Mapping vowel color and morphology: A cross-cultural analysis of vocal timbres in four yodeling traditions

Lawrence Shuster and Yannick Wey

A close up, comparative analysis of timbre in four samples of yodeling is presented. Yodeling, as a way of singing with asemantic syllables in different voice registers, relies strongly on timbre for musical expression. Timbre emerges as an integral structural variable whereas in many other European musical traditions, timbre fulfills a secondary, supporting role. This cross-cultural examination includes yodeling from Austria, the Central African Republic, Gabon, and the United States, and focuses on two aspects of timbre: vowel space and spectral morphology. Results show how various, non-related practices of yodeling rely on timbral features of interlocking low harmonics and a division of vowel space between physical voice registers (head voice, chest voice). Under close inspection, highly heterogeneous vowel spaces and timbral morphologies are revealed among the relevant practices. Culturally-based spectral expressions observed contrast with earlier studies that had focused on commonalities such as a rough division between grave and acute vowels.

Introduction

Of the four fundamental parameters of musical sound — pitch, duration, loudness, and timbre (Cohen & Katz 1979), timbre is to this date the least understood. At the same time, an understanding of timbre is crucial for the growth of musical genres, styles, and varieties, as it forms the basis for recognizing the characteristic sound of instruments and musicians. Listeners can distinguish musical instruments and singing styles very quickly by their timbral properties (Fales 2002). In her article "The Paradox of Timbre," Cornelia Fales diagnosed a "pitch-centrism" and a "timbre deafness" or, as later clarified, a "timbre amnesia" in African music research (Fales 2002:56) — a persistent "focus on melody" (this statement can also likely relate to rhythm), "where the dominant parameter is timbre." Whereas in western art music timbre fulfills a decorative, ornamental role subordinated to the primary structural parameters of pitch and rhythm, in many music traditions timbre emerges as an important formative variable, equal in status to pitch and rhythm and sometimes even more important. Similarly, previous research on yodeling of the Alpine region has focused on pitch, which led to the discovery of non-equal tempered tunings in Central Europe. Pitch has been studied in Swiss (Zemp 2015, Wey 2020) and Austrian (Ambrazevicius 2020) yodeling but studies on timbre are missing. This imbalance (Dreier 2007) assumes that pitch is "governed by law while timbre is governed by taste" (Fales 2002, 56). Therefore, variations of pitch will be judged as correct or incorrect, while differences in timbre will be perceived as a matter of taste, sounding pleasing or displeasing.

In addition to this habitual "pitch-centrism," there are technological reasons for the deficit in timbre research, as explained by Graf in a report on the state of sonographic analysis of timbre (Graf 1975). Phonographic recordings as well as the sonograms used for analysis in earlier studies (early and mid-20th century) only displayed a narrow spectrum of overtones, limiting the possibilities for displaying timbre. Therefore, a primary reason for the lack of understanding regarding timbre is that while it is always at the forefront of our musical experience along with pitch and rhythm, without the assistance of technology it was difficult to explore.

To understand the generation of vocal timbre and its constitutional role in the formation of music without lyrics, we analyze the timbre of yodeling tunes from different cultural backgrounds. We understand yodeling as a way of singing on non-lexical syllables in different voice registers (chest voice, head voice). Although not based on words from a dictionary, yodeling applies syllables, which usually consist of either just one vowel or a consonant followed by a vowel.

The primary motivation for our study is twofold: first, to develop new analytical tools and techniques for modeling and interpreting the sonic design of musical spectra with a particular emphasis on vowel tone color; second, to consider Cogan's theory of binary oppositions (Cogan 1984) as a point of departure in developing several extensions to Cogan's initial theoretical formulations, as well as introducing new analytical strategies for modeling aspects of spectral morphology, the latter made possible through expansions of traditional set class theory and the use of transformational networks. As this is a data-driven study, our analytical method is devised so that it can be reverse engineered, enabling our results to be replicated, or at least closely approximated, by using the original sound sample, various free software applications, and the data archives provided. Our approach to timbre is both specific and highly constrained as we focus exclusively on vowel tone color in unaccompanied solo vocal performances. Onset attack points, and considerations involving the use of consonants, are not included in our study. While we readily acknowledge the importance of these and other considerations, we assert that timbre is such a complex, multifaceted phenomenon that it can never be adequately understood nor explored with any single theoretical approach or analytical strategy. Instead, we tease out and explore particular features and attributes of sonic design while intentionally neglecting or veiling others, anticipating that a piecing together of analytical insights and perspectives in the end might help elucidate a more comprehensive model of musical timbre and its manifold attributes. As such, our study presents an expanded view of vowels and tone color but a offers a somewhat restrictive view of timbre, limited in scope to consideration of unaccompanied solo vocal performances only.

Cross-cultural comparative studies on yodeling

Yodeling has served as the subject of several recent studies involving cross-cultural comparisons of musical characteristics. In some cases, yodeling is seen as an example of a prehistoric remnant of music and/or assumed to share a common evolutionary path (Grauer 2006, Fenk-Oczlon 2009, Nikolsky 2020). We refer to some of these theses insofar as they include yodeling in their reasoning and the implications of their observations of timbre, without prejudice to such far-reaching conclusions. Drawing a connection between Central Africa and the Alps, Grauer (2006, 17) presents similarities between two recorded examples from the compilation *Voices of the World* (1999) — a Ju'hoansi[1] healing song and an Appenzell[2] Zäuerli (yodel). Both include "wide intervals, relaxed, open voices, and nonsense vocables" and "the cantometric profiles [...] would be almost identical." Grauer compares both of these styles, along with several others, based on their cantometrics profiles (Grauer 2006, 7). However, Grauer (2006) does not mention timbre specifically in his text; the main argument for his thesis — that a "pigmy/bushmen-style" persists across different cultures and serves as a window into a far musical past — is the use of the "hocket" (Grauer 2006, 13) technique, an alternation of sequences in different voices.

One can easily point to significant differences in the polyphonic structure and the meter of the different musical traditions mentioned. The Appenzell Zäuerli from Switzerland are led by a single melody that is often sung without any accompaniment. During group performances, a second voice adds another melodic line, which follows the first closely, often in relations of thirds or sixths, in most cases below the first voice. The other singers form a choir and hold long notes in the low register, resembling a drone. In contrast, as Fürniss demonstrates in her analysis of Aka and Baka yodeling, the voices may be based on a single set of sequences but interweave to a polyphony where there is no hierarchy from higher to lower voices. However, there are further arguments for yodeling as a window into a prehistoric past. Referring to Nordic *kulning*, Nikolsky states: "The surviving Nordic tradition of kulning provides the gist of the Neolithic pastoral music-making." He derives this assumption from the idea that *kulning* developed in parallel with animal husbandry (Nikolsky 2020, 24). Categorizing yodel as a remnant of prehistoric times is not a new hypothesis; it was widespread in the German anthropological schools of the interwar period and was formulated for example by Robert Lach (1928). In 1939, Wolfgang Sichardt published his dissertation, *Der alpenländische Jodler und der Ursprung des Jodelns* ("The Alpine Yodel and the Origin of Yodeling"). He identified five historical layers based on musical features of recorded yodel melodies. Nikolsky's (2020, 24) emphasis "that matriarchy influenced early pastoralism" and that therefore *kulning* developed in Nordic culture, where herding has been primarily in the responsibility of women, coincides with Sichardt's idea of yodelers as originally "matriarchal planters" ("mutterrechtliche Pflanzer", Sichardt 1939, 167), a thesis he derived from the now obsolete "Kulturkreislehre."

These cases illustrate why cross-cultural comparisons of yodeling, however valuable from an analytical point, can easily lead to interpretations that should be scrutinized further. The discussion of musical similarities between yodeling cultures is itself complex and has sparked lively debates. After Fürniss (2006) referred to observations by Olivier & Fürniss (1997) that showed different concepts between *Aka* and *Ju'hoansi* multipart singing, Grauer (2008, 2009) wrote two papers accusing both of "playing down" the importance of aural similarities and arguing for the common ancestry of the two musical traditions. Furthermore, the argument that musical similarities indicate a common evolution must be extended by the possibility that these similarities are also manifestations of a common present or recent past.

1. The Ju'honasi live primarily on the Kalahari Desert in Namibia. In some examples of the literature, they are referred to by the outdated name "Bushmen."
2. Appenzell: A region in the northeast of Switzerland.

Yodeling practices exist with very different social functions, ranging from hunting and hunting-related rituals (Fürniss 2006) to herding or signaling (Tiderman-Österberg 2020) as well as other forms of communication (Fink-Mennel 2007, 93), protection from natural disasters (Haid 2008, 62), or therapy (Walcher 2014). There have been attempts to determine the "original" function of yodeling, but, considering the lack of historical sources in orally transmitting societies, there is no way to define which function originated in which period. In all cases, yodeling today also serves for entertainment purposes; the entertainment purpose might be the only quasi-universal aspect of yodeling today. The wide range of functions may however count as a caution against evolutionary comparisons and equivalences (as proposed by Fenk-Oczlon 2017; Grauer 2006; and Nikolsky 2020); the often-used argument that social context can be neglected for the purpose of analysis may hold for a straight-forward acoustical analysis but becomes a liability when the results are subject to interpretation and framed in theories of musical development, evolution, and functions of music.

THE USE OF VOWELS IN YODELING

There are various possibilities in categorizing vowels. Whether, for example, /e/ and /ɛ/ are differentiated, depends on the case. Quoting different vowel categorizations below, we indicate in each case which alphabet was used for categorization. For the presentation of our own results, we always use the International Phonetic Alphabet (IPA). The level of detail with which we distinguish between the vowels seeks to be appropriate for the given representations of data, ranging from categorical perception (in transcriptions) to acoustical measurements (see table 2).

Thanks to a large corpus of published transcriptions, Austrian yodeling provided a readily accessible source to study the use of vowels. Graf (1965, 16) compiled 858 transcribed vowels from two large collections of Austrian yodels[3] and counted their appearance in the low, middle, and high registers.[4] The letters "a," "å," and "o" appear almost exclusively in the lower register; "i" in the higher register. The use of "u" is ambiguous. There may be unknown deviations from the actual singing since this summary is based on transcriptions carried out mostly without the help of audio recordings.[5]

To compare the selections of vowels cross-culturally, Fenk-Oczlon and Fenk (2009, 2) "demonstrate a close match between vowel pitch and musical pitch in nonsense syllables of Alpine yodelers and in the yodeling of African Pygmies." They compared 15 monophonic yodel transcriptions from the Pommer collection and referred to previous studies on yodeling of "African pygmies" that reported a "strong relationship between vowel timbre and musical pitch in meaningless syllables."

---

3. The collections are Josef Pommer: *444 Jodler und Juchezer*, Vienna 1901, and Georg Kotek and Raimund Zoder: *Im Heimgarten*, Vienna 1950.
4. Graf (1965) does not specify how he defined the different registers. They do not concur with the change between head voice and chest voice, because these registers are not indicated in the relevant notations.
5. Even though the phonograph was available by the turn of the 20th century and employed for folk music research in parts of the Austro-Hungarian Empire, the collector Josef Pommer rejected it and argued against its value for music research.

To this date, the most comprehensive study of yodel timbre concerns the Aka community in Central Africa (Fürniss 1992), based on audio recordings collected by Simha Arom (1976). The results demonstrate not only the quantitative distribution of vowels in the given sample but also account for the overtone spectrum of these vowels in comparison with their overtone spectrum in the French language. Fürniss (1992, 59) distinguishes between a female and a male voice; however, the differences are not significant, as summarized later: "these results are valid for both the singers of different sexes, although individual variations can be found" (Fürniss 1992, 94; translated from the German by the present author). Furthermore, the two singers seem to have strictly avoided a crossover regarding the use of vowels in both the high and low registers. This contrasts with Graf's findings depicted above, where the use of certain vowels dominates the high or low register but crosses over to another pitch register in a smaller number of cases. Still not revealed is how strongly yodel syllables are connected to spoken language. Fürniss (1992, 30) lists the vowels of the Aka language, /I, e, ɛ, a, o, ɔ, u/; these do not include vowels /y, ʉ, ø/ from the table of yodel vowels. Graf included the Austrian-German dialect /å/.

