Sound Symbolism of Vowel Metaphors in Choral Pedagogy*

Olivia A. Colace

A thesis submitted in partial fulfillment of the requirements for

the degree of Bachelor of Arts in Linguistics

Bryn Mawr College

December 2024

Abstract

Singers face the daunting task of learning how to precisely control parts of their body that they cannot see. As a result, vocal pedagogy relies heavily on metaphorical language. This thesis will investigate 10 such adjectives used to describe, teach, and elicit vowel sounds in choral singing: "bright", "dark", "wide", "narrow", "open", "closed", "warm", "harsh", "rich", and "thin." This thesis seeks to discover if these vowel metaphors have specific acoustic and thus corresponding articulatory correlates in a sound symbolic capacity. A survey was conducted to assess judgments of sung vowels with respect to each vowel metaphor. Judgments were analyzed with respect to F1, F2, F3, and musician identity. Significant acoustic correlates were found for all 10 adjectives. The results of this study have important implications for vocal pedagogy, the perception of singing, and sound symbolism at large.

*I am grateful to my thesis advisor, Prof. Noah Elkins, for his guidance and support. Thank you to my second reader, Prof. Rikker Dockum, for the helpful, thoughtful feedback. Much love and thanks to my peer reviewers Kit Briggs and Jared Saef and all my classmates. I am absolutely indebted to Ella Foster-Molina for her superior data-wrangling skills and Tao Wang for showing me the way of the regression table. Major thanks are due to Prof. Kirby Conrod for helping me foster the birth of this project amidst some challenging circumstances, and to Prof. Amanda Payne for helping me plant the seeds of what would become this thesis all the way back in my sophomore year. Thank you to Prof. Nate Zullinger for all the books and participants and most of all, for the music. My deepest thanks and love to the Tri-Co Linguistics Department, the Bi-Co Chamber Singers, and all my friends and family who have supported me throughout this journey, and of course for taking my survey. ;)

Table of Contents

1 Introduction

Singing is a unique sort of science. Singers must have exquisite control over many mechanisms they cannot see and do not have precise anatomical knowledge about. While a teacher can tell a pianist to use a different finger to play a note, a voice teacher can't go inside a singer's mouth to adjust the movements of the vocal tract. As a result, voice teachers and choral directors tend to use more metaphorical language to guide singers. This thesis seeks to discover if there is a correlation between these unique metaphors used to describe, teach, and elicit vowel sounds in choral singing and the realized vowel sounds. Ten metaphorical adjectives will be investigated in the present study: "bright", "dark", "wide", "narrow", "open", "closed", "warm", "harsh", "rich", and "thin". Vowel features will be measured by F1, F2, and F3. By assessing individuals' judgments of vowel sounds relative to these adjectives via survey, this thesis will determine if vowel metaphors systematically correlate to acoustic and corresponding articulatory features of vowels in a sound symbolic capacity.

Ultimately, all ten vowel metaphors were found to have correlations with one or more formants. This adds substantially to our knowledge of examples of sound symbolism, moving beyond simply identifying the relationship but characterizing acoustic mechanisms. Additionally, the findings of this study have important implications for vocal pedagogy. This thesis is structured as follows: §2 reviews relevant background about sound symbolism, metaphor, choral singing and pedagogy, and perception of music. §3 describes the survey and subsequent analytical methodology of the study. §4 reports the results. §5 discusses implications and limitations of the findings. §6 discusses applications to choral pedagogy and suggests directions for future study.

2 Background

This section summarizes relevant background that is vital to the logic and importance of the present study. §2.1 discusses important fronts in sound symbolism, §2.2 reviews theories of metaphor, §2.3 discusses specifics of choral music and pedagogy, and §2.4 explores musical perception.

2.1 Sound symbolism

2.1.1 Overview

If indeed intuitions about vowel sounds in choral music are as strong as they appear to be, sound symbolism may be the predominant phenomenon at play. Sound symbolism is the systematic, non-arbitrary relationship between sound and meaning (Kawahara, 2020). One notable example is the *kiki/bouba* effect, which has been proven to be robust across language and culture (Ćwiek et al., 2022). While *kiki* is overwhelmingly associated with a jagged, angular shape and *bouba* overwhelmingly with a bulbous, rounded shape, the numerous variables at play make it difficult to pinpoint exactly which phonetic features are most salient in establishing the shape sound symbolic relationship. Other studies have worked towards a more detailed solution: Knoeferle et al. (2017) found that vowels with a lower F2 (i.e. more back) and higher F3 (i.e. more rounded) are more likely to correspond to a rounder shape, and vice versa for jagged, angular shapes. Furthermore, shape sound symbolism of vowels has been found to be robust in young children (Spector & Maurer, 2013). Size sound symbolic relationships are also well documented — for vowels, high F1 (low vowels), low F2 (back vowels), and increased duration is associated with increased size (Knoeferle et al., 2017; Rojczyk, 2011). Additionally, in the English lexicon itself, the presence of certain phonemes (most significantly /i/, /ɪ/, /ɑ/, and /t/) in a size adjective can predict if it is "large" or "small" (Winter & Perlman, 2021). This finding is argued to hold crosslinguistically, reflecting a universal statistical tendency in language.

However, there is disagreement in the literature over the crosslinguistic validity of sound symbolism. Some find that crosslinguistic validity holds (Ćwiek et al., 2022; D'Anselmo et al., 2019; Kantartzis et al., 2011; Nygaard et al., 2009; Shinohara & Kawahara, 2010; Svantesson, 2017; Yoshida, 2012), while others have discovered language-specific effects influenced by lexical factors, prominence, phoneme inventories, and more (Haynie et al., 2014; Shih et al., 2019). While a more complex picture has emerged, sound symbolism is undoubtedly a broadly prevalent phenomenon (somewhat) inherent to human perception.

2.1.2 Physical explanations for sound symbolism

One hypothesis that is perhaps widely applicable to sound symbolism is the frequency code (Ohala, 1983, 1984, 1995). Briefly, lower frequencies tend to be judged as large, while higher frequencies are judged as small. This connects to the fact that the larger a resonator is, the lower the frequency, and vice versa. This is also relevant in considering the size of the vocal tract and vocal folds in different individuals.

As mentioned above, low vowels (high F1) tend to be judged as larger, while high vowels (low F1) are judged as smaller (Kawahara, 2024; Knoeferle et al., 2017; Shinohara & Kawahara, 2010). The size associations of high and low vowels are thought to have something to do with the size of the oral aperture – oral aperture is physically smaller for high vowels and larger for low vowels, thus providing a basis for the association (Kawahara, 2024). Additionally, low vowels have an inherently lower fundamental frequency (F0) than high vowels (Whalen & Levitt, 1995), meaning the large/small association is consistent with the frequency code. Note however that F1 values here are not consistent with the frequency code.

Similarly, front vowels (high F2) are judged to be smaller than back vowels (low F2) (Knoeferle et al., 2017; Shinohara & Kawahara, 2010) – this may have to do with the fact that the resonating cavity is smaller when the tongue is farther forward in the mouth (i.e. during the production of front vowels), thus contributing to the size judgments. Another explanation could lie in the frequency code – a higher F2 frequency (front vowels) is judged as small, while a lower F2 frequency (back vowels) is judged as larger (Kawahara, 2024).

Perhaps more simply, rounded vowels are likely judged to be round because of the iconicity of lip rounding. Conversely, absence of lip-rounding thus has associations with jagged shapes. There is no evidence suggesting an acoustic basis for this relationship (Kawahara, 2024)

Fascinatingly, color-sound symbolic relationships have been documented as well: [a] is judged to be red, [i] is judged to be yellow/green (Mok et al., 2019; Moos et al., 2014). This likely is rooted in the fact that red light has a lower frequency than yellow/green light, which corresponds with F0 and F2 values of the vowels. Additionally, in synesthetes, [a] is associated with darkness and [i] with lightness (Asano & Yokosawa, 2011). Kawahara (2024) argues this "makes physical sense, because high frequency waves have higher energy." This finding is especially relevant for the present study to see if the same principle holds true for "dark" and "bright" (≈"light") in a choral context.

2.1.3 Sound symbolism in music

Sound symbolism is noted to exist in music as well. People consistently identify the major mode as "happy" and the minor mode as "sad" (Gerardi & Gerken, 1995; Kastner & Crowder, 1990). This association is hypothesized to be connected with the resolution of a "tension" chord such as an augmented chord (composed of the root note, a major third, and a sharpened fifth) (Cook, 2007).[1] Downward half-step resolution of any tone results in a major chord, while upward resolution results in a minor chord. The rest of the logic comes from the frequency code, as introduced above. A low or falling pitch (downwards resolution) is associated with a large size (or resonator), thus establishing a relationship with dominance and strength, translating into "happiness" in the major mode. Conversely, a high or rising pitch (upwards resolution) is associated with a small size (or resonator), thus providing an association with weakness and defeat, or "sadness" in the minor mode. Since this sound symbolic phenomenon is well documented in music, it is reasonable to look for other instances of sound symbolism in music. Furthermore, frequency code may affect the perception of music in other ways.

A sound symbolic relationship between a word such as "bright" and acoustic and thus articulatory features of vowels would allow a teacher to effectively instruct a singer to make a particular adjustment to the vowel sound (vowel modification will be discussed more in depth in §2.3.2). Some adjectives used to describe vowels in choral music are also size/shape adjectives, particularly "tall", "wide", "narrow", "round", "open", and "closed." This raises the question if the sound symbolic association between these words and vowel sounds in choral music environments matches those found in the literature. The iconicity of these words is also not to be ignored, as they can be easily used to describe mouth shape — in this way, some sound symbolic relationships may have anatomical roots, as some described in §2.1.2 do. It is unclear how other

[1] A tension chord is any chord that "wants" to be resolved to a major or minor chord, i.e. creates anticipation of a subsequent major or minor chord.

types of adjectives related to light, taste, temperature, and more, as described above, potentially come to have sound symbolic relationships with vowel sounds, and whether that process would be specific to choral pedagogy or not. The continued investigation of sound symbolism and its potential origins is important in order to understand human perception (Kawahara, 2020; Spector & Maurer, 2013).

2.2 Metaphor

Much of the language surrounding music is by necessity metaphorical – how does one describe what music sounds like? How music makes you feel? How do we distinguish music from merely sound? Metaphors are already inherent to everyday life – they shape how people think, behave, and experience life (Lakoff & Johnson, 1980; Thibodeau et al., 2017).

Metaphor is "understanding and experiencing one kind of thing in terms of another" (Lakoff & Johnson, 1980). Lakoff and Johnson make an important distinction between what they call conceptual metaphors and linguistic metaphors. A conceptual metaphor (denoted in all capital letters) is the mapping itself between two domains, while a linguistic metaphor is how the mapping surfaces in language. A classic example of a conceptual metaphor is TIME IS MONEY: in English, one can *spend*, *lose*, *budget*, *invest* time, etc. (Lakoff & Johnson, 1980). Metaphors are not merely stylistic: they are embedded in the very fabric of language.

Metaphor is used in music to describe musical "space", pitch relationships (governed by PITCH RELATIONSHIPS ARE RELATIONSHIPS IN VERTICAL SPACE) (Zbikowski, 1998), musical "motion" (describing movement of pitch, rhythm, etc. governed by MOVING MUSIC, MUSCICAL LANDSCAPE, and MUSIC AS MOVING FORCE) (Johnson & Larson, 2003), and more. To explain lighting metaphors such as "scintillating" being used to describe the sounds of instruments such as guitars and keyboards, Kolodgy (2008) proposes the conceptual metaphors PITCH RANGE IS INTENSITY OF LIGHT and RELATIONSHIPS IN LUMINOSITY ARE RELATIONSHIPS ON A VERTICAL SCALE. These mappings may be relevant in the usage of "bright" and "dark" in choral pedagogy. Similarly, MUSIC IS TEMPERATURE (Antovic, 2015) gives precedence for the usage of "warm" in vocal pedagogy.

By extension of musical discourse being largely governed by conceptual metaphors, the same must be true for vocal pedagogy (discussed in more detail below in §2.3.1). Perhaps a conceptual metaphor mapping is established first, which is then indexed with acoustic sound symbolic information – a destination for indexing is a perquisite for any indexing to occur.

2.3 Choral music and pedagogy

Choral music is vocal music intended to be sung by multiple people on each part (soprano, alto, tenor, and bass, with fewer or additional divisi as the composer wishes). Western choral tradition emphasizes "classical" repertoire written by European composers from the Renaissance to the 20th century.

2.3.1 Vocal pedagogy

Singing, like speaking, is composed of three main elements – respiration, phonation, and articulation. Singers must have excess air pressure in the lungs in order to support the increased range and duration of pitches compared to speaking. The oscillation of vocal folds transform the airstream into sound waves, while the vocal tract modifies and filters these waves to yield the final sound (Sundberg, 1977, 1999). Some think of singing as "intensified speech" (Appelman, 1986). Singers are taught how to breathe, how to sing notes in tune and with rhythmic accuracy, how to modify diction of segments for voice performance (particularly vowels), how to "shape" vocal lines with stress and dynamics, how to sing with other stylistic elements like legato and vibrato, which is a controlled, periodic oscillation in pitch.[2] All of these elements distinguish singing from speech. Most of all, singers must be taught to create "good" sound quality – what this means in practice will be discussed more later in this section.

Singers must be taught to exquisitely control each of the above elements, even though most parts of the body that control them are invisible (e.g. the lungs, diaphragm, larynx, vocal folds, glottis, usually tongue, etc.). While some direct instruction can be given (e.g. lower your jaw, round your lips, etc.), due to the largely invisible nature of the vocal apparatus, vocal pedagogy by necessity relies heavily on metaphor, gestures, and mental visualization. Additionally, studies have shown that singers lack anatomical knowledge of the vocal tract (Kwak et al., 2014; Rodríguez Marconi et al., 2018; Sielska-Badurek et al., 2017), even as they advance in their careers. This reflects how anatomy is not typically a part of vocal pedagogy. The usage of metaphorical terms including "open", "narrow", "full", "broad", "deep", and "short" to describe and teach vowel sounds in singing is documented as far back as 1771 (Bayly, 1771).

Choral singers are more aware (consciously or not) of vowel sounds and how they're articulated than the average person (Marczyk et al., 2022), but they are usually not formally trained in linguistics. Any instruction in linguistics is often limited and occasionally incorrect. Many articulatory descriptions even in music journals are outright incorrect, for example Cappadonia (1962), Gregg and Scherer (2006), and Jeffers (1988). These include drawing incorrect vowel charts, claiming the tongue is low for all vowels, some incorrect IPA glosses, and incorrect claims about rounding. This lack of linguistic knowledge means that vocal pedagogy is usually quite divorced from articulatory phonetics and linguistics as a whole – both teachers and singers "have been conditioned to think of sounds as 'dark' or 'bright'" in the absence of phonetic knowledge. Thus, words like "dark" and "bright" have only "persist[ed] as a teaching tool" because "the singer has no [articulatory] point of reference for identifying a change in sound" (Appelman, 1986). Students "[remember] the preferred vowel color through repetition," perhaps indexing a word such as "bright" or "dark" with the preferred sound, rather than understanding and replicating the articulation that creates a particular vowel sound.

[2] Vibrato, as opposed to a "straight tone" which has no fluctuation in pitch for a given note, is a characteristic element of opera, while it is generally more stylistic in choral singing.

Ultimately, the goal of a choral musician is to achieve the “best” or most “beautiful” sound — while musicians may have intuitions about what this means, there is no one formula to describe “beauty.” What is considered “beautiful” is also subject to the style of singing, culture, personal preference, and other contexts. Appelman (1986) argues that “all phases of voice culture have been based upon teacher preference of sound. Teachers of voice have assembled their pedagogical skills by imitation from many different sources and by widely differing means.” This elusive and subjective quest for beauty (i.e. sound preference) drives teachers to elicit particular vowel sounds from singers (Cappadonia, 1962; Hosbach-Cannon et al., 2020).

2.3.2 Vowel migration/modification

Secondary to preferences for sound and beauty in singing and imitation of other singers, vowel spaces in singing can be markedly different from those in speaking (Sundberg, 1999). Such differences in vowel spaces have been demonstrated experimentally (Sundberg, 1974). With this group of singers, formant values were generally found to be lower than in speaking (Fig. 1). While vowel spaces of course depend on the speaker/the size of the vocal tract, “vowel migration” is explicitly taught to singers (Appelman, 1986; Sundberg, 1999). Note that vowel migration or modification is distinct from outright replacement of a vowel. This aspect of vocal pedagogy is thought to originate from the Italian teachers of *bel canto* (“beautiful singing”) in the 18th century (Appelman, 1986), a key way of teaching vowel migration being metaphors such as “bright” and “dark”. Appelman argues that “what most teachers hear as a change in vowel color [i.e. quality] is actually a migration of the phoneme from its true acoustic center,” i.e. a change in formant values. This lends credence to the idea that the acoustic effects of teaching metaphors like “bright” and “dark” can be measured through formants.

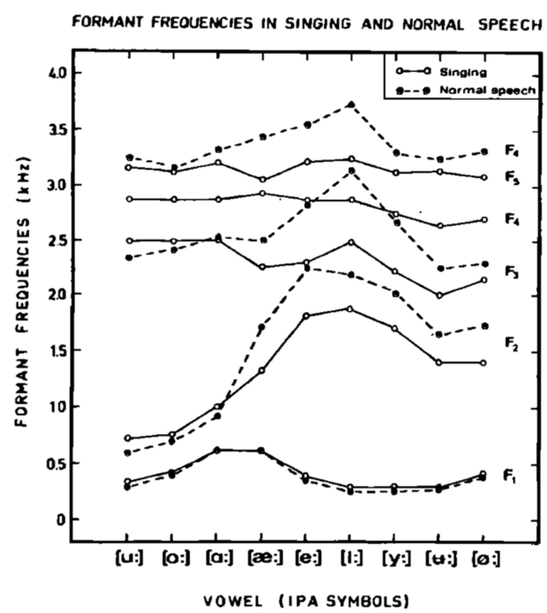

Figure 1. Average formant frequencies of long Swedish vowels of normal male speech (dashed lines) vs those of four professional bass-baritone singers (solid lines) (Sundberg, 1974).

Vowel modification may serve other stylistic purposes like legato, which is achieved by singing notes in a smooth, connected way – it is harder to create legato if vowels are very different from each other. Additionally, there may be some functional reasons to modify vowels – at higher pitches, if a singer doesn't modify the vowel, the F0 will be above the F1. Therefore, the formants will be misaligned with the fundamental and its harmonics, so the sound is inadequately amplified. Vowel modification brings the F1 closer to the F0 and therefore creates better amplification of the fundamental and harmonics (shown visually in Fig. 2), which creates a louder and "better" sound (Sundberg, 1977, 1999).

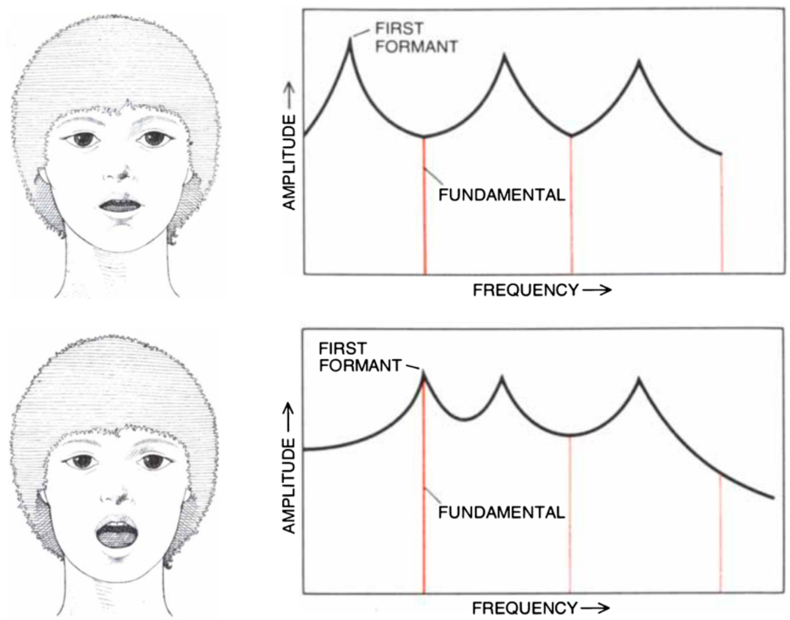

Figure 2. Effect of no vowel modification (top panel) or vowel modification (bottom panel) on formant alignment when singing at higher pitches (Sundberg, 1977). Raising F1 value (i.e. lowering the vowel) brings it closer to F0, allowing for better amplification of the fundamental.

2.4 The perception of music

2.4.1 Physical aspects of musical perception

Musical perception is obviously bound by the limits of what the ear can detect. Human ears are incredibly sensitive to minute changes in sound pressure level (dB), frequency (Hz), and time (milliseconds) (Rasch & Plomp, 1999). Additionally, humans have significant range in aural perception – the human ear can hear from about 20-20,000 Hz (Purves et al., 2001). If we take most singing to lie between the pitches of C2 and C6, the fundamental frequencies (=pitch) of singing generally range from 65.4 Hz to 1046.5 Hz (Byrd, 2007). Harmonic frequencies (integer multiples of the fundamental, also known as partials) are salient up to 10,000 Hz (Pierce, 1999). Singers sing at pitches that individuals would almost never produce when speaking, on both ends of the spectrum. When speaking, adult males generally center around 110 Hz and adult females around 200 Hz. The highest pitches achieved in singing vs speaking diverge dramatically (Table 1) (Sundberg, 1999). Note that formants frequencies are independent of pitch in sung vowels (Pierce, 1999).

