral skewness was significantly higher for the digitized wax cylinder recording condition during in-context and steady-state singing. As reported in Section 4.5, spectral skewness was also found to be higher in the digitized wax cylinder recording condition for female chest voice singing, but results were not statistically significant. This discrepancy could be due to the relatively low SIL's found in the female chest voice singing task and warrants further investigation. If, for example, future study of historical wax cylinder recordings of female chest voice (below F_4) reveal spectral skewness measurements closer to 0 or more similar to values found in recordings of male participants for this study, it could validate claims by scholars such as Crutchfield (2012, pp. 614-615) who suggest that female chest voice techniques have changed since 1900.

It seems that increased spectral skewness in digitized wax cylinder recordings is an effect of wax cylinder phonograph technology and, furthermore, that the measurement holds promising value as an analytical tool to study registration techniques in historical singers. Furthermore, the frequency range of the wax cylinder technology seems to contribute to changes in spectral skewness. While further investigation of this hypothesis is necessary, it is clear that a higher spectral skewness—which indicates a higher concentration of spectral information in lower frequencies and a steep decline in the spectral energy of higher frequencies—represents a consistent pattern in this study and can be considered to be an effect of the wax cylinder phonograph system.

5.4.6 Spectral Kurtosis (M4)

As was discussed in Section 1.4.3, spectral kurtosis describes the "peakedness" of a spectrum's center of gravity and the presence of outliers in the spectrum. Results from this study indicate that spectral kurtosis tends to be significantly higher in digitized wax cylinder recordings during in-context singing and steady-state singing; however, it does not seem to be significantly changed during the female chest voice task (see Figure 99).

It is curious that spectral kurtosis was not significantly altered in digitized wax cylinder recordings during the female chest voice task. As noted in Section 5.4.4, the SFC was completely attenuated in digitized wax cylinder recordings of the female chest voice task, as were all frequencies below 440 Hz. Therefore, it is

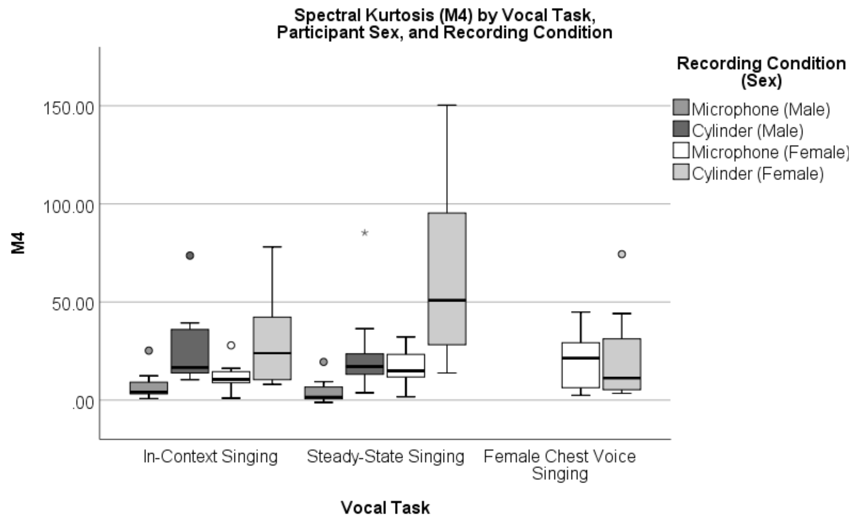


Figure 99. Spectral kurtosis (M4) differences by task, participant sex, and recording condition.

relatively unexpected that spectral kurtosis would not be significantly affected by recording condition for that task.

Further contextualization of this measurement is necessary. *Praat*'s documentation states that spectral kurtosis is a measure of how the shape of the spectrum around the spectral mean is different from the shape of a Gaussian curve. Section 4.5 reported that there were no significant differences between spectral skewness or spectral kurtosis as measured from different recording conditions for the female chest voice task. Additionally, spectral skewness for female participants during the female chest voice task was higher than the same measurement on the same pitch (C_4) for male participants during the steady-state singing task. In the same way, mean spectral kurtosis and spectral skewness measurements from microphone recordings were similar for female participants during steady-state singing and female chest voice tasks. These three factors together with past literature that suggests spectral kurtosis is positively correlated with CQ (Awan, 2015) may be an indicator that modern professional operatic sopranos and mezzo sopranos sing in chest voice with a lighter mechanism. Again, if this is true, it could validate claims made by some scholars that female chest voice singing has changed over time.

As stated in Section 4.6.6, participant S3 was found to have higher spectral kurtosis measurements in both recording conditions during steady-state singing ($M4_{Mic}$ = 71.17; $M4_{Cyl}$ = 150.33) than all of the other participants. The corresponding LTAS (Figure 100) indicated that $2f_o$ was the most prominent during both recording conditions. In comparison to participant S2 (Figure 101), whose spectral kurtosis measurement remained relatively constant between recording conditions, it seems likely that the *Praat* algorithm is highlighting the L_1-L_2 and L_2-L_3 in S3's spectra with higher spectral kurtosis measurements.

Overall, it was interesting to note that spectral kurtosis measurements from the microphone recording condition did not correspond with group mean measurements as studied in amateur singers (Awan, 2015) or in individuals after voice therapy (Tanner et al., 2005). These results may simply indicate that the populations studied were different, but it may also point to the presence of more energy in the f_o for professional opera singers. Further investigation is necessary, but it does seem that spectral kurtosis—as a measure of "peakedness" around the spectral mean—is higher in wax cylinder recordings and can be used to characterize an effect of the wax cylinder phonograph technology.

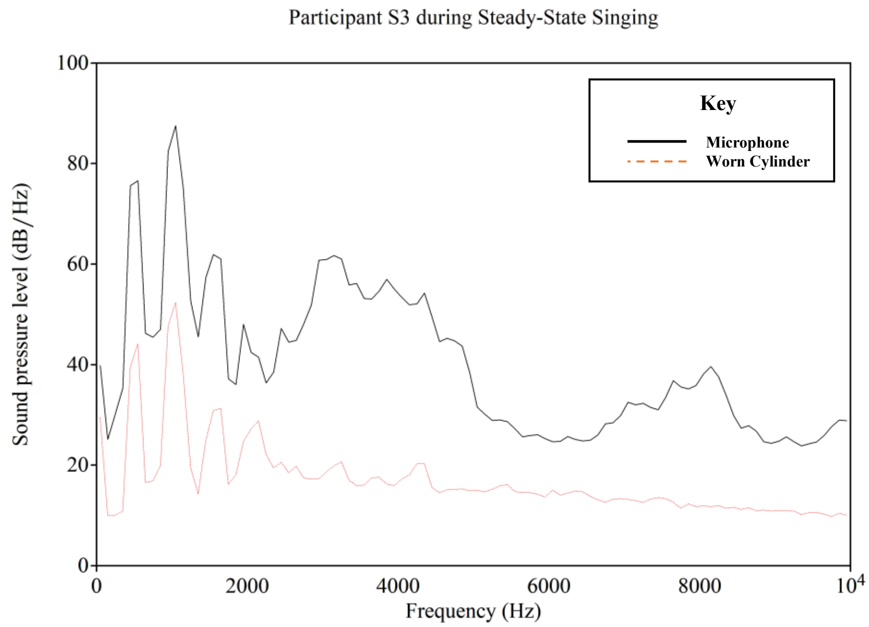


Figure 100. LTAS (200 Hz frequency resolution) of participant S3 during steady-state singing. Notice the relative level differences between f_o, $2f_o$, and $3f_o$, which may have contributed to higher-than-expected spectral kurtosis measurements for this participant.

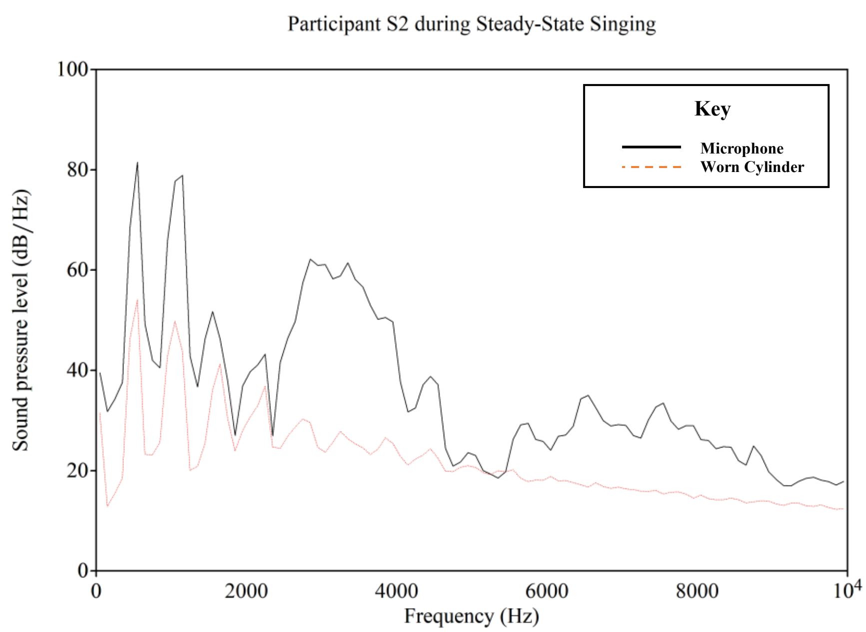


Figure 101. LTAS (200 Hz frequency resolution) of participant S2 during steady-state singing. Notice the relative level differences between f_o, $2f_o$, and $3f_o$, compared to those of participant S3 above.

5.5 Regression Analysis

Regression analysis of metrics that demonstrated significant differences based on recording condition resulted in the creation of five statistically significant regression models. Future study using these models will have to consider the effect of a singer's distance from the horn, the type of horn used, the type of task performed by the original singer, and the filtering effects of the room in which original cylinder recordings were recorded. That is to say, while the regression models created from this study for L_1-L_2 differences and spectral skewness for different tasks may result in a more objective analytical technique, representing the methodology used in this study, they cannot be interpreted as true representations of the voice metrics from original wax cylinder recordings for significantly different recording configurations.

It is of particular relevance to this study that regression models for the aforementioned regression equations accounted for anywhere from 33.3% to 65.1% of the variance in the regression models. This result seems to point to L_1-L_2 differences (during in-context singing, steady-state singing, female chest-voice singing) and spectral skewness (during in-context singing and steady-state singing) as reasonable metrics with which to characterize and study historical singing. Furthermore, regression analysis in combination with analysis of the main effect of recording condition on spectral measurements seems to indicate that L_1-L_2 difference, HNR, and the four spectral moments of the LTAS are effects of wax cylinder phonograph technology during specific tasks. These regression models will be used in future studies to analyze digitized wax cylinder recordings in the University of California Santa Barbara Library's online Audio Cylinder Archive (UCSB Library, 2005) and the Library of Congress's National Jukebox. Furthermore, it is possible that continued analysis of spectral measurements by using participant sex as a dummy variable or by creating regression models that only study one sex may limit the effect of variance on spectral measurements and may be more stable than the ones created in this study. As with any regression analysis, extracting more samples from these recordings may result in more accurate models.

5.6 Case Study

A primary goal of this dissertation was to quantitatively study the effects of wax cylinder phonograph technology on the voice output signal of singers and suggest an objective analytical method with which to study historical recordings of singers. This section presents two case studies of historical singers and modern counterparts (male and female) to demonstrate how the results of this study can be used to understand historical wax cylinder recordings.