The investigation of vowels in yodeling so far broadly sums up that the use of vowels corresponds with the voice register (high/low or head voice/chest voice), and thereby with certain tessitura. Voice register and pitch range are not to be confused; the point of switching between head voice and chest voice can vary strongly between singers or even during a performance by the same singer. In the studies discussed, vowels are generally used as tokens for timbre, and vice versa.[6] Research so far has worked almost exclusively on the level of alphabetic vowels (except for Fürniss 1992). A categorization based on a limited set of vowels gives a reduced and well readable picture of timbre. As shown in the previous sections, studies of the use of vowels have shown overlaps between different yodeling cultures. The vowels are discriminated on different scales: 'a-e-i-o-u', a subset of the IPA (Fürniss 1992), or local dialects (Graf 1969); but we have to ask how similar these sounds are, which are all transcribed for example as /i/. We have to address this question with an analysis that makes use of technology and methods developed in recent years. Furthermore, the analysis of vowels often relied on notation alone, and as for Austria, the data cited are based on Pommer's transcription from the turn of the 20th-century, based on the aural impression without recordings.

Approach and Audio Samples

We conduct a detailed analysis of the coloring and the application of vowels, in four examples of yodeling in different musical cultures. Although based on different music-aesthetical principles and rooted in different traditions, all of these examples share a common vocal technique. We ask the following questions: What characteristics define the timbre of different yodeling practices? How does timbre change between the vocal

6. Fenk-Oczlon (2017, 1) notes: "Timbre is, beyond question, the primary parameter that allows us to discriminate between different vowels, but vowels also have intrinsic pitch, intensity, and duration." While the idea of an "intrinsic pitch" is plausible based on the quantitative distributions shown by Graf (1975) and Fürniss (1992), it remains unexplained what is meant by the intrinsic intensity and durations of vowels.

registers? How does timbre, beyond vowel designations, vary among and within the samples?

For a timbral analysis, samples of monophonic singing had to be selected, so that the timbre of the voice does not interfere with other voices and musical instruments. This greatly limited the options for sample selection, as most yodeling is performed by groups (in the case of Central European Alpine, Aka, and Baka yodeling), or accompanied by instruments such as the guitar (in the case of North American cowboy yodeling). The focus on monophonic recordings means that these are out of the context of the (usually group) performances. The following transcriptions of the audio samples are necessarily descriptive, as they account for the pitch and the timbre of the voice. Additionally, we indicate the use of head voice by using squared note heads (a notation also used for this purpose by Fink-Mennel 2007 and Wise 2012). For readability, we chose standard staff notation. However, solo yodeling does not generally conform to equal-tempered tuning (even in the European context) (Wey 2020) and the pitch designation is therefore approximate.

Sample 1: solo phrase from Bobangi — *Chant pour célébrer la prise de gibier*[7] (Aka, Central African Republic)

The description of the song (Arom 1994) reads: "Bobangi (track 6) is performed by six women.[8] Here the voices take on the characteristic guttural quality also found in other forms of Pygmy music. The listener will notice the use of the yodel, particularly by the soloist. This piece stands in sharp contrast to the preceding ones by virtue of the importance assumed by the 'blocks of silence' separating the different verses."

Aka music is rooted in hunting and recalls the hunter-gatherer life of the past: "In spite of the progressive loss of hunting rituals, the dances that belonged to them are often maintained and perpetuated in a context of entertainment when men and women meet to dance together" (Fürniss 2006, 167). Yodel functions as a method to attract game: "we yodel in the forest to eat. The yodel is pleasing to the forest spirits and then they give us game," explained "a Biyaka woman" to Robert Farris Thompson (quoted after Rouget 2011, 103). The yodeler represents one of several voices in a polyphonic setting, where voices are primarily distinguished by contour and timbre, rather than pitch collection or rhythm (Roeder 2019, 1). Fürniss (2006) explains how "the *dìyèí*, literally 'yodel,' is sung above all the other parts by women. It consists of melodies of mainly wide intervals and uses specific vowels correlated to the two yodel registers: the low yodel register with open vowels as /e/, /a/, and /o/, and the high yodel register with closed vowels as /i/, /y/, and /u/" (Fürniss 2006, 11).[9] Aka *dìyèí* is closely embedded in the musical structure of the other voices in polyphonic settings: Elements are combined, transposed, and varied to create a yodeling voice switching between chest voice and head voice (Fürniss 2006). Aka learn yodel by

7. "Song for celebrating the catching of game" (translation by the present author).
8. Unfortunately, the names of the performers have not been published in the CD booklet.
9. Fürniss clarifies that the vowels are not written in IPA and that "e is pronounced like in 'bed,' o like in 'four'."

imitation; they "hardly ever refer to the parts and their patterns explicitly. Indeed, they are immanent concepts that are never taught to the musicians as such" (Fürniss 2006, 176).

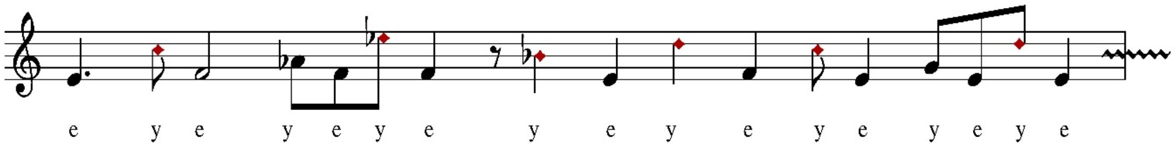


Figure 1. Opening solo yodel phrase from "Bobangi." From the CD "Aka Pigmy Music" (track 6) (1973/1994: Audivis, Unesco, recordings and commentary by Simha Arom).

Sample 2: phrase from a recording labeled "yodel etudes" (Baka, Gabon)

In the CD booklet, Sallée (2001, 10) writes about the "yodel etudes": "The yodel consists of alternately changing register and vocal emission. The voice breaks in a constant back-and-forth figure, shaping broad, jagged melodic contours. The lower register, sung in chest voice, has a round, full timbre; the high register is a pinched *falsetto*, transparent and flute-like. Note the choice of open vowels in the lower register (ah, short e as in bed), and closed vowels for the upper (e as in see)." The editor already described the selection of vowels as clearly divided into voice registers, using only one vowel for head voice and two for chest voice. It remains open whether their registers coincide with pitch ranges.

Baka yodel is, at least in some communities, part of a ritual called *yeli*. Traditionally, a group of women performed *yeli* to invoke good luck for their men going on an elephant hunt (Rouget 2011, 93; Weig 2018, 199). As Rouget explains, "the *yeli* involves neither dancing nor men. The word itself refers to a kind of choral singing, polyphonic, of course, and, in particular, yodeled" (Rouget 2011, 93).[10] *Yeli* distinguishes itself from other singing practices because it includes neither words nor dancing (Rouget 2011, 94). As Doerte Weig explains, *yeli* has not been performed in recent history: "no one recalled elephant hunting rituals, and especially the women spoke of this with great sorrow emphasizing that the *yeli* had not taken place for many years" (Weig 2018, 199). Elephant hunting has long been in conflict with international environmental protection efforts. A 2020 UN report found that "that armed eco-guards engage in violence and threats of violence against the indigenous Baka people in the Messok Dja area [Congo, near Gabon and Cameroon]" (UNDP 2020, 7), and since the elephant hunt has been abandoned by the local communities (UNDP 2020, 29), "the destruction of their fields and crops by elephants" has emerged as a new problem.

Abandoned as a ritual, yodeling has turned into a form of entertainment, but also a way of communicating, as further recounted by Weig: "Singing in the forest is mostly a way of entertainment, for example, whilst the women are digging for manioc or collecting nuts during the day. If there is another group of Baka walking through the forest, then calling out, yodeling briefly, is a way of locating each other or communicating information about something important that happened back in the village or another location" (Weig 2018, 197). The yodeling voice resonated with the natural environment of the forest, "in the

10. For yodel as a mostly female practice in Baka music see Sallée (2001, 10): "the women are the best interpreters of the yodeling vocal technique."

cathedral-like acoustics of the forest canopy, the yodel [...] resonates especially well and can be heard far away." (Rouget 2011, 103).[11]

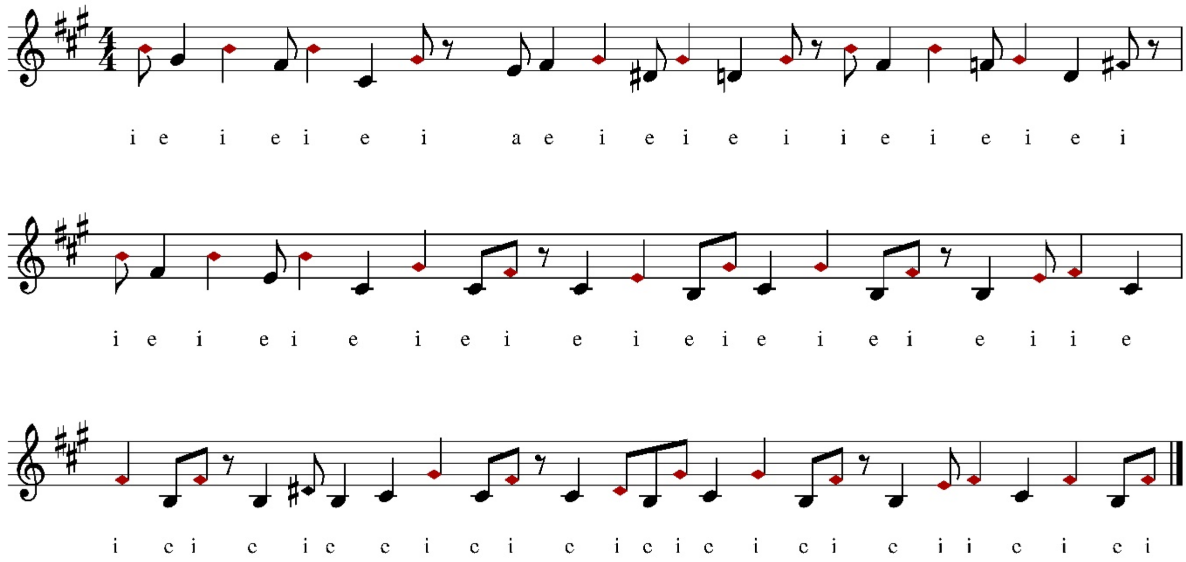


Figure 2. Excerpt from a "yodel etude", Baka, Gabon. From the CD *Gabon. Music of the Bibayak pygmies. Epic cantors* (track 2). Radio France/Harmonia Mundi 2001.

Sample 3: Paul Fetz — Juz (Vorarlberg/Austria)

Fink-Mennel (2007, 120) notes of this recording: "singer: Paul Fetz (1914–2003). Recorded in Egg-Grossdorf, 6 March 1997 by Evelyn Fink" (translated from the German by the present author). The *Juz* is a local name for a short yodel. According to Fink-Mennel (2007, 97), the musical form is free and different from some other Austrian yodels, not divided into a stable number of bars. It belongs to "the elementary forms of singing and is used for communication, organization or entertainment sometimes over long distances" (Fink-Mennel 2007, 93, translated from the German by the present authors) and can be distinguished and understood by its tone sequence. Yodeling in the Alpine region, which mainly covers Switzerland and Austria, regularly takes the form of question and answer in two phrases. This is justified with the adoption of song forms, but also insinuates the hypothesis of its origin as communication between different singers.

11. Similarly described by Weig (2018:198): "The practice of yodelling is like an echo in that the yodel resonates especially well under the dense and closed canopy of the tropical forest."

Figure 3. "Juz" from Vorarlberg/Austria. From the CD "Johlar und Juz — Registerwechselnder Gesang im Bregenzerwald" (track 26). Supplement to the book with the same title by Evelyn Fink-Mennel, Graz 2007.

Sample 4: Don Walser — One Last Yodel (Texas, U.S.)

North American cowboy yodel is an understudied subject, with notable exceptions such as Wise's *Yodeling and Meaning in American Music* (2016). Songs usually alternate between verses with text and short yodeling segments sung only on yodel syllables. Despite a highly successful career as musician, Don Walser (1934–2006) has so far received considerably less attention than did yodelers like Wilf Carter (alias Montana Slim, 1906–94) and especially Jimmie Rodgers (1897–1933).