	Maximum pitch/F0 (Hz)
Adult male (speaking)	200
Adult female (speaking)	350
Soprano (singing)	1400
Alto (singing)	700
Tenor (singing)	523
Baritone (singing)	390
Bass (singing)	350

Table 1. Maximum pitch in Hz for each voice type speaking and singing (Sundberg, 1999).

Looking for salient variables in musical perception comes down to reverse engineering – investigators must look for differences between sounds that are perceived differently in some respect. Interestingly, this is a major tenet in how musical halls are designed – acoustic engineers must start with perceptive realities and work backwards (Schroeder, 1999). This thinking will be mirrored in the present study: judgments of sounds with known acoustic characteristics are compared to reveal the salient acoustic variables.

2.4.2 Musical perception and cognition

Auditory processing is malleable and dynamic – an enormous amount of information must be processed simultaneously in the perception of music. Perhaps unsurprisingly, from a neurobiological perspective, it is difficult to draw a definitive line between the processing of musical versus non-musical sounds. It is possible to make predictions, but there are no fixed thresholds for what is "musical" and what is not (Weinberger, 1999). However, studies have shown that similar to how linguistic perceptual ability is influenced by native language, musical experience and exposure affects perception of music and auditory processing in a neurologically measurable way (Kraus & Banai, 2007; Mandikal Vasuki et al., 2016). Musicians were found to perform better on auditory tasks than non-musicians, suggesting that musicians have "an enhanced ability to detect statistical regularities in auditory stimuli" (Mandikal Vasuki et al., 2016).

Some elements of musical processing like sensitivity to pitch and rhythm seem to be innate in humans, as informed by their presence in infancy (Dowling, 1999). However, beyond this foundation, musical perceptual abilities must be cultivated and are highly dependent on exposure to and experience with music. Dowling (1999) explains: "Perceptual learning with the music of a culture provides the listener with a fund of implicit knowledge of the structural patterns of that music, and this implicit knowledge serves to facilitate the cognitive processing of music conforming to those patterns." It is perhaps difficult to go through life without a baseline level of exposure to music (dependent on culture, of course), thus so-called non-musicians are expected to have some level of musical discernment. However, this should not discount the increased perceptual abilities of musicians.

In order to communicate with other musicians, or simply to communicate about music in general, these perceptions must be translated into language – for example, the vowel metaphors under investigation in this study. In accordance with the perceptual realities described above, the present study will look for differential judgments between musicians (specifically choral singers) and non-musicians, as well as assessing for variation in judgments due to varying familiarity with individual vowel metaphors.

2.4.3 Timbre and expression

Musicians have many words to describe what essentially comes down to sound quality – common ones include the "color" of a sound, or perhaps more generally the "timbre" of a sound. Timbre is an elusive parameter that is generally defined as the difference between two sounds when pitch and intensity are otherwise the same (Rasch & Plomp, 1999). Some timbres are inherent, like that of an instrument or of an individual singer, but timbre can also be manipulated within these confines. However, features of timbre are broad and hard to define. While frequency, formants, and bandwidth fundamentally shape sound, it is unclear if/when these are the determining factors in creating a timbre or not. Temporal factors such as attack transients and decay patterns have been demonstrated to be important, but other variables are questionable – individuals have been shown to be able to successfully identify instruments (including voices) in very low quality recordings, thus suggesting that specific spectral qualities can't be the determined factor because they are completely distorted in such recordings (Risset & Wessel, 1999).

Additionally, musical expression is a similarly elusive yet uniquely beautiful thing about the human production of music. Music has such an ability to evoke emotions and other sensations in those who play and listen to it, which is tightly linked to the expressivity of the musician. Timbre is an important aspect of expression, but any and every aspect of musical production can be carefully manipulated in a musician's expressive interpretation of a piece. Despite the broad domain of musical expression, experiments have found that singers' intended emotions are identified correctly by listeners 60-80% of the time (Sundberg, 1999). Children have been found to share this ability to perceive emotions in music with their accuracy improving with age, showing that emotion identification in music can be learned (Cunningham & Sterling, 1988; Dolgin & Adelson, 1990; Terwogt & Van Grinsven, 1991).

Vowel quality is often cited as a large (and unique) part of vocal timbre and thus musical expression. This study will further demystify timbre and expression by identifying specific acoustic parameters (F1, F2, F3) that create certain timbres as approximated by various vowel metaphors.

2.4.4 Evidence to suggest a correlation between metaphors and vowel acoustics

Bloothooft & Plomp (1984) showed that elicitation of different vocal timbres or “modes” produced similar acoustic results across singers. They analyzed vowels sung by professional singers in nine different “modes”: “neutral”, “light”, “dark”, “free”, “pressed”, “soft” (pianissimo), “loud” (fortissimo), “straight” (without vibrato), and “extra vibrato”. The first five are outright metaphors, while the latter four have specific and consistent definitions as noted. These words were similarly taken from singing pedagogy, with all the singers having familiarity with them. The singers were allowed to interpret the terms as they liked.

Bloothooft & Plomp were not investigating acoustic correlates of these terms, but rather spectral variance due to different singers, vowels, and modes of singing. With regard to the “modes”, they measured the “extent to which singers realized the mode of singing of the individual vowels in an individual way”, calculated by the difference between the total spectral variance and the sum of the spectral variances due to the main effects and the two-way interactions of all factors (vowel, singer, mode).

Overall, they found that the least spectral variance results from the realization of mode of singing. In other words, the words used to elicit different modes of singing produced acoustically similar results across different singers and vowels (for treble voices, variance = 20-70 dB^2, bass voices, 52-78 dB^2),. This supports the notion that metaphors such as “dark” have a consistent acoustic correlate that is presumably generalizable in singing. To the best knowledge of the author, the present study is the first to seek to discover these semantic-acoustic, sound symbolic relationships.

3 Methods

This study aims to draw correlations between acoustic features of vowels (F1, F2, and F3) and 10 different metaphorical adjectives used in choral pedagogy: “bright”, “dark”, “wide”, “narrow”, “open”, “closed”, “warm”, “harsh”, “rich”, and “thin”. These adjectives were chosen based on personal experiences of the author as a choral singer and the presence of these words in choral, speech pathology, and otolaryngology journals and other literature to describe vowel sounds in the context of singing (to name a few, Appelman, 1986; Bayly, 1771; Bloothooft & Plomp, 1984; Cappadonia, 1962; Fillebrown, 1911; Gregg & Scherer, 2006; Hosbach-Cannon et al., 2020; Jeffers, 1988; Sundberg, 1977; Yanagisawa et al., 1990). Notice that these adjectives are approximately arranged in hypothesized opposite pairs (Table 2) – while these words may be considered opposites in other modes of usage, this may or may not be the case with respect to the perception of sung vowels. Ultimately, opposite relationships held for all pairs in some regard, which will be discussed more at length in §5.1.1.

bright	dark
wide	narrow
open	closed
warm	harsh
rich	thin

Table 2. Adjectives arranged with their hypothesized opposite pair.

3.1 Vowel tokens

10 sung vowel tokens were recorded with an h2n recorder in a soundproof environment. All tokens were sung by the author (mezzo-soprano) at G4 (392 Hz) at the same volume and intensity and without vibrato. The recorded vowels were in the acoustic spaces of [i], [ɪ], [ɛ], [æ], [ə], [u], [o], and [ɑ] (Table 3, Fig. 3). [i̞] and [ʉ] were deliberately sung in a way that is generally not desirable when singing, with their counterparts [i] and [u] being the appropriate choral sung vowels, as judged by the author. The audio files were processed in Logic Pro X to remove the onset and offset of phonation and to normalize the length of each token to 1.0 seconds. Formants of the vowel tokens were extracted manually in Praat using the Burg algorithm (Boersma & Weenink, 2024).

Token number	IPA	F1	F2	F3
1	[æ]	1116	1533	2725
2	[ɑ]	722	1174	2737
3	[ə]	859	1367	2669
4	[ɛ]	534	1572	2727
5	[i]	401	2096	2722
6	[i̞]	398	2247	2727
7	[ɪ]	394	1895	2698
8	[o]	399	1138	2728
9	[u]	398	938	2547
10	[ʉ]	389	1172	2689

Table 3. F1, F2, and F3 values of vowel tokens.

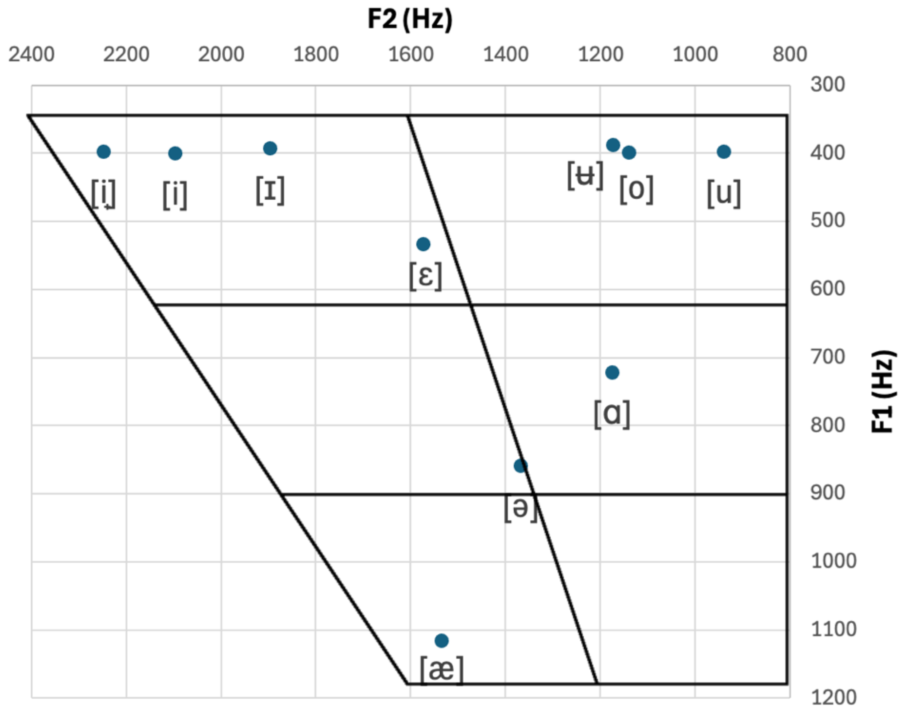

Figure 3. Vowel chart as approximated by F1 (vowel height) and F2 (frontness/backness) of vowel tokens.

Despite [o] and [ʉ] (Tokens 8 and 10, respectively) being very similar in F1-3 values, they sound very different. Since monopthongal [o] doesn't exist in the author's variety of English (Western American), attempts to get that vowel quality yields very peripheral results which can be very acoustically similar to [u]. Other peculiarities in formant values appear to be due to speaker effects and deliberate vowel modification for singing, as discussed in §2.3.2.

3.2 The survey

This study was conducted via survey on Qualtrics.com. The survey assessed participants' judgments about recorded sung vowels in relation to the 10 adjectives listed above ("bright", "dark", "wide", "narrow", "open", "closed", "warm", "harsh", "rich", and "thin").

3.2.1 Participants

Participants were recruited via email and word of mouth. Choral singers, other musicians (non-singers), and non-musicians alike who are fluent readers of English were recruited. Before beginning the survey, participants completed informed consent (Appendix A) and attested they were 18 years of age or older. Incomplete responses (< 90%) were omitted, with $n = 71$ responses ultimately analyzed. Participants were on average 32.3 years old (median = 21 years, SD = 19.7). 63.4% of participants were women, 28.2% men, and 8.4% neither. 80.3% of participants were white, 8.4% were East Asian or Southeast Asian, 5.6% were mixed race, 2.8% were South Asian, 1.4% were Latinx, and 1.4% were Black or African American. First languages were English (84.5%), English and another language (12.7%), and Mandarin (2.8%). Participants were clustered in the Northeast, with the most participants in Pennsylvania (n = 25) (Fig. 4).

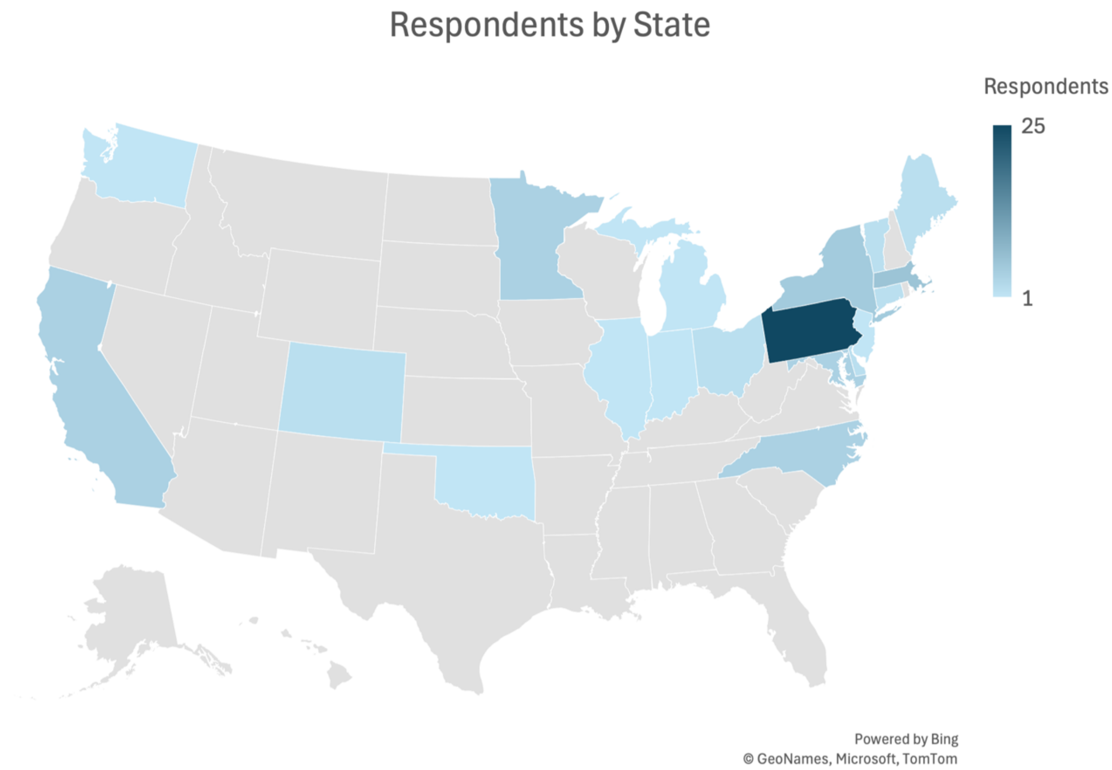

Figure 4. Geographic distribution of respondents by state. States with no respondents are colored gray.

Participants were told they would be listening to sung vowels but were not told that all were sung by the same person, somewhat in the manner of a matched-guise test (Loureiro-Rodríguez & Acar, 2022). This was done to attempt to control for biases they may have, thus preventing them from artificially narrowing their perceptions based on potential perceived innate qualities of the singer.

3.2.2 Collecting judgments

There was 1 experimental section in the survey. The experimental section asked participants to judge how "dark" etc. a vowel token was using a slider (e.g. "not dark at all" to "very dark") (see Appendix A). For each vowel token, all participants were asked to judge each descriptor on individual scales in separate questions. Both measures were taken to not presuppose opposites and allow each adjective to be assessed in isolation against all vowel tokens. All questions were presented in a random order within the experimental section. Participants were shown a random subset of 30 of the 100 total questions.

3.2.3 Demographics and musical background

Survey demographics evaluated age, location, language background, race/ethnicity, as reported above in §3.2.1, and musical background of participants (Appendix A). Musical background was assessed on a score from 0-5, with non-musicians having a score of 0 and the most experienced musicians having a score of 5. The formula is included in Appendix A. Choral singers and other musicians were considered separately. The maximum score was capped at 5 regardless of the raw calculated score. This prevents there being a large distinction between individuals who have been a musician for 20 vs 30 years, 40 vs 50, etc., since these individuals will be assessed in comparison to others with a much wider range of experience, including no experience at all. For the same reason, for the purposes of this study, the distinction between individuals who have for example a doctorate vs a master's in music is not important, and thus this level of education was not assessed. The questions were chosen to gain a broad understanding of an individual's experience while being able to quantify it in a straightforward way. Finally, participants were asked to rate their familiarity with each adjective in a random order using a slider. This accounts for variation in pedagogy (including self-teaching, exposure that non-musicians may have, etc.) and allows for familiarity and musical experience to be tested separately as factors that may influence participants' perceptions. Participants are expected to make more consistent judgments as musical background score and familiarity increases.

3.2.4 Analysis

All statistical analyses and plots were created in R using RStudio with the tidyverse packages (R Core Team, 2021; Wickham et al., 2019). Participants who identified themselves as both choral singers and another type of musician were treated as choral singers. Participants who identified as both musicians and non-musicians were treated as musicians – they explained that they felt they were inexperienced or were a musician in the past. These differences are meant to be accounted for in the musical background score.

Six linear models were formulated to assess significant interactions between the judgments for each metaphor and F1, F2, and F3 and to reveal musician effects. Model 1 predicts judgment values by formant frequencies of F1, F2, and F3. This serves to analyze overall trends. Model 2 is the same as Model 1 but adds musician identity (choral musician, non-choral musician, or non-musician) as a predictor, illuminating if musician identity affects the judgments overall for all 3 formants. Model 3 predicts judgment values by each individual vowel token and musician identity. Model 3 is a robustness check for Model 2. Model 4 is the full linear model, using the interaction terms between each formant (F1, F2, F3) and musician identity to predict judgment values. This model looks for musician effects specific to a particular formant. Model 5 looks at only the significant interactions but keeps other formants as predictors. Model 6 removes insignificant formants. Models 5 and 6 help highlight significant results and are robustness checks for Model 4.

To assess whether musician identity affected consistency of judgments (i.e. if singers and/or musicians make more consistent judgments than non-musicians), Kruskal-Wallis tests were run comparing the standard deviation in judgments of each musician category against each other. Standard deviations were calculated with respect to each vowel token.

To assess whether judgments were more consistent as familiarity with terms increased, linear models were run to assess the relationship between familiarity with the terms and standard deviations of vowel judgments relative to each metaphor. As above with musician identity, standard deviations were calculated with respect to each vowel token. Familiarity scores were rounded to the nearest 10. Model F1 predicts judgment standard deviation by familiarity. Models F2 and F3 are robustness checks that add tokens as predictors.

Musical background scores were categorized 0-1, 1-2, 2-3, 3-4, 4-5. All non-musicians (n = 24) have a score of 0. Most musician respondents were students (n = 27 out of 47 total musicians), so the majority of scores lie in the 2-3 range (among musicians, mean = 2.4, SD = 1.6) (Fig. 5). Due to this inadequate score distribution, no meaningful statistics were able to be produced with these data.

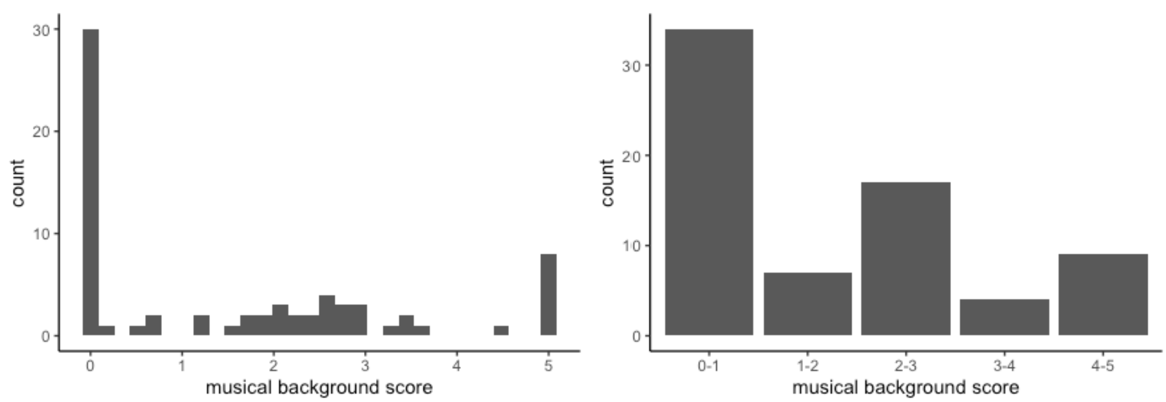

Figure 5. Musical background score distribution. The left displays the distribution of the continuous score. The right displays the distribution of the binned score. Since the scores are not in a normal distribution, no statistics could be generated based on musical background.