Comparative LTAS's were created from two recordings: one token from participant B2's *messa di voce* task and one second from a digitized wax cylinder recording of baritone Antonio Scotti (1866-1936) singing "Bella sicome un angelo" from G. Donizetti's (1797-1848) *Don Pasquale* (1907) which was retrieved from UCSB's Cylinder Audio Archive. Upon inspecting the resultant spectra visually (Figure 102), it was clear that the f_o was attenuated in both samples. Furthermore, it seemed that participant B2's LTAS included a spectral

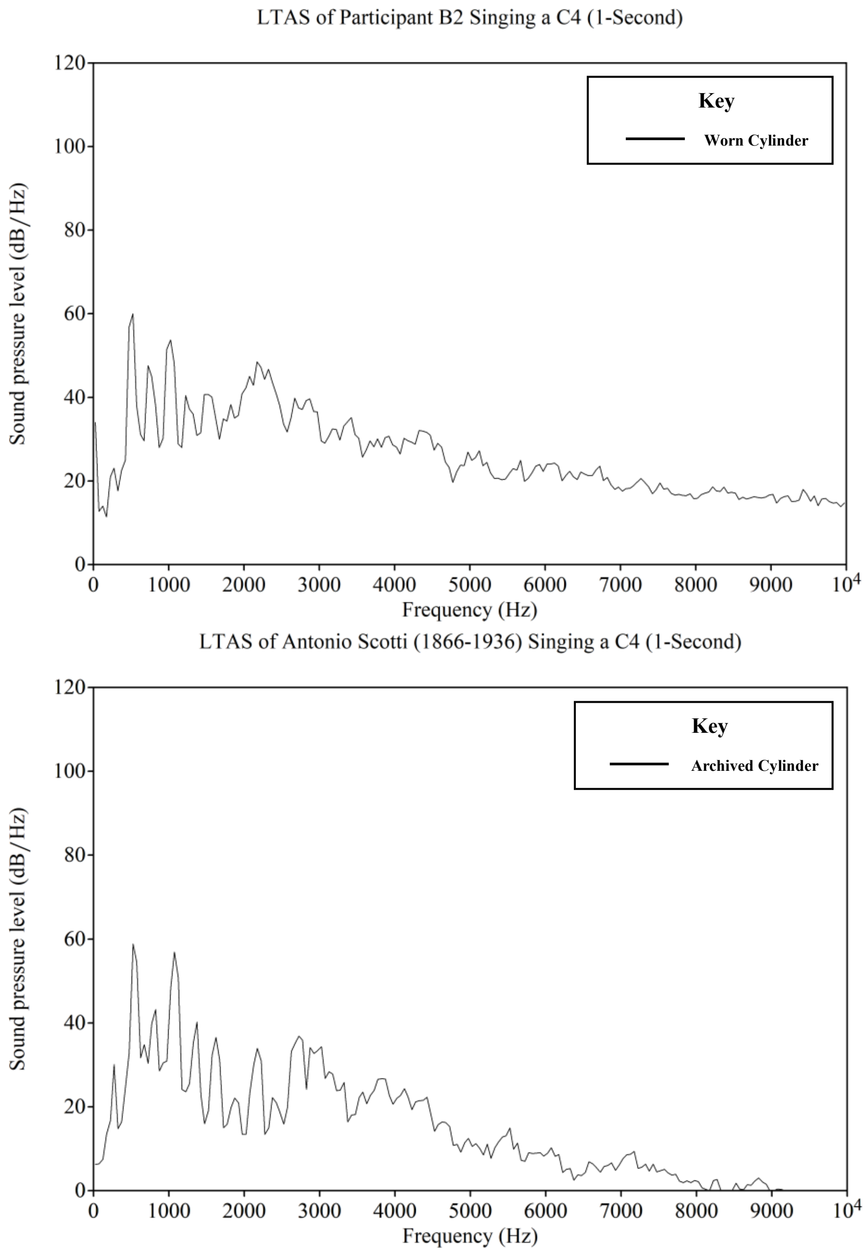


Figure 102. Two LTAS (100 Hz frequency resolution) of Participant B2 and Antonio Scotti (1866-1936) singing the pitch C4 for 1 second. Note the significant attenuation of the fundamental frequency in both spectra and the more prominent spectral peak around 2.25 kHz for the modern singer.

peak near 2.25 kHz which was higher in SIL compared to a comparable spectral peak in Scotti's recording that was present near 2.9 kHz. This observation might indicate that Scotti's singer's formant cluster was closer to that of a modern tenor's and could provide clues about differences between modern and historical operatic male singing.

Further spectral analysis indicated that the spectral skewness for participant B2 (2.98) was lower than the sample from Scotti (3.73). In comparison, the same token for participant B2 recorded on a microphone had a spectral skewness of 1.85. Given that this study reported spectral skewness is significantly higher in digitized wax cylinder recordings, it seems reasonable to infer that the increased prominence of the spectral peak between 2-3 kHz in participant B2's spectrum compared to Scotti might be a difference between professional operatic baritones in the early 20th-century compared to the early 21st-century.

It is interesting to note that, while the first harmonic in the sample of Scotti seemed to be attenuated, it was higher in relation to the second harmonic than in the sample of participant B2. This observation was reflected in the L_1-L_2 difference measurements of Scotti (-29.60 dB) and participant B2 (-34.05 dB). It is possible that the difference between the two samples is a result of each singer's individual vocal characteristics. For example, an L_1-L_2 difference measurement of -29.60 dB is well within the range of wax cylinder recordings of low male voices analyzed in this study (-24.70 dB to -32.17 dB). It is also worth mentioning that the average L_1-L_2 difference of participant B2's tokens during the *messa di voce* task was -30.92 dB. In other words, while the spectra of the two singers may seem to have relative level differences between the first and second spectral peaks, it is possible that the difference is simply a function of the samples analyzed. This comparison emphasizes the need to use multiple samples and singers whenever possible to draw conclusions about singing style and timbre.

The vibrato rate and extent of both singers were analyzed using *VoceVista 3.3.7*. Scotti's vibrato rate (6.8 Hz) was found to be faster than participant B2's (5.3 Hz). Additionally, his vibrato extent (43 cents) was found to be more than 50% lower than participant B2 (87 cents). This difference agrees with past studies that have found a decrease in vibrato rate and an increase in vibrato extent over time (Ferrante, 2011; F. K. Gable, 1992; Seashore, 1937a); however, it is important to note that this study has presented evidence that vibrato extent measurements from digitized wax cylinder recordings might be overestimated. As such, the difference between singers from different time periods is relevant, but the actual measurement might require the proverbial asterisk.

Similar analysis was conducted of participant M1 singing approximately 1 second from the *messa di voce* task (C_5) and contralto, Marie Delna (1875-1934) singing "Mon coeur s'ouvre à sa voix" from *Samson et Dalila* by C. Saint-Saëns (1835-1921) (Figure 103) on the pitch $C\#_5$. Unfortunately, for this analysis, it was not possible to find a sample completely without orchestral accompaniment for the historical recording and, as such, results should be interpreted with

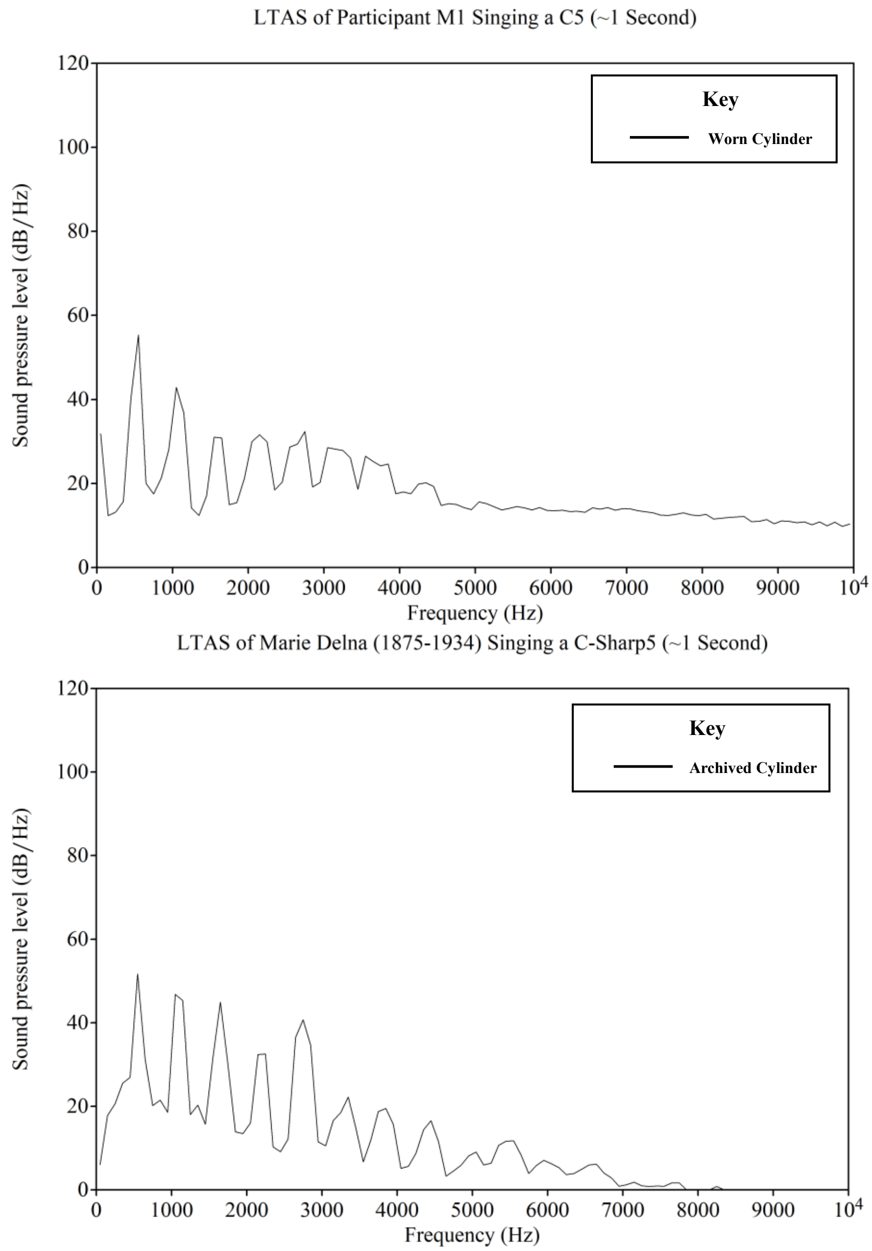


Figure 103. Two LTAS (100 Hz frequency resolution) of Participant M1 and Marie Delna (1875-1934) singing the pitch C5/C#5 for 1 second. Note the difference between the levels of the first and second spectral peaks between the singers.

caution. Future study will undoubtedly include only recordings without accompaniment so as to limit the influence other instruments on the spectrum.

Nonetheless, spectral analysis lead to some interesting differences between the modern and historical professional opera singers.

As stated in this document, the lower portion of this spectrum does not seem to be significantly altered by the wax cylinder phonograph system, because it is above the second resonance of the phonograph horn (~404 Hz). Therefore, it was interesting to observe a difference between L_1-L_2 difference measurements between participant M1 (14.59 dB) and Delna (4.60 dB). This result in combination with the lower spectral skewness found in Delna's sample (1.84) compared to participant M1 (7.86) seem to point to the possibility that Delna was singing with a higher intensity or was closer to the horn. In fact, it is interesting to note that Delna's spectral skewness measurement was lower than that of any of the mezzo soprano wax cylinder recordings made for this study. At the very least, it seems likely that Delna was singing differently than what the modern listener would expect based on these two acoustic measurements.