Several publications state that yodeling emerged in American country music when blues musicians adapted the "Swiss yodel" (Abbot & Seroff 1993; Wise 2008; Coltman 1976, 91). However, the source of inspiration was most likely Tyrolean, i.e. Austrian yodel. Most famous in the U.S. was the Rainer family, who toured and influenced singers in various countries (Hupfauf 2016).[12] Although inspired by performances of the Rainer family and possibly other touring groups from Austria, the explanation by "acculturation" of "Swiss" yodel (Abbot & Seroff 1993) understates how much American singers transformed the sound of yodel and developed it into a recognizable feature of their own music. Connections with instrumental music are established with the banjo, and with the guitar in the American South, where yodel songs were influenced by African-American blues musicians (Coltman 1976, 92). Inspired by Austrian immigrants, blues musicians in the American South yodeled and developed a style that can easily be differentiated from its Alpine counterparts. Wise (2010, 12) describes how, in Jimmie Rodger's case, the yodeling segments in his songs became iconic of his personal vocal style: "following the same incipits based on a few basic melodic shapes and rhythmic profiles and intoned with the same syllables, he makes his yodels instantly recognizable and familiar, regardless of the style of the song." Furthermore, the texts of songs that include yodeling vastly differ between cowboy yodel songs and those from Alpine Europe. For example, a phrase like the one in Jimmie Rodger's famous "T for Texas" that reads "I'm gonna buy me a shotgun with

12. The Rainer family toured the United States from 1839 to 1843 (Wise 2016, 87). In the following decade, imitations and parodies the Tyrolean singers spread and were adopted by some minstrel shows, the yodeling sometimes referred to as "Tyrolean warbling" (Hupfauf 2016: 198).

a great long shiny barrel," would not fit in a Swiss or Austrian yodel song, most of which revolve around topics of (longing for) love and nature.[13]

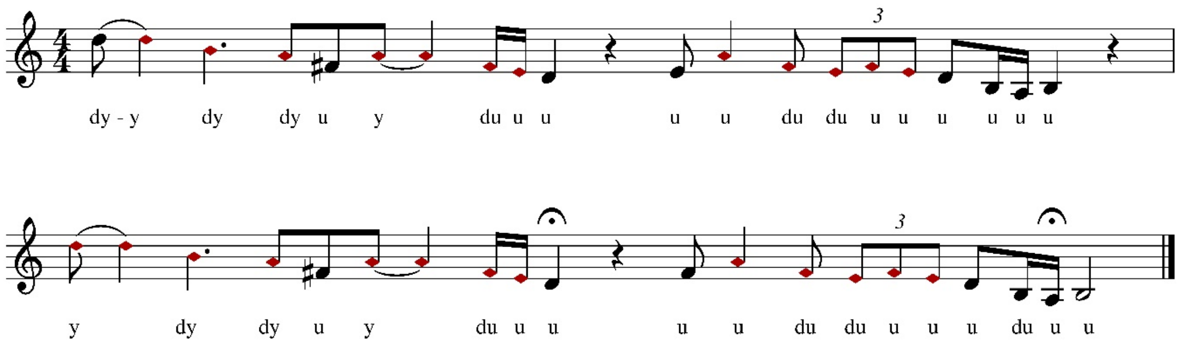


Figure 4. "One Last Yodel" by Don Walser. From the album "Just Me and My Old Guitar" (track 17) Mark Rubin / Rubinchik Recordings, 2011.

Analytical method and data retrieval

Parsing the timbral surface: spectral segmentations

The initial step of this analytical method involves parsing the timbral surface into a succession of spectral segmentations, the boundaries of which are defined by the respective duration of each pitch within the succession of fundamental tones that comprise the melody. Figure 5 demonstrates a sample spectral segment which represents the second of 21 spectral segments excerpted from the first phrase of the Baka performance. Observe that each spectral segment appears as a highlighted, vertical slice of the available spectrum as demonstrated in the main display of the spectrogram. Each segment contains the fundamental pitch (G4) and all accompanying overtones. The horizontal boundaries of each respective vertical spectral segment are coextensive with the duration of the fundamental pitches that form the melody, while the vertical boundaries, or spectral compass, of the segment are defined by the lowest (fundamental) and highest sounding frequencies. The respective durations of each segment can be easily approximated by referencing the time ruler along the bottom horizontal in the main display of the figure. Because each spectral segment corresponds to a discrete pitch within the succession of fundamental pitches that comprise the melody, and each spectral segment is characterized by well-defined attack point onsets, the boundaries of each segment can be easily ascertained by eye.

13. The same can be observed in comparison to Aka music: Fürniss (2006, 13) translates the text of an Aka song with yodel part as "The hair of my pubes is dense," which would fit neither Alpine European nor cowboy yodel songs.

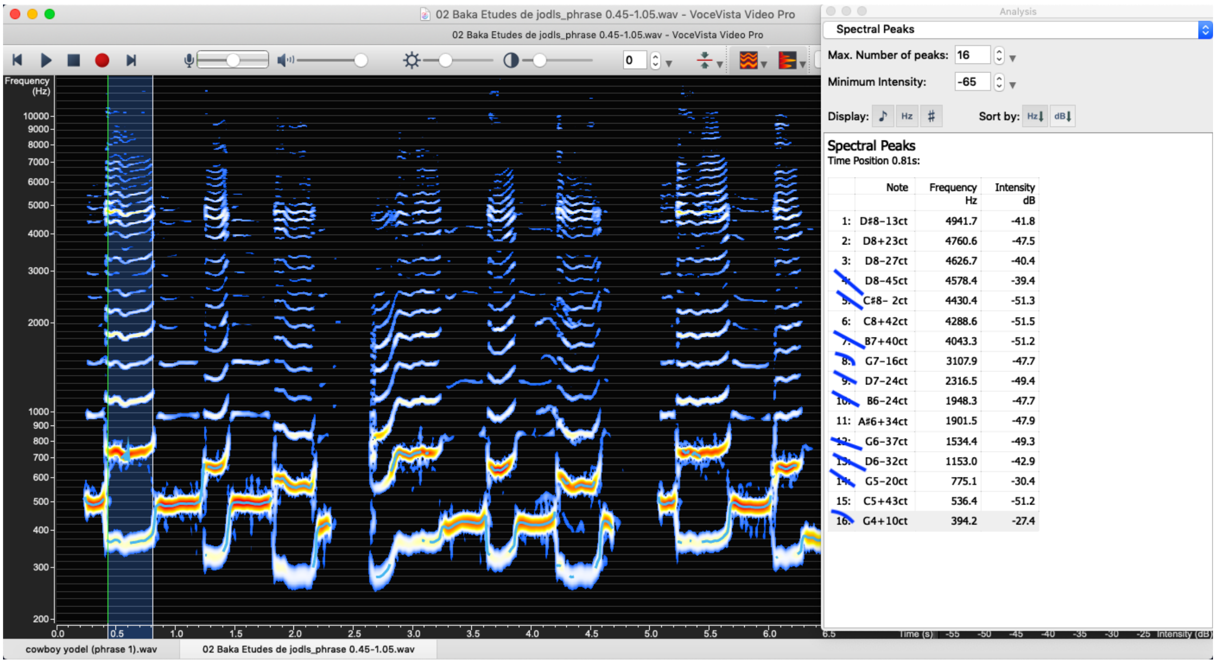


Figure 5. Spectral segmentations in the example of the Baka recording, segment no. 2.

Measuring harmonics and dBFS intensity values

Having defined the spectral segments which provide the basic units of analysis, we now explore ways in which to characterize them and explore correspondences between them. The relative intensity, or amplitude, of the individual harmonics contained in each spectral segment can be determined through a fast Fourier transform (FFT) and displayed as a function of frequency. Each fundamental pitch and its accompanying harmonics are measured in terms of their respective amplitudes using the dBFS scale (decibels relative to full-scale); a digital measurement of amplitude where the value "0" is assigned to the loudest sound in each signal and all other measurements are shown as negative integers in comparison. Once the user has defined a given spectral segment, the software provides an automated analysis revealing both the average and peak amplitudes, frequency, pitch class, and octave registration, for each harmonic within the segment as shown in figure 5.

Observe that in addition to those harmonics associated with the fundamental pitch G4, crossed out in blue, other harmonics also appear in the harmonics inventory which are not associated with the fundamental G4. These harmonics belong to the preceding spectral segment with the overlap resulting from reverberation, echo, delay, and other acoustic artifacts stemming from the recording process and/or the performance environment. While we acknowledge that such phenomena are very much part of the performance and certainly contribute to our perception of timbre, these instances of segmental overlap do not adversely affect our perception of vowel color associated with the fundamental pitch which represents the primary target of this study. Accordingly, these harmonics are simply not included in the analysis.

To make the dBFS values demonstrated in figure 5 meaningful we attempt to differentiate those harmonics which feature in our perceptual experience from those that don't, as only a portion of the visual information displayed in the spectrograph can be heard or experienced. Since the dBFS values are context specific, we need to establish a general threshold for salience with the understanding that these values themselves are, at best, averaged approximations. The various perceptual and cognitive considerations provide a kind of filtering process which assist in distilling those features which play a prominent role in our listening, and provide some measure of phenomenological accountability, permitting us to ground our analytical assertions in perception, as opposed to many studies which interpret spectrographic images primarily through visual analysis alone.

To confirm the salience of various sonic elements, verification protocols were implemented which involved the tedious process of sifting through the timbral surface using a variety of filters available within the VoceVista Video software. These filters enable the user to single out individual frequencies, as well as combinations of frequencies, for aural confirmation. In order to account for the possibility of masking, the various harmonics are then variously combined to confirm their aural salience, while those deemed imperceptible are eliminated. As a result of this procedure, -55dBFS has been established as the threshold for salience for our analysis of this recording, meaning that we assume the vast majority of diligent listeners can, after careful and repeated listening, confirm the presence of those sonic details asserted in our analysis. In the same way that visual acuity varies greatly between people, aural acuity varies tremendously as well. Adjusting this salience threshold (-55dBFS) is tantamount to modulating the degree of magnification with which the timbral surface is viewed, with incremental increases or decreases in threshold revealing detailed views of the timbral surface accordingly.

Defining spectral sets and classes

In an unpublished paper presented at the Annual Meeting of the Society for Music Theory in 2012, Lawrence Shuster introduced a new theoretical formulation for defining spectral sets and classes. Spectral sets tell us which harmonics are perceptually available within a sound at a specific level of magnification (dBFS value). Once the respective values for each harmonic in each spectral segment have been compiled, and the aural salience of constituent elements has been confirmed, those harmonics remaining are considered members of the corresponding spectral set associated with each segment.

The analytical notation for a spectral set is [H1, H2, H3], where square brackets indicate an ordered collection of pitch classes and their corresponding octave registrations, as informed by their respective order within the harmonic series, with the fundamental pitch always appearing in the first position. More generally, the notion of spectral class simply generalizes the notation of the spectral set by indicating only the respective position of each harmonic within the harmonic series, and not the specific pitch class nor the corresponding octave registration. Consideration of spectral class allows us to compare sonic contexts that have different fundamental pitches and appear enclosed in parenthesis, as opposed to square brackets. In uncommon instances where the respective harmonic assigned to an order position is absent, a dash (-) serves as a placeholder.

While successful in enumerating the perceptible harmonics appearing within each spectral segment, as well as in indicating their associated octave registrations and positions within the harmonic series, many other important details are intentionally omitted, such as information regarding intonation, duration, consideration of attack onsets, as well as the respective intensity of each harmonic, for each spectral set. Like the use of pc sets in set class theory, spectral sets also provide us with a stripped-down reduction of all qualities and characteristics not pertaining to those of pitch class, octave registration, and position in the harmonic series. Of course, the most significant distinction between pc sets and spectral sets is reflected in the way such sets are perceived; in pc sets we can hear each constituent element as a distinct pitch with a specific frequency, whereas with spectral sets, we hear the composite timbre generated by the combination of all harmonics in a given set as opposed to experiencing each harmonic individually.

Pitch drift and variable intonation often emerge as important expressive devices characteristic of performance practice in numerous musical traditions. As a result of these considerations, when there emerges a significant deviation from the pitch classes indicated in the transcription and their actual acoustical counterparts as evidenced in the recording and acoustical measurements, we will refer to each fundamental pitch and their associated harmonics by their actual frequency measurements. It should be observed, however, that this does not mean that such cases are considered discrepant by the performers or cultural participants, only that the respective sizes and distributions of intervallic spaces are different. To represent these insights and make use of the available software applications, we measure pitch, frequency, and duration using tools and strategies associated with western music simply as a means of representation. If anything, the complexities involved in such nuanced, subtle participatory discrepancies attest to the sophistication underlying the organization of sounds.

The following table demonstrates the 21 spectral sets for the first phrase segment of the Baka performance. In each case, the fundamental pitch and octave registration is provided along with the associated order positions of the harmonics series as well. The complete data are included in the supplement.