4 Results

The results of Model 4 are summarized in Table 4 on the following page. The summaries of all 6 models are included for each metaphor in Appendix B. Scatter plots visualizing these data are included in Appendix C. Supplemental data assessing consistency of judgments based on musician identity and familiarity are included in Appendix D.

	bright	dark	wide	narrow	open	closed	warm	harsh	rich	thin
F1	47.7*** (9.53)	-37.9*** (11.2)	41.4** (12.5)	-6.05 (12.4)	59.0*** (10.9)	-53.1*** (11.5)	-41.0*** (11.8)	69.0*** (14.8)	-44.2*** 12.1	8.72 (12.1)
F2	30.1*** (6.59)	-24.1** (7.53)	11.6 (8.33)	12.3 (8.44)	0.366 (7.19)	-10.8 (8.07)	-29.5*** (8.05)	39.8*** (6.97)	-21.9** (7.05)	13.6. (7.98)
F3	61.9 (51.2)	-22.1 (54.2)	-48.7 (66.1)	-1.66 (65.7)	70.9 (53.5)	-107. (59.7)	-48.9 (61.8)	46.1 (51.5)	61.1 (62.5)	-61.4 (73.6)
mus	-136 (230)	-102 (227)	-130 (300)	86.2 (274)	-604 (300)	-768** (293)	-311 (269)	7.92 (256)	173 (261)	122 (313)
non.mus	170 (201)	173 (209)	-175 (244)	177 (248)	-83.4 (222)	-116 (243)	-232 (253)	132 (224)	309 (256)	0.201 (279)
F1*mus	-13.4 (18.0)	20.4 (17.9)	9.67 (20.9)	-14.9 (21.8)	8.66 (18.2)	-26.0 (22.4)	39.7* (19.2)	-69.3** (23.1)	28.0 (18.5)	-33.4 (21.3)
F1*non.mus	-20.7 (14.8)	28.4 (17.4)	17.6 (18.4)	-48.4** (17.8)	16.0 (16.8)	-25.9 (17.4)	29.8 (18.4)	-30.0 (18.7)	23.3 (18.0)	-44.7* (18.2)
F2*mus	-7.76 (11.8)	10.4 (12.2)	-17.5 (14.3)	20.6 (14.0)	-31.3* (12.5)	9.52 (13.5)	2.58 (12.5)	-3.64 (11.5)	-8.03 (12.0)	2.27 (13.9)
F2*non.mus	-25.5* (10.6)	14.8 (11.1)	-19.7. (11.8)	-14.2 (12.2)	-2.61 (11.6)	11.5 (11.9)	9.07 (11.5)	-8.64 (10.5)	5.26 (11.3)	7.05 (12.8)
F3*mus	56.6 (89.7)	30.0 (88.8)	57.3 (116)	-40.2 (107.9)	241* (115)	282* (113)	108 (105)	11.6 (99.1)	-65.2 (101)	-41.1 (121)
F3*non.mus	-44.6 (77.8)	-78.1 (82.0)	71.6 (94.9)	-46.3 (96.2)	29.0 (86.7)	41.4 (94.3)	76.2 (98.1)	-39.5 (86.6)	-121 (98.1)	4.79 (108)

Table 4. Summary of full linear model (Model 4) for judgments in relation to all metaphors. Estimate coefficients are reported as average change in judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

4.1 Results by metaphor

4.1.1 "bright"

F1 and F2 were found to positively correlate with judgments of vowel "brightness" (Table 4, Table B1, Figures C1 and C2). In other words, the lower and more front a vowel is, the "brighter" it is judged to be. Additionally, there is weak ($p < 0.1$) evidence to suggest that F3 is positively associated with "brightness", i.e. less rounding. Non-musicians were found to be less sensitive (i.e. weaker positive correlation) to F2 than musicians, choral and otherwise, suggesting that F2 is the most salient acoustic variable for musicians in making a "brightness" judgment (Table B1, Figure C2B-D).

4.1.2 "dark"

F1 and F2 were found to negatively correlate with judgments of vowel "darkness" (Table 4, Table B2, Figures C4 and C5), notably the exact opposite of the findings for "bright". Thus, the higher and more back a vowel is, the "darker" it is judged to be. No significant effects were found due to musician identity.

4.1.3 "wide"

F1 was found to have a positive correlation with judgments of vowel "wideness" (Table 4, Table B3, Figure C7). The lower a vowel is, the "wider" it is judged to be. Additionally, there is weak ($p < 0.1$) evidence to suggest that non-musicians positively associate F2 with judgments of "wideness" (Table B3, Figure C8D).

4.1.4 "narrow"

F1 was found to have a negative correlation and F2 a positive correlation with judgments of vowel "narrowness" (Table 4, Table B4, Figures C10 and C11). Thus, "narrowness" is characterized by higher and more front vowels. While these results did not emerge in the full model (Model 4), these relationships are suggested by Models 1 and 2. Interestingly, non-musicians were noted to be more sensitive (i.e. stronger negative correlation) to F1 with respect to "narrowness", while singers were not sensitive to F1 at all (Table B4, Figures C10B and C10D). These findings are notably opposite those of "wideness" with respect to F1.

4.1.5 "open"

F1 and F3 were found to positively correlate with judgments of vowel "openness", though only the F1 correlation emerged in the full model (Table 4, Table B5, Figures C13 and C15). This suggests "openness" corresponds with more front and less rounded vowels. Musicians were found to be more sensitive to F2 (negative correlation) and F3 (stronger positive correlation) than both singers and non-musicians with respect to "openness" (Table B5, Figures C14C and C15C).

4.1.6 "closed"

F1 was found to negatively correlate with judgments of vowel "closedness" (Table 4, Table B6, Figure C16), i.e. the higher a vowel is, the more "closed" it is judged to be. Musicians (non-singers) were found to positively associate F3 with judgments of "closedness" (Table B6, Figure C18C), suggesting that "closedness" is associated with less rounded vowels. These findings notably are opposite those of "open", with the caveat that the F3 finding only held for musicians for "closeness".

4.1.7 "warm"

F1 and F2 were found to negatively correlate with judgments of vowel "warmness" (Table 4, Table B7, Figures C19 and C20), notably very similar to the findings for "dark". Thus, the higher and more back a vowel is, the "warmer" it is judged to be. Conversely, musicians (non-singers) were found to positively associate F1 with judgments of vowel "warmness" (Table B7, Figure C19C).

4.1.8 "harsh"

F1 and F2 were found to positively correlate with judgments of vowel "harshness" (Table 4, Table B8, Figures C22 and C23), notably very similar to the findings for "bright" and opposite the findings of "warm". In other words, the lower and more front a vowel is, the "harsher" it is judged to be. Conversely, musicians (non-singers) were found to negatively associate F1 with judgments of vowel "harshness" (Table B8, Figure C22C).

4.1.9 "rich"

F1 and F2 were found to negatively correlate with judgments of vowel "richness" (Table 4, Table B9, Figures C25 and C26), again similar to the findings for "dark" and "warm". The higher and more back a vowel, the "richer" it is judged to be. No significant effects were found due to musician identity.

4.1.10 "thin"

F2 was found to positively correlate with judgments of vowel "thinness" (Table 4, Table B10, Figure C29). There is weak evidence ($p < 0.1$) to suggest that F1 negatively correlates with judgments of vowel "thinness". However, non-musicians do significantly negatively associate F1 with judgments of vowel "thinness" (Table B9, Figure 28D). In other words, the more front and perhaps higher a vowel is, the "thinner" it is judged to be.

4.2 Exposure effects

Neither familiarity with each vowel metaphor nor musician identity were found to affect consistency of judgments (Table D1, Figure D1), despite choral singers being on average 32%

(±8%) more familiar with each vowel metaphor than other musicians and non-musicians (Table D2) .

4.3 Summary

Table 5 summarizes the correlations found for each vowel metaphor. Figure 6 shows the approximate acoustic destination of each vowel metaphor on a vowel chart.

	F1	**F2**	**F3**	**musician identity effects**
bright	+	+		non.mus less sensitive to F2
dark	-	-		
wide	+			
narrow	-	+		non.mus more sensitive, singers n.s. to F1
open	+		+	mus more sensitive to F2 (-) and F3
closed	-			mus more sensitive overall and to F3 (+)
warm	-	-		mus (+) F1
harsh	+	+		mus (-) F1
rich	-	-		
thin		+		singers n.s., non.mus sensitive to F1 (-)

Table 5. Summary of findings. “+” indicates a positive correlation, “-” indicates a negative correlation. Blank cells indicate no significant correlation. “More sensitive” = stronger correlation (same direction), “less sensitive” = weaker correlation (same direction). “n.s.” – not sensitive/not significant, i.e. no correlation. “mus” = musicians who are not choral singers. “non.mus” = non-musicians.

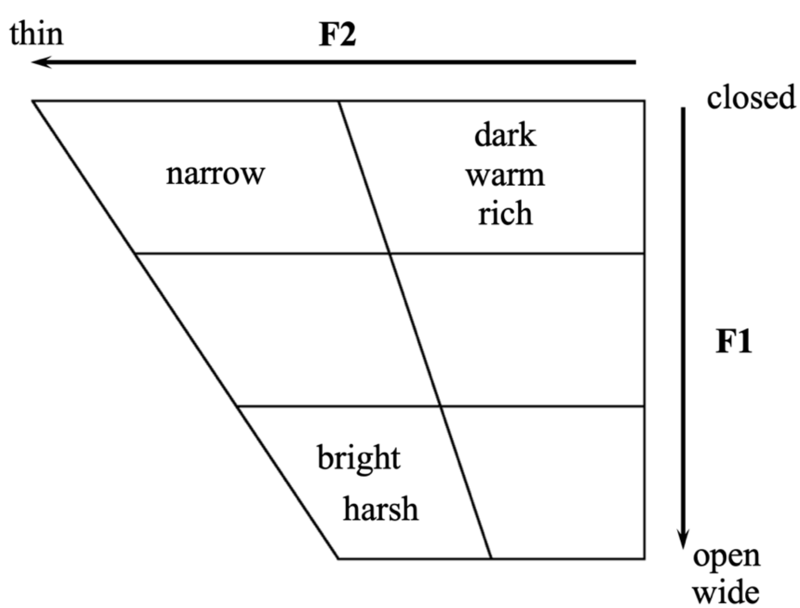

Figure 6. Vowel metaphors plotted on a vowel chart by their approximate acoustic destination.

5 Discussion

5.1 Vowel metaphors have acoustic and articulatory correlates

This study demonstrates sound symbolic, systematic correlations between "warm", "dark", "wide", "narrow", "open", "closed", "warm", "harsh", "rich", and "thin" and formant frequencies. This is the first study to look for sound symbolic acoustic correlates of metaphors used in vocal pedagogy. Note that none of these terms necessarily designate discrete acoustic space – this is distinct from the acoustic space of a vowel where you stray outside of it and consequently land in the acoustic space of another vowel. Rather, these findings illustrate acoustic *destinations* of vowel metaphors used in choral pedagogy. To make a vowel "warmer" for example, it is raised and moved further back – but a vowel doesn't have to necessarily be a high back vowel to be judged as "warm". These are dynamic scales and judgments.

5.1.1 Opposite pairs

Notably, all five hypothesized opposite pairs proved to be opposites (with some caveats). "Bright" (positively correlated with F1 and F2; front low) is opposite to "dark" (negatively correlated with F1 and F2; high back). The personal intuition of the author is that front vowels are "bright" while back vowels are "dark" – it's fascinating and unexpected that vowel height also seems to play a role for these two metaphors. There was weak evidence to suggest "bright" is associated with the absence of rounding, but as will be discussed below in §5.3, F3 values were not very informative.

"Wide" is opposite to "narrow" with respect to F1 – "wide" is associated with lower vowels (high F1), while "narrowness" judgments increase with increasing vowel height (low F1). However, "narrowness" judgments are also positively correlated with F2, so "wide" and "narrow" are not true opposites. There is physical logic in the opposite correlations with F1 – the oral aperture is physically larger ("wider") for low vowels (high F1) and smaller ("narrower") for high vowels (low F1). It is interesting that only "narrowness" judgments corresponded with F2, as F2 also affects the size of the resonating cavity – it is unclear why "wideness" judgments do not also depend on F2.

Next, "open" and "closed" are also opposites with respect to F1 – "openness" is associated with low vowels, "closed" is associated with high vowels. "open" also was found to be associated with the absence of rounding, and there was weak evidence to suggest the same for "closed". Fundamentally, "open" and "closed" are true opposites here – again, F3 values were not very informative, as will be discussed below in §5.3. Here, vowel height is taken to be more salient than absence of rounding. These results are particularly interesting not only because they mirror those of "wide" and "narrow", but also because "open" and "closed" are terms used in vowel phonetics – open vowels are high vowels and closed vowels are low vowels. This study finds that "open" and "closed" share their phonetic meanings. These terms may have arisen in these two domains by similar means – again, as is the case with "wide" and "narrow", the oral aperture is physically larger ("open") for low vowels (high F1) and smaller ("closed") for high vowels (low F1). Humans, whether singers or linguists, can certainly notice these physical realities and thus index vocabularies accordingly.

"Warm" and "harsh" are opposites with respect to F1 and F2 – these findings are analogous to those of "bright" and "dark", with "bright" corresponding to "harsh" (low front) and "dark" to "warm" (high back). Outside of the domain of singing, darker light might be described as "warm", while brighter light could be described as "harsh". Thus, there is a basis for the acoustic correlations to be the same in the realm of choral singing. These findings crucially demonstrate a level of semantic consistency across domains.

Finally, "rich" and "thin" were found to be opposites with respect to F2 – "richness" judgments negatively correlate with F2 ("richer" = more back), while "thinness" judgments negatively correlate with F2 ("thinner" = more forward). This again might have something to do with the size of the resonating space – a larger resonating space (i.e. back vowels) might be construed as "rich", while a smaller one (i.e. front vowels) might be construed as "thin". The same logic might apply to vowel height/oral aperture. However, "richness" negatively correlated with F1, and no relationships were found between "thin" and F1. One might expect a "richer" vowel to have a larger oral aperture, but the opposite was found. The reason for this result is unclear, and it is also unclear why "thinness" judgments do not depend on F1. Also of note is that the correlates of "rich" are the same as "dark" and "warm" – this makes sense on the surface since these three terms are perhaps semantically related to each other outside the domain of singing.

5.1.2 Exposure effects, or lack thereof

Surprisingly, very few differences were observed due to musician identity. It seems that the correlates of the vowel metaphors investigated in this study are not domain-specific to choral singing. There are a few exceptions, which will be discussed below. Singers did not agree on judgments for a particular vowel any more than other musicians or non-musicians, even though singers were on average 32% (±8%) more familiar with each vowel metaphor. Similarly, increasing familiarity did not correspond with increasing agreement on judgments of particular vowel tokens. This suggests that these terms and their acoustic correlates are not specific to the domain of choral singing. These findings reject the model that these vowel metaphors become indexed over time with their acoustic destinations through repetition of the desired sound from a teacher. This is surprising because it is unclear how individuals learn what the correlates are and how the terms come to be indexed with the correlates. These findings are also strange given that the increased aural perceptual abilities of musicians is well-documented (Dowling, 1999; Kraus & Banai, 2007; Mandikal Vasuki et al., 2016). Two potential explanations are proposed at present: 1) since vowels are present in speech and not specific to singing, these terms are indexed in the general population, or 2) participants inaccurately reported their familiarity with terms, perhaps due to subconscious familiarity acquired gradually through cultural exposure.

Despite musician identity not having an overall effect on judgments, there were a few peculiarities in judgments relative to specific formants. The only result that emerged specific to musicians was the positive correlation between "brightness" judgments and F2 – non-musicians were not as sensitive, suggesting that sensitivity to F2 for "brightness" is domain-specific. However, musicians (non-singers) displayed increased sensitivity to several formants for "open" and "closed" and had reverse associations with respect to F1 for "warm" and "harsh". Singers and non-musicians made comparable judgments in these cases, which is surprising. It's not wholly unexpected that other musicians would make different judgments than singers, but it's surprising that this was not accompanied by different judgments by non-musicians. Additionally, non-musicians were more sensitive to F1 for "narrow" and were sensitive to F1 for "thin", while musicians (singers and non-singers) were not. Of particular note is that singers were not found to be sensitive to F1 at all with respect to "narrowness" or "thinness". Again, one might expect non-musicians to be *less* sensitive to acoustic cues, but not *more* sensitive – why would lack of experience with and exposure to music pedagogy result in more and stronger correlations? These results may be artifacts of limitations of the present study, but remain rather a mystery and points for future research.

5.2 Establishing sound symbolic relationships

Since each vowel metaphor investigated in this study was found to have a systematic acoustic correlate, they are all indeed examples of sound symbolism. Many different types of sound symbolism are at play among these ten vowel metaphors. "Bright" and "dark" are examples of light-sound symbolism, "warm" is temperature-sound symbolism, and "harsh" and "rich" are sensory-sound symbolism, for lack of a better term. "Wide", "narrow", "open", "closed", and "thin" are all size-sound symbolism (though an argument could be made for shape). The findings of these latter five are consistent with existing literature on size-sound symbolic relationships – high F1 (low vowels) and low F2 (back vowels) are associated with increased size, while low F1 (high vowels) and high F2 (front vowels) are associated with decreased size (Kawahara, 2024; Knoeferle et al., 2017; Shinohara & Kawahara, 2010). The iconicity of oral aperture, the effects of vowel height on oral aperture, and the effects of vowel frontness/backness on resonant space all provide logical pathways to these sound symbolic relationships, as discussed above in §5.1.1. "Bright", "dark", "warm", "harsh", and "rich" are not consistent with the literature above because they are not examples of size-sound symbolism.

Recall that in in synesthetes, [a] is associated with darkness and [i] with lightness (Asano & Yokosawa, 2011). It is difficult to directly compare the results of this study with Asano & Yokosawa because their experiment was based on orthographical cues and not acoustic ones – but taking [a] to be a back vowel and [i] to be a front vowel, this finding is consistent with the results of this study with respect to “bright” and “dark”, and by extension “warm”, “harsh”, and “rich”, as discussed in §5.1.1.

The ability of light to reach different parts of the oral cavity may play a role in light-sound symbolic relationships and related sensory-sound symbolism. In typical lighting, light will fall primarily on the front and bottom of the oral cavity when opened, making these parts of the oral cavity literally “brighter.” Conversely, less light can reach the back and roof of the oral cavity, making these regions literally “darker”. Additionally, conceptual metaphor mappings such as MUSIC IS TEMPERATURE (Antovic, 2015) and PITCH RANGE IS INTENSITY OF LIGHT (Kolodgy, 2008) may contribute to temperature- and light-sound symbolism.

5.3 Limitations

This study has a number of limitations. First, the participant pool inherently limited many aspects of analysis. Most participants were white women from the Northeast US, therefore representing limited backgrounds that therefore could influence vowel judgments. Additionally, there was an unequal distribution of singers, other musicians, and non-musicians, and an unequal range of experience within the musicians. This meant the sample sizes were unequal, leading to unequal and rather low statistical power. This led to an inadequate distribution of musical background scores, which meant musical background score was unable to be used as a parameter in this study.

Furthermore, inaccuracies of self-reporting may have influenced the results of this study. Participants may have judged tokens relative to each other or in general, which also may have influenced judgments. Additionally, participants may have had varying interpretations of what is “very” and “not at all” in relation to each metaphor, even if they were responding to the same acoustic cues as other participants. Finally, as previously discussed, participants may not have accurately reported their familiarity with each vowel metaphor.

Survey questions were randomized unequally, with each question shown to anywhere between 13 and 31 participants (mean = 21, SD = 3.6). This means unequal numbers of responses were analyzed per token and per metaphor, with varying distributions of musician type responding to each question. This makes it difficult to find meaningful statistical relationships when parameters are subcategorized as they were during analysis.

There were several problems with the vowel tokens used in the survey. While the vowel tokens were attempted to be sung at the exact same pitch and intensity, there were inevitably inconsistencies due to human error. Formant values may have been artificially altered because the author sang the tokens and therefore had a preconceived notion of what the formant values should be, rather than singing in a free and realistic way. F3 values in particular were limited in range and did not consistently correlate with roundness, therefore reducing its significance as a variable. Finally, only 10 tokens were included in order to keep the survey at a reasonable length for participants, but this limits the number of data points. Future study should include more vowel tokens, potentially mirroring the methodology of Bloothooft & Plomp (1984) where multiple tokens of the same vowel are elicited from singers using the vowel metaphors under investigation. This would allow not only for a more detailed acoustic map across the vowel space, but also would provide the opportunity to determine if participant judgments of the vowel metaphors align with the singers' interpretation of them.

Several participants reported that they perceived the vowel tokens as electronic. This is likely because the onset and offset of tokens were removed and the tokens were perhaps too short in length (Pierce, 1999; Risset & Wessel, 1999). These measures were taken to attempt to remove additional variables that could affect perception. Additionally, it is difficult to normalize the length of a token without removing the onset and offset. Onsets in particular are critical in influence the perceived timbre – thus, removing the onset shifts the focus on the vowel quality. However, this clearly reduced participants' ability to perceive the audio as a human voice. Vowel tokens in future experiments should be longer and perhaps fade in and out to mimic human onset and offset, which still allows for normalization of the tokens.

Finally, many points of interest in the data were unable to be investigated due to time constraints. For one, tokens 8 and 10 ([o] and [ʉ], respectively) were extremely close in all three formant values, but sounded very different – how did judgments differ between these two tokens? Although familiarity did not affect judgments overall, how did familiarity affect judgments by token? While familiarity did not affect consistency of judgments overall, judgments appeared to be more consistent for some individual vowel tokens based on familiarity or musician identity. Is there a common variable that influences judgment consistency? Are formants influencing consistency, or might it be some other vowel feature that was not measured? Last, the relative strength of the correlations was unable to be assessed – the data suggest that more intricate relationships exist. Additionally, the differences caused by varying musical background may surface in the relative strength of correlations, which may be why seemingly no differences were observed due to exposure/musical background in the above analysis.

6 Conclusion

6.1 Applications

6.1.1 Pedagogy and vocal injury

Even though the sound symbolism of vowel metaphors does not appear to be domain-specific to vocal pedagogy, the findings of this study illustrate the efficacy of this aspect of vocal pedagogy, complementing the work of Bloothooft and Plomp (1984). By using the terms investigated in this study, teachers can in theory effectively elicit specific acoustic results. More data is needed to understand the efficacy of these terms in practice, even though specific correlates have been revealed.