The LTAS of the modern recording of M1 is also interesting to observe. As seen in Figure 104, there is a relatively large spectral peak near 3.25 kHz. This peak corresponds with a lower spectral skewness in the modern recording of M1 (1.57) and a higher spectral mean (Microphone: 1167.90 Hz, Cylinder 643.45 Hz), even compared to Delna (1006.34 Hz). These differences may suggest either that the recording technique for Delna was different and was able to capture her voice more accurately, or that the two voices were fundamentally different. The vibrato rate and extent of both of these singers was also found to differ (Delna:

6.5 Hz, 56 Cents; M1: 5.6 Hz, 72 Cents), which also may indicate that the voices were technically different.

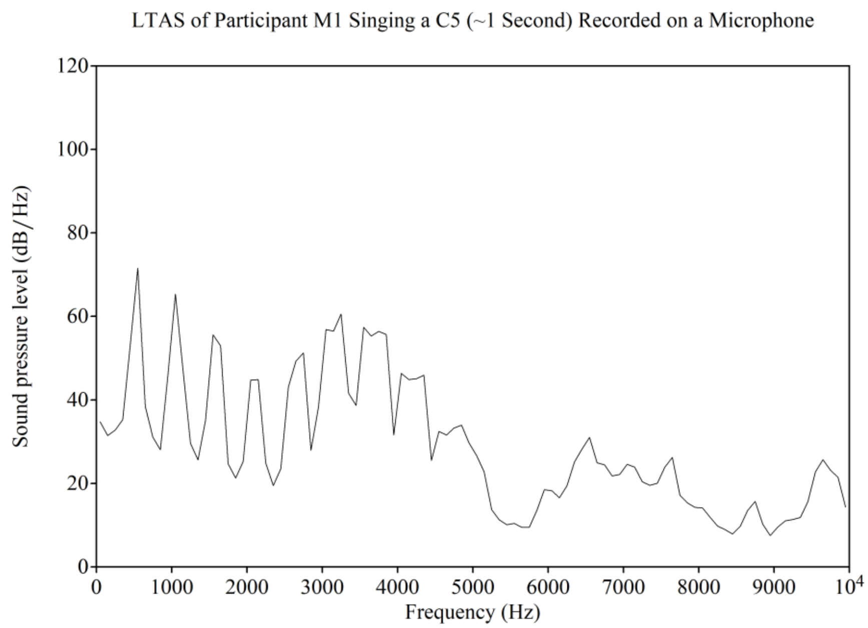


Figure 104. LTAS (100 Hz frequency resolution) of participant M1 singing the *messa di voce* task (1-second). Note the higher spectral peak near 3.25 kHz, which is both higher than the spectral peak observed in the recording of Delna, and seemingly lost in the digitized wax cylinder recording of participant M1.

It is also interesting to observe that harmonic content in Delna's recording continues to nearly 7 kHz compared to participant M1 whose signal fades into the noise floor at approximately 5 kHz. It seems likely based on the results from this study that the recording of Delna is not as worn as the M1 recording. That is to say, the historical recording in this comparison might be a *more* accurate representation of the singer's voice than the modern facsimile created for this study. This possibility does not invalidate the cylinder recordings made during this study, but it does suggest that extant digitized wax cylinder recordings may be relatively true representations of what singers sounded like at the turn of the 20th-century. A logical next step to understand this phenomenon better is to study professional voice teachers' perception of the differences between singers recorded by microphones and singers recorded on wax cylinder phonographs.

5.7 Limitations

There are a number of variables that could not be controlled in this research design. Most of these variables were related to the inability to replicate exact recording conditions from the wax cylinder phonograph era. A clear limitation of this study is the use of brown wax-cylinder replicas. Historical recordings were inscribed onto organic material, which degrades over time. This study used brown wax cylinder replicas to record modern samples. As such, the media was similar to those the earliest sound engineers would have used, but was not identical.

Modern sound engineers use digital signal processing methods to modify audio signal. By "mixing" the recorded audio in post-production, it is changed to fit commercial or artistic standards/aesthetics. Historical recording engineers also used methods to modify their final product; however, they often relied on different horns and changing the position of singers in relation to the horn (National Phonograph Company, 1909; Torick, 1977). Due to this practice, archived historical recordings are always filtered by the room and by the placement of the singer in the room. It is impossible to control for this variable other than maintaining a constant distance from the microphone array and phonograph horn as was done for the current study.

In the same way, this study performed data collection in a sound treated room at New York University, but, obviously, archived digitized wax cylinder recordings were originally created in different spaces. Each room has its own frequency response and reverberation characteristics which affect the final recording that the listener hears. As such, future inquiry using the results from this study should bear in mind that the frequency responses reported in this study contain the filtering effects of the research lab used. Likewise, any comparison between archived digitized wax cylinder recordings must consider that the recordings undoubtedly contain the filtering effects of the room in which the singer was originally recorded.

As was previously mentioned, early recording engineers often used different horns for their recordings. These horns each have their own distinct frequency response—thereby filtering the audio signal (Burkowitz, 1977; D. Davis, Patronis, & Brown, 2013; Zakaria, 2016). While the Phase Two: Singer Recordings (Section 3.4.2) controlled for this variable by using only one period replica horn, it is necessary to note that there is not documentation available regarding the use of horns for specific historical recordings. Therefore, the analysis of historical recordings in Section 5.6 and any future study of archived digitized wax cylinder recordings are limited due to the unknown filtering effects of the period horns used at the time of recording.

Unfortunately, due to technical issues during data collection (see Section 3.5.1), it was not possible to analyze sound pressure level (SPL) measurements between recording conditions or between participants. This error is unfortunate given that SPL seems to play a key role in non-linear aspects of the wax cylinder phonograph system. The spectral and temporal measurements reported in this study do not rely on SPL between participants and were chosen to circumvent issues that arose during data collection.

5.8 Future Research

There are a number of exciting possibilities for future investigation following the results from this study. The most immediate of which is to test the regression equations on historical recordings to attempt to correct spectral measurements and, by doing so, to account for some of the effects of the wax cylinder phonograph system on historical recordings of singers.

Past study culminated in the creation of phonograph horn filters (Zakaria, 2016). Anecdotally, while this current study has shown that the wax cylinder phonograph system significantly changes certain spectral measurements, the cylinder recordings created do not sound wholly different than their equivalent microphone recordings. Furthermore, playback on a period phonograph as reported in Section 3.4.3 sounded perceptually similar to microphone recordings regardless of wear. It is possible that the phonograph horn could contribute to the reduction of high frequency noise and result in a "warmer" timbral characteristic.

As such, convolving Zakaria's playback horn impulse responses with digitized wax cylinder recordings of professional opera singers might explain a technical solution achieved in the early 20th-century by cylinder phonograph manufacturers (i.e. using the horn as a filter).

In the same way, it is apparent that a crucial next step in this study of historical recordings of singers must be related to the perception of these recordings by voice professionals. While this study has shown that the wax cylinder phonograph system significantly alters spectral measurements of the voice output signal, it does not convey the anecdotal evidence that the modern professional opera singers' voices were perceptually similar across recording condition. This result was unexpected. If, indeed, recordings of singers on digitized wax cylinders—even without a playback horn—sound perceptually similar to identical recordings from flat-response microphones, then our field may have to consider the possibility that Western operatic singing has changed in just over 100 years. Furthermore, it would be reasonable to evaluate historical vocal pedagogy texts through this lens. If the technical goals of Western classical singing teachers were different only 100 years ago, then the methods as laid out in historical texts would have been designed to achieve different vocal sounds than are desired in the modern day. This possibility could free the modern-day voice teacher or vocal coach to discover new methods to help singers create new, exciting, and efficient sounds.

An earlier iteration of this study intended to study the contact quotient (CQ) of singers in order to compare vocal fold contact with spectral measurements on flat-response microphone recordings and digitized wax cylinder recordings. Unfortunately, logistical issues during data collection in combination with recent research concerning the accuracy of EGG measurements led to the exclusion of CQ measurements from this study.

It is important to note that more recent studies indicate that CQ data analyzed using the electroglottograph (EGG) signal cannot be used to demonstrate degrees of vocal fold closure (Herbst, Schutte, Bowling, & Svec, 2017). The authors state that it is necessary to study the results from data derived from videokymography in order to make inferences about the closed quotient. While it might be interesting to attempt to study the relationship between CQ and spectral measurements as extracted from digitized wax cylinder phonograph recordings, it might be particularly helpful to attempt to replicate or expand upon past research that has studied the relationship between EGG profile and spectral moments of the LTAS (Awan, 2015).

From an analytical perspective, future singing voice research would be well-served to investigate spectral moments of the LTAS and how they might be altered by style, register, and timbre. The field of vocal pedagogy has used formant tuning analysis—a micro-analysis of the relationship between harmonics and vocal tract resonances—for the better part of twenty years. It is possible that spectral moments might be better suited to explain more nuanced differences between style, vocal register, and timbre that have yet to be explained by traditional analytical methods.

Results from this study seem to indicate that spectral mean may be altered by the task performed by individuals. If this pattern were to hold with the speaking voice, the result could be clinically relevant given that past studies have tied some or all of these measurements to vocal effort, vocal strain, and the effects of voice therapy (Anand et al., 2018; Harwardt, 2011; Tanner et al., 2005). It is also possible that using a long-term average spectrum (LTAS) as the basis for spectral moment analysis could result in the loss of transient data. As such, using a series of short-term Fourier transforms (STFT) and subsequently averaging spectral moment measurements calculated from those spectra might lead to slightly different spectral analyses. While it was not the purpose of this study to measure how spectral moment measurements change based on task, pitch, duration of the sample, and method of calculation, the possibility that these measurements could change in such a way that would alter the results of hypothesis testing points toward the need for future study.

5.9 Conclusion

This chapter has discussed the results of this study and suggested reasons for the changes between recording conditions that were observed. Overall, it seems that any analysis of digitized wax cylinder recordings of singers must take into account the possibility that the first harmonic will not be present in the audio sample if the f_o is below 523 Hz. The absence of the first harmonic will skew any measurement that takes the relative SIL of f_o into account. It is also likely that the absence of f_o will alter the perception of the quality of the sound produced by historical singers, but further study is necessary.

Any analysis of historical singers must take into account the raised noise floor and lower SNR as discussed in Sections 5.2.2 and 5.4.2. It is also possible that the increase in noise above 3 kHz in worn cylinder recordings may contribute to the listeners perception of singers when a low proportion of the singer's radiated sound spectrum is below that level as is found in recordings of female singers. This study has also confirmed that the harmonic-to-noise ratio is lower in digitized wax cylinder recordings and suggests that cylinder wear can account for noise above 3 kHz. Furthermore, these findings may contribute to changes in spectral moment measurements such as spectral standard deviation and spectral mean—the dispersion of frequency information in the spectrum and the average frequency of the spectrum. This study suggests that spectral moments of the LTAS can be used to characterize differences in the shape of the spectrum between recording conditions. They also point to spectral moments as potential descriptors of differences between singers from different time periods, but future study is necessary to confirm or refute that possibility.

Overall, it is clear from the findings of this study that the spectrum is significantly changed in terms of decreased frequency range, increased frequency variance, increased concentration of frequency information below the spectral mean, and increased “peakedness” around the spectral mean. It is also clear that the frequency range of the wax cylinder phonograph technology significantly limits the recording of fundamental frequencies throughout the male singing range and the lower female singing range. That being said, this study validates past research related to vibrato rate and extent, while noting that vibrato extent measurements might be slightly overestimated in digitized wax cylinder recordings.

When considered as a whole, the spectral effects of the wax cylinder phonograph system as well as its relationship with vibrato measurements discussed in this document begin to elucidate how our field can analyze historical recordings of singers, but further study is necessary. Specifically, a number of participants for this study were surprised to discover that their voice did not sound wholly different once the wax cylinder recording was digitized, suggesting the possibility that operatic singing may have changed drastically sometime after 1920. This project has brought to light the need to study how voice teachers perceive singers recorded on this antiquated technology, how voice professionals perceive vibrato, and how spectral moment measurements can be altered by the task performed and by the length of the analysis window. Through this type of marriage between quantitative and musicological methods, it will be possible to understand the history of singing in greater detail and to augment our interpretation of an art form that continues to evolve.