1	**[B4, B5, F#6, -, D#7, F#7]**
	(123-56)
2	**[G4, G5, D6, G6, B6, D7, G7, B7, C#8]**
	(123456789)
3	**[B4]**
	(1)
4	**[F#4, F5, C6, F6, A6, C#8, D8]**
	(1234567)
5	**[B4, B5, F#6]**
	(123)
6	**[C#4, C#5, G#5, C#6, F6, G#6, B6, C#7]**
	(12345678)
7	**[G#4]**
	(1)
8	**[E4, E5, B5, E6, A6, B6, -, D8]**
	(123456-8)
9	**[F#4, F#5, C#6, F#6, A#6, C#7, D#7, F#7, G#7, B7]**

	(123456789T)
10	[G#4, G#5]
	(12)
11	[D#4, D#5, A#5, D#6, G6, A#6, C#7
	(1234567)
12	[G#4, G#5]
	(12)
13	[D4, D5, A5, D6, D7]
	(12345)
14	[G#4, G#5]
	(12)
15	[B4, B5, F#6]
	(123)
16	[F#4, F#5, C#6, F#6, A#6, C#7]
	(123456)
17	[B4]
	(1)
18	[F4, E5, B5, E6, G#6, B6]
	(123456)
19	[G4, G5]
	(12)
20	[D4, D5, A5, D6, F#6, A6, C7, D7, E7, F#7, G7]
	(123456789TE)
21	[F#4]
	(1)

Table 1. Spectral sets and classes inventory for the Baka performance.

Available types of relationships between spectral sets

Available types of relationships among spectral sets include consideration of the fundamental pitch and corresponding octave registration, the cardinality (the number of harmonics in the set), the order positions of the harmonic series included in the set, and most significantly, the set's associated lower order formant frequencies (F1 and F2) and the specific IPA vowel type and color they define.

The various forms and types of relationships can be conceptualized in terms of a hierarchy of correspondence, organized from strongest to weakest. If two sets share the same vowel, fundamental, cardinality, and order positions of the harmonic series, then they have identical vowel colors. If two sets have different cardinalities but share everything else, they are related by the inclusion relation as either literal or abstract subsets or supersets. The analytical notation is T0(y) where y is equal to the difference in cardinality between the two sets. If two sets have different fundamentals and different pitch classes but the same number of harmonics, the same order positions of the harmonic series, and the same vowel type, then they are related by transposition, or Tx, where x is equal to the number of semitones between fundamental pitches. If two sets meet the above requirements except for different cardinalities, they are related as abstract, or transposed, subsets or supersets.

The analytical notation is Tx(y), where x is equal to the level of transposition in semitones and y equals the amount of offset between cardinalities. If two sets do not share the same vowel color, they do not share this specific type of timbral correspondence predicated on vowel color, despite the fact they may be related by various forms of affiliation by pitch and frequency, as well as other types of timbral correspondences. While considerations involving pitch class, octave registration, cardinality, and order positions within the harmonic series all play a significant role in contributing to the formation of a spectral segment's unique timbral profile, they do so in a secondary capacity. Of the various forces shaping vocal timbre, it is a set's corresponding vowel type and the specific formant frequency bandwidths that define it which play the most significant role in terms of timbral shaping. In speech and singing, the harmonics generated by the vocal cords are further shaped by the resonant frequency bandwidths, or formants, of the vocal tract and oral cavity which act as filters, intensifying some harmonics while dampening others. Phonologists have established that spoken vowels are distinguished by the ratios expressed between the first three resonant frequency bandwidths, or formants, typically abbreviated as (F1, F2…). however, in most cases only the first two formants are required for vowel disambiguation (Ashby 2005).

If formant frequencies are primarily responsible for disambiguating a vowel and defining its associated tone color, then the remaining non-formant frequency harmonics further modify these primary vowel timbres to produce a variety of available shadings, hues, and intensities. Timbral equivalencies are predicated on vowel-formant classification, whereas frequency and pitch class relationships are construed solely based on octave registration, pitch class, and respective order positions with the harmonic series. Spectral sets which share the same vowel also share the same formant frequencies, the latter of which establishes the primary tone color of the vowel.

Measuring vowel formant frequencies

To situate the specific vowel sound within the greater spectrum of available vowel colors of the International Phonetic Alphabet (IPA), we must first disclose the lower order formant frequencies necessary for vowel disambiguation in each spectral set. One application available in the PRAAT linguistics software enables an analyst to select and analyze a given spectral segment using wideband spectrography which reveals the distribution and positioning of the lower order formant frequency bandwidths (F1 to F5).

By applying this technology, we can demonstrate close approximations for the first and second formant frequency bandwidths for each spectral set in each of the four analytical samples. If a listener can identify the vowel sounds associated with a given spectral set, then they are hearing the formant bandwidths which ultimately define each vowel and inform its unique timbral coloring. Figure 6 demonstrates the first and second formant frequencies of the Baka example, isolated within the spectrum using PRAAT. The specific formant frequency measurements for each spectral set in the four analytical samples are included in the attached datasets available in the appendix of this study. Observe in the following figure that with the exception of the first formant frequency (appearing as the lowest horizontal red line on the graph), the formant frequency bandwidths visualized in the four remaining horizontal red lines, representing formant frequency bandwidths F2, F3, and F4 respectively, are neither consistent, stable, nor uniform, but are apparently in a state of nearly continuous micro fluctuations. In measuring the formant frequencies for F1 and F2 within each successive formant frequency bandwidth, the data sample was collected from the approximate center of successive segmentations where the signal was generally in its most stable, constant state. Once again, it is the spatial ratio expressed between formant frequencies F1 and F2 from which we infer the corresponding vowel type and quality.

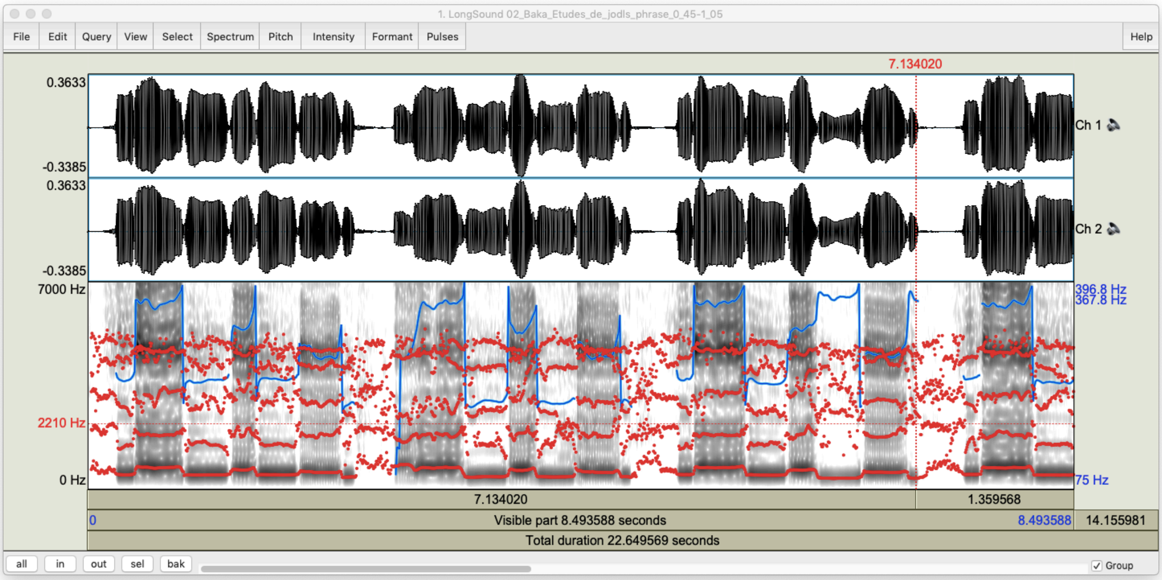


Figure 6. Measuring formant frequency bandwidths in the Baka sample using PRAAT.

Mapping vowel formants in IPA vowel space

Once the corresponding frequency values for the formants located in the first and second order positions have been obtained (F1 and F2), we can now isolate these frequency bandwidths within the visual display of the spectrogram and determine which harmonics are intensified as a result of their alignment with these formant frequency bandwidths and which are not. Moreover, we can not only demonstrate the specific harmonics that are intensified because of their frequency alignment within these formant frequencies; using the VoceVista Video Pro software, we can map these formant frequencies directly onto the IPA vowel space and plot the exact position of each vowel sound with the available spectrum of IPA vowel colors with a high degree of technical accuracy. Figure 7 demonstrates the projection of the first and second formant frequencies associated with spectral set 2 of the Baka sample.

Observe that the distribution of available vowels within the greater IPA space are neither uniform nor symmetrical. Rather, each syllable, as indicated on the graph, may be conceptualized as a specific vowel with a specific locus on the graphic display. But rarely do instances of specific IPA vowels map onto their corresponding vowel labels in a direct, one to one correspondence. Instead, we hear the designated IPA vowel plot, as defined by the specific intersection of formants F1 and F2, and the wide variety of available instances of these vowels existing at varying distances, and subject to varying degrees of gravitational attraction, from the given reference vowel within the space. Where one vowel ends and another begins challenges our sonic perceptions in the same manner as distinguishing at which point does blue end and purple begin. The individual IPA vowel space plots for each spectral set, in each of the four analytical samples, are included in the corresponding data set in the appendix.

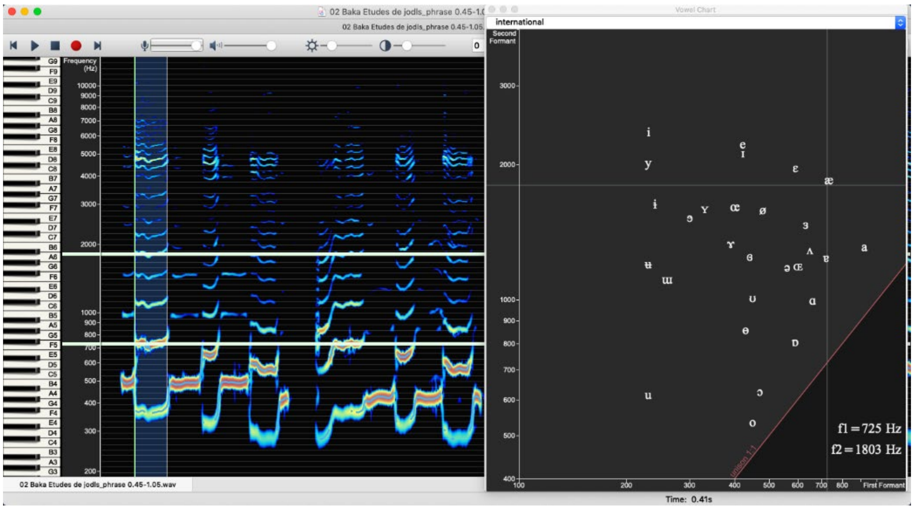


Figure 7. Sample vowel plot within the IPA vowel space for the Baka sample.