Due to the intense and sometimes unnatural use of the voice, singers are at an increased risk of vocal injury (Achey et al., 2016; Benninger, 2011; Koufman et al., 1996). Vocal injuries can have profound and lasting effects. Singers are barred from their livelihood if they are unable to audition or perform due to an injury, and therefore lose income (Hosbach-Cannon et al., 2020). Individuals may also experience social and emotional effects of voice changes or even loss due to a vocal injury. A common way singers injure themselves is by using improper technique in response to vague instruction. Ambiguous metaphors can easily lead to misunderstandings and thus put the singer at risk of straining body parts to mimic the sound. However, explicit articulatory instruction is usually not an option due to lack of anatomical and articulatory knowledge, both on the part of the teacher and the student (Kwak et al., 2014; Rodríguez Marconi et al., 2018; Sielska-Badurek et al., 2017).

The findings of this study may help to reduce the risk of vocal injury because understanding what the true acoustic and thus the corresponding articulatory targets of choral vowel metaphors could help singers achieve the correct sounds in an articulatorily specific and therefore healthier way. Combining the metaphorical and articulatory approaches can be compared to the indexing of vowel metaphors – but by explicitly teaching the articulatory targets, this indexing process would be deliberate, clear, and precise.

While all the vowel metaphors investigated in this study were found to have acoustic and corresponding articulatory correlates, identifying terms that have poor articulatory correlates is also important so that they can be phased out of use, helping instruction to be more targeted and less likely to lead to vocal injury.

6.1.2 Implications for sound symbolism

This study is consistent with previous findings of correlates of size-sound symbolism and perhaps light-sound symbolism (Asano & Yokosawa, 2011; Kawahara, 2024; Knoeferle et al., 2017; Shinohara & Kawahara, 2010). Interestingly, the sound symbolic relationships discovered in this study do not appear to be domain-specific to singing. As previously discussed, this may be due to the fact that vowels themselves are not domain-specific. In other words, the universality of vowel sounds in spoken languages outweighs the domain-specificity of choral pedagogy. Thus, the findings of this study suggest somewhat generalized (at least in the US) sound symbolism of the vowel metaphors investigated in this thesis. No claims can be made about crosslinguistic validity. This adds to the body of literature that has found universal sound symbolic tendencies (Ćwiek et al., 2022; D'Anselmo et al., 2019; Kantartzis et al., 2011; Nygaard et al., 2009; Shinohara & Kawahara, 2010; Svantesson, 2017; Yoshida, 2012). Future studies should assess cross-cultural and cross-linguistic validity of the sound symbolic relationships discovered in this study.

Additionally, the convergence of acoustic destinations for "bright" and "harsh" and "dark", "warm", and "rich" shows that their semantic relationships were maintained across domains. These relationships may have held due to seemingly universal sound symbolic tendencies.

6.2 Future study

As discussed in §5.3, this study has many limitations and leaves much room for future research. Only so many vowel metaphors could be investigated in one study – other important ones in choral and vocal pedagogy include “tall”, “full”, “round”, and “pure”/“muddy”, with the final two potentially being another opposite pair. All of these vowel metaphors should be candidates for future study.

6.2.1 Other contributing factors

While this study only investigated the effect of F1/F2/F3 values on vowel quality, there are many other factors at play in the perception of vowels. This includes but is not limited to volume, intensity, length, pitch, the individual voice type/vocal tract size, perceived qualities of the singer, and so on. While these factors were attempted to be controlled for in this experiment, it is still possible that they contributed to participants’ perception of the vowels. Additionally, breath control, diaphragm control, jaw opening, laryngeal position, epiglottic diameter, velum position, and more anatomical modes all can influence singing and thus may have an influence on perceived vowel quality. Future studies should consider evaluating these variables in relation to vowel judgments.

While a teacher may use a metaphor like “bright” explicitly in reference to a vowel, in the experience of the author, these words can also be used in reference to the “sound” of the singing voice in general. Since pitch can only be sustained on sonorous segments, the perception and description of vowels in choral music might reasonably be generalized to the singing voice as a whole, including sonorous segments other than vowels. However, one can argue that the role of consonants (regardless of sonority) should not be ignored, even if they constitute a lower cumulative duration in signing than vowels – many sound symbolic relationships exist with consonants, including that voiceless fricatives (e.g. [θ], [s]) are small (Coulter & Coulter, 2010), nasals (e.g. [m], [n], [ŋ]) are large (Berlin, 2006; Lapolla, 1994), and more (Kawahara, 2024). Diction is emphasized in vocal pedagogy – singers manipulate the articulation of a consonant by changing its duration, intensity, etc. for intelligibility and other stylistic purposes. Consonants offer a wealth of material for future study of perception of singing.

Another factor to consider is the role of visual cues in singing. In the absence of being able to see the mechanisms that control the singing voice, external gestures and facial configurations are a key strategy in vocal pedagogy. The present study assessed vowel judgments based only on acoustic cues – do visual cues influence vowel judgments as well? Would the incorporation of visual cues have altered the findings of this study, for example, if participants were shown a video of someone singing rather than an audio file?

6.2.1.1 Velar raising

The role of the velum in singing is a surprisingly unclear one. In speech, velum height is the greatest in obstruents, followed by high vowels, then low vowels (Bell-Berti et al., 1979). Velum raising is regular but very small when speaking (Blaylock et al., 2016). In singing though, singers are taught to raise the velum ("lift the soft palate") to perhaps an unnatural degree with respect to speaking, particularly when singing higher pitches (Appelman, 1986; Cappadonia, 1962). Studies show that this velar raising does in fact take place in professional singer populations (Aura et al., 2019; Yanagisawa et al., 1990).

However, research doesn't agree on what velar raising actually does. Mechanically, raising the velum should close the velopharyngeal port, thus decreasing the size of the vocal tract and eliminating air diversion to the nasal cavity. There are conflicting results over whether the velopharyngeal port is open during singing (Austin, 1997; Birch et al., 2002; Gramming et al., 1993; Tanner et al., 2005), which is generally associated with a lowered velum. Yanagisawa et al. (1990) on the other hand discovered that the velopharyngeal port can remain open even when the velum is raised. The increased control singers seem to have over the velum creates some unexpected results, allowing for more variation in the size of the velopharyngeal port opening. These conflicting results may also be a result of differing singing styles. Additionally, velar raising may raise the larynx, thus reducing the length of the vocal tract and reducing air pressure which assists with producing higher pitches (Aura et al., 2019).

Regardless of what the mechanism is, singers must be explicitly taught in order to achieve these effects. However, once again, the mechanism is invisible to the naked eye – teaching must be indirect and may yield inconsistent results. Velar raising remains an intriguing and difficult area of research in speech and voice science. More research is needed to investigate the effect of velar raising on vowel quality and perception.

6.3 Final remarks

This thesis shows that vowel metaphors used in choral and vocal pedagogy do in fact have specific acoustic (F1, F2, F3) and corresponding articulatory (vowel height, frontness/backness, roundness) correlates, thus demonstrating sound symbolism. Some minor differences were observed due to musician identity. No effects were observed based on familiarity with the vowel metaphors. These findings also show how conceptual metaphor mappings can elicit specific acoustic and articulatory results. Importantly, this study reports seemingly universal instances of sound symbolism that was presumed to be domain-specific. As the first study to investigate sound symbolism in choral pedagogy, this thesis paves the way for future research to investigate other sound symbolic language in music and other variables that maybe be important in the perception of sung vowels.

Appendix A

This appendix includes supplemental information about the experimental survey described in §3.

A.1 Survey Consent

Perception of Sung Vowels

Hi! Thank you for your interest in my project. Please read the following information carefully so that you understand what the survey is about and what it will involve. At the end, there are checkboxes to indicate your consent to participate in the project if you wish to do so.

What is this survey about?
This survey is for undergraduate thesis research about sung vowels and how people perceive them.

Who made this survey?
This survey was made by Olivia Colace (Bryn Mawr College '25) for her senior thesis in Linguistics. The thesis advisor is Prof. Noah Elkins (Haverford College).

What will the survey involve?
If you decide to take part in this project, you will be asked to complete a two-part survey. The first part will ask you to listen to some audio files of a person singing and describe how you perceive them. Headphones are recommended. The second part will ask you some questions about yourself: some basic demographic information such as your age, gender, ethnicity, and where you're from, as well as describing your musical background. The survey should take roughly 5-10 minutes to complete.

What are the risks and benefits of participating in this project?
There are minimal personal risks or benefits to you for taking this survey.

Will you collect any personal information?
The only information about you that I'll ask for is basic demographic information and some information about your musical background at the end of the survey. All of the demographic questions have a "do not wish to answer" response, which you may select if you don't want to share that information with me. The information that I will collect about you should not be enough to identify you personally, but in any case, please do not include any personally identifying information in any of your answers.

What will you do with any personal information?
I will not be asking for personally identifying information. Your responses will be anonymized and taken together with the responses of everyone else to report the findings of this study. My professor and classmates may also assist in looking at your responses. In my thesis, I may use quotes that you write in the comments section of the survey, but I will do my best to ensure that the quotes are edited to maintain your privacy. My thesis presentation will be public and the final thesis will be available online publicly.

How will you keep my information secure?
This is an anonymous survey. The data is being collected using a secure (encrypted) connection to the host survey service provider. Results are stored in a password protected account accessible by only the researcher and system administrators. While no absolute guarantees can be made regarding security, these measures provide safeguards against outside agents accessing the electronic data.

Who can I contact if I have questions or concerns?
Please email Olivia Colace if you have any questions or concerns: ocolace@brynmawr.edu

If I don't want to participate, how do I opt out?
You may close the survey window or navigate way from the webpage at any time in order to opt out. If you decide to opt out in the middle of the study and want to ensure that any data already collected is deleted, you may email Olivia Colace at ocolace@brynmawr.edu.

Has this research been reviewed by an ethics committee?
Yes, this research has been determined exempt from IRB review by the Haverford College IRB.

Who can participate in this survey?
Anyone who is a fluent reader of English and 18 years old or older can participate in this survey.

Thank you for reading! If you would like to participate in the survey, please attest to the following statements by checking each box:

- I have read and understood the information above.
- I am 18 years of age or older.
- I agree to participate in the survey.

A.2 Experimental questions

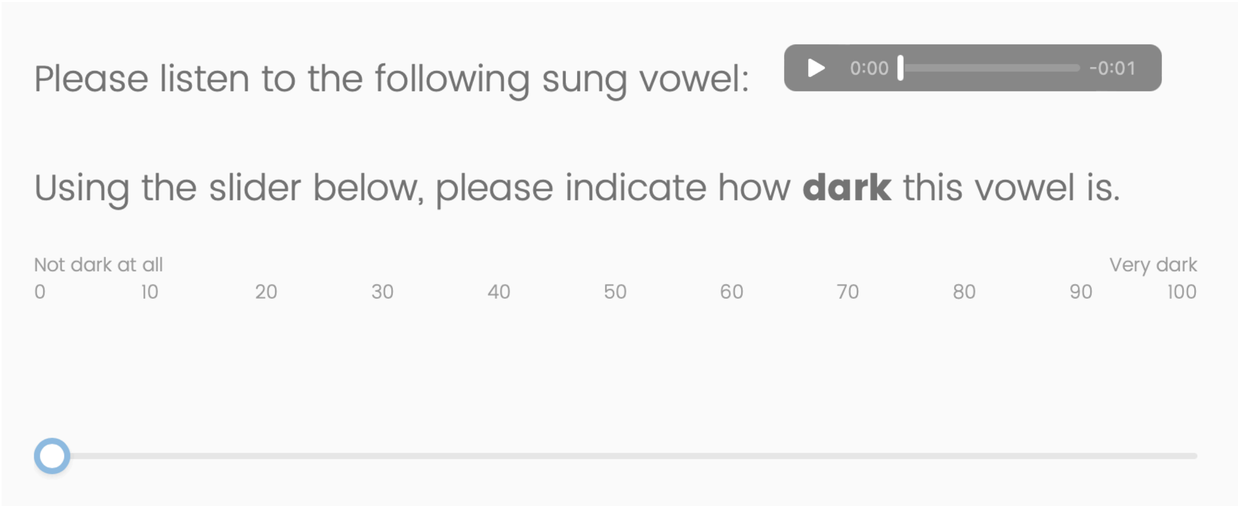

Figure A1. Example experimental survey question, repeated for each vowel token and for each adjective (bright, dark, wide, narrow, open, closed, warm, harsh, rich, thin).

A.3 Demographic questions

1. How old are you?
2. Which of the following best describes your gender?
    a. Woman
    b. Man
    c. Neither of these
    d. Do not wish to answer
3. On a survey like the US census, which of the following options would you choose to best describe your race/ethnicity? Select all that apply.
    a. White
    b. Black
    c. African American
    d. Indigenous
    e. Asian
    f. East Asian
    g. Southeast Asian
    h. South Asian
    i. Latino/a/x
    j. Native Hawaiian or Pacific Islander
    k. Arab
    l. Other (with free response text)
    m. Do not wish to answer
4. What US zip code have you spent the most time in over the last 5 years?

5. What is/are your first language(s)?

A.4 Musical background questions + scoring

1. I am… (select all that apply)
    a. a choral musician/singer
    b. a musician (non-choral/not a singer)
    c. not a musician **(automatic 0, skip to next section of survey)**
2. *if "a musician (non-choral/not a singer)" is selected:* What instrument(s) do you play?
3. I consider myself to be… (select all that apply)
    a. an amateur musician (i.e. self-taught) **(+0.75)**
    b. a student musician (i.e. taking lessons, music classes, in music school) **(+1.25)**
    c. a professional musician **(+3)**
    d. a music teacher **(+1)**
    e. other (with free response text)
4. How many years have you been a musician? **+(ans/10)**
5. How often do you practice, teach, or have rehearsal?
    a. a few times a week **(-0)**
    b. a few times a month **(-1)**
    c. a few times a year **(-2)**
    d. not at all **(-3)**

A.5 Familiarity questions

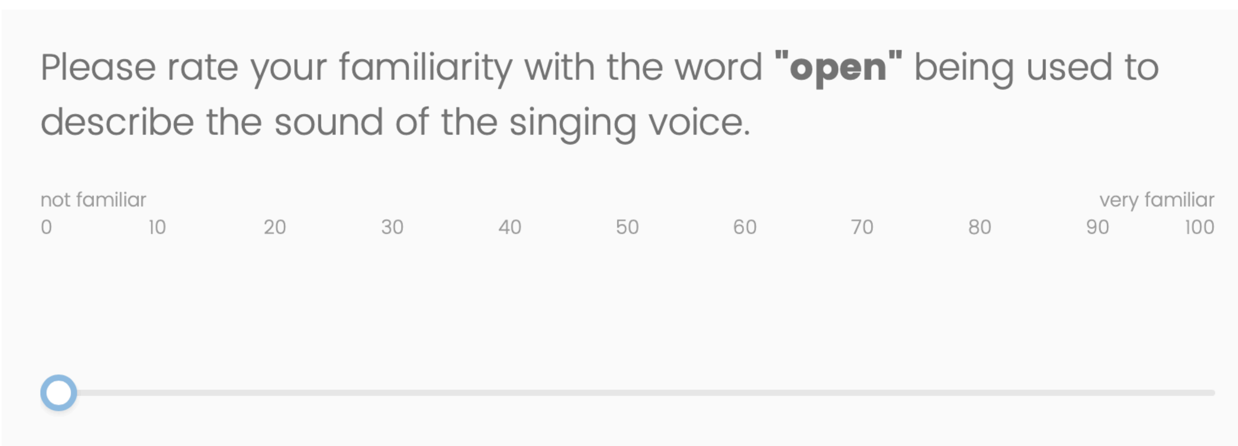

Figure A2. Example familiarity survey question, repeated for each adjective (bright, dark, wide, narrow, open, closed, warm, harsh, rich, thin).

Appendix B

This appendix includes regression tables displaying all six linear models for each metaphor.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
F1	38.0*** (6.65)	37.8*** (6.67)		47.7*** (9.53)	38.3*** (6.57)	
F2	20.8*** (4.61)	21.1*** (4.63)		30.1*** (6.59)	29.8*** (6.18)	31.5*** (6.40)
F3	58.5. (34.5)	57.1. (34.7)		61.9 (51.2)	58.7. (34.1)	
mus		-2.94 (4.29)	-2.95 (4.20)	-136 (230)	1.41 (15.5)	0.0368 (16.9)
non.mus		0.802 (3.76)	0.663 (3.68)	170 (201)	41.3** (14.2)	39.1* (15.5)
F1*mus				-13.4 (18.0)		
F1*non.mus				-20.7 (14.8)		
F2*mus				-7.76 (11.8)	-3.04 (9.89)	-3.03 (10.8)
F2*non.mus				-25.5* (10.6)	-27.9** (9.41)	-26.0* (10.3)
F3*mus				56.6 (89.7)		
F3*non.mus				-44.6 (77.8)		

Table B1. Linear models for judgments in relation to “bright”. Estimate coefficients are reported as average change in brightness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	-23.2** (7.27)	-23.7** (7.28)		-37.9*** (11.2)		
F2	-16.6*** (4.80)	-16.7*** (4.80)		-24.1** (7.53)		
F3	-40.6 (35.1)	-41.7 (35.1)		-22.1 (54.2)		
mus		5.63 (4.12)	4.61 (4.10)	-102 (227)		
non.mus		0.880 (3.83)	0.745 (3.80)	173 (209)		
F1*mus				20.4 (17.9)		
F1*non.mus				28.4 (17.4)		
F2*mus				10.4 (12.2)		
F2*non.mus				14.8 (11.1)		
F3*mus				30.0 (88.8)		
F3*non.mus				-78.1 (82.0)		

Table B2. Linear models for judgments in relation to “dark”. Estimate coefficients are reported as average change in darkness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	49.6*** (8.00)	49.2*** (8.01)		41.4** (12.5)	49.9*** (7.99)	
F2	0.465 (5.23)	0.445 (5.26)		11.6 (8.33)	10.1 (7.43)	41.2*** (12.0)
F3	-5.56 (42.3)	-6.28 (42.4)		-48.7 (66.1)	-9.29 (42.3)	
mus		3.17 (5.19)	3.75 (5.22)	-130 (300)	26.4 (19.2)	22.4 (20.9)
non.mus		-2.63 (4.19)	-3.17 (4.20)	-175 (244)	23.8 (15.6)	20.6 (17.0)
F1*mus				9.67 (20.9)		
F1*non.mus				17.6 (18.4)		
F2*mus				-17.5 (14.3)	-15.1 (12.2)	-12.0 (13.3)
F2*non.mus				-19.7. (11.8)	-17.1. (9.81)	-15.5 (10.7)
F3*mus				57.3 (116)		
F3*non.mus				71.6 (94.9)		

Table B3. Linear models for judgments in relation to "wide". Estimate coefficients are reported as average change in wideness judgment per kHz with standard error reported in parentheses. "mus" = musicians who are not choral singers. "non.mus" = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
F1	-27.5*** (8.09)	-27.5*** (8.12)		-6.05 (12.4)	-5.20 (11.9)	-9.53 (11.5)
F2	11.5* (5.46)	11.6* (5.48)		12.3 (8.44)	12.3* (5.41)	
F3	-9.37 (42.4)	-11.1 (42.6)		-1.66 (65.7)	-18.2 (42.2)	
mus		0.403 (5.15)	-0.830 (5.10)	86.2 (274)	12.7 (12.4)	12.1 (12.5)
non.mus		3.18 (4.33)	1.98 (4.29)	177 (248)	30.1** (10.5)	29.3** (10.6)
F1*mus				-14.9 (21.8)	-21.6 (20.4)	-20.9 (20.5)
F1*non.mus				-48.4** (17.8)	-47.1** (16.8)	-46.0** (17.0)
F2*mus				20.6 (14.0)		
F2*non.mus				-14.2 (12.2)		
F3*mus				-40.2 (107.9)		
F3*non.mus				-46.3 (96.2)		

Table B4. Linear models for judgments in relation to “narrow”. Estimate coefficients are reported as average change in narrowness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	67.6*** (7.24)	67.3*** (7.26)		59.0*** (10.9)	66.2*** (7.17)	
F2	-7.77 (4.98)	-7.93 (4.99)		0.366 (7.19)	1.32 (7.09)	-7.40 (8.41)
F3	126** (38.8)	123** (39.1)		70.9 (53.5)	61.1 (52.2)	151* (61.4)
mus		3.90 (4.52)	2.93 (4.48)	-604 (300)	-631* (291)	-714* (349)
non.mus		-0.0869 (4.10)	-0.516 (4.07)	-83.4 (222)	-137 (214)	-195 (256)
F1*mus				8.66 (18.2)		
F1*non.mus				16.0 (16.8)		
F2*mus				-31.3* (12.5)	-32.5** (12.0)	-37.8** (14.4)
F2*non.mus				-2.61 (11.6)	-4.98 (11.3)	-6.89 (13.5)
F3*mus				241* (115)	253* (110)	287* (132)
F3*non.mus				29.0 (86.7)	53.7 (82.4)	76.4 (98.6)

Table B5. Linear models for judgments in relation to “open”. Estimate coefficients are reported as average change in openness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	-66.6*** (8.02)	-67.1*** (8.02)		-53.1*** (11.5)	-67.0*** (7.88)	
F2	-4.15 (5.32)	-4.21 (5.31)		-10.8 (8.07)	-4.19 (5.22)	
F3	-54.1 (42.2)	-48.7 (42.3)		-107. (59.7)	-123* (54.2)	-183** (57.5)
mus		-7.51 (4.90)	-6.92 (4.80)	-768** (293)	-823** (264)	-830** (306)
non.mus		-2.31 (4.12)	-2.81 (4.02)	-116 (243)	-184 (213)	-156 (247)
F1*mus				-26.0 (22.4)		
F1*non.mus				-25.9 (17.4)		
F2*mus				9.52 (13.5)		
F2*non.mus				11.5 (11.9)		
F3*mus				282* (113)	302** (97.7)	305** (113)
F3*non.mus				41.4 (94.3)	67.7 (79.1)	57.5 (91.8)