BIBLIOGRAPHY

Alipour, F., Scherer, R. C., & Finnegan, E. (2012). Measures of spectral slope using an excised larynx model. *Journal of Voice, 26*(4), 403-411. doi:10.1016/j.jvoice.2011.07.002

Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2018). Objective indices of perceived vocal strain. *Journal of Voice*, doi:10.1016/j.jvoice.2018.06.005

Awan, S. N. (2015). An examination of the relationship between electroglottographic contact quotient, electroglottographic decontacting phase profile, and acoustical spectral moments. *Journal of Voice - Official Journal of the Voice Foundation, 29*(5), 519-529. Retrieved from https://search.proquest.com/docview/1727348569

Awan, S. N., & Frenkel, M. L. (1994). Improvements in estimating the harmonics-to-noise ratio of the voice. *Journal of Voice, 8*(3), 255-262. doi:10.1016/S0892-1997(05)80297-8

Björkner, E. (2006). *Why so different?: Aspects of voice characteristics in operatic and musical theatre singing.* KTH School of Computer Science and Communication. Stockholm, Sweden.

Björkner, E. (2008). Musical theater and opera Singing—Why so different? A study of subglottal pressure, voice source, and formant frequency characteristics. *Journal of Voice, 22*(5), 533-540. doi:10.1016/j.jvoice.2006.12.007

Bloothooft, G., & Plomp, R. (1986). Spectral analysis of sung vowels, III: Characteristics of singers and modes of singing. *The Journal of the Acoustical Society of America, 79*(3), 852-864. doi:10.1121/1.393423

Boersma, P.Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Paper presented at the *Institute of Phonetic Sciences 17,*

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [computer software]

Boren, B., & Roginska, A. (2011). (2011). Multichannel impulse response measurement in matlab. Paper presented at the *131st Convention of the Audio Engineering Society,*

Boren, B., Roginska, A., & Gill, B.Maximum averaged and peak levels of vocal sound pressure. Paper presented at the *135th Convention of the Audio Engineering Society, 1* , 254.

Bozeman, K. (2013). *Practical vocal acoustics: Pedagogic applications for teachers and singers*. Hillsdale, NY:

Burkowitz, P. K. (1977). Recording, art of the century? *Journal of the Audio Engineering Society, 25*(10-11), 873-879.

Cesari, U., Iengo, M., & Apisa, P. (2013). Qualitative and quantitative measurement of the singing voice. *Folia Phoniatrica Et Logopaedica, 64*(6), 304-309. doi:10.1159/000348734

Chamoux, H.(2004).The archeophone phonograph pages: A universal turntable for edison cylinder. Retrieved from https://www.archeophone.org/warcheophone_specifications.php

Coffin, B. (2002). *Historical vocal pedagogy classics*. Lanham, MD: The Scarecrow Press.

Crutchfield, W. (2012). Vocal performance in the nineteenth century: An overview. In C. Lawson, & R. Stowell (Eds.), *The cambridge history of musical performance* (). New York, NY: Cambridge University Press.

Crutchfield, W. (1983). Vocal ornamentation in verdi: The phonographic evidence. *19th-Century Music, 7*(1), 3-54. doi:10.2307/746545

Davis, D., Patronis, E., & Brown, P. (2013). Voices out of the past. *Sound system engineering* (). Burlington, MA: Focal Press.

Davis, M. F. (2014). The psychoacoustics of audio and electroacoustics. In T. D. Rossing (Ed.), *The springer handbook of acoustics* (pp. 783-787). New York: Springer.

Doscher, B. (1994). *The functional unity of the singing voice* (2nd ed.). US: Scarecrow Press. Retrieved from http://lib.myilibrary.com?ID=392722

Duvvuru, S., & Erickson, M. (2013). The effect of change in spectral slope and formant frequencies on the perception of loudness. *Journal of Voice : Official*

Journal of the Voice Foundation, 27(6), 691-697. doi:10.1016/j.jvoice.2013.05.004

E Mendoza, N Valencia, J Muñoz, & H Trujillo. (1996). *Differences in voice quality between men and women: Use of the long-term average spectrum.* . United States: doi:10.1016/S0892-1997(96)80019-1

Earthworks, I.(2018).Earthworks M30 high definition microphone: Data sheet. Retrieved from https://earthworksaudio.com/wp-content/uploads/2018/07/M30-Data-Sheet-2018.pdf

Elliott, M. (2008). *Singing in style: A guide to vocal performance practices*. New Haven, CT: Yale University Press.

Erickson, M. L. (2004). Sound source identification: The possible role of timbre transformations. *Music Perception, 21*(4), 587-610. doi:10.1525/mp.2004.21.4.587

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Farina, A.Advancements in impulse measurements by sine sweeps. Paper presented at the *122nd Convention of the Audio Engineering Society,* , 1-21.

Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice, 16*(4), 480-487. doi:10.1016/S0892-1997(02)00123-6

Ferrante, I. (2011). Vibrato rate and extent in soprano voice: A survey on one century of singing. *The Journal of the Acoustical Society of America, 130*(3), 1683-1688. doi:10.1121/1.3621017

Fesler, J. C. (1983). Electrical reproduction of acoustically recorded cylinders and disks, part 2*. *Journal of the Audio Engineering Society, 31*(9), 674-694.

Feynberg, R. Y. (2014). *Modern contact transfer: Assessing its effect on wax cylinder audio.* New York University. New York, NY.

Flynn, A., Trudeau, J., & Johnson, A. M. (2018). Acoustic comparison of lower and higher belt ranges in professional broadway actresses. *Journal of Voice,* doi:10.1016/j.jvoice.2018.10.006

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America, 84*(1), 115-123. doi:10.1121/1.396977

Gable, Frederick K. (tr.).(2009).Vibrato as expressive ornament: Translation of summary chapter from das vibrato in der musik des barock (1988) by greta moens-haenen. Retrieved from https://www.york.ac.uk/music/conferences/nema/haenen/#introduction

Gable, F. K. (1992). *Some observations concerning baroque and modern vibrato.* . Claremont, Calif: Claremont Graduate School. doi:10.5642/perfpr.199205.01.09

Genelec. (2007). *Data sheet: Genelec 8030A active monitoring system.*

Guzman, M. (2013). Influence on spectral energy distribution of emotional expression. *Journal of Voice - Official Journal of the Voice Foundation, 27*(1), 12-e10. Retrieved from https://search.proquest.com/docview/1283954194

Guzman, M. A., Dowdall, J., Rubin, A. D., Maki, A., Levin, S., Mayerhoff, R., & Jackson-Menaldi, M. C. (2012). Influence of emotional expression, loudness, and gender on the acoustic parameters of vibrato in classical singers. *Journal of Voice, 26*(5), 675.e-675.e11. doi:10.1016/j.jvoice.2012.02.006

Hakes, J., Shipp, T., & Doherty, E. T. (1988). Acoustic characteristics of vocal oscillations: Vibrato, exaggerated vibrato, trill, and trillo. *Journal of Voice, 1*(4), 326-331. doi:10.1016/S0892-1997(88)80006-7

Hallqvist, H., Lã, F. M. B., & Sundberg, J. (2017). Soul and musical theater: A comparison of two vocal styles. *Journal of Voice, 31*(2), 229-235. doi:10.1016/j.jvoice.2016.05.020

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America, 101*(1), 466-481. doi:10.1121/1.417991

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America, 106*(2), 1064-1077.

Harwardt, C.Comparing the impact of raised vocal effort on various spectral parameters. Paper presented at the *Interspeech 2011,* , 2941-2944.

Hayes, A., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*(4), 709-722. doi:10.3758/BF03192961

Henderson, W. J. (1938). *The art of singing*. Cambridge, MA: Da Capo Press.

Henrich, N. (2007). *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception.* Université Pierre et Marie Curie - Paris VI). Retrieved from http://tel.archives-ouvertes.fr/docs/00/12/31/33/PDF/NH-PhDthesis-301101.pdf

Henrich, N., Wolfe, J., & Smith, J. (2011). Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *The Journal of the Acoustical Society of America, 129*(2), 1024-1035. doi:10.1121/1.3518766

Henrich, N., Wolfe, J., & Smith, J. (2014). Vocal tract resonances in singing: Variation with laryngeal mechanism for male operatic singers in chest and falsetto registers. *The Journal of the Acoustical Society of America, 135*, 491-501. doi:10.1121/1.3518766

Herbst, C. T., Hertegard, S., Zangger-Borch, D., & Lindestad, P. (2017). Freddie mercury-acoustic analysis of speaking fundamental frequency, vibrato, and subharmonics. *Logopedics, Phoniatrics, Vocology, 42*(1), 29-38. doi:10.3109/14015439.2016.1156737

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769-778. doi:10.1044/jshr.3704.769

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research, 38*(6), 1212-23.

Howard, D. M., & Angus, J. A. (2009). *Acoustics and psychoacoustics* (4th ed.). Burlington, MA: Focal Press.

Howell, I.(2015).Unmasking dame nellie melba's vibrato: The challenge of drawing technical conclusions from historical recordings. Retrieved from http://vocped.ianhowell.net/melba/

Howell, I. (2016). *Parsing the spectral envelope: Toward a general theory of vocal tone color.* New England Conservatory. Boston, MA. Retrieved from http://vocped.ianhowell.net/wp-content/uploads/HOWELL-Parsing-the-spectral-envelope-PROQUEST-FINAL-Corrected-Oct-26-2016.pdf

Howell, I. (2017). Necessary roughness in the voice pedagogy classroom: The special psychoacoustics of the singing voice. *VoicePrints: Journal of the New York Singing Teachers' Association*, , 4-7.

Howes, P. (2004). The relationship between measured vibrato characteristics and perception in western operatic singing. *Journal of Voice - Official Journal of the Voice Foundation, 18*(2), 216-230. doi:10.1016/j.jvoice.2003.09.003

Hunter, E. J., Spielman, J., Starr, A., & Popolo, P. (2007). Acoustic voice recording, "I am seeking recommendations for voice recording hardware…". *Perspectives on Voice and Voice Disorders, 17*(1), 7-14. doi:10.1044/vvd17.3.7

IASA Technical Committee.(2009).Guidelines on the production and preservation of digital audio objects. Retrieved from www.iasa-web.org/tc04/audio-preservation

Johnson, A. M., & Kempster, G. B. (2011). Classification of the classical male singing voice using long-term average spectrum. *Journal of Voice, 25*(5), 538-543. doi:10.1016/j.jvoice.2010.05.009

Johnson-Read, L., Chmiel, A., Schubert, E., & Wolfe, J. (2015). Performing lieder: Expert perspectives and comparison of vibrato and singer's formant with opera singers. *Journal of Voice, 29*(5), 645.e1-645.e32. doi:10.1016/j.jvoice.2014.10.020

Kardach, J., Wincowski, R., Metz, D. E., Schiavetti, N., Whitehead, R. L., & Hillenbrand, J. (2002). Preservation of place and manner cues during simultaneous communication: A spectral moments perspective. *Journal of Communication Disorders, 35*, 533-542.

Katz, M. (2010). *Capturing sound: How technology has changed music*. Oakland, CA: University of California Press.

Kauffman, D. (1992). Portamento in romantic opera. *Performance Practice Review, 5*(2), 139-158. doi:10.5642/perfpr.199205.02.03

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820-857. doi:10.1121/1.398894

Klinger, B.(2007).Cylinder records: Significance, production, and survival. Retrieved from http://loc.gov/rr/record/nrpb/pdf/klinger.pdf

Knight, R. M. (1988). *The development of the baritone voice in operas of selected composers from 1750-1830.* Northwestern University. Evanston, IL. Available from ProQuest Dissertations Publishing. Retrieved from https://search-proquest-com.proxy.library.nyu.edu/docview/303709341

Ladofoged, P. (2005). *Vowels and consonants*. Malden, MA: Blackwell Publishing.

Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C. (2011). Spectral- and cepstral-based measures during continuous speech: Capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice, 25*(5), e22-e232. doi:10.1016/j.jvoice.2010.06.007

Marshall, D.(2011).Digital audio effects. Retrieved from http://users.cs.cf.ac.uk.proxy.library.nyu.edu/Dave.Marshall/CM0268/PDF/10_CM0268_Audio_FX.pdf

Maryn, Y., Roy, N., Corthals, P., Van Cauwenberge, P., & De Bodt, M. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America*, *126*(5), 2619-2634. doi:10.1121/1.3224706

McCoy, S. (2004). *Your voice: An inside view* (1st ed.). United States: Inside View Press. Retrieved from http://catalog.hathitrust.org/Record/004968416

Meulengracht-Madsen, H. (1976). On the transcription of old phonograph wax records. *Journal of the Audio Engineering Society*, , 27-32.

Miller, D. G. (2008). *Resonance in singing: Voice building through acoustic feedback.* Princeton, NJ: Inside View Press.

Miller, D. G., & Horne, R. (2008). *VOCEVISTA visual feedback for instruction in singing.*

Miller, D. G., & Schutte, H. K. (1990). *Formant tuning in a professional baritone.* . New York, N.Y: Raven Press. doi:10.1016/S0892-1997(05)80018-9

Miller, D. G., & Schutte, H. K. (2005). 'Mixing' the registers: Glottal source or vocal tract? *Folia Phoniatrica Et Logopaedica*, *57*(5-6), 278-291. doi:10.1159/000087081

Miller, R. (1997). *National schools of singing: English, french, german, and italian techniques of singing*. New York, NY: Schirmer Books.

Miller, R. (1977). *English, french, german and italian techniques of singing: A study in national tonal preferences and how they relate to functional efficiency*. Lanham, MD: Scarecrow Press.

Miller, R. (1996). *The structure of singing: System and art in vocal technique*. Boston, MA: Schirmer, Cengage Learning.

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, *54*(3), 610-619. doi:10.1121/1.1913640

Morange, S., Dubois, D., & Fontaine, J. (2010). Perception of recorded singing voice quality and expertise: Cognitive linguistics and acoustic approaches. *Journal of Voice, 24*(4), 450-457. doi:10.1016/j.jvoice.2008.08.006

Morrison, B. L. (2013). *The chest voice function in the classically trained soprano: A survey of selected vocal pedagogy treatises from the seventeenth century through the twentieth century and recording analysis from 1900 to the present with discussion of the implications for the modern vocal pedagogue.* University of Oklahoma. Norman, OK.

Nandamudi, S. (2017). *Aerodynamics of vocal vibrato.* Bowling Green State University. Bowling Green, OH. Available from Dissertations & Theses @ Bowling Green State University. Retrieved from https://search-proquest-com.proxy.library.nyu.edu/docview/1973499828

National Phonograph Company. (1903). *Edison phonographs catalog*. Orange, NJ: National Phonograph Company.

National Phonograph Company. (1909). *How to make records at home*. Orange, NJ: National Phonograph Company.

Nelson-Strauss, B., Gevinson, A., Brylawski, S., & Loughney, P. (2012). *The library of congress national recording preservation plan.* . Washington, D.C: Council on Library and Information Resources. Retrieved from http://purl.fdlp.gov/GPO/gpo37398

Neumann, K., Schunda, P., Hoth, S., & Euler, H. A. (2005). The interplay between glottis and vocal tract during the male passaggio. *Folia Phoniatrica Et Logopaedica*, *57*(5-6), 308-327. doi:10.1159/000087084

Nix, J., Perna, N., James, K., & Allen, S. (2016). Vibrato rate and extent in college music majors: A multicenter study. *Journal of Voice - Official*

Journal of the Voice Foundation, 30(6), 75-e41. Retrieved from https://search-proquest-com.proxy.library.nyu.edu/docview/1862316779

Osborne, R. (2012). *Vinyl: A history of the analogue record*. Burlington, VT: Ashgate.

Owen, T., & Fesler, J. (1983). Electrical reproduction of acoustically recorded cylinders and disks. *Journal of the Audio Engineering Society, 31*(4), 266-275.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184.

Philip, R. (1992). *Early recordings and musical style: Changing tastes in instrumental performance, 1900-1950*. Cambridge, UK: Cambridge University Press.

Prame, E. (1994). Measurements of the vibrato rate of ten singers. *The Journal of the Acoustical Society of America, 96*(4), 1979-1984. doi:10.1121/1.410141

Reed, V. W. (1983). *The development and use of the tenor voice: 1600 to the present.* Columbia University. New York, NY.

Rothman, H. B., Diaz, J. A., & Vincent, K. E. (2000). Comparing historical and contemporary opera singers with historical and contemporary jewish cantors. *Journal of Voice - Official Journal of the Voice Foundation, 14*(2), 205-214. doi:10.1016/S0892-1997(00)80028-4

Scherer, K. R., Sundberg, J., Fantini, B., Trznadel, S., & Eyben, F. (2017). The expression of emotion in the singing voice: Acoustic patterns in vocal performance. *The Journal of the Acoustical Society of America, 142*(4), 1805-1815. doi:10.1121/1.5002886

Schutte, H. K., Miller, D. G., & Duijnstee, M. (2005). Resonance strategies revealed in recorded tenor high notes. *Folia Phoniatrica Et Logopaedica, 57*(5-6), 292-307. doi:10.1159/000087082

Scotti, A. (1907). *Bella sicome un angelo from don pasquale.* . West Orange, NJ: Edison Records.

Seashore, C. E. (1937a). The psychology of music. VI. the vibrato (1) what is it? *Music Educators Journal, 23*(4), 3-31+33. Retrieved from http://catalog.hathitrust.org.proxy.library.nyu.edu/Record/001657431

Seashore, C. E. (1937b). The psychology of music. VII. the vibrato: (2) what makes it good or bad? *Music Educators Journal, 23*(5), 30-31. doi:10.2307/3384578

Seashore, C. E. (1937c). The psychology of music. VIII. the vibrato: (3) how can we approach an ideal vibrato? *Music Educators Journal, 23*(6), 28-29. doi:10.2307/3384721

Sjölander, P., & Sundberg, J. (2004). Spectrum effects of subglottal pressure variation in professional baritone singers. *The Journal of the Acoustical Society of America, 115*(3), 1270-1273. doi:10.1121/1.1646403

Stan, G., Embrechts, J., & Archambeau, D. (2002). Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society, 50*(4), 249-262. Retrieved from http://www.aes.org/e-lib/inst/browse.cfm?elib=11083

Stark, J. (1999). *Bel canto: A history of vocal pedagogy*. Buffalo, NY: University of Toronto Press.

Steane, J. B. (1993). *The grand tradition: Seventy years of singing on record: 1900-1970*. Portland, OR: Amadeus Press.

Steane, J. B. (1996). *Singers of the century*. Portland, OR: Amadeus Press.

Steane, J. B. (2003a). *Singers of the century*. Portland, OR: Amadeus Press.

Steane, J. B. (2003b). *Singers of the century*. Portland, OR: Amadeus Press.

Sundberg, J. (1994). Acoustic and psychoacoustic aspects of vocal vibrato. *Dept. for Speech, Music and Hearing Quarterly Progress and Status Report, 355*(2-3), 45-68.

Sundberg, J. (1987). *The science of the singing voice*. Dekalb, Ill: Northern Illinois Univ. Pr.

Sundberg, J. (2015). Respiratory and acoustical differences between belt and neutral style of singing. *Journal of Voice - Official Journal of the Voice Foundation, 29*(4), 418-425. Retrieved from https://search.proquest.com/docview/1708747060

Sundberg, J., Andersson, M., & Hultqvist, C. (1999). Effects of subglottal pressure variation on professional baritone singers' voice sources. *The*

Journal of the Acoustical Society of America, 105(3), 1965-1971. doi:10.1121/1.426731

Sundberg, J., & Högset, C. (2001). Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. *Logopedics Phoniatrics Vocology, 26*(1), 26-36. doi:10.1080/140154301300109107

Sundberg, J., La, F. M. B., & Gill, B. P. (2011). Professional male singers' formant tuning strategies for the vowel /a/. *Logopedics Phoniatrics Vocologgy, 36*(4), 156-167.

Sundberg, J., & Nordenberg, M. (2006). Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech. *The Journal of the Acoustical Society of America, 120*(1), 453-457. doi:10.1121/1.2208451

Svec, J. G., Sramkova, H., & Granqvist, S.Basic requirements on microphones for voice recordinggs. Paper presented at the *Maveba 2009,* , 157-160.

Švec, J. G., & Granqvist, S. (2018). Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research, 61*(3), 441-461. doi:10.1044/2017_JSLHR-S-17-0095

Tanner, K., Roy, N., Ash, A., & Buder, E. H. (2005). Spectral moments of the long-term average spectrum: Sensitive indices of voice change after therapy? *Journal of Voice, 19*(2), 211-222. doi:10.1016/j.jvoice.2004.02.005

Titze, I. (2000). *Principles of voice production* (2nd ed.). Iowa City, IA: National Center for Voice and Speech.

Titze, I. (2008). Nonlinear source-filter coupling in phonation: Theory. *The Journal of the Acoustical Society of America, 123*(5), 2733-2749. doi:10.1121/1.2832337

Titze, I., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., . . . Wolfe, J. (2015). *Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization.* . United States: Acoustical Society of America. doi:10.1121/1.4919349

Titze, I., & Winholtz, W. S. (1993). Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research, 36*(6), 1177-1190.

Torick, E. L. (1977). The musician and the record. *Journal of the Audio Engineering Society, 25*(10/11), 880-884.

UCSB Library.(2005).UCSB cylinder audio archive. Retrieved from http://cylinders.library.ucsb.edu/

Välimäki, V., González, S., Kimmelma, O., & Parviainen, J. (2008). Digital audio antiquing-signal processing methods for imitating the sound quality of historical recordings. *Journal of the Audio Engineering Society, 56*(3), 115-139. Retrieved from http://www.aes.org/e-lib/inst/browse.cfm?elib=14378

Vennard, W. (1968). *Singing, the mechanism and the technic* (Fifth ed.). New York, NY: Carl Flescher. Retrieved from http://catalog.hathitrust.org/Record/000004888

Xu, Y. (2013). *ProsodyPro — A tool for large-scale systematic prosody analysis.* Laboratoire Parole et Langage, France. Retrieved from https://www.openaire.eu/search/publication?articleId=od________322::4eddbac39884255979522d73418af97b

Xu, Y., Kelly, A., & Smillie, C. (2013). Emotional expressions as communicative signals. *Prosody and iconicity* (pp. 33) Benjamins. Retrieved from http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&xri:pqil:res_ver=0.2&res_id=xri:ilcs-us&rft_id=xri:ilcs:rec:abell:R04890967

Zakaria, M. A. (2016). *The edison standard phonograph: Digitally recreating three phonograph playback horns for transferred cylinders.* New York University. New York, NY.

APPENDIX A

PARTICIPANT RECRUITMENT LETTER

NEW YORK UNIVERSITY
A private university in the public service

Music and Performing Arts' Professions
New York University
Steinhardt School of Culture, Education, and Human Development
35 West 4th Street, Suite 1077
New York, NY 10012
jdg467@nyu.edu

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

To Whom It May Concern:

I am a Ph.D. Candidate in the Program of Voice Performance at New York University's Steinhardt School of Culture, Education, and Human Development. My research interest involves the study of historical recording technology in order to understand how historical operatic singing has changed throughout the twentieth century.