RESULTS

Mapping spectral transformations between sets

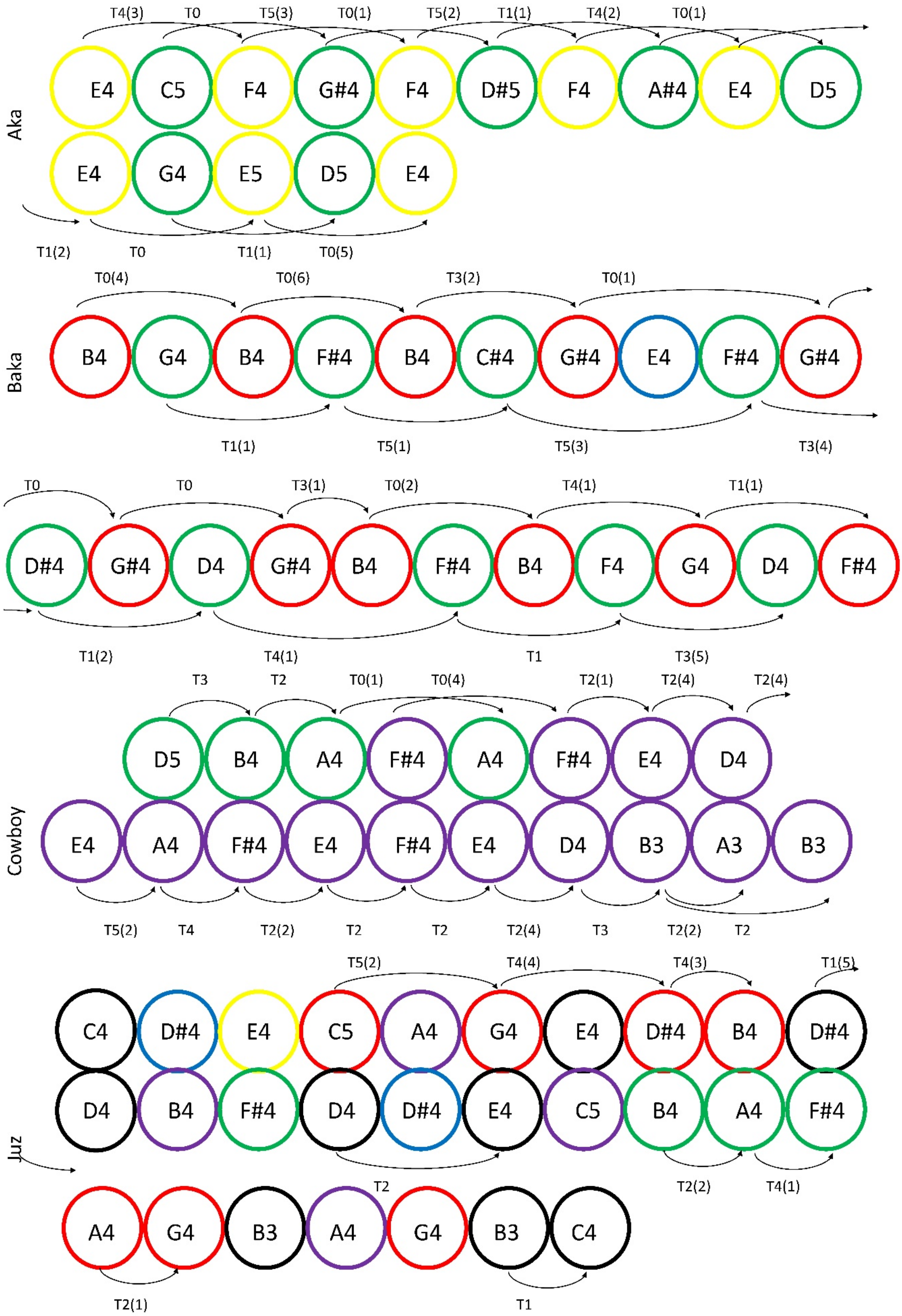
Our consideration of tone color began with the delineation of spectral segments and the characterization of the internal elements within these segments as spectral sets. Now, we will examine the way successive spectral sets are connected and transformed using a combination of transposition and/or the subset/superset relations defined above. Figure 8 demonstrates the succession of spectral sets in each of the four analytical samples. In reading the graph, the following criteria apply: each circle represents a numbered spectral set; and sets sharing the same vowel color are related by the available spectral transformations adumbrated previously: identity (T0); literal subset/superset (T0(y)); transposition (Tx); transposed subset/superset (Tx(y). If two spectral sets do not share the same primary vowel color, they do not share a timbral correspondence, although they may be related by other pitch and frequency correspondences.

The Baka excerpt consists of a large introductory phrase segment containing 21 spectral sets (see transcription). Pitch drift and flexible intonation result in two available forms of some pitches: C#4, D#4 (D4) E4, F#4 (F4) G#4 (G4), and B4. Only two vowels (/i/ and /e/) are employed with a single instance of a third vowel (/a/) in set 8. The distribution of syllables features a strict alternation of syllables beginning with i in spectral set 1. The alternation of the two vowel colors, juxtaposed in strict alternation, provides a sense of circularity and continuance. The same primary vowel colors in the same order each time, but with variances in terms of cardinality, height, fundamental, pitch class and octave registration, provide a variety of available shadings. There are two instances where the strict alternation is temporarily suspended: between sets 7 and 9, where the single occurrence of the *e* syllable denotes the juncture between phrase segments, and sets 14 and 15, which also break the pattern with the repetition of an I-syllable.

The Aka excerpt has 15 spectral sets and, like the Baka excerpt, employs two contrasting vowel colors distributed in strict alternation. Odd-numbered spectral sets use the vowel /e/, whereas even-numbered sets employ /y/. In addition to sharing the same vowel, observe that the odd-numbered sets also share a common fundamental pitch (E4) which functions as a type of melodic pedal point throughout. As a result of pitch drift, a characteristic feature of this performance practice, the pitch E4 in sets 3, 5, and 7 appears notated as F4, the actual sounding pitch. As before, variances in terms of set cardinality, pitch, register, and intonation produce multiple available shadings of these two primary vowel colors.

The excerpt from "One Last Yodel" consists of a single repeated phrase. Melodic contour, durational accent, and the presence of a rest divide the phrase into two corresponding parts designated subphrase A and subphrase B. The former has eight spectral sets, the latter ten. Throughout the yodel, Walser employs only two vowel colors. The y-syllable colors the beginning of subphrase A (sets 1–5) whereas the remainder employs the darker vowel /u/ exclusively for the sustained portion of the phrase and cadence. The pitch collection consists of a B minor pentatonic scale: A3 B3 D4 E4 F#4 A4 B4. Together, there are a total of 16 spectral transformations expressed between consecutive sets sharing the same vowel color. Of these, ten transformations feature transposition at T2 which are either crisp with identical cardinalities or feature subset/superset relationships involving small distinctions in set density. The T2 transpositions reflect the stepwise motion of the melody and promote a consistent degree of change or transformational distance between successive sets; the use of the same vowel color establishes timbral uniformity and consistency.

Strongly contrasting with all three of the previous samples is the performance of "Juz," by Paul Fetz. The performance consists of an 8-measure symmetrical period; our analysis will consider the initial four measure antecedent phrase (mm. 1–4, see figure 3) which divides at the midpoint into two subphrases, designated subphrase A (mm. 1–2) and subphrase B (mm. 3–4); the former containing 13 spectral sets; the latter 14. The sheer breadth and scope of the available timbral palette makes for a striking contrast with the three previous samples. The performance of "Juz" features six vowel colors, (/o/, /i/, /u/, /y/, /a/, and /e/). Each of these is available in numerous shadings, hues, and intensities, based on discrepancies involving cardinality, vowel height, pitch, frequency, as well as other nuanced acoustical considerations. Also, unlike the other excerpts, there is no patterned distribution of vowel colors; instead, they appear juxtaposed and freely mixed, resulting in a kaleidoscopic succession of vowel colors, creating diverse pathways through the spectrum of available vowel colors. Due to the low incidence of vowel repetition, there are few timbral pockets where the same vowel remains in play for any duration, excepting sets 8–9 and 21–22 where the acute vowel /i/ occurs twice in succession.



Colors for vowels: yellow=[e], green=[y], red=[i] blue=[a], purple=[u], black=[o]

Figure 8. Spectral transformations for each of the four analytical samples.

Characterizing spectral morphology

So far, we have adopted a detailed, taxonomic approach to the analysis of vowel tone color — one that examines the vertical organization of harmonics above a common fundamental, organizes these into sets, and assigns descriptive labels for the purpose of inventory and comparison. Now we will engage a different perspective, one that attempts to understand how the sonic details characteristic of each successive moment change over time. A central feature of Cogan's early analytical approach involved the use of 13 sonic oppositions, or "features," borrowed from the linguist Roman Jakobson's research in phonology. The collection of 13 oppositions defines a unique feature space. Instances of these various oppositions result in feature complexes or "vectors." Cogan segments timbral surfaces into successions of discrete moments, each of which is described as a combination of features and assigned a corresponding value: "+" positive or "-" negative. Low levels of activity are assigned negative status; high activity levels, positive. Where there is a combination of both, the designation "+/-" or "mixed" is applied. In addition, where one of the 13 oppositions is not relevant, he uses zero to identify them as "neutral."[14]

The premise of this type of feature-vector analysis is that by viewing an object from multiple vantage points, each emphasizing a different feature or opposition, one could begin to make sense of an object's essential structure. When it comes to feature analysis, the more the better. Each distinctive feature or oppositional pairing reveals a single cross-section. When these individual sections are multiplied and combined, the resolution of the composite image comes into increasing exposure. The more features or oppositions employed, the more analytical perspectives are afforded, resulting in the underlying structure becoming increasingly articulate. Cogan states that the number of oppositional pairings is not fixed, and others can be formulated or deleted from the table depending on the context. Once the initial inventory of binary values has been accomplished, Cogan simply adds the corresponding values in each column, the total values of which are considered as indicative of the corresponding degree of structural intensity characteristic of each discrete sonic moment.

Cogan's method provides a simple but elegant method for characterizing and modeling aspects of spectral morphology, given the severe lack of available software technologies at the time. While successful in terms of its general descriptive power facilitating a means in which to inventory, catalogue, and describe the variety of spectral contexts available within the range of our sonic experience, a more sophisticated methodology is required if issues of spectral morphology and transformation are to be pursued further and with greater scientific accuracy and awareness. A central issue of feature analysis in general is that the various oppositional pairings are rarely equal in terms of their respective perceptual weight and significance of how they impact our sonic experience. Some features play a much stronger, more formative role, and impact our perceptions in more meaningful ways than others, depending on the context. Other complexities arise when one considers that it is nearly impossible to reverse engineer Cogan's analytical model in any way and arrive at anything vaguely resembling the original

14. The fact that Cogan's approach allows for mixed (+/-) and neutral values means it is not strictly binary.

sound source. Certainly, one important measure indicating the general utility of any analytical tool used for timbral analysis is reflected in how much of the original acoustic signal can be obtained by reverse engineering the analytical process.

The primary characteristic that distinguishes our approach to morphology as distinct from Cogan's table of binary oppositions is that in our methodology the different features of sonic design are conceptualized not as simple binary oppositions but rather as a spectrum of incremental gradations or values, each of which can be measured and evaluated. We can thus provide a more detailed characterization of how various sonic elements interact on the timbral surface and contribute to spectral morphology.[15] We readily acknowledge that our analytical approach is far from comprehensive and that there are numerous other contributing factors involved in our perception and experience of spectral morphology, but we distinguish these four analytical parameters in particular as important constituents that are universal to sonic design and morphology in all styles and traditions of unaccompanied vocal music. We therefore provide a foundation on which more sophisticated future studies can extend: 1) vowel intonation and modulation within the available space of the IPA; 2) fluctuations in the number of constituent harmonics in each spectral set (cardinality); 3) fluctuations of the respective height or compass of spectral segments as defined by their lowest sounding element, the fundamental pitch, and the frequency of the highest perceptible harmonic; and 4) fluctuations in the respective intensity values (dBFS) of each constituent harmonic, in each spectral set, across the performance. Observe that our consideration of spectral set cardinality encapsulates Cogan's narrow/wide oppositional pairing due to the fact that as set cardinality increases, so does a set's respective spectral width (in terms of its visual display in the spectrogram). In addition, Cogan's opposition rich/sparse is also summarized because as set cardinality increases so does the set's inherent richness as determined by the number of sounding components within a spectral segment. Finally, Cogan's soft/loud opposition is enveloped within our fourth analytical parameter which charts fluctuations in the respective amplitude of harmonics across the performance.

Thus far, we have established that the performer's selection of the corresponding vowel is tantamount to selecting the primary timbral color by which it is distinguished. If the available primary timbral colors are delineated by the available vowel sounds within the greater IPA vowel space, then distinctions in terms of fundamental pitch, octave registration, cardinality, order positions within the harmonic series, among other considerations result in establishing the extreme variety of distinctive spectral shades, hues, and intensities available for each vowel instance.

Phonologists class vowels into three categories of intensity: acute vowels such as /i/ and /e/ which are bright, and rich in overtones; neutral vowels such as /a/; and grave vowels

15. Denis Smalley's concept of *spectromorphology* bears some similarities to aspects of our analytical method in that we both consider the way spectral intensities change over time as descriptive of spectral morphology (Smalley 1997). Whereas our focus involves applied analysis of vocal music, Smalley's motivation concerns the development of listening strategies and compositional techniques for computer and electro-acoustic music. Smalley (1997) characterizes his work as follows: Spectromorphology is not a compositional theory or method, but a descriptive tool based on aural perception. It is intended to aid listening and seeks to help explain what can be apprehended in over four decades of electroacoustic repertory: how composers conceive musical content and form — their aims, models, systems, techniques, and structural plans.

which are dark, such as /o/ and /u/. The vowels which comprise the IPA vowel space can be conceptualized as a spectrum of available vowel colors and intensities arranged from the darkest, grave vowels on one end to the brightest, acute vowels on the other. The individual IPA vowel plots for each spectral set in each of the four analytical examples can be combined and characterized as successive coordinated transformational pathways through the available IPA vowel space.