Table B6. Linear models for judgments in relation to "closed". Estimate coefficients are reported as average change in closedness judgment per kHz with standard error reported in parentheses. "mus" = musicians who are not choral singers. "non.mus" = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	-21.5** (7.87)	-21.3** (7.85)		-41.0*** (11.8)	-41.9*** (11.3)	-35.3** (11.9)
F2	-26.7*** (4.98)	-26.3*** (4.98)		-29.5*** (8.05)	-25.8*** (4.92)	
F3	-0.538 (41.6)	-7.82 (41.7)		-48.9 (61.8)	-8.55 (41.3)	
mus		8.08 (4.46)	7.87. (4.31)	-311 (269)	-17.1 (10.9)	-19.9. (11.9)
non.mus		4.14 (4.14)	5.34 (4.04)	-232 (253)	-12.4 (10.5)	-14.4 (11.4)
F1*mus				39.7* (19.2)	44.1* (17.4)	47.7* (19.0)
F1*non.mus				29.8 (18.4)	29.3. (17.1)	34.2. (18.6)
F2*mus				2.58 (12.5)		
F2*non.mus				9.07 (11.5)		
F3*mus				108 (105)		
F3*non.mus				76.2 (98.1)		

Table B7. Linear models for judgments in relation to “warm”. Estimate coefficients are reported as average change in warmness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	38.4*** (8.12)	39.1*** (8.18)		69.0*** (14.8)	67.3*** (14.1)	50.0** (16.8)
F2	36.4*** (4.52)	36.1*** (4.55)		39.8*** (6.97)	36.3*** (4.50)	
F3	39.2 (36.4)	43.8 (37.0)		46.1 (51.5)	43.3 (36.4)	
mus		-2.60 (4.34)	-2.31 (4.23)	7.92 (256)	32.2** (11.8)	30.9* (14.6)
non.mus		-3.10 (3.85)	-3.77 (3.76)	132 (224)	11.0 (10.1)	4.09 (12.4)
F1*mus				-69.3** (23.1)	-66.8** (21.1)	-64.0* (26.1)
F1*non.mus				-30.0 (18.7)	-28.4 (17.8)	-15.8 (21.8)
F2*mus				-3.64 (11.5)		
F2*non.mus				-8.64 (10.5)		
F3*mus				11.6 (99.1)		
F3*non.mus				-39.5 (86.6)		

Table B8. Linear models for judgments in relation to “harsh”. Estimate coefficients are reported as average change in harshness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	-27.8*** (7.42)	-28.0*** (7.51)		-44.2*** 12.1		
F2	-22.8*** (4.74)	-22.7*** (4.79)		-21.9** (7.05)		
F3	5.72 (40.7)	5.00 (41.0)		61.1 (62.5)		
mus		0.265 (4.30)	0.606 (4.24)	173 (261)		
non.mus		3.06 (4.03)	3.13 (3.95)	309 (256)		
F1*mus				28.0 (18.5)		
F1*non.mus				23.3 (18.0)		
F2*mus				-8.03 (12.0)		
F2*non.mus				5.26 (11.3)		
F3*mus				-65.2 (101)		
F3*non.mus				-121 (98.1)		

Table B9. Linear models for judgments in relation to “rich”. Estimate coefficients are reported as average change in richness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
F1	-13.4. (8.04)	-13.6. (8.07)		8.72 (12.1)	9.81 (11.6)	2.33 (11.5)
F2	15.7** (5.47)	15.6** (5.50)		13.6. (7.98)	16.1** (5.43)	
F3	-72.0 (46.9)	-72.7 (47.1)		-61.4 (73.6)	-65.5 (46.5)	
mus		-4.49 (5.12)	-4.57 (5.20)	122 (313)	15.7 (12.4)	15.2 (12.6)
non.mus		-2.07 (4.51)	-1.75 (4.66)	0.201 (279)	24.5* (10.8)	22.4* (11.0)
F1*mus				-33.4 (21.3)	-35.5. (20.0)	-35.0. (20.3)
F1*non.mus				-44.7* (18.2)	-45.7** (17.0)	-43.8* (17.2)
F2*mus				2.27 (13.9)		
F2*non.mus				7.05 (12.8)		
F3*mus				-41.1 (121)		
F3*non.mus				4.79 (108)		

Table B10. Linear models for judgments in relation to “thin”. Estimate coefficients are reported as average change in thinness judgment per kHz with standard error reported in parentheses. “mus” = musicians who are not choral singers. “non.mus” = non-musicians. Asterisks indicate significance (*** = 0.001, ** = 0.01, * = 0.05, . = 0.1).

Appendix C

This appendix includes scatter plots of vowel judgments for each metaphor vs formant values, shown overall and separated by musician identity.

C.1 “bright”

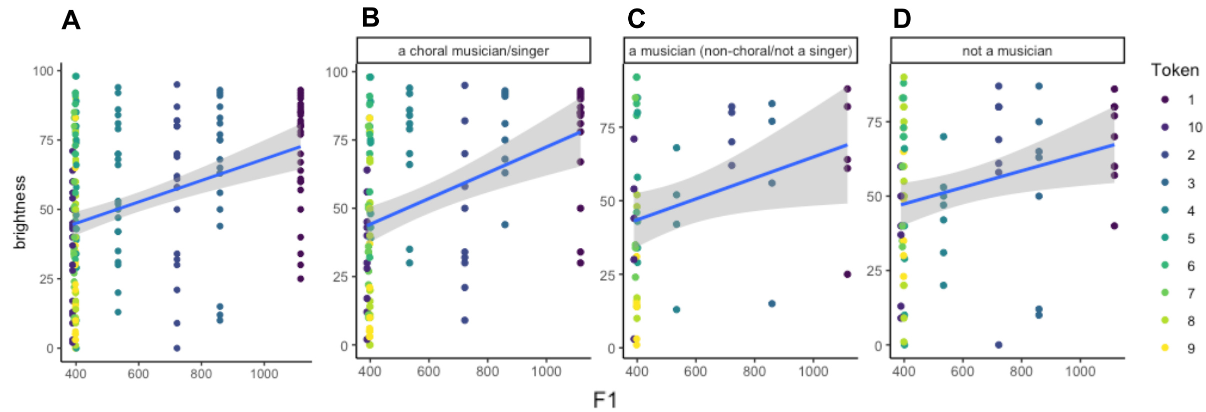

Figure C1. Linear regressions between “bright” judgments and F1 for A) all participants (R^2 = 0.12, $p < 0.001$), B) choral singers (R^2 = 0.16, $p < 0.001$), C) other musicians (R^2 = 0.097, $p < 0.05$), and D) non-musicians (R^2 = 0.084, $p < 0.05$).

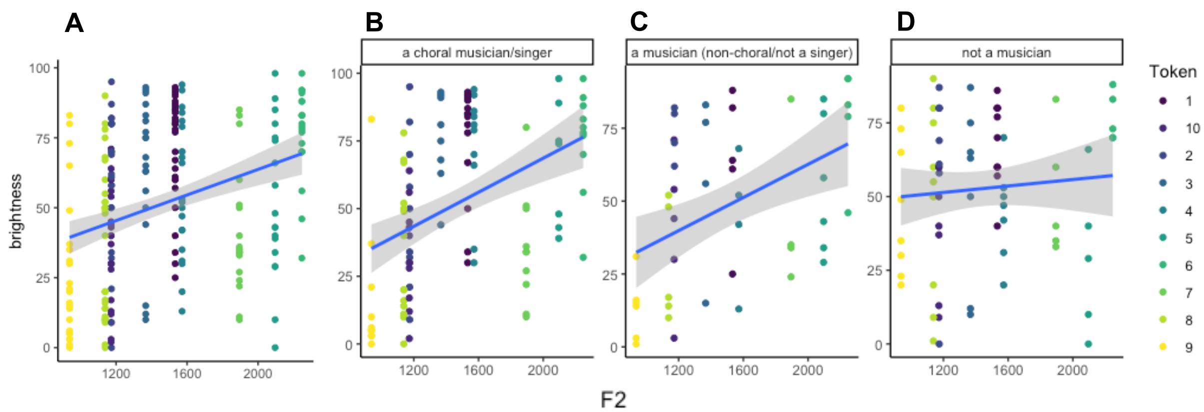

Figure C2. Linear regressions between “bright” judgments and F2 for A) all participants (R^2 = 0.11, $p < 0.001$), B) choral singers (R^2 = 0.18, $p < 0.001$), C) other musicians (R^2 = 0.20, $p < 0.01$), and D) non-musicians (R^2 = 0.01, $p > 0.1$).

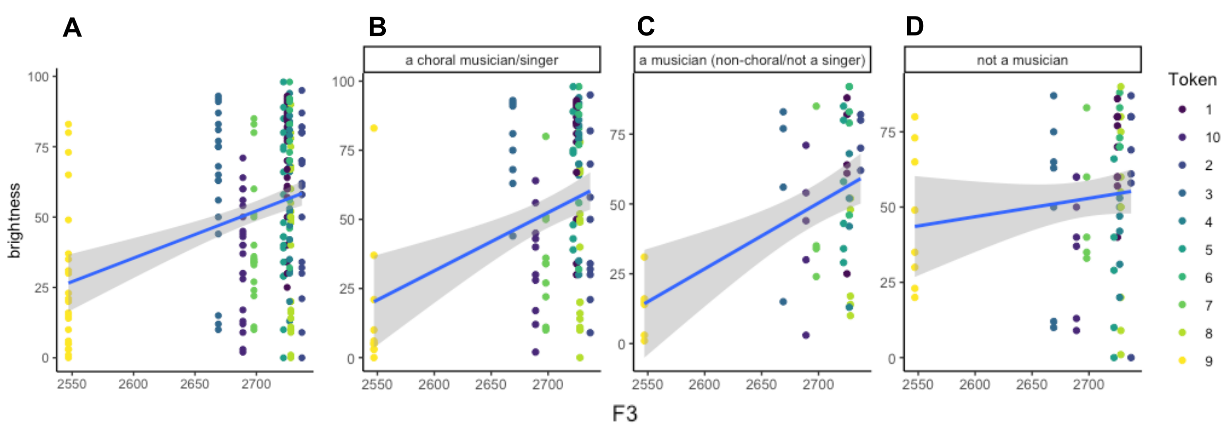

Figure C3. Linear regressions between “bright” judgments and F3 for A) all participants (R^2 = 0.11, $p < 0.001$), B) choral singers (R^2 = 0.14, $p < 0.001$), C) other musicians (R^2 = 0.25, $p < 0.001$), and D) non-musicians (R^2 = 0.02, $p > 0.1$).

C.2 "dark"

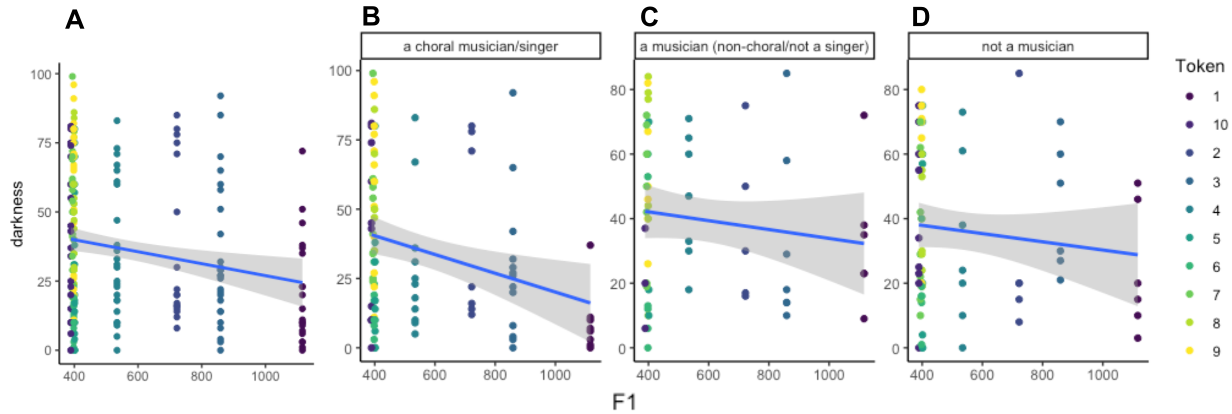

Figure C4. Linear regressions between "dark" judgments and F1 for A) all participants ($R^2 = 0.04$, $p < 0.01$), B) choral singers ($R^2 = 0.08$, $p < 0.01$), C) other musicians ($R^2 = 0.02$, $p > 0.1$), and D) non-musicians ($R^2 = 0.014$, $p > 0.1$).

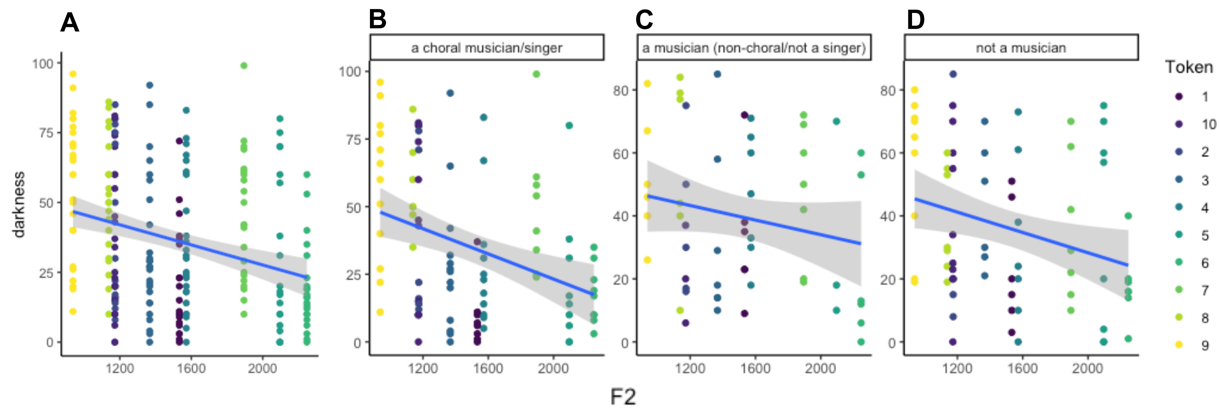

Figure C5. Linear regressions between "dark" judgments and F2 for A) all participants ($R^2 = 0.08$, $p < 0.001$), B) choral singers ($R^2 = 0.12$, $p < 0.001$), C) other musicians ($R^2 = 0.04$, $p > 0.1$), and D) non-musicians ($R^2 = 0.08$, $p < 0.05$).

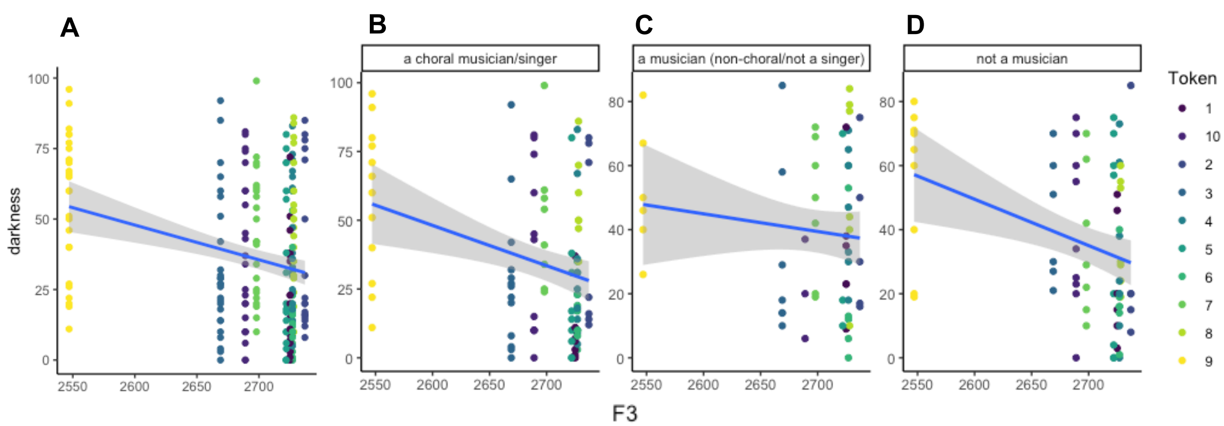

Figure C6. Linear regressions between "dark" judgments and F3 for A) all participants ($R^2 = 0.07$, $p < 0.001$), B) choral singers ($R^2 = 0.09$, $p < 0.01$), C) other musicians ($R^2 = 0.016$, $p > 0.1$), and D) non-musicians ($R^2 = 0.12$, $p < 0.01$).

C.3 "wide"

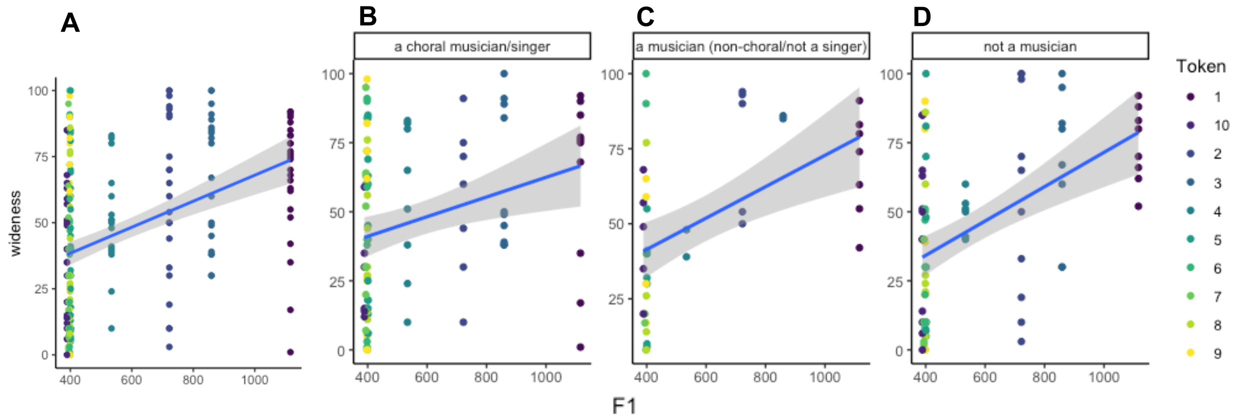

Figure C7. Linear regressions between "wide" judgments and F1 for A) all participants (R^2 = 0.17, $p < 0.001$), B) choral singers (R^2 = 0.09, $p < 0.01$), C) other musicians (R^2 = 0.27 , $p < 0.001$), and D) non-musicians (R^2 = 0.23, $p < 0.001$).

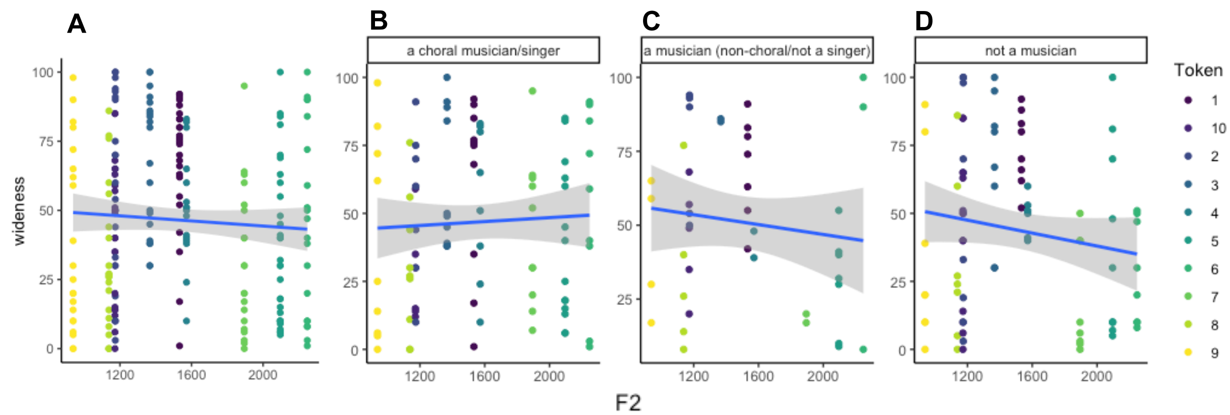

Figure C8. Linear regressions between "wide" judgments and F2 for A) all participants (R^2 = 0.004, $p > 0.1$), B) choral singers (R^2 = 0.003, $p > 0.1$), C) other musicians (R^2 = 0.02, $p > 0.1$), and D) non-musicians (R^2 = 0.03, $p > 0.1$).

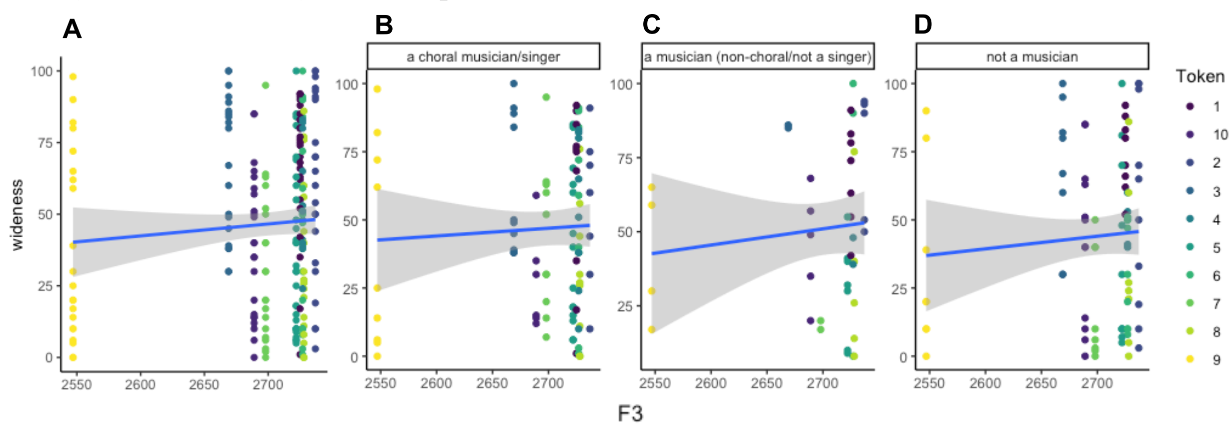

Figure C9. Linear regressions between "wide" judgments and F3 for A) all participants (R^2 = 0.005, $p > 0.1$), B) choral singers (R^2 = 0.003, $p > 0.1$), C) other musicians (R^2 = 0.01, $p > 0.1$), and D) non-musicians (R^2 = 0.01, $p > 0.1$).