This study involves recording professional operatic singers between the ages of twenty-five and sixty-five on modern and historical recording equipment as well as measuring how much their vocal folds close while singing. The entire process is non-invasive and will allow me to study how the oldest commercial recording technology impacts the voices it records. Furthermore, this study will lead to future research that focuses on both vocal performance practice and the perception of historical voice recordings.

In order to qualify for this study, it is asked that you fit the following criteria:

- You should be an opera singer (you sing primarily operatic or Western classical repertoire).
- You should be between the ages of twenty-five to sixty-five (25-65).
- You should not have had a history vocal injury or pathology in the past two years.

AND

- You should earn a majority of your income (greater than or equal to 50%) as an opera singer.

OR

- You have sung a principal role in at least one AGMA production in the last two years.

Your participation is appreciated and vital to the progress of this line of research. Unfortunately, many studies in vocal pedagogy and voice science are limited due to the difficulty of finding professional operatic singers to record. In this research project, however, it is necessary to be able to compare professional operatic singers to the greats of the past.

I am happy to speak with you in person or virtually should you have any questions about my research. I can be reached via cellphone at 239.896.0339 or via email at jdg467@nyu.edu.

Regards,

Joshua Glasner, M.M., Ph.D. Candidate
Adjunct Professor of Voice/Steinhardt Fellow
New York University

APPENDIX B

PARTICIPANT CONSENT FORM

NEW YORK UNIVERSITY
A private university in the public service

Music and Performing Arts' Professions
New York University
Steinhardt School of Culture, Education, and Human Development
35 West 4th Street, Suite 1077
New York, NY 10012
jdg467@nyu.edu

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

Consent Form for IRB Study #IRB-FY2017-747

You have been invited to take part in a research study to learn more about how wax cylinder technology impacts voice recordings and how historical voices functioned. This study will be conducted by Joshua Glasner, STEINHARDT - Music & Performing Arts Professions, Steinhardt School of Culture, Education, and Human Development, New York University, as a part of his Doctoral Dissertation. His faculty sponsor is Professor Aaron Johnson, School of Medicine, New York University.

If you agree to be in this study, you will be asked to do the following:

- Complete a questionnaire about your background (age, gender, singing experience, etc.).
- Sing a collection of scales and vocalises into a microphone and wax cylinder recording equipment.
- Wear a non-invasive device called an Electroglottograph (EGG) that measure vocal fold contact while singing.

You will be audio recorded. You may review these recordings and request that all or any portion of the recordings be destroyed.

Participation in this study will involve Approximately 15 minutes of your time: 10 minutes to complete the questionnaire and 2.5 minutes to sing the vocal scales/tasks. It is

possible that you may have to sing the scales/tasks a second time should the recording equipment malfunction. There are no known risks associated with your participation in this research beyond those of everyday life.
Although you will receive no direct benefits, this research may help the investigator understand how wax cylinder technology impacts voice recordings and how historical voices functioned.

Confidentiality of your research records will be strictly maintained by assigning code numbers to each participant so that data is never directly linked to individual identity and keeping all data on a secured Google Drive account that will only be accessible to the investigator. Your information from this study will not be used for future research.

Participation in this study is voluntary. You may refuse to participate or withdraw at any time without penalty. For interviews, questionnaires, or surveys, you have the right to skip or not answer any questions you prefer not to answer.

If there is anything about the study or your participation that is unclear or that you do not understand, if you have questions or wish to report a research-related problem, you may contact Joshua Glasner at (239) 896-0339, jdg467@nyu.edu, 35 West 4th St., Suite 1077, New York, NY 10012, or the faculty sponsor, Aaron Johnson at (646) 754-1207, Aaron.Johnson@nyumc.org, 345 East 37th Street, Suite 306 New York, NY 10016.

For questions about your rights as a research participant, you may contact the University Committee on Activities Involving Human Subjects (UCAIHS), New York University, 665 Broadway, Suite 804, New York, New York, 10012, at ask.humansubjects@nyu.edu or (212) 998-4808. Please reference the study # (#IRB-FY2017-747) when contacting the IRB (UCAIHS).
You have received a copy of this consent document to keep.

Agreement to Participate

____

Subject's Signature & Date

APPENDIX C

MICROPHONE AND MONITOR FREQUENCY RESPONSE CHARTS

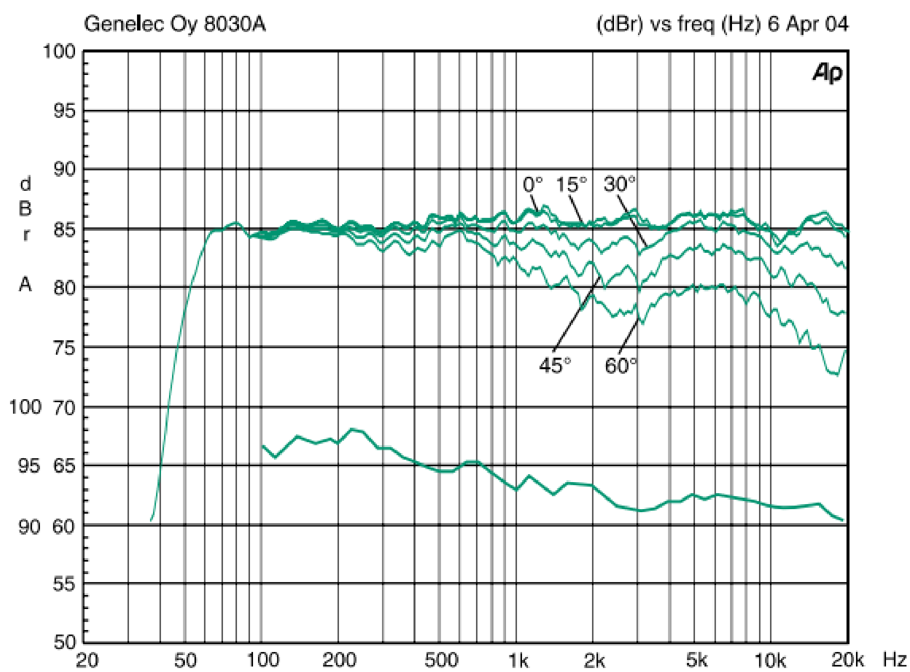
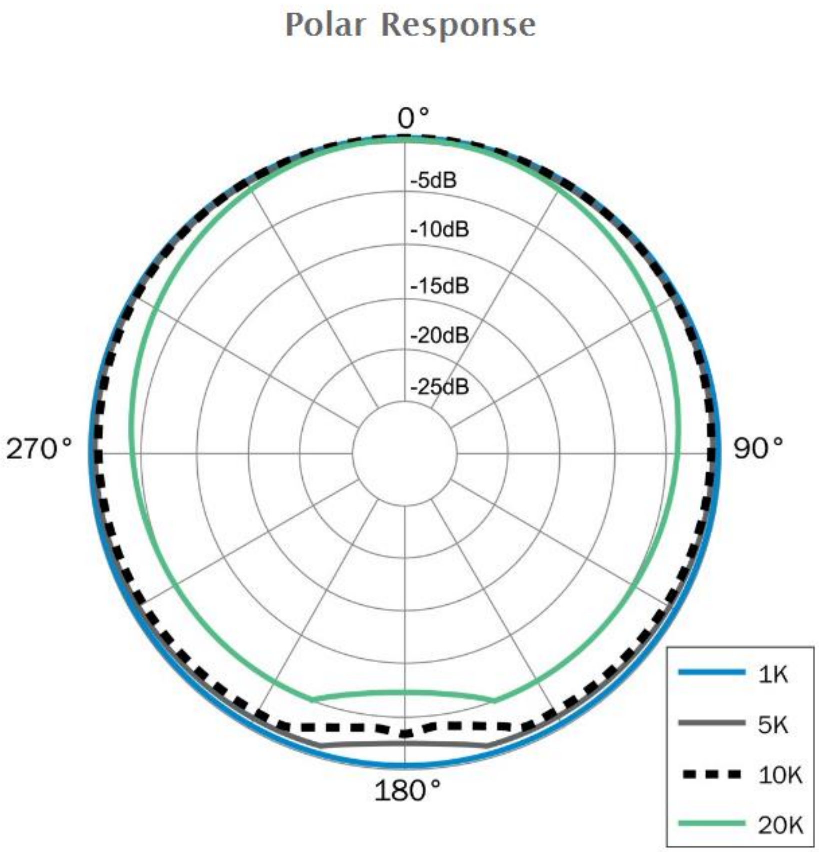
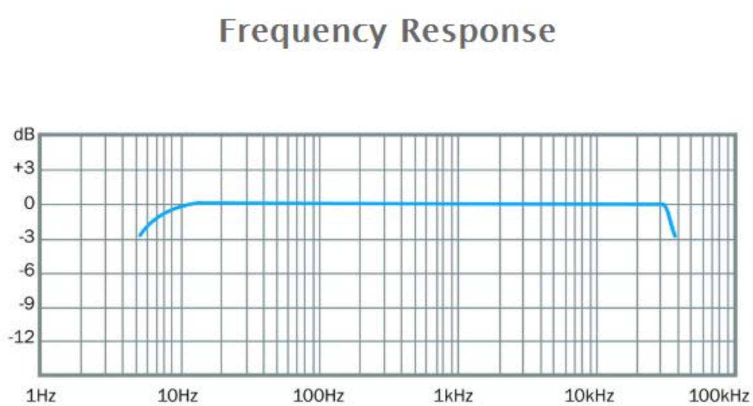


Figure 3: The upper curve group shows the horizontal directivity characteristics of the 8030A measured at 1 m. The lower curve shows the systems power response.

Horizontal directivity frequency response of the Genelec 8030A active nearfield monitor.

(Genelec, 2007)





Published polar response and frequency response of the Earthworks M30 omni measurement microphone.

(Earthworks, 2018)

APPENDIX D

MATLAB AND PRAAT SCRIPTS

Generating Digital Sine-Tones

```
%Matlab File: tone.m
%Modified from Zakaria (2016)

T = 1; %Period
Fs = 44100; %Sampling Rate
N = T * Fs;
t = 0 : 1/Fs : T; %Create array
Fo = 440; %Fo is the desired sine tone frequency
y = sin(Fo*2*pi*t); %Create sine tone
plot(t,y); %Plot sine wave
soundsc(y,Fs); %Play digital sine tone
audiowrite(['sine',num2str(Fo),'.wav'],y,Fs,'BitsPerSample',24); %Write to .wav file
```

Generating Digital White Noise

```
%Matlab File: whiteNoise.m
%Modified from Zakaria (2016)

fs = 44100; %sampling rate
dur = 10; %duration
t = (0:1/fs:dur-1/fs)'; %create array
x = rand( size(t) ) * 2 - 1; %generate white noise
audiowrite(['WhiteNoise',num2str(dur),'s.wav'], x, fs, 'BitsPerSample', 24); %write to .wav file
soundsc(x,44100); %play signal
```

Generating Digital 30 Exponential Sine-Sweep

```
%Matlab File: sweepsine.m
%Modified from Zakaria (2016)

freq1 = 20; %start frequency
freq2 = 20000; %end frequency
fs = 44100; %sampling rate
endTime = 30; %length of sweep
N = endTime * fs;
inst_f = linspace(freq1, freq2, N);
phi = 2 * pi * cumsum(inst_f) / fs;
sweep = sin(phi); %create exponential sine sweep
soundsc(sweep,fs); %play signal
audiowrite('sweep30.wav',sweep,fs,'BitsPerSample',24); %write output to .wav file
```

(Zakaria, 2016)