We can not only characterize the exact location of any vowel within the space with technical precision and map distances between instances of successive vowels, but also approximate fluctuations in the respective degrees of vowel intensity, thus providing an important tool for characterizing the spectral morphology of successive vowel instances as change over time. If we can identify the vowel sound, we are hearing the primary tone color for an unaccompanied vocal utterance, be it speech sound or song. We perceive the transformations of successive vowel sounds throughout the performance; we can map each instance as a discrete point within the IPA vowel space and demonstrate the variety of transformational pathways each performance sample inscribes within the available IPA vowel space, the resultant image considered descriptive of spectral morphology and the successive transformations of vowel intensities over time.

The locations and movements of the melody within two-dimensional vowel space, defined by the two relevant formants, can be traced in video clips made for each sample. In the video, the spectrograph with the current segment marked is shown on the left side, the formants with the intersection of the vertical (formant 1) and horizontal line (formant 2) indicating the exact location for the given segment. We can then isolate and project these formant-boosted frequencies onto the IPA vowel space model to determine their precise position of each vowel sound within the IPA vowel space as determined by the frequencies of the first and second order formant frequencies.

Sample	**Link**
Don Walser — One Last Yodel	https://youtu.be/SMEsxYtLeXQ
Aka — Bobangi	https://youtu.be/9zOa97LosLQ
Baka — Etude des Jodls	https://youtu.be/dMxKXF22MVg
Paul Fetz — Juz	https://youtu.be/1kDbmhqjDiE

Table 2. Links to the videos depicting the formants and the vowels moving along with the melodies.

Moreover, we can also consider not only the distribution of vowels — we can also model the varying fluctuations in vowel intensity between them, in terms of the three regions of vowel intensity: the grave, neutral, and acute regions of vowel space. From figure 9 it is clear that the four yodeling practices differ clearly in terms of how their respective vowels are positioned within the IPA vowel space. Although some vowels, such as /u/, /i/, and /e/, occur in multiple samples, they still show up in different areas. The acute vowels /e/ and /i/ are located in the high frequency range, in which the example of the Baka "Yodel Etude" moves (green). The Aka piece, "Bobangi," moves also in two separate vowels, /y/ and /e/, but with decisively lower formant frequencies. Note in the following table 3 that in "Bobangi," e is the brighter vowel (compared to /y/) but in the Baka "Etude des Jodls," it is the graver vowel (compared to /i/). In contrast to both are the darker timbres in Don Walser's "One Last Yodel," in which he vocalizes mostly with /u/, found in blue in the lower left corner of the graph. The variable selection of vowels of the Austrian "Juz" is represented by the scattered purple dots across vowels /y/, /u/, /a/, and /o/. In all four samples, vowels are chosen in correlation (but not necessarily exact consistency) with the change between head voice and chest voice. The "Baka" and "Cowboy" samples use grave vowels for lower and acute vowels for higher notes. The Aka sample shows the reverse division; "Juz" mixes types of vowels with registers.

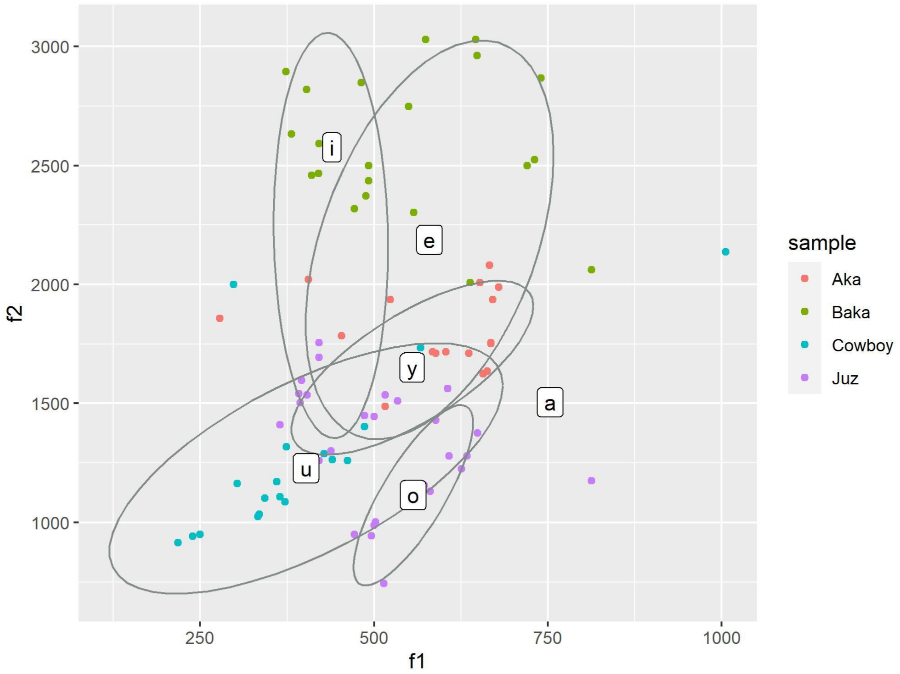


Figure 9. Locations of the spectral segments in vowel space in all four samples. Vowel labels: mean values for the given vowel across samples. Ellipses: one standard deviation.

Sample	Chest Voice	Head Voice
Aka	**e**	**y, (e)**
Baka	**e**	**i**
Cowboy	**u**	**u, y**
Juz	**o, a, e**	**u, i, y**

Table 3. Vowels that are relatively grave within the respective sample are in blue, while those that are relatively acute are red. Vowels are written in the international phonetic alphabet. (Rarely used vowels are in brackets.)

Modeling selected features of spectral morphology

Our second consideration, spectral density (i.e., cardinality), indicates the respective weight of a spectral set contingent on the number of elements in the set. Increases in set cardinality represent progressive or increasing intensity fluctuation values, while the reverse represents decreasing, or recessive intensity fluctuations. Mapping successive changes in set cardinality across the performance reveals the organization of an important consideration regarding spectral morphology. Similarly, consideration of our third analytical parameter, spectral height, refers to the overall compass of the spectral set as measured from the lowest sounding frequency (i.e., the fundamental), to the highest sounding harmonic frequency. In unaccompanied vocal music there is a direct correlation between the number of sounding harmonics (cardinality) and the corresponding spectral compass of the set, given that the spectral sets usually increase or decrease membership in alignment with the respective order position of harmonics within the harmonic series. A spectral set with a cardinality of five will typically consist of the first five order positions of the corresponding harmonic series; a seven-note set will use the order positions 1–7 of the corresponding series, and so forth.

On rare occasions when one or more harmonics are missing due to masking or other interference, an incomplete set results. Whereas complete sets sound full, incomplete sets sound progressively hollow as the number of missing elements increases. In unaccompanied vocal music there is usually a direct reciprocal correspondence between the number of sounding harmonics and the corresponding spectral height of the set, since spectral sets usually increase or decrease members in alignment with order position of harmonics in the harmonic series. Both cardinality and spectral compass are different measures for a similar aspect of timbre and are interdependent, as demonstrated in example 11. Consequently, we can rely on one of the measures in the following.

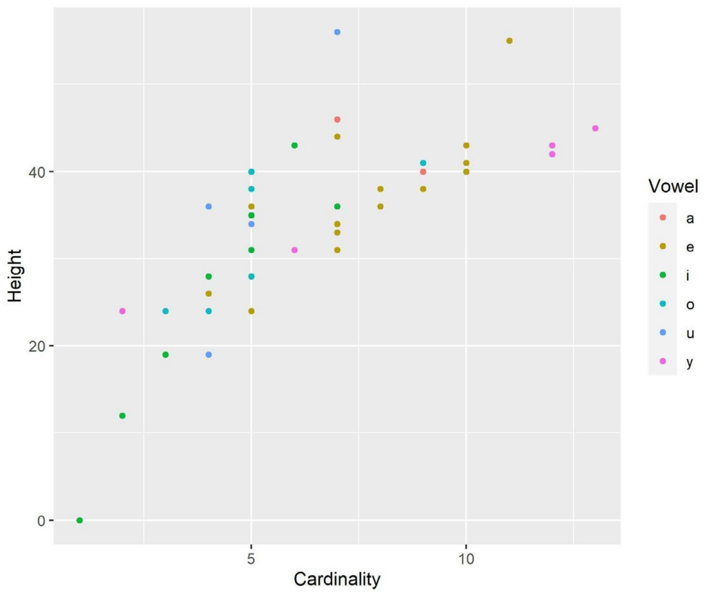


Figure 10. Cardinality (x-axis) and height (y-axis) based on all samples, divided into vowels (color). Multiple values at the same location are represented by only one dot.

In the following example 12 we can observe the intersections between vowels corresponding with cardinality and spectral height: /y/ is expressed with very high overtones which result in a high cardinality/height. However, the spread for vowels is large and there is significant overlap among all vowels. This means that aspects such as pitch and loudness also have an impact on cardinality and height. Nonetheless, the graph indicates that vowels /y/ and /e/ result in higher cardinalities and heights than /i/ and /o/. We can further demonstrate how spectral height fluctuates in both agreement and contrast with vowels. The change between high and low registers may coincide with fluctuations in the spectral height, but this is not consistently the case. Lines in example 12 are colored to represent vowels and show the change between the current vowel and the next vowel.

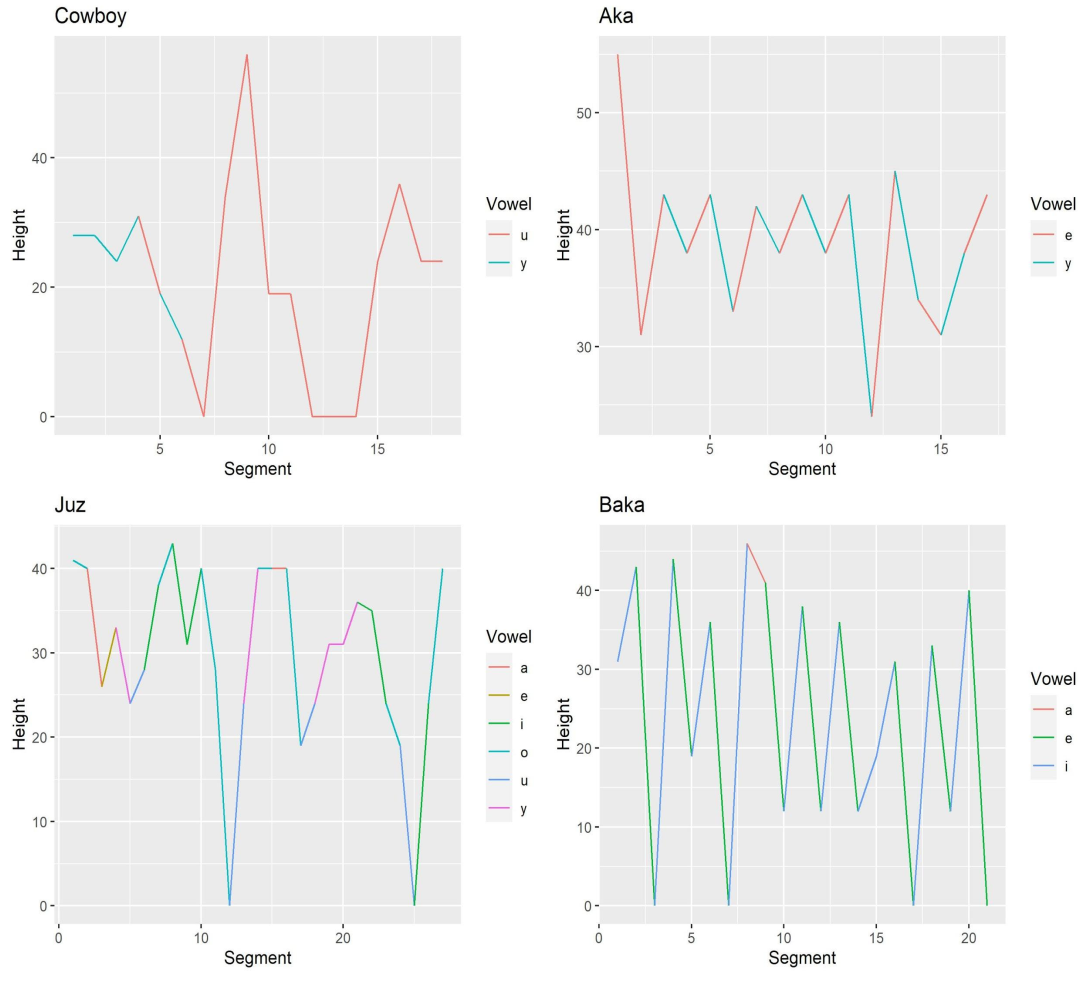


Figure 11. Spectral compass for the melodic flow in all four samples.