C.4 “narrow”

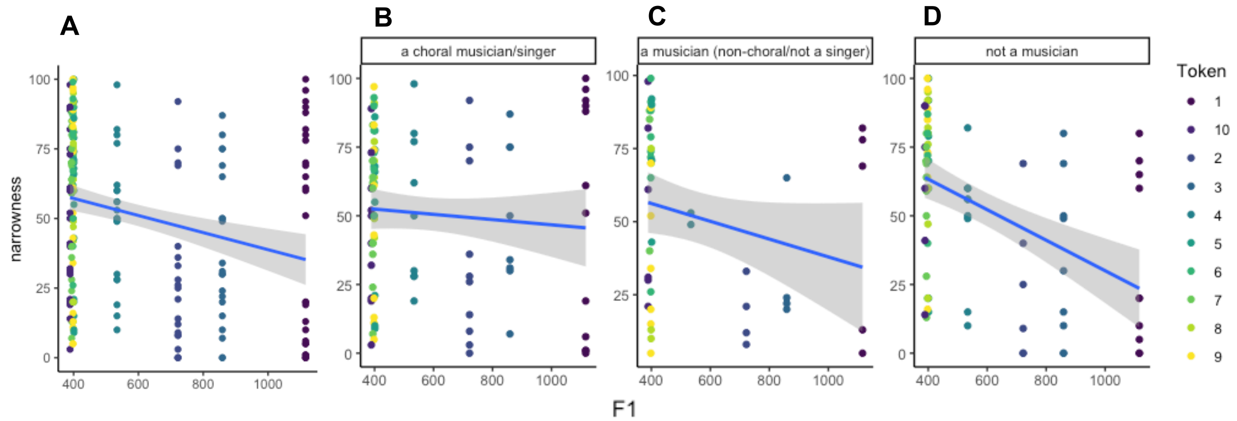

Figure C10. Linear regressions between “narrow” judgments and F1 for A) all participants (R^2 = 0.07, $p < 0.001$), B) choral singers (R^2 = 0.01, $p > 0.1$), C) other musicians (R^2 = 0.06, $p < 0.1$), and D) non-musicians (R^2 = 0.21, $p < 0.001$).

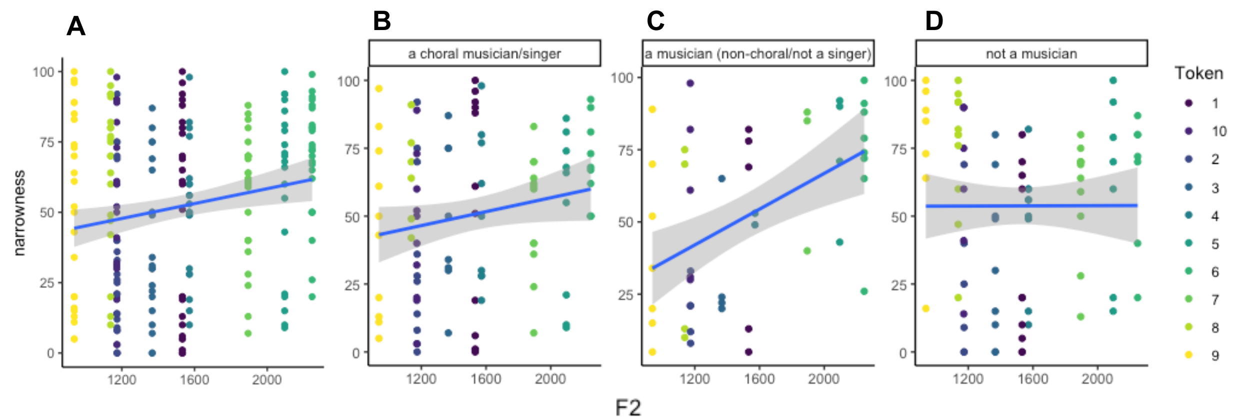

Figure C11. Linear regressions between “narrow” judgments and F2 for A) all participants (R^2 = 0.03, $p < 0.01$), B) choral singers (R^2 = 0.03, $p < 0.1$), C) other musicians (R^2 = 0.23, $p < 0.001$), and D) non-musicians ($R^2 = 1.0 \times 10^{-5}$, $p > 0.1$).

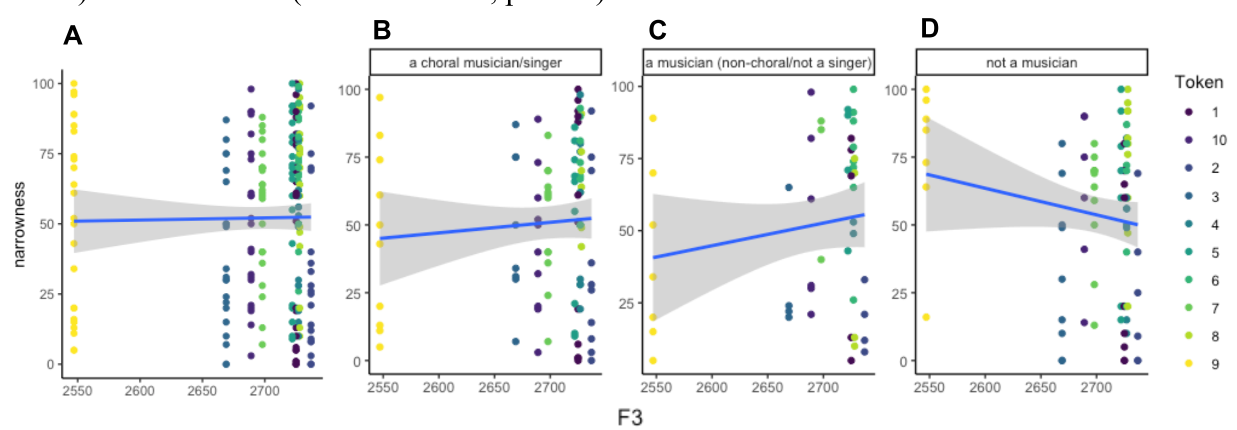

Figure C12. Linear regressions between “narrow” judgments and F3 for A) all participants ($R^2 = 2.0 \times 10^{-4}$, $p > 0.1$), B) choral singers (R^2 = 0.005, $p > 0.1$), C) other musicians (R^2 = 0.03, $p > 0.1$), and D) non-musicians (R^2 = 0.03, $p > 0.1$).

C.5 “open”

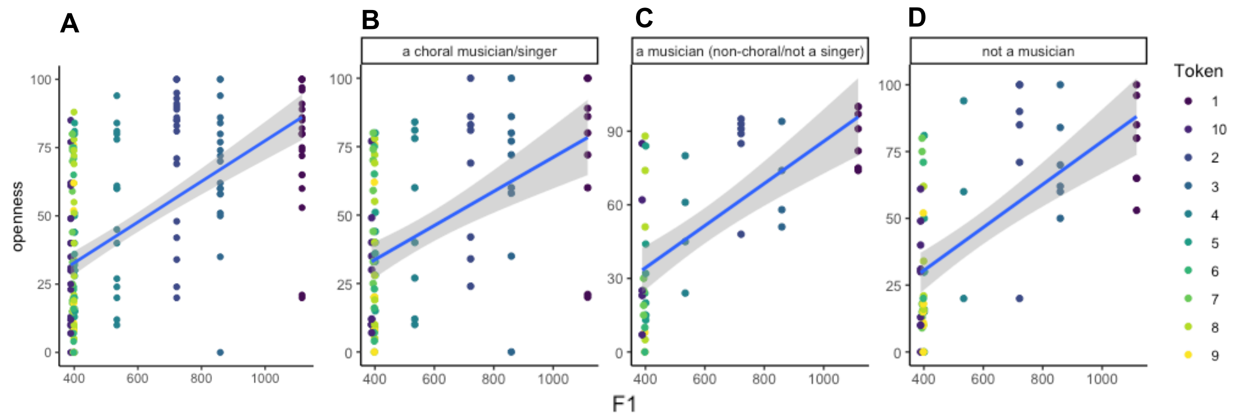

Figure C13. Linear regressions between “open” judgments and F1 for A) all participants (R^2 = 0.36, p < 0.001), B) choral singers (R^2 = 0.25, p < 0.001), C) other musicians (R^2 = 0.48, p < 0.001), and D) non-musicians (R^2 = 0.43, p < 0.001).

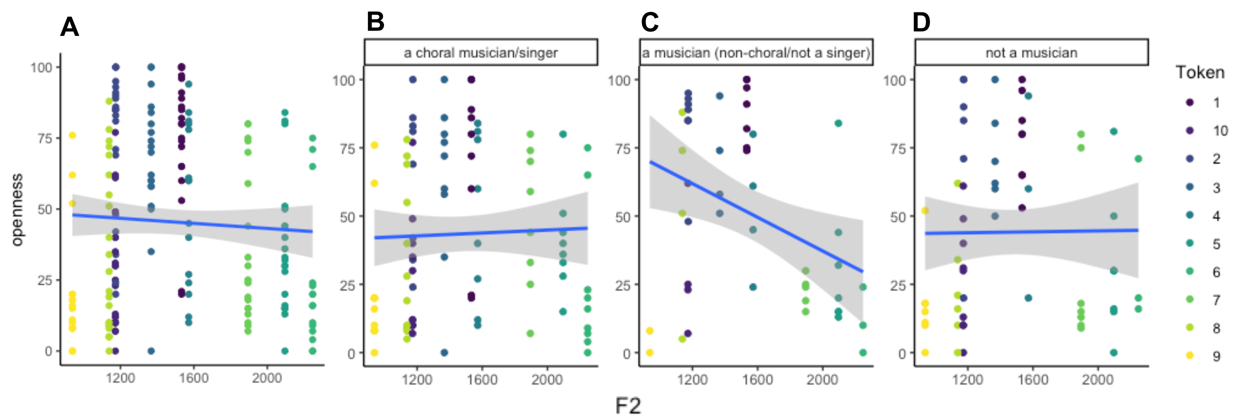

Figure C14. Linear regressions between “open” judgments and F2 for A) all participants (R^2 = 0.003, p > 0.1), B) choral singers (R^2 = 0.001, p > 0.1), C) other musicians (R^2 = 0.13, p < 0.05), and D) non-musicians ($R^2 = 1.0 \times 10^{-4}$, p > 0.1).

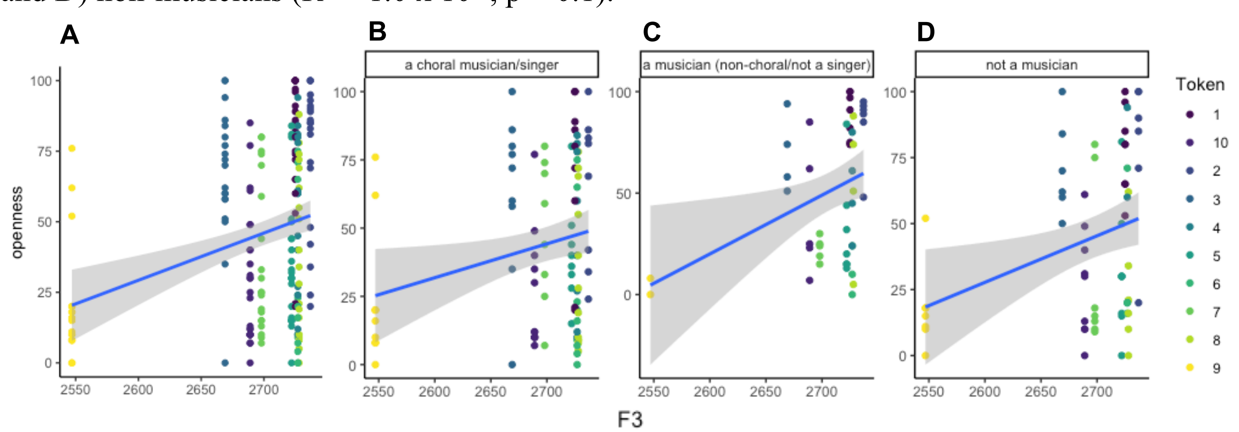

Figure C15. Linear regressions between “open” judgments and F3 for A) all participants (R^2 = 0.08, p < 0.001), B) choral singers (R^2 = 0.05, p < 0.05), C) other musicians (R^2 = 0.12, p < 0.05), and D) non-musicians(R^2 = 0.10, p < 0.05).

C.6 "closed"

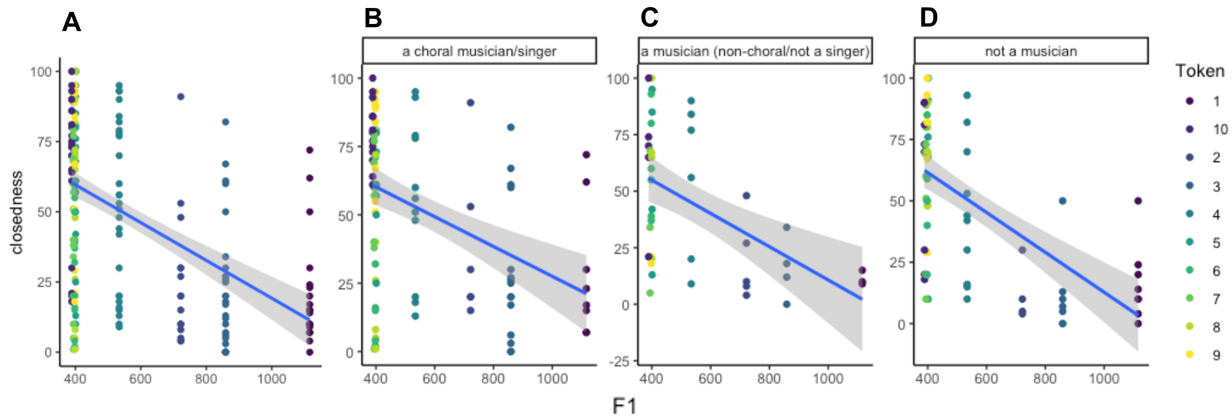

Figure C16. Linear regressions between "closed" judgments and F1 for A) all participants (R^2 = 0.26, $p < 0.001$), B) choral singers (R^2 = 0.18, $p < 0.001$), C) other musicians (R^2 = 0.27, $p < 0.001$), and D) non-musicians (R^2 = 0.39, $p < 0.001$).

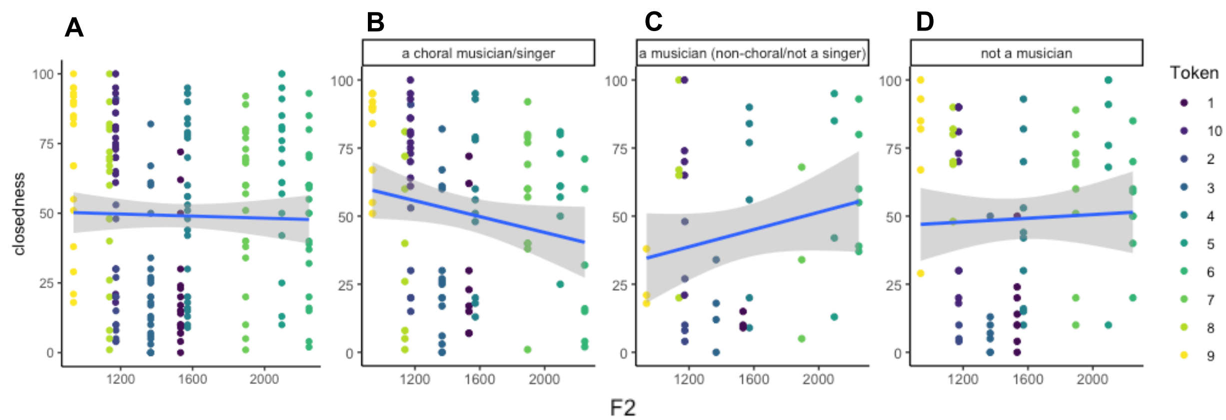

Figure C17. Linear regressions between "closed" judgments and F2 for A) all participants (R^2 = 1.0 x 10^{-4}, $p > 0.1$), B) choral singers (R^2 = 0.04, $p < 0.1$), C) other musicians (R^2 = 0.05, $p > 0.1$), and D) non-musicians (R^2 = 0.002, $p > 0.1$).

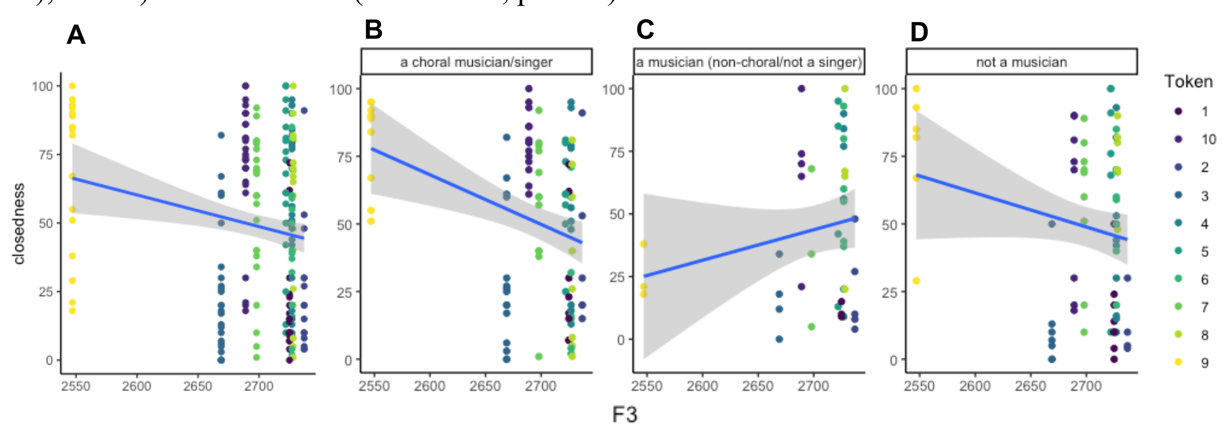

Figure C18. Linear regressions between "closed" judgments and F3 for A) all participants (R^2 = 0.04, $p < 0.01$), B) choral singers (R^2 = 0.10, $p < 0.01$), C) other musicians (R^2 = 0.04, $p > 0.1$), and D) non-musicians (R^2 = 0.04, $p < 0.1$).

C.7 "warm"

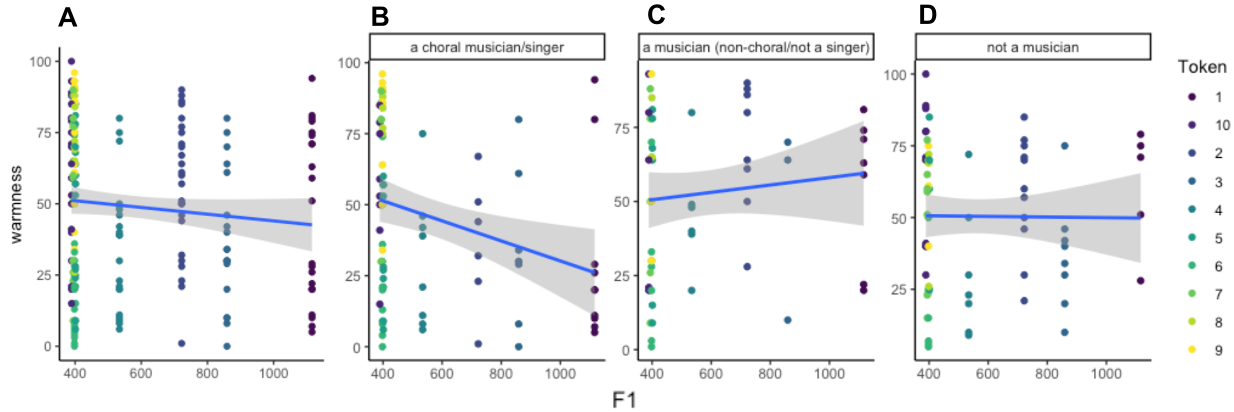

Figure C19. Linear regressions between "warm" judgments and F1 for A) all participants ($R^2 = 0.01$, $p > 0.1$), B) choral singers ($R^2 = 0.09$, $p < 0.01$), C) other musicians ($R^2 = 0.01$, $p > 0.1$), and D) non-musicians ($R^2 = 1.0 \times 10^{-4}$, $p > 0.1$).