Generating Frequency-Modulating Sine Wave (Vibrato)

```
%Matlab File: vibrato.m
%From Marshall (2011)

function y=vibrato(x,SAMPLERATE,Modfreq,Width)
ya_alt=0;
Delay=Width; % basic delay of input sample in sec
DELAY=round(Delay*SAMPLERATE); % basic delay in # samples
WIDTH=round(Width*SAMPLERATE); % modulation width in # samples
if WIDTH>DELAY
  error('delay greater than basic delay !!!');
  return;
end

MODFREQ=Modfreq/SAMPLERATE; % modulation frequency in # samples
LEN=length(x);        % # of samples in WAV-file
L=2+DELAY+WIDTH*2;    % length of the entire delay
Delayline=zeros(L,1); % memory allocation for delay
y=zeros(size(x));     % memory allocation for output vector

for n=1:(LEN-1)
   M=MODFREQ;
   MOD=sin(M*2*pi*n);
   ZEIGER=1+DELAY+WIDTH*MOD;
   i=floor(ZEIGER);
   frac=ZEIGER-i;
   Delayline=[x(n);Delayline(1:L-1)];
   %---Linear Interpolation-----------------------------
   y(n,1)=Delayline(i+1)*frac+Delayline(i)*(1-frac);

   %---Allpass Interpolation------------------------------
   %y(n,1)=(Delayline(i+1)+(1-frac)*Delayline(i)-(1-frac)*ya_alt);
   %ya_alt=ya(n,1);
end
```

Generating a Frequency-Modulating Sine Tone (Vibrato)

```
%Call Function vibrato.m
%Matlab File: vibrato_eg.m
%Modified from Marshall (2011)

clear all;
close all;
infile = 'sine440.wav'; %call the .wav sample
[ x, Fs] = audioread(infile); % read in .wav sample

%set parameters for vibrato
```

```
%change these to experiment with vibrato

Modfreq = 6.5; %modulating Frequency
Width = 0.0008; % 0.8 Milliseconds
yvib = vibrato(x, Fs, Modfreq, Width); %generate vibrato
audiowrite('out_vibrato.wav', yvib, Fs); %write output .wav files

% plot the original and equalised waveforms

figure(1)
hold on
plot(x(1:500),'r'); %plots original waveform
plot(yvib(1:500),'b'); %plots frequency-modulated waveform
title('Vibrato First 500 Samples');

soundsc(yvib, Fs); %play the frequency-modulated sine tone
```

(Marshall, 2011)

***ProsodyPro* Praat Script with Alterations**

```
# runs Bio-Informational Dimensions (BID) as described in (Xu, 2013)
# including additional spectral moment measurements
# complete script found on http://www.homepages.ucl.ac.uk/~uclyyix/ProsodyPro/
# retrieved January 2019

procedure BID_measures starttime endtime
	titleline$ = name$
	energytitle$ = "Energy_Profile__250Hz"
	bidline$ = label$
	energyline$ = ""
	select Sound 'name$'
	Extract part... starttime endtime Rectangular 1 no

	select Pitch 'name$'
	median_pitch = Get quantile... starttime endtime 0.5 Hertz
	if median_pitch = undefined
	median_pitch = 100
	endif
	call Energy_bands
	call Formant_dispersion
	call Voice_quality
	select Pitch 'name$'
	Remove

	if !hasBIDtitle
	titleline$ = "'titleline$'
	h1-h2	h1*-h2*	H1-A1	H1-A3	cpp	center_of_gravity	standard_deviation
	skewness	kurtosis	harmonicity	Hammarberg_index	energy_below_500Hz
	energy_below_1000Hz	F_dispersion1_3	F_dispersion1_5
	median_pitch	jitter	shimmer
	'energytitle$"newline$'"
	fileappend 'name$'.BID 'titleline$'
	hasBIDtitle = 1
	endif
	select Sound 'name$'_part
	To Spectrum... yes
	center_gravity = Get centre of gravity... 2
	standard_deviation = Get standard deviation... 2
	skewness = Get skewness... 2
	kurtosis = Get kurtosis... 2
	max_frequency = Get highest frequency
	hammarberg_index = Get band energy difference... 2000 5000 0 2000
	energy500Hz = Get band energy... 0 500
	energy1000Hz = Get band energy... 0 1000
```

```
	total_energy = Get band energy... 0 max_frequency
	energy500Hz = energy500Hz / total_energy
	energy1000Hz = energy1000Hz / total_energy
	plus Sound 'name$'_part
	Remove
	mean_h1_h2 = mean_h1_h2_'m'
	mean_h1_H2 = mean_h1_H2_'m'
	mean_h1_A1 = mean_h1_A1_'m'
	mean_h1_A3 = mean_h1_A3_'m'
	mean_cpp = mean_cpp_'m'
	bidline$ = "'bidline$'					'mean_h1_h2'
	'mean_h1_H2'					'mean_h1_A1'
	'mean_h1_A3'					'mean_cpp'
	'center_gravity'				'standard_deviation'
	'skewness'					'kurtosis'
	'harmonicity'					'hammarberg_index'
	'energy500Hz'					'energy1000Hz'
	'dispersion1_3'					'dispersion1_5'
	'median_pitch'					'jitter'
	'shimmer'					'energyline$''newline$'"
	fileappend 'name$'.BID 'bidline$'
endproc
```

APPENDIX E

VOCAL TASK LISTS (PREPARATION DOCUMENTS)

NEW YORK UNIVERSITY
A private university in the public service

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

Historical Recordings: Vocal Task List (Soprano)

For the following tasks, please sing each phrase in your "best professional voice."

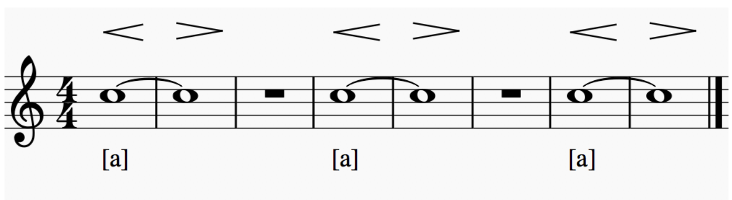
1. **Sing the following phrase on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the scale twice with approximately two seconds between each repetition.

If you are comfortable singing a C6, please sing this phrase instead:

2. **Sing a *messa di voce* on the following pitch on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the task two more times with approximately two seconds between each repetition.



3. **Sing in chest voice on the following pitch on the vowel [a] and sustain it for 5 seconds. Repeat the task two more times with approximately two seconds between each repetition.**

4. **Please sing the following phrase of "Caro mio ben" (G. Giordani).**

NEW YORK UNIVERSITY
A private university in the public service

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

Historical Recordings: Vocal Task List (Mezzo Soprano)

For the following tasks, please sing each phrase in your "best professional voice."

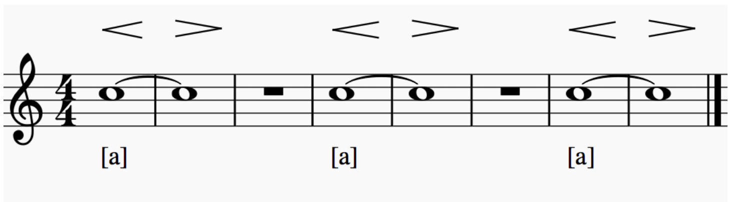
1. **Sing the following phrase on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the scale twice with approximately two seconds between each repetition.

If you are comfortable singing a A5, please sing this phrase instead:

2. **Sing a *messa di voce* on the following pitch on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the task two more times with approximately two seconds between each repetition.



3. **Sing in chest voice on the following pitch on the vowel [a] and sustain it for 5 seconds. Repeat the task two more times with approximately two seconds between each repetition.**

4. **Please sing the following phrase of "Caro mio ben" (G. Giordani).**

NEW YORK UNIVERSITY
A private university in the public service

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

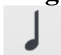
Historical Recordings: Vocal Task List (Tenor)

For the following tasks, please sing each phrase in your "best professional voice."

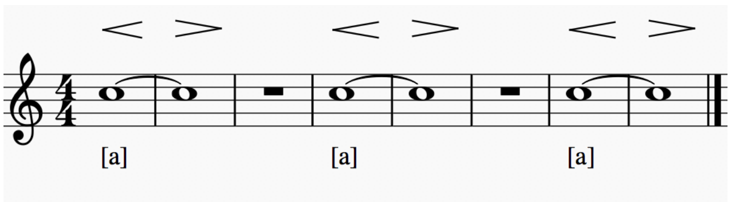
1. **Sing the following phrase on the vowel [a] at a tempo of**  **= 100.**

   Repeat the scale twice with approximately two seconds between each repetition.

If you are comfortable singing a C5, please sing this phrase instead:

2. **Sing a *messa di voce* on the following pitch on the vowel [a] at a tempo of**  **= 100.**

   Repeat the task two more times with approximately two seconds between each repetition.



3. **Sing in *falsetto* on the following pitch on the vowel [a] and sustain it for 5 seconds. Repeat the task two more times with approximately two seconds between each repetition.**

4. **Please sing the following phrase of "Caro mio ben" (G. Giordani).**

NEW YORK UNIVERSITY
A private university in the public service

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

Historical Recordings: Vocal Task List (Baritone)

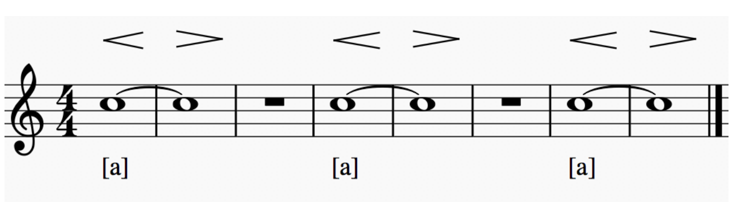
For the following tasks, please sing each phrase in your "best professional voice."

1. **Sing the following phrase on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the scale twice with approximately two seconds between each repetition.

2. **Sing a *messa di voce* on the following pitch on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the task two more times with approximately two seconds between each repetition.



3. **Sing in *falsetto* on the following pitch on the vowel [a] and sustain it for 5 seconds. Repeat the task two more times with approximately two seconds between each repetition.**

4. **Please sing the following phrase of "Caro mio ben" (G. Giordani).**

NEW YORK UNIVERSITY
A private university in the public service

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

Historical Recordings: Vocal Task List (Bass-Baritone)

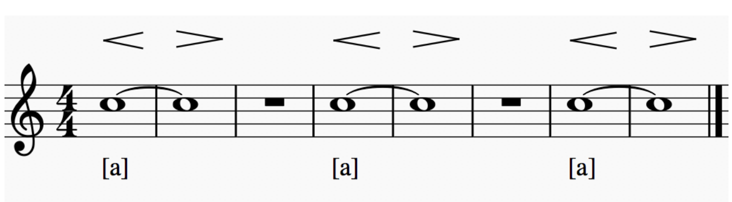
For the following tasks, please sing each phrase in your "best professional voice."

1. **Sing the following phrase on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the scale twice with approximately two seconds between each repetition.

2. **Sing a *messa di voce* on the following pitch on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the task two more times with approximately two seconds between each repetition.



3. **Sing in *falsetto* on the following pitch on the vowel [a] and sustain it for 5 seconds. Repeat the task two more times with approximately two seconds between each repetition.**

4. **Please sing the following phrase of "Caro mio ben" (G. Giordani).**

NEW YORK UNIVERSITY
A private university in the public service

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

Historical Recordings: Vocal Task List (Bass)

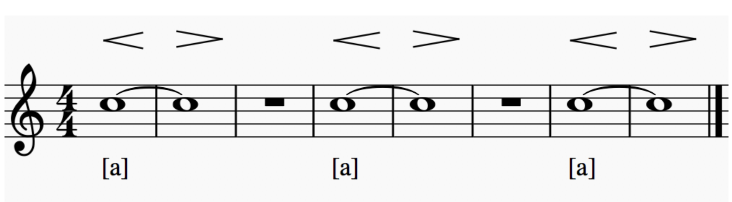
For the following tasks, please sing each phrase in your "best professional voice."