Thus far, our appreciation of spectral sets has been primarily static: identifying the various harmonics which comprise a set and characterizing these as spectral sets for the purpose of inventory and comparison. Now, we consider how the intensity values of each harmonic in each spectral set change over time. Figure 12 shows the changes in intensity during melodic progression for order positions 1–3 of the harmonic series.

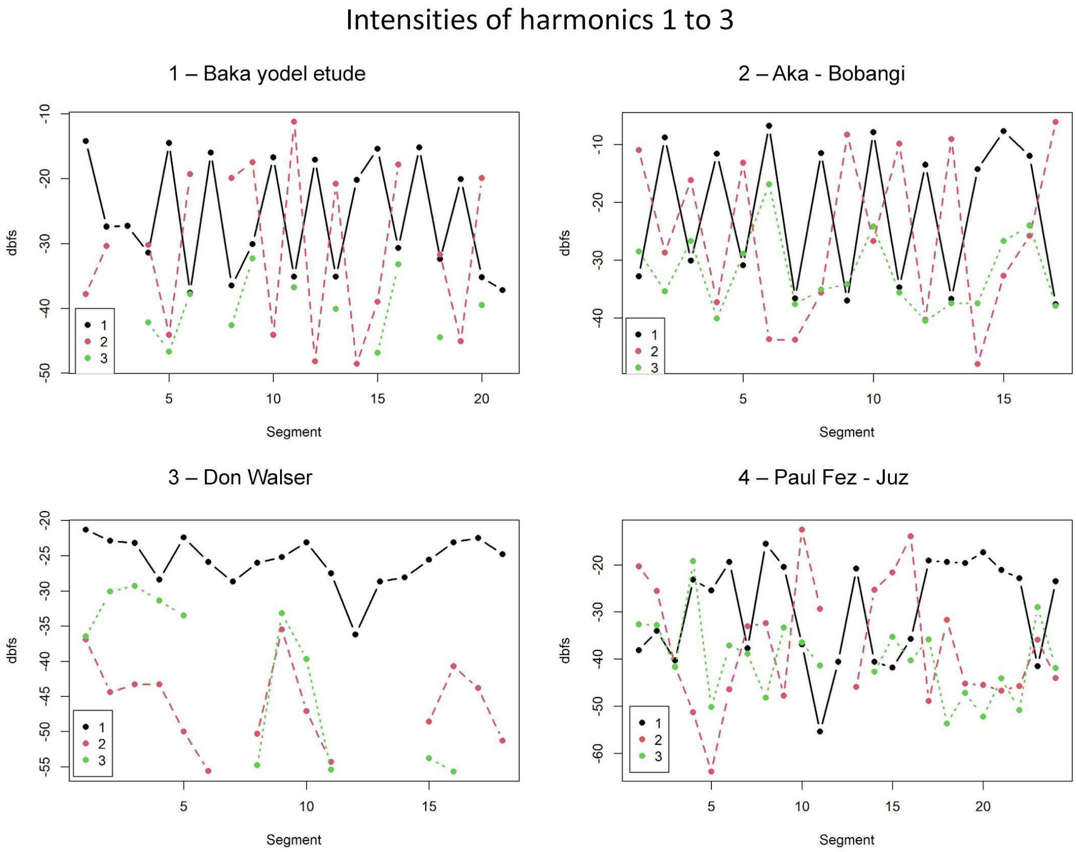


Figure 12. The intensities of harmonics one (black), two (red), and three (green) over the course of the melody (notes numbered on the x-axis).

The intensities of harmonics one and two need to be discussed, as they alternate as the most intense frequency in three of the samples. This implies that the fundamental frequency (harmonic 1, black in figure 12) is not always the most intense frequency but is still, in all present cases, the perceived pitch. This problem will not be exclusive to yodeling but is prominent in three samples, the Baka yodel etude, Bobangi, and Juz. In contrast, in "One Last Yodel," the fundamental throughout coincides with the loudest frequency. However, the sample demonstrates another feature in that harmonic three is more intense than harmonic two in the first few segments. Harmonic three follows neither but outlines the morphology that is the tendency of all subsequent upper harmonics, which is why it is represented in this figure. The switch between the two lowest harmonics occurs in different ways. In both the Aka and Baka samples, there are sections where the switch occurs between each segment. This coincides with a near-binary change of vowels (/i/ and /e/, or /e/ and /y/), demonstrating the large distance between the two voice registers interlocking. This strong contrast between two sets of vowels and their rapid changes is found in neither the Alpine European ("Juz") nor the American ("One Last Yodel") sample.

The most comprehensive representation of spectral morphology can be viewed in 3D space. We designed three-dimensional surfaces, where intensity (z-axis) is shown based on time (x-axis) and harmonics (y-axis). Figure 13 shows a screenshot of the surfaces of all four samples. These are the most comprehensive visual representations of the present data. Colors in the 3D graphs are like the color coding commonly used in spectrographs, with short-wave colors for higher intensities and long-wave colors for lower intensities.[16] This shows how individual harmonics peak in different notes and that there are vast differences in the overtone spectrum depending on the individual segment. Cardinality fluctuates between two and twelve audible harmonics.

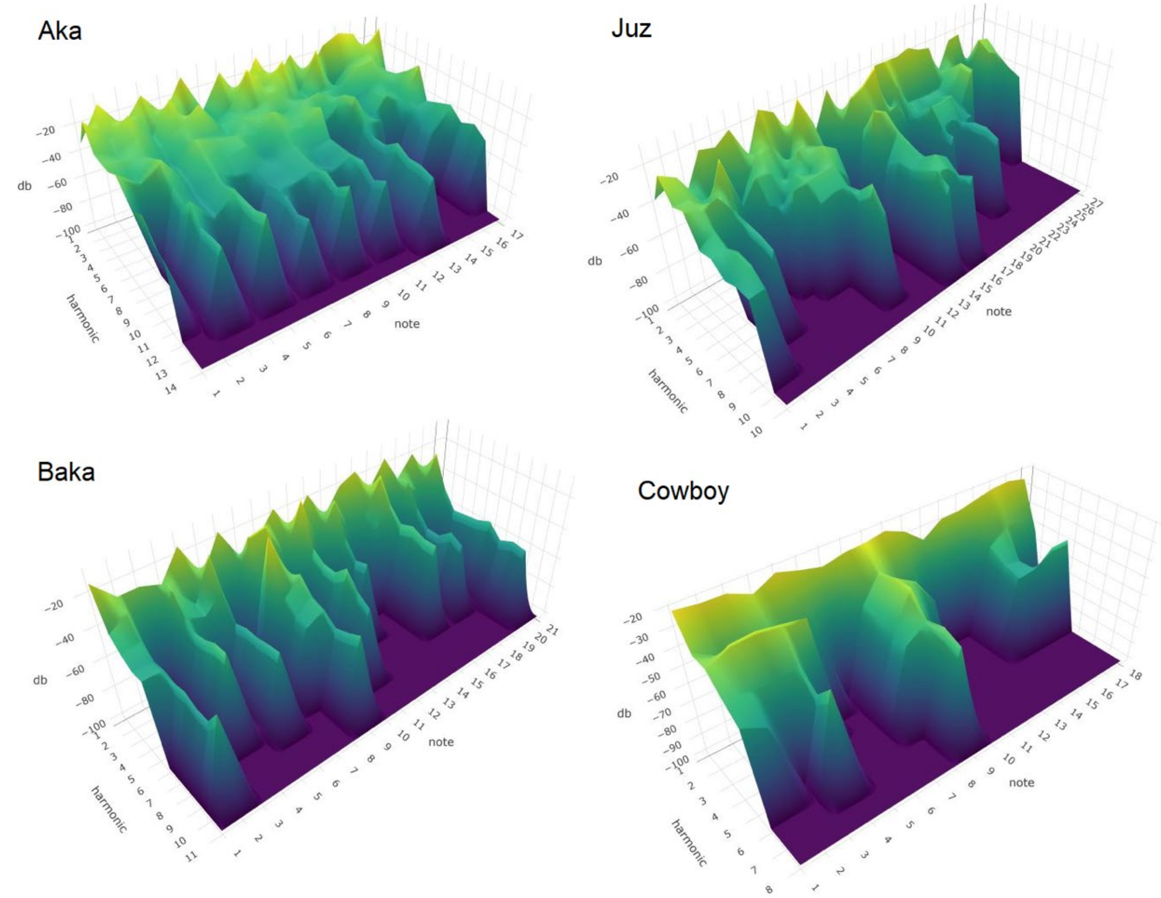


Figure 13. 3D surfaces of the timbral morphology in the four examples.

Three-dimensional graphs of the spectral morphology for each sample can be accessed through the links in the table below. The graphs are interactive, allowing the viewer to zoom, rotate, and move the object, thereby inspecting segments and harmonics in detail. The surfaces are self-contained HTML files and are included in the supplementary data.[17] There, one can inspect the details of spectral morphology far beyond binary oppositions.

Conclusions

Of all parameters of musical production and design, timbre is the least understood and least studied. Yet it is at the fore of most people's initial response to a musical stimulus, alongside pitch and rhythm. The general deficit of studies focusing on timbre in the history of western analytical circles stems primarily from two considerations. First, in the vast

16. We substituted missing values with a baseline of -100 (i.e., values below the threshold) for technical reasons.
17. Supplementary data: https://doi.org/10.5281/zenodo.5607754.

majority of western art music timbre functions in a secondary, ornamental capacity largely subjugated to the privileged positions accorded to pitch and rhythm. Since western music did not emphasize timbre as a formative variable of musical production, other musical cultures that did were historically deemed primitive or otherwise inferior and underdeveloped. And second, while performers and listeners are keenly sensitive to even the most subtle contrasts in timbre, due to the lack of both appropriate software technologies as well as conceptual models, timbre remains one of the few unexplored facets of musical experience and design.

This study proposed a small assortment of analytical tools and strategies useful for characterizing various facets of sonic design and musical timbre in unaccompanied vocal music. The analytical method developed herein provides a means by which to inventory, catalogue, and compare the wide diversity of sonic contexts in various global yodeling traditions. Additionally, it provides the analyst with the requisite tools to map and model the manner in which various sonic phenomena ultimately responsible for timbral definition change over time. We consider the latter to be descriptive of spectral morphology. The utility of our analytical toolkit is that it provides an ethnomusicologist with average technical skills a systematic method to explore the timbral organization of all forms of unaccompanied vocal music, regardless of historical period, culture, genre, or style, using computer software readily available online.

The initial stage of examination involved the parsing of the timbral surface into a succession of spectral segments; each segment contains a fundamental pitch, all occurring overtones, and the associated vowel. The next step involved aural verification protocols to confirm the aural salience of constituent elements, and establishment of the perceptual threshold used to indicate the approximate aural acuity of individual spectral elements. All sounding phenomena existing above the established perceptual threshold are bundled together and characterized as a spectral set, representing the basic building blocks of timbral surfaces. The next step explored a variety of timbral correspondences available between spectral sets predicated on their associated vowel types as well as their respective set cardinality, height, fundamental pitch, and octave registration. Consideration of the variances observed in the latter account for the wide diversity of timbral shades, hues, and intensities available for each spectral set sharing the same basic vowel color. It was demonstrated that the transformations that connect successive spectral sets with the same primary vowel coloring are related by transposition and/or the inclusion relation as subsets/supersets. Instances of specific vowel sounds were examined to extract their low order formant frequencies, which were then mapped onto their overlapping harmonic frequencies in the visual display of the spectrograph and projected within the spectrum of available vowel colors contained in the IPA vowel space.

Preliminary inquiries involving spectral morphology were then explored based on the transformational pathways inscribed within the IPA vowel space by the succession of individual vowel plots for each spectral set in each of the four analytical samples. Additional considerations modeled morphology not only in terms of vowel succession, but also in terms of fluctuations in the respective intensity of each harmonic in each set over time, as well as changes in spectral density, height, frequency, and octave registration.

There are many more nuanced, detailed considerations involving spectral morphology that we do not include. We focused on these specific features of sonic design to provide a foundation from which more detailed, specialized studies may extrapolate.