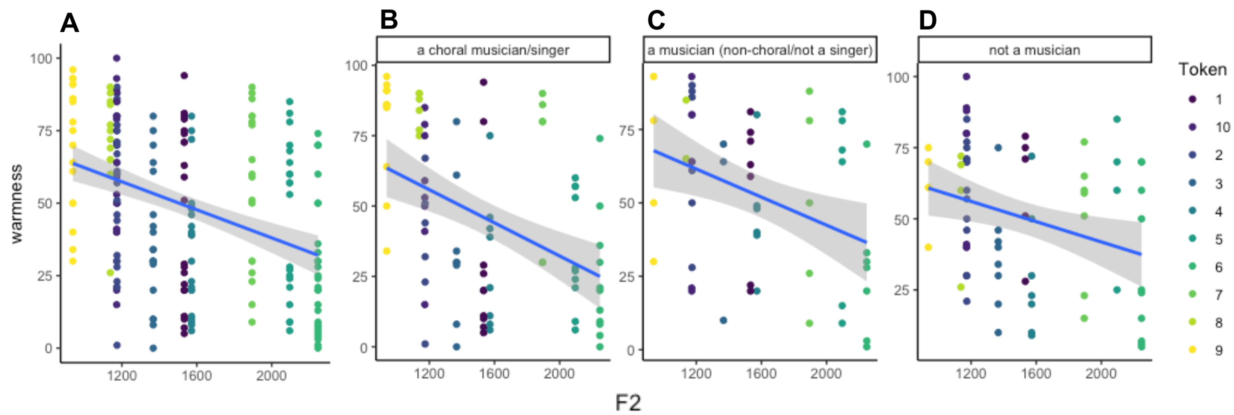

Figure C20. Linear regressions between "warm" judgments and F2 for A) all participants ($R^2 = 0.14$, $p < 0.001$), B) choral singers ($R^2 = 0.18$, $p < 0.001$), C) other musicians ($R^2 = 0.14$, $p < 0.01$), and D) non-musicians ($R^2 = 0.09$, $p < 0.05$).

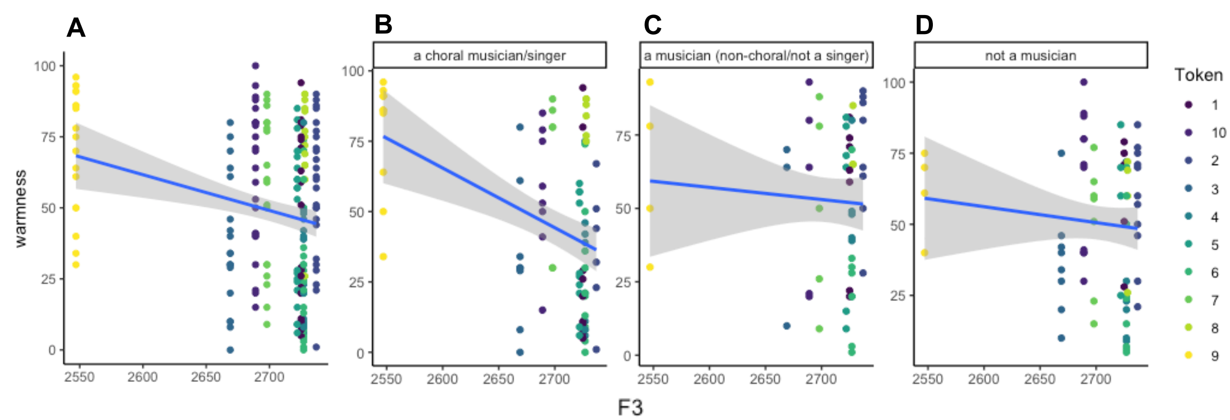

Figure C21. Linear regressions between "warm" judgments and F3 for A) all participants ($R^2 = 0.05$, $p < 0.001$), B) choral singers ($R^2 = 0.17$, $p < 0.001$), C) other musicians ($R^2 = 0.005$, $p > 0.1$), and D) non-musicians ($R^2 = 0.01$, $p > 0.1$).

C.8 " harsh"

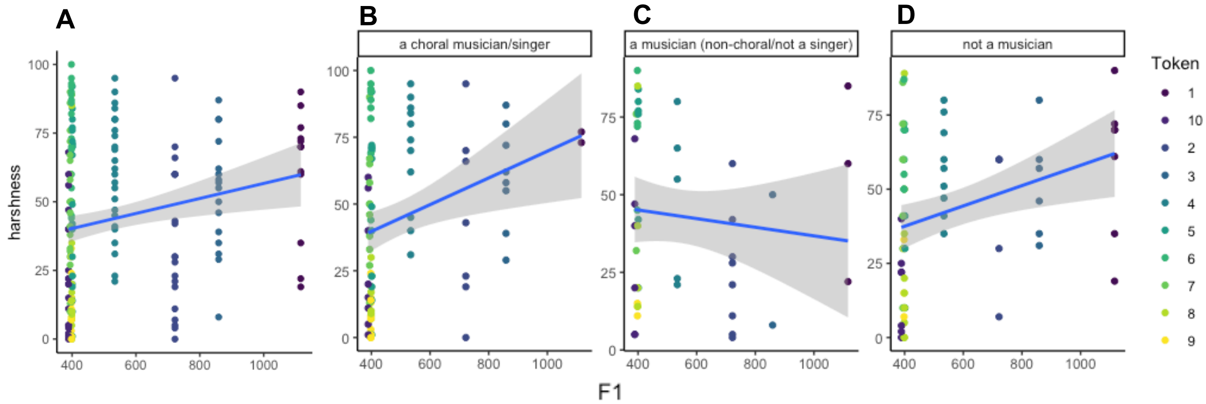

Figure C22. Linear regressions between "harsh" judgments and F1 for A) all participants (R^2 = 0.04, $p < 0.01$), B) choral singers (R^2 = 0.079, $p < 0.01$), C) other musicians (R^2 = 0.01, $p > 0.1$), and D) non-musicians (R^2 = 0.11, $p < 0.01$).

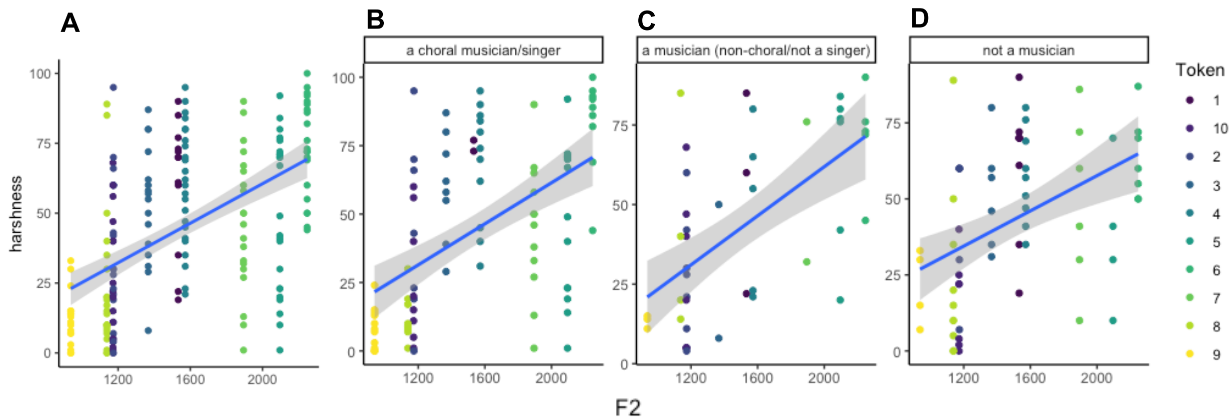

Figure C23. Linear regressions between "harsh" judgments and F2 for A) all participants (R^2 = 0.27, $p < 0.001$), B) choral singers (R^2 = 0.28, $p < 0.001$), C) other musicians (R^2 = 0.36, $p < 0.001$), and D) non-musicians (R^2 = 0.2, $p < 0.001$).

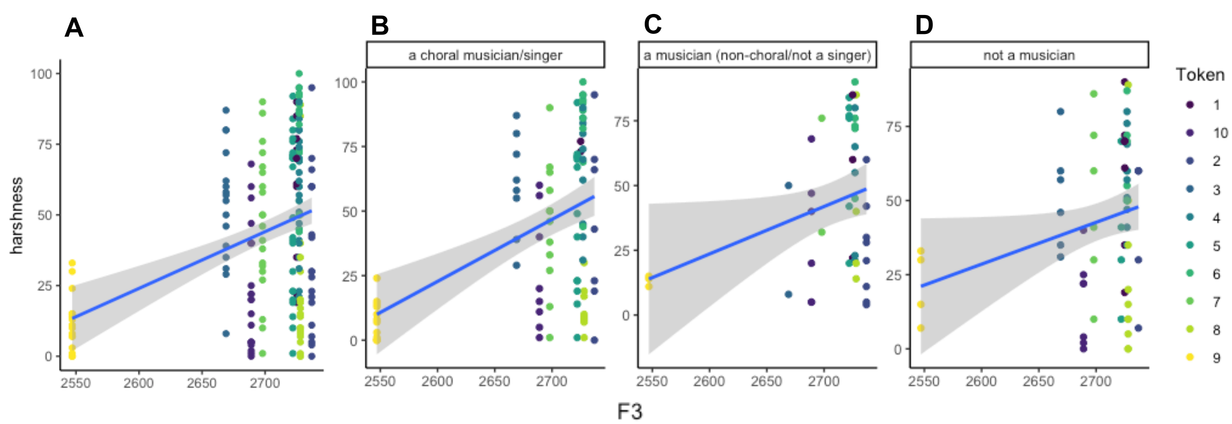

Figure C24. Linear regressions between "harsh" judgments and F3 for A) all participants (R^2 = 0.14, $p < 0.001$), B) choral singers (R^2 = 0.21, $p < 0.001$), C) other musicians (R^2 = 0.10, $p < 0.04$), and D) non-musicians (R^2 = 0.06, $p < 0.05$).

C.9 “rich”

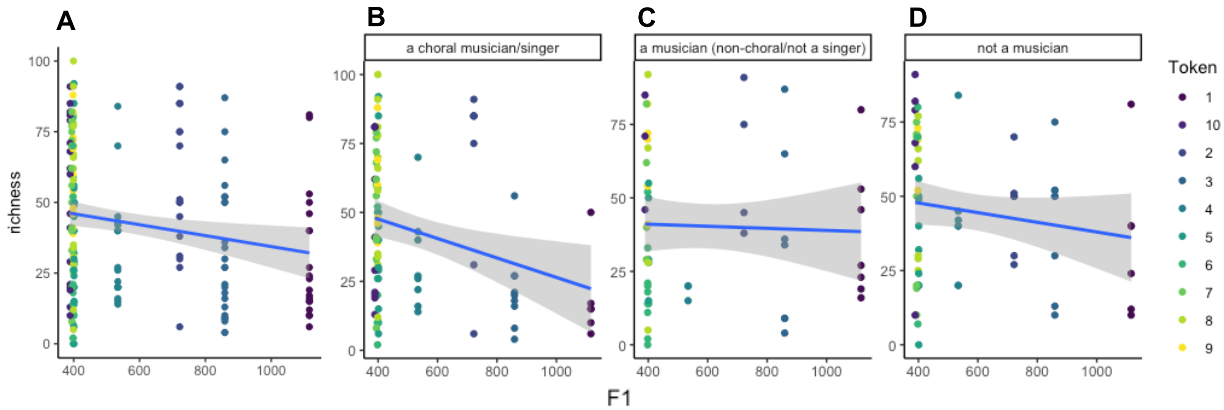

Figure C25. Linear regressions between “rich” judgments and F1 for A) all participants (R^2 = 0.03, $p < 0.05$), B) choral singers (R^2 = 0.08, $p < 0.01$), C) other musicians (R^2 = 0.001, $p > 0.1$), and D) non-musicians (R^2 = 0.03, $p > 0.1$).

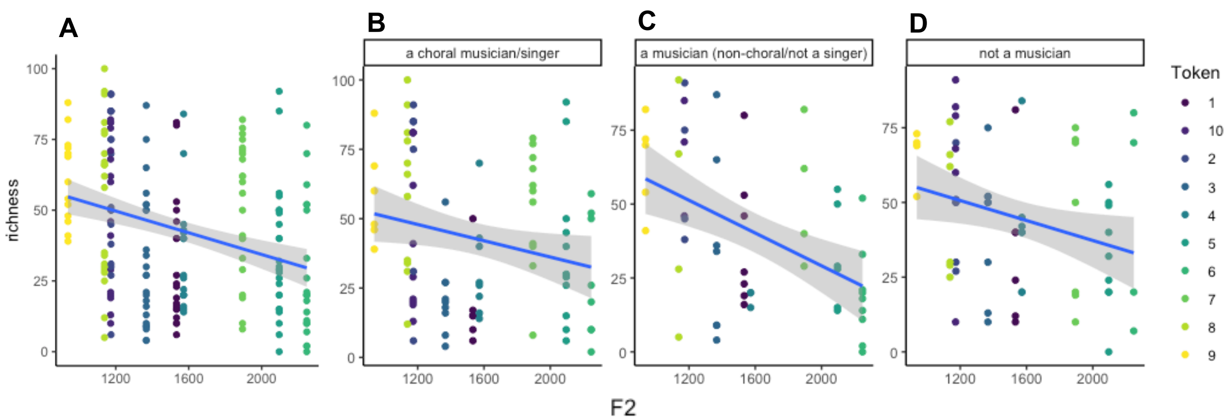

Figure C26. Linear regressions between “rich” judgments and F2 for A) all participants (R^2 = 0.10, $p < 0.001$), B) choral singers (R^2 = 0.05, $p < 0.05$), C) other musicians (R^2 = 0.22, $p < 0.001$), and D) non-musicians (R^2 =0.08, $p < 0.05$).

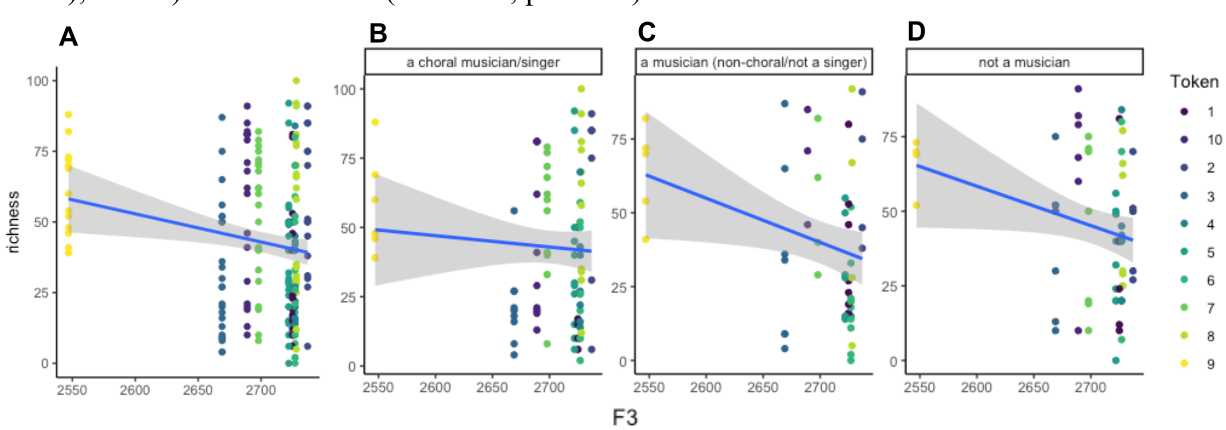

Figure C27. Linear regressions between “rich” judgments and F3 for A) all participants (R^2 = 0.03, $p < 0.01$), B) choral singers (R^2 = 0.005, $p > 0.1$, C) other musicians (R^2 = 0.09, $p < 0.05$), and D) non-musicians (R^2 = 0.06, $p < 0.05$).

C.10 “thin”

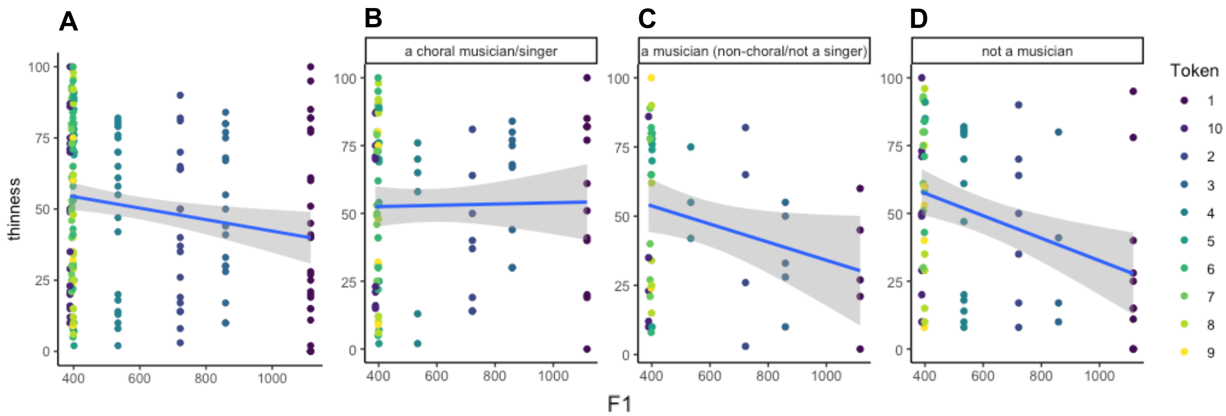

Figure C28. Linear regressions between “thin” judgments and F1 for A) all participants (R^2 = 0.03, $p < 0.01$), B) choral singers ($R^2 = 5.0 \times 10^{-4}$, $p > 0.1$), C) other musicians (R^2 = 0.08, $p < 0.05$), and D) non-musicians (R^2 = 0.13, $p < 0.01$).

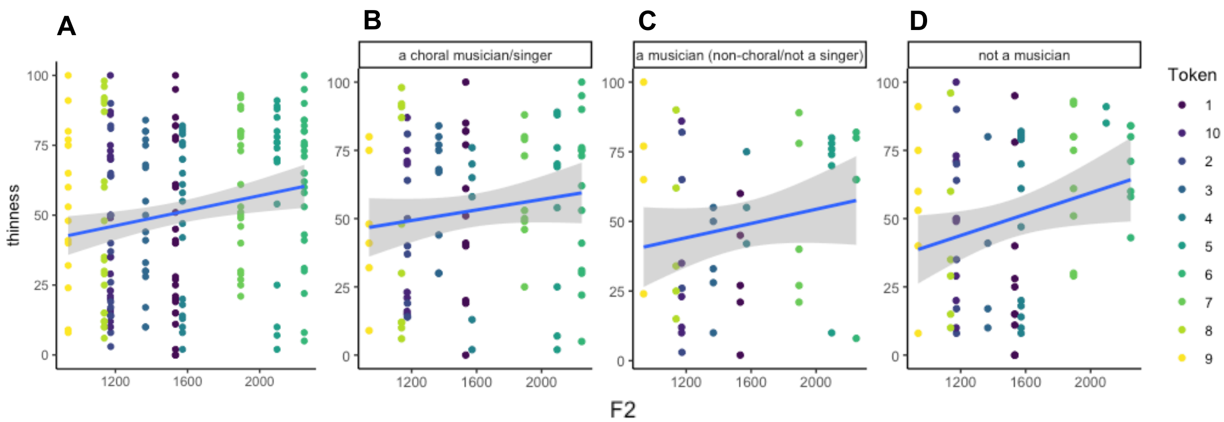

Figure C29. Linear regressions between “thin” judgments and F2 for A) all participants (R^2 = 0.04, $p < 0.01$), B) choral singers (R^2 =0.02, $p > 0.1$), C) other musicians (R^2 = 0.04, $p > 0.1$), and D) non-musicians (R^2 = 0.06, $p < 0.05$).

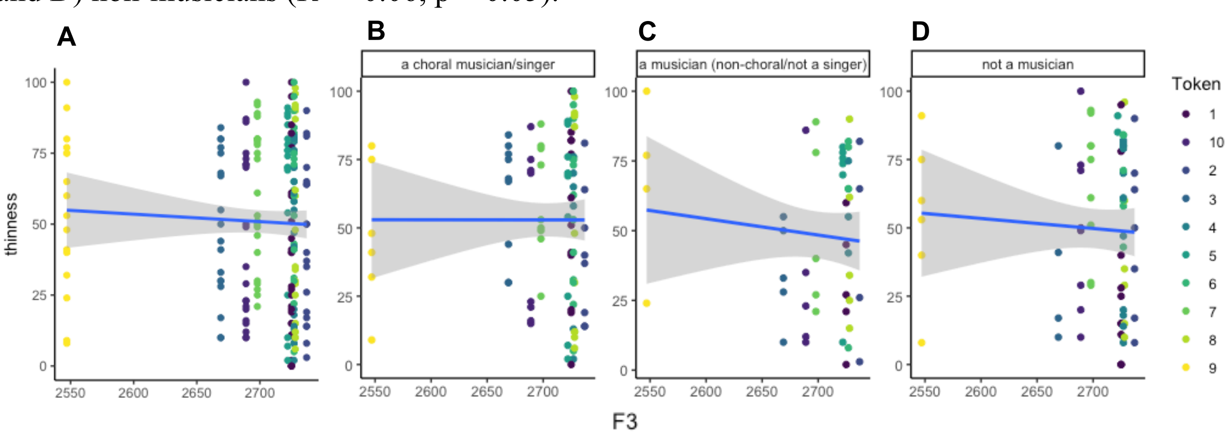

Figure C30. Linear regressions between “thin” judgments and F3 for A) all participants (R^2 = 0.002, $p > 0.1$), B) choral singers ($R^2 = 3.2 \times 10^{-7}$, $p > 0.1$), C) other musicians (R^2 = 0.01, $p > 0.1$), and D) non-musicians (R^2 = 0.004, $p > 0.1$).

Appendix D

This appendix includes supplemental data about effect of familiarity and musician identity on consistency of judgments.