1. **Sing the following phrase on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the scale twice with approximately two seconds between each repetition.

2. **Sing a *messa di voce* on the following pitch on the vowel [a] at a tempo of ♩ = 100.**

   Repeat the task two more times with approximately two seconds between each repetition.



3. **Sing in *falsetto* on the following pitch on the vowel [a] and sustain it for 5 seconds. Repeat the task two more times with approximately two seconds between each repetition.**

4. **Please sing the following phrase of "Caro mio ben" (G. Giordani).**

APPENDIX F

PARTICIPANT QUESTIONNAIRE

NEW YORK UNIVERSITY
A private university in the public service

Joshua Glasner, Ph.D. Candidate

Sponsoring Committee:
Aaron Johnson, Ph.D., Dissertation Chair, Faculty Sponsor
Agnieszka Roginska, Ph.D., Committee Member
Ronald C. Scherer, Ph.D., Committee Member
Ana Flavia Zuim, Ph.D., Committee Member

Historical Recordings: Questionnaire

Voice Type: ____________ **Age:** ____________

Years of Voice Study: _____

Please answer the following questions. All answers are confidential and anonymous. Feel free to ask for clarification if needed.

1. How many years have you performed Opera or Western Classical music full-time?

2. What is your primary style/genre of singing? (e.g. Opera, Concert/Oratorio, Choral)

---

3. What is your highest musical degree earned?

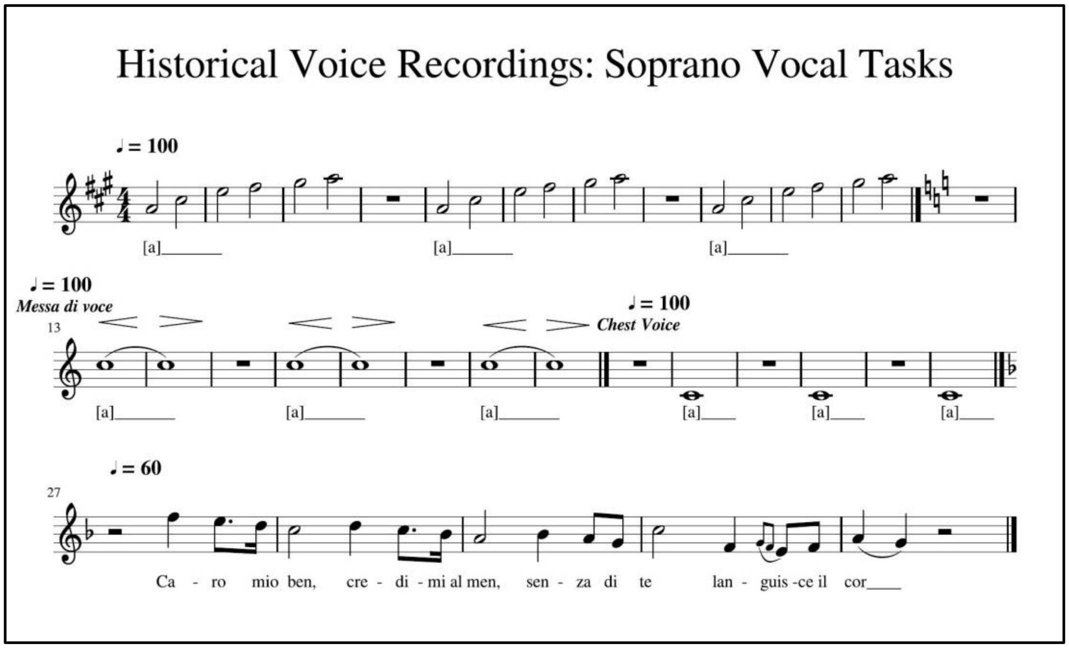
4. Have you had a vocal injury or pathology in the past two years? **Y/N**
5. Have you ever listened to a singer whose performance was recording before 1920? **Y/N**
6. If you answered yes for question 5, please describe what you remember of the singer's voice in one or two sentences.

7. This study uses an Electroglottograph (EGG) to measure when the vocal folds close. It works by placing two electrodes on either side of the larynx (thyroid lamina) and an imperceptible electrical current being passed through your vocal folds. Are you comfortable with a researcher manually placing the electrodes on your neck? **Y/N**

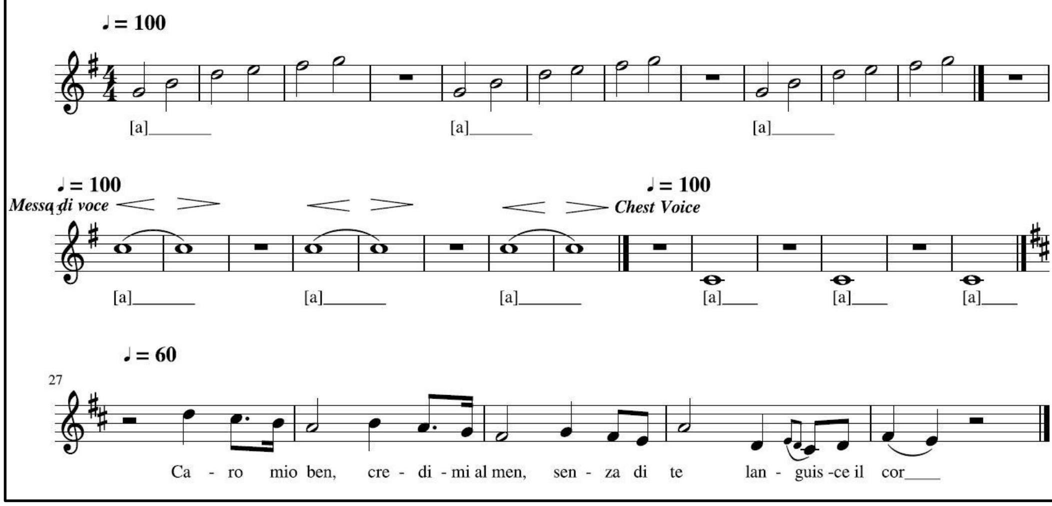
Please let me know if you have any questions about this study, its purpose, or any of the research methods employed. If you have any concerns, please do not hesitate to let me know. It is very important to me that you are comfortable and enjoy your time participating in this exciting research study.

APPENDIX G

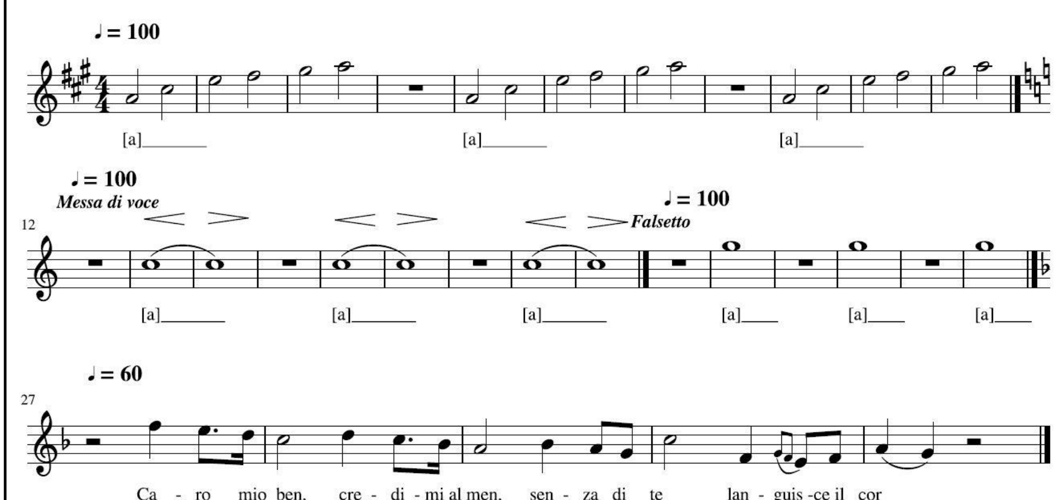
VOCAL TASK LISTS (PERFORMANCE SHEETS)



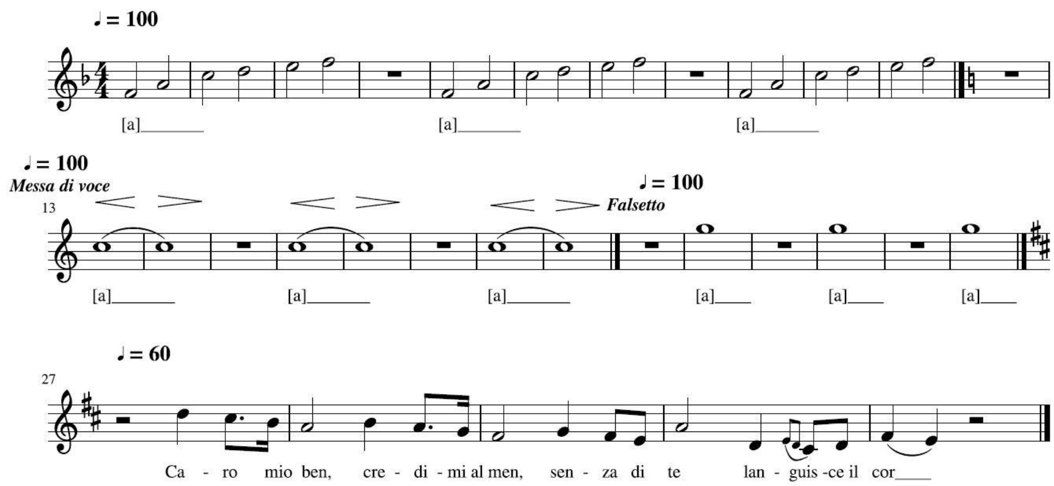
Historical Voice Recordings: Mezzo Soprano Vocal Tasks



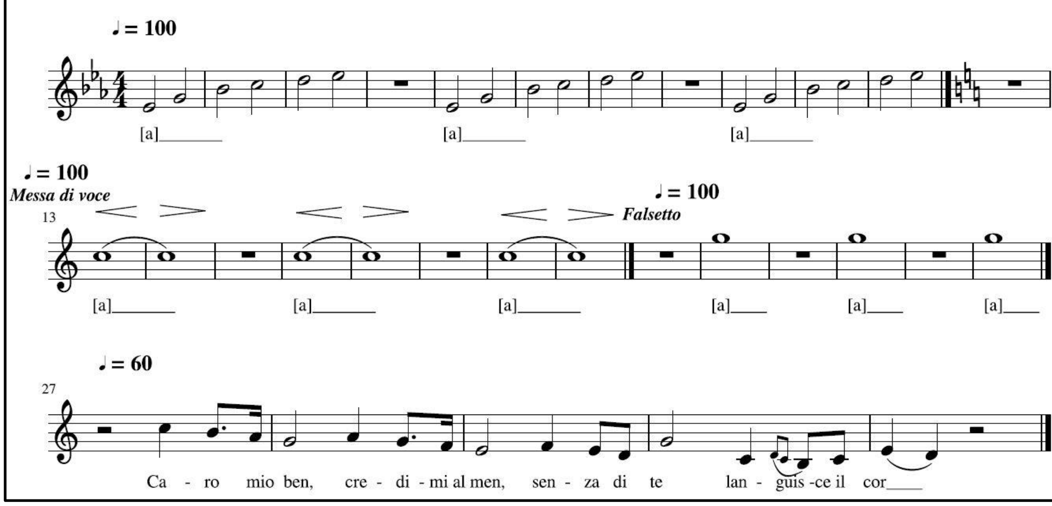
Historical Voice Recordings: Tenor Vocal Tasks



Historical Voice Recordings: Baritone Vocal Tasks



Historical Voice Recordings: Bass-Baritone Vocal Tasks



Historical Voice Recordings: Bass Vocal Tasks