The detailed analysis and comparisons between the four samples reveal insights into differing and overarching features of timbre in various yodeling practices. Our analysis demonstrates that while in the Aka and Baka samples, both performances feature the rapid binary alternation of available vocal registers — upper "head" voice juxtaposed with the lower "chest" voice — yet each is distinguished by their respective palette of available vowel colors. Consideration of the process by which performers select vowels for specific instances are fascinating cultural and performance-practice questions simply beyond the limited range of this brief analytical study. Regardless of how, or why, or when the specific vowels occur, our method provides a manner in which to characterize and map the unique timbral designs that result as a consequence of the performer's selection process.

The choice of vowels is structural and plays a defining role, but is only an approximate indicator of timbre. Vowels that, for example, transcribe to /e/ in the phonetic alphabet may be shaped by different formant ranges, as demonstrated in figure 9 for the Aka and Baka samples. While the transcription suggests a congruent tone color, the present analysis reveals distinctions, albeit not perceived by our categorical hearing. A melodic transcription of the samples "Bobangi" and "Etude des Jodls" could be interpreted as highly similar because of the almost binary alteration between high and low notes. However, this would be a pitch-centered analysis; our analysis of the timbre shows a crucial difference in that the use of the (relatively) acute and the grave vowel, and their associated vowel colors, are reversed.

Historically, cowboy yodelers were inspired by traveling Austrian singing groups, but they developed their own technique and style. In terms of tone color, we assume that the coloring of the vowels was adapted to their own language and the expression of their own folk songs. This partly explains the independence of the two samples from the USA and Austria. Furthermore, the "Juz" from the Vorarlberg region does not represent the same kind of yodeling as that of the aforementioned singing groups from the 19th and 20th centuries, but rather is an expression of a regional tradition. While Paul Fetz uses a broad palette of vowels and consonants, Don Walser stays in a relatively consistent timbral space, and switches to chest voice only for brief segments. In this particular example of a cowboy yodel, fast changes between voice registers are not as significant as they are in some well-known songs of the genre. Instead, the singer employs a mostly grave timbre to shape a lyrical tune.

As for future directions, our results add to the ongoing discussion of the history, functions, and aesthetics of yodeling as a distinct singing technique present in various musical traditions. Although ours is a cross-cultural, comparative study, our scope is fairly restricted, including only four representative samples of the numerous yodeling traditions currently practiced worldwide. Our foundational study is best conceptualized as a springboard from which future, more specialized studies may further develop and extrapolate, refining the analytical method and associated tools we have developed and expanding the breadth and scope of analytical samples employed. An automation of some of the steps used here is desirable to enable the study of larger samples. However, it would have to account for all the selective processes described to retrieve viable data. We were restrained to monophonic examples which are not representative of mostly multi-part vocal music. An extrapolation of our methods to multi-part music and even combinations of vocal and instrumental music will be ultimately needed to further our understanding of timbre as a musical condition of equal importance to pitch and rhythm. The majority of the analytical techniques and strategies developed here have their origin in more universal, acoustic studies of vowel color and transformation in speech sound and music. Indeed, by restricting our analytical investigation to consideration of those timbral features and qualities which are common to all styles of unaccompanied vocal music, we are able to develop a basic matrix for cross-cultural comparative study. Consequently, the results of our analytical study become truly meaningful and significant only when combined with other complementary approaches and ultimately contextualized within the broader purview of pertinent historical, ethnographic, and other cultural considerations.

REFERENCES

Abbott, Lynn, and Doug Seroff. 1993. "America's Blue Yodel." *Musical Traditions* 11: 2–11.

Ambrazevičius, Rytis. (forthcoming). "Some Features of Austrian Yodeling." *Proceedings of the Xth Symposium on Traditional Polyphony.*

Ashby, Michael. 2005. *Introducing Phonetic Science*. Cambridge: Cambridge University Press.

Cogan, Robert. 1984. *New Images of Musical Sound.* Cambridge MA: Harvard University Press.

Cohen, Dalia, and Ruth Katz. 1979. "The Interdependence of Notation Systems and Musical Information." *Yearbook of the International Folk Music Council* 11: 100–113. https://doi.org/10.2307/767567.

Coltman, Robert. 1976. "Roots of the Country Yodel: Notes Toward a Life History." *John Edwards Memorial Foundation Quarterly* 12(41): 91–94.

Dreier, Wolfgang. 2007. "Jodeln nach Noten? — Schade drum! Annäherungen an ein hochkomplexes Musikkonzept am Beispiel historischer Aufnahmen." *Vierteltakt* 3: 12–16. http://wolfgang-dreier.at/wp-content/papercite-data/pdf/dreier2007a.pdf.

Fales, Cornelia. 2002. "The Paradox of Timbre." *Ethnomusicology* 46(1): 56–95. https://doi.org/10.2307/852808.

Fenk-Oczlon, Gertraud, and August Fenk. 2009. "Musical Pitch in Nonsense Syllables: Correlations with the Vowel System and Evolutionary Perspectives." *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland*, 110–13. https://jyx.jyu.fi/bitstream/handle/123456789/20863/urn_nbn_fi_jyu-200941l246.pdf.

Fenk-Oczlon, Gertraud. 2017. "What Vowels Can Tell Us about the Evolution of Music." *Frontiers in Psychology* 8(1581). https://doi.org/10.3389/fpsyg.2017.01581.

Fink-Mennel, Evelyn. 2007. *Johlar und Juz. Registerwechselnder Gesang im Bregenzerwald.* Graz and Feldkirch: Wolfgang Neugebauer Verlag.

Fürniss, Susanne. 1992. *Die Jodeltechnik der Aka-Pygmäen in Zentralafrika*, Berlin: Dietrich Reimer Verlag.

———. 2006. "Aka Polyphony: Music, Theory, Back and Forth." *Analytical Studies in World Music*, edited by Michael Tenzer, 163–204. https://halshs.archives-ouvertes.fr/halshs-00453689/document.

———. 2016. "What Is a Part? Polyphony between Perception and Conception." *Res Musica* 8: 73–84. https://halshs.archives-ouvertes.fr/halshs-02387100/document.

Graf, Walter. 1965. "Naturwissenschaftliche Gedanken über das Jodeln. Die phonetische Bedeutung der Jodelsilben." *Schriften des Vereins zur Verbreitung naturwissenschaftlicher Kenntnisse in Wien* 105: 2–25.

Graf, Walter. 1975. "Sonographische Untersuchungen, mit Beispielen aus dem deutschsprachigen Volksgesang." In *Handbuch des Volksliedes*, edited by Rolf Bredning, Lutz Röhrich, and Wolfgang Suppan, 583–623. Munich: Wilhelm Fink.

Grauer, Victor. 2006. "Echoes of Our Forgotten Ancestors." *The World of Music* 48(2): 5–59.

———. 2008. "Some Notable Features of Pygmy and Bushmen Polyphonic Practice, with Special Reference to Survivals of Traditional Vocal Polyphony in Europe." *Proceedings of the Fourth International Symposium on Traditional Polyphony,* 15–19 September 2008: 147–57. https://drive.google.com/open?id=1VfVfV_-3vGPePop4nq1WOg92KvFIvHzw.

———. 2009. "Concept, Style, and Structure in the Music of the African Pygmies and Bushmen: A Study in Cross-Cultural Analysis." *Ethnomusicology* 53/3: 396–424.

Haid, Gerlinde. 2008. "Klänge gegen Naturgefahren." *alpine space – man & environment* 4: 55–72.

Hupfauf, Sandra. 2016. *Die Lieder der Geschwister Rainer und «Rainer Family» aus Zillertal (1822–1843): Untersuchungen zur Popularisierung von Tiroler Liedern in Deutschland, England und Amerika* (Schriften zur musikalischen Ethnologie 5), Innsbruck 2016.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge University Press.

Kisliuk, Michelle. 1998. *Seize the Dance! BaAka musical life and the ethnography of performance.* New York: Oxford University Press.

Lach, Robert. 1928. "Die Tonkunst in den Alpen." In *Die Österreichischen Alpen. Eine zusammenfassende Darstellung*, edited by Hans Leitmeier, 332–80. Leipzig and Vienna: Franz Deuticke.

Nikolsky, Aleksey. 2020. "The Pastoral Origin of Semiotically Functional Tonal Organization of Music." Frontiers in Psychology 11:1358. http://dx.doi.org/10.3389/fpsyg.2020.01358.

Olivier, Emmanuelle (ed.). 1997. *Namibie. Chants des Bushmen Ju'hoansi.* [S.l.]: Ocora.

Plantenga, Bart. 2004. *Yodel-Ay-Ee-Oooo: The Secret History of Yodeling Around the World.* New York: Routledge.

Reuter, Christoph. 2004. "Von der 'Physik der Klangfarben' zur 'Psychologie der Klangfarben." In *Musikpsychologie. Jahrbuch der Deutschen Gesellschaft für Musikpsychologie 17: Musikalische Begabung und Expertise*, edited by Klaus-Ernst Behne, Günther Kleinen, and Helga de la Motte-Haber, 109–125. Göttingen: Hogrefe. http://dx.doi.org/10.23668/psycharchives.2987.

Roeder, John. 2019. "Timely Negotiations: Formative Interactions in Cyclic Duets." *Analytical Approaches to World Music* 7(1): 1–19. http://aawmjournal.com/articles/2019a/Roeder_AAWM_Vol_7_1.pdf.

Rouget, Gilbert. 2011. "Musical Efficacy: Musicking to Survive — the Case of the Pygmies," translated by Margaret Buckner, *Yearbook for Traditional Music* 43: 89–121. https://doi.org/10.5921/yeartradmusi.43.0089.

Shuster, Lawrence. 2012. "Spectral Sets, Transformations, and Morphologies in Tuvan Khoomei." Paper presented at the Society for Music Theory 35th Annual Meeting, New Orleans, 1–4 November 2012.

Sichardt, Wolfgang. 1939. *Der alpenländische Jodler und des Ursprung des Jodelns.* Berlin: Hahnefeldt.

Stomberg, Isak, and Jennie Tiderman-Österberg. 2021. "Duets With Nature: How Natural Acoustics Affect the Experience of Performing Nordic Herding Music in Outdoor Settings." *AAWM Music and Nature* 1: 11–67.

Tiderman-Österberg, Jennie. 2020. "Why Sweden's Ancient Tradition of Calling Home the Herds Is Women's Work." *Smithsonian Magazine*, 28 September 2020, https://www.smithsonianmag.com/smithsonian-institution/why-swedens-ancient-tradition-calling-home-herds-womens-work-180975904/.

United Nations Development Programme — OAI, Social and Environmental Compliance Unit. 2020. *Investigating allegations of non-compliance with UNDP social and environmental commitments relating to the following UNDP activities: Integrated and Transboundary Conservation of Biodiversity in the Basins of the Republic of Congo, TRIDOM II. (October 2017 – March 2023).* https://info.undp.org/sites/registry/secu/SECU_Documents/SECU0009_Draft%20Investigation%20Report_For%20Public%20Comment0ddf041323354a9ca3864d50de9970b7.pdf.

Walcher, Maria. 2014. "Singen und Jodeln — Anmerkungen zu den Potentialen von traditioneller Musik im Heilungsprozess." Paper presented at the Tag der Erfahrungsmedizin, Basel, 20. September 2014.

Weig, Doerte. 2018. "Resonating with Different Worlds: How Baka Musical Practices Generate Sociality, Identities and Connection to Ritual Spirits." In *Making Music, Making Society*, edited by Josep Martí and Sara Revilla Gútiez, 191–216. Newcastle: Cambridge Scholars.

Wey, Yannick. 2019. *Transkription wortloser Gesänge: Technik und Rückwirkungen der Verschriftlichung des Jodelns und verwandter Gesänge im deutschsprachigen Alpenraum*. Innsbruck: Innsbruck University Press. http://dx.doi.org/10.15203/3187-81-8.

———. 2020. "Transformations of Tonality: A Longitudinal Study of Yodeling in the Muotatal Valley, Central Switzerland." *Analytical Approaches to World Music* 8(1): 144–63. http://www.aawmjournal.com/articles/2020a/Wey_AAWM_Vol_8_1.html.

Wise, Timothy. 2008. "Lullabies, Laments, and Ragtime Cowboys: Yodeling at the Turn of the Twentieth Century." *American Music* 26(1): 13–36. https://doi.org/10.2307/40071687.

———. 2012. "Jimmie Rodgers and the Semiosis of the Hillbilly Yodel." *The Musical Quarterly* 93(1), 6–44. https://doi.