	Model F1	Model F2	Model F3
bright	-0.0330 (0.0551)	-0.0363 (0.0579)	0.192 (0.146)
dark	-0.130 (0.12857)	0.00230 (0.0378)	-0.130 (0.129)
wide	0.0348 (0.0589)	0.0241 (0.0595)	-0.126 (0.155)
narrow	-0.0375 (0.0542)	-0.0379 (0.0549)	-0.220 (0.213)
open	0.0304 (0.0630)	0.0172 (0.0664)	0.128 (0.156)
closed	-0.00304 (0.0493)	0.000720 (0.0509)	0.193 (0.159)
warm	-0.00720 (0.0534)	-0.00919 (0.0559)	-0.122 (0.162)
harsh	0.00307 (0.0615)	0.0261 (0.0586)	0.470 (0.302)
rich	-0.00506 (0.0460)	0.000428 (0.0515)	-0.0661 (0.179)
thin	0.00277 (0.0644)	-0.00908 (0.0636)	-0.168 (0.147)

Table D1. Linear regression table showing standard deviation of judgment values predicted by self-reported familiarity scores with each term. No significant relationships were found.

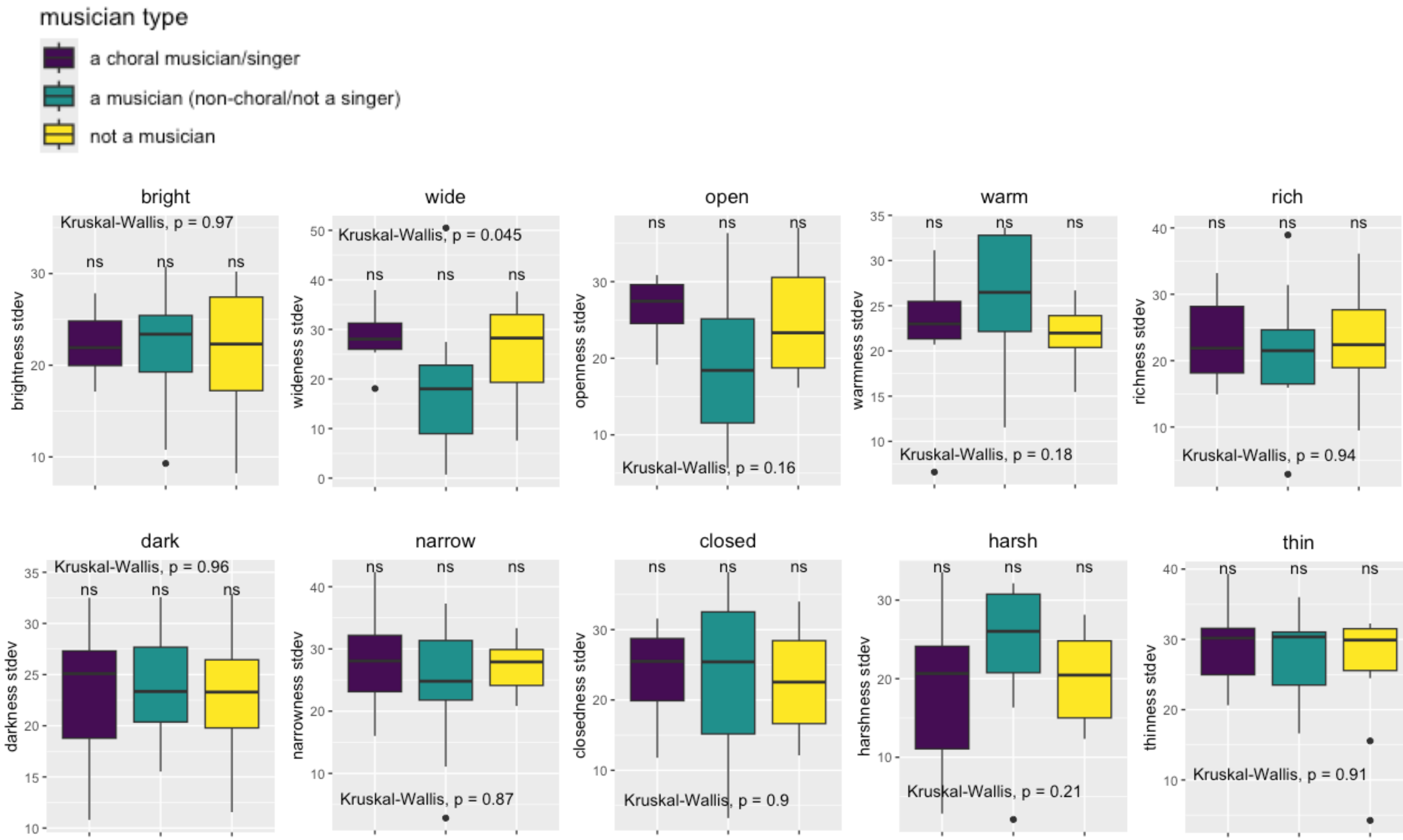

Figure D1. Standard deviations of judgments with respect to each vowel metaphor, separated by musician type. Kruskal-Wallis tests reveal no significant differences in standard deviation between the types of musicians for any vowel metaphor ($p > 0.1$ for all vowel metaphors). This shows that variation in judgements was effectively the same across musician types.

	Choral singers	Musicians (non-choral)	Non-musicians	Avg. non-singers	Difference between singers and avg. non-singers
bright	82.90323	53.25	47.65217	50.451085	32.452145
dark	71.35484	22.125	26.25	24.1875	47.16734
wide	67.70968	34.1875	34.82609	34.506795	33.202885
narrow	54.70968	31.0625	28.69565	29.879075	24.830605
open	81.80645	56	44.52174	50.26087	31.54558
closed	74.83871	39.1875	24.52174	31.85462	42.98409
warm	77.06452	59.5	43.79167	51.645835	25.418685
harsh	65.67742	42.75	40.47826	41.61413	24.06329
rich	70.64516	49.875	45.69565	47.785325	22.859835
thin	62.12903	29.5	27.95652	28.72826	33.40077
				avg:	**31.7925225**
				stdev:	**8.13316873**

Table D2. Average familiarity scores for each vowel metaphor across musician types. Non-singers are musicians (non-choral) and non-musicians. The final column reports on average how much greater the familiarity scores of singers were than the rest of the participants.

References

Achey, M. A., He, M. Z., & Akst, L. M. (2016). Vocal Hygiene Habits and Vocal Handicap Among Conservatory Students of Classical Singing. *Journal of Voice*, *30*(2), 192–197. https://doi.org/10.1016/j.jvoice.2015.02.003

Antovic, M. (2015). *Metaphor in Music or Metaphor About Music: A Contribution to the Cooperation of Cognitive Linguistics and Cognitive Musicology* (SSRN Scholarly Paper 2566258). https://doi.org/10.2139/ssrn.2566258

Appelman, D. R. (1986). *The Science of Vocal Pedagogy*. Indiana University Press. https://publish.iupress.indiana.edu/projects/the-science-of-vocal-pedagogy

Asano, M., & Yokosawa, K. (2011). Synesthetic colors are elicited by sound quality in Japanese synesthetes. *Consciousness and Cognition*, *20*(4), 1816–1823. https://doi.org/10.1016/j.concog.2011.05.012

Aura, M., Geneid, A., Bjørkøy, K., Rantanen, M., & Laukkanen, A.-M. (2019). The Nasal Musculature as a Control Panel for Singing—Why Classical Singers Use a Special Facial Expression? *Journal of Voice*, *33*(4), 510–515. https://doi.org/10.1016/j.jvoice.2017.12.016

Austin, S. F. (1997). Movement of the velum during speech and singing in classically trained singers. *Journal of Voice*, *11*(2), 212–221. https://doi.org/10.1016/S0892-1997(97)80080-X

Bayly, A. (1771). *A Practical Treatise on Singing and Playing with Just Expression and Real Elegance: Being an Essay on I. Grammar. II. Pronunciation; Or, The Art of Just Speaking. III. Singing -- Its Graces -- Their Application. -- On Cathedral Compositions*. J. Ridley.

Bell-Berti, F., Baer, T., Harris, K. S., & Niimi, S. (1979). Coarticulatory effects of vowel quality on velar function. *Phonetica*, *36*(3), 187–193.

Benninger, M. S. (2011). The professional voice. *The Journal of Laryngology & Otology*, *125*(2), 111–116. https://doi.org/10.1017/S0022215110001970

Berlin, B. (2006). The First Congress of Ethnozoological Nomenclature. *Journal of Royal Anthropological Institution*, *12*, 23–44. https://doi.org/10.1111/j.1467-9655.2006.00271.x

Birch, P., Gümoes, B., Stavad, H., Prytz, S., Björkner, E., & Sundberg, J. (2002). Velum Behavior in Professional Classic Operatic Singing. *Journal of Voice*, *16*(1), 61–71. https://doi.org/10.1016/S0892-1997(02)00073-5

Blaylock, R., Goldstein, L., & Narayanan, S. S. (2016). Velum Control for Oral Sounds. *Interspeech 2016*, 1084–1088. https://doi.org/10.21437/Interspeech.2016-1408

Bloothooft, G., & Plomp, R. (1984). Spectral analysis of sung vowels. I. Variation due to differences between vowels, singers, and modes of singing. *The Journal of the Acoustical Society of America*, *75*(4), 1259–1264. https://doi.org/10.1121/1.390732

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version 6.4.22) [Computer software]. http://www.praat.org/

Byrd, D. (2007). *A Table of Musical Pitches*. School of Informatics & Jacobs School of Music, Indiana University. https://homes.luddy.indiana.edu/donbyrd/Teach/MusicalPitchesTable.htm

Cappadonia, A. C. (1962). The Formation of Vowel Sounds in Singing. *The Choral Journal*, *2*(5), 21–27.

Cook, N. D. (2007). The Sound Symbolism Of Major And Minor Harmonies. *Music Perception: An Interdisciplinary Journal*, *24*(3), 315–319. https://doi.org/10.1525/mp.2007.24.3.315

Coulter, K. S., & Coulter, R. A. (2010). Small Sounds, Big Deals: Phonetic Symbolism Effects in Pricing. *Journal of Consumer Research*, *37*(2), 315–328. https://doi.org/10.1086/651241

Cunningham, J. G., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and Emotion*, *12*(4), 399–413. https://doi.org/10.1007/BF00992362

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., … Winter, B. (2022). The *bouba/kiki* effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841), 20200390. https://doi.org/10.1098/rstb.2020.0390

D'Anselmo, A., Prete, G., Zdybek, P., Tommasi, L., & Brancucci, A. (2019). Guessing Meaning From Word Sounds of Unfamiliar Languages: A Cross-Cultural Sound Symbolism Study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00593

Dolgin, K. G., & Adelson, E. H. (1990). Age Changes in the Ability to Interpret Affect in Sung and Instrumentally-Presented Melodies. *Psychology of Music*, *18*(1), 87–98. https://doi.org/10.1177/0305735690181007

Dowling, W. J. (1999). The Development of Music Perception and Cognition. In D. Deutsch (Ed.), *The Psychology of Music (Second Edition)* (pp. 603–625). Academic Press. https://doi.org/10.1016/B978-012213564-4/50016-0

Fillebrown, T. (1911). *Resonance in Singing and Speaking*. Oliver Ditson Company.

Gerardi, G. M., & Gerken, L. (1995). The Development of Affective Responses to Modality and Melodic Contour. *Music Perception: An Interdisciplinary Journal*, *12*(3), 279–290. https://doi.org/10.2307/40286184

Gramming, P., Nor, L., Sundberg, J., & Elliot, N. (1993). Does the nose resonate during singing? *Quarterly Progress and Status Report, Speech Transmission Lab*, *34*(4), 035–042.

Gregg, J. W., & Scherer, R. C. (2006). Vowel Intelligibility in Classical Singing. *Journal of Voice*, *20*(2), 198–210. https://doi.org/10.1016/j.jvoice.2005.01.007

Haynie, H., Bowern, C., & LaPalombara, H. (2014). Sound Symbolism in the Languages of Australia. *PLOS ONE*, *9*(4), e92852. https://doi.org/10.1371/journal.pone.0092852

Hosbach-Cannon, C. J., Lowell, S. Y., Colton, R. H., Kelley, R. T., & Bao, X. (2020). Assessment of Tongue Position and Laryngeal Height in Two Professional Voice

Populations. *Journal of Speech, Language and Hearing Research (Online)*, *63*(1), 109–124. https://doi.org/10.1044/2019_JSLHR-19-00164

Jeffers, R. (1988). *Translations and Annotations of Choral Repertoire: Vol. 1: Sacred Latin Texts*. Earthsongs.

Johnson, M. L., & Larson, S. (2003). "Something in the Way She Moves"-Metaphors of Musical Motion. *Metaphor and Symbol*, *18*(2), 63–84. https://doi.org/10.1207/S15327868MS1802_1

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese Sound-Symbolism Facilitates Word Learning in English-Speaking Children. *Cognitive Science*, *35*(3), 575–586. https://doi.org/10.1111/j.1551-6709.2010.01169.x

Kastner, M. P., & Crowder, R. G. (1990). Perception of the Major/Minor Distinction: IV. Emotional Connotations in Young Children. *Music Perception: An Interdisciplinary Journal*, *8*(2), 189–201. https://doi.org/10.2307/40285496

Kawahara, S. (2020). Sound symbolism and theoretical phonology. *Language and Linguistics Compass*, *14*(8), e12372. https://doi.org/10.1111/lnc3.12372

Kawahara, S. (2024). *Phonetic bases of sound symbolism: A review*. OSF. https://doi.org/10.31234/osf.io/fzvsu

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-05965-y

Kolodgy, P. (2008). *Lighting Metaphors in Music Criticism* [Undergraduate]. Swarthmore College.

Koufman, J. A., Radomski, T. A., Joharji, G. M., Russell, G. B., & Pillsbury, D. C. (1996). Laryngeal Biomechanics of the Singing Voice. *Otolaryngology–Head and Neck Surgery*, *115*(6), 527–537. https://doi.org/10.1016/S0194-59989670007-4

Kraus, N., & Banai, K. (2007). Auditory-Processing Malleability: Focus on Language and Music. *Current Directions in Psychological Science*, *16*(2), 105–110. https://doi.org/10.1111/j.1467-8721.2007.00485.x

Kwak, P. E., Stasney, C. R., Hathway, J., Minard, C. G., & Ongkasuwan, J. (2014). Knowledge, Experience, and Anxieties of Young Classical Singers in Training. *Journal of Voice*, *28*(2), 191–195. https://doi.org/10.1016/j.jvoice.2013.08.006

Lakoff, G., & Johnson, M. L. (1980). *Metaphors We Live By*. University of Chicago Press.

Lapolla, R. J. (1994). An experimental investigation into phonettc symbolism as it relates to Mandarin Chinese. In *Sound Symbolism* (pp. 130–147). Cambridge University Press.

Loureiro-Rodríguez, V., & Acar, E. F. (2022). The Matched-Guise Technique. In R. Kircher & L. Zipp (Eds.), *Research Methods in Language Attitudes* (1st ed., pp. 185–202). Cambridge University Press. https://doi.org/10.1017/9781108867788.016

Mandikal Vasuki, P. R., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*, *342*, 112–123. https://doi.org/10.1016/j.heares.2016.10.008

Marczyk, A., Belley, É., Savard, C., Roy, J.-P., Vaillancourt, J., & Tremblay, P. (2022). Learning transfer from singing to speech: Insights from vowel analyses in aging amateur singers and non-singers. *Speech Communication*, *141*, 28–39. https://doi.org/10.1016/j.specom.2022.05.001

Mok, P. P. K., Li, G., Li, J. J., Ng, H. T. Y., & Cheung, H. (2019). Cross-modal association between vowels and colours: A cross-linguistic perspective. *The Journal of the Acoustical Society of America*, *145*(4), 2265. https://doi.org/10.1121/1.5096632

Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *I-Perception*, *5*(2), 132–142. https://doi.org/10.1068/i0626

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181–186. https://doi.org/10.1016/j.cognition.2009.04.001

Ohala, J. J. (1983). The phonological end justifies any means. *Proceedings of the 13th International Congress of Linguists*, 232–243.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16. https://doi.org/10.1159/000261706

Ohala, J. J. (1995). The frequency code underlies the sound-symbolic use of voice pitch. In J. Nichols, J. J. Ohala, & L. Hinton (Eds.), *Sound Symbolism* (pp. 325–347). Cambridge University Press. https://doi.org/10.1017/CBO9780511751806.022

Pierce, J. R. (1999). The Nature of Musical Sound. In D. Deutsch (Ed.), *The Psychology of Music (Second Edition)* (pp. 1–23). Academic Press. https://doi.org/10.1016/B978-012213564-4/50002-0

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). *Neuroscience* (2nd ed.). Sinauer Associates. https://www.ncbi.nlm.nih.gov/books/NBK10924/

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rasch, R., & Plomp, R. (1999). The Perception of Musical Tones. In D. Deutsch (Ed.), *The Psychology of Music (Second Edition)* (pp. 89–112). Academic Press. https://doi.org/10.1016/B978-012213564-4/50005-6

Risset, J.-C., & Wessel, D. L. (1999). Exploration of Timbre by Analysis and Synthesis. In D. Deutsch (Ed.), *The Psychology of Music (Second Edition)* (pp. 113–169). Academic Press. https://doi.org/10.1016/B978-012213564-4/50006-8

Rodríguez Marconi, D., Morales Cárdenas, C., Gaete Antilen, L., Garrido Ormeño, M., & Pardo Reyes, C. (2018). Level of physiological, anatomical and pathological knowledge of the singing voice on the part of professional and amateur singers. *Revista CEFAC*, *20*(5), 621–631. https://doi.org/10.1590/1982-021620182054118

Rojczyk, A. (2011). Sound symbolism in vowels: Vowel quality, duration and pitch in sound-to-size correspondence. *Poznań Studies in Contemporary Linguistics*, *47*(3), 602–615. https://doi.org/10.2478/psicl-2011-0030

Schroeder, M. R. (1999). Concert Halls: From Magic to Number Theory. In D. Deutsch (Ed.), *The Psychology of Music (Second Edition)* (pp. 25–46). Academic Press. https://doi.org/10.1016/B978-012213564-4/50003-2

Shih, S. S., Ackerman, J., Hermalin, N., Inkelas, S., Jang, H., Kavitskaya, D., Kawahara, S., Oh, M., Starr, R. L., & Yu, A. (2019). *Cross-linguistic and language-specific sound symbolism: Pokemonastics*. http://user.keio.ac.jp/~kawahara/pdf/PokemonasticsShihEtAl.pdf

Shinohara, K., & Kawahara, S. (2010). A Cross-linguistic Study of Sound Symbolism: The Images of Size. *Annual Meeting of the Berkeley Linguistics Society*, 396–410. https://doi.org/10.3765/bls.v36i1.3926

Sielska-Badurek, E., Osuch-Wójcikiewicz, E., Sobol, M., Kazanecka, E., & Niemczyk, K. (2017). Singers' Vocal Function Knowledge Levels, Sensorimotor Self-awareness of Vocal Tract, and Impact of Functional Voice Rehabilitation on the Vocal Function Knowledge and Self-awareness of Vocal Tract. *Journal of Voice*, *31*(1), 122.e17-122.e24. https://doi.org/10.1016/j.jvoice.2016.01.011

Spector, F., & Maurer, D. (2013). Early Sound Symbolism for Vowel Sounds. *I-Perception*, *4*(4), 239–241. https://doi.org/10.1068/i0535

Sundberg, J. (1974). Articulatory interpretation of the "singing formant." *Journal of the Acoustical Society of America*, *55*(4), 838–844. https://doi.org/10.1121/1.1914609

Sundberg, J. (1977). The Acoustics of the Singing Voice. *Scientific American*, *236*(3), 82–91.

Sundberg, J. (1999). The Perception of Singing. In D. Deutsch (Ed.), *The Psychology of Music (Second Edition)* (pp. 171–214). Academic Press. https://doi.org/10.1016/B978-012213564-4/50007-X

Svantesson, J.-O. (2017). Sound symbolism: The role of word sound in meaning. *WIREs Cognitive Science*, *8*(5), e1441. https://doi.org/10.1002/wcs.1441

Tanner, K., Nelson, R., Merrill, R. M., & Power, D. (2005). Velopharyngeal Port Status During Classical Singing. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1311–1324.

Terwogt, M. M., & Van Grinsven, F. (1991). Musical Expression of Moodstates. *Psychology of Music*, *19*(2), 99–109. https://doi.org/10.1177/0305735691192001

Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How Linguistic Metaphor Scaffolds Reasoning. *Trends in Cognitive Sciences*, *21*(11), 852–863. https://doi.org/10.1016/j.tics.2017.07.001

Weinberger, N. M. (1999). Music and the Auditory System. In D. Deutsch (Ed.), *The Psychology of Music (Second Edition)* (pp. 47–87). Academic Press. https://doi.org/10.1016/B978-012213564-4/50004-4

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*(3), 349–366. https://doi.org/10.1016/S0095-4470(95)80165-0

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the {tidyverse}. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, *6*(1). https://doi.org/10.5334/gjgl.1646

Yanagisawa, E., Kmucha, S. T., & Estill, J. (1990). Role of the Soft Palate in Laryngeal Functions and Selected Voice Qualities. *Annals of Otology, Rhinology & Laryngology*, *99*(1), 18–28. https://doi.org/10.1177/000348949009900104

Yoshida, H. (2012). A Cross-Linguistic Study of Sound Symbolism in Children's Verb Learning. *Journal of Cognition and Development*, *13*(2), 232–265. https://doi.org/10.1080/15248372.2011.573515

Zbikowski, L. M. (1998). Metaphor and Music Theory. *Music Theory Online*, *4*(1). https://www.mtosmt.org/issues/mto.98.4.1/mto.98.4.1.zbikowski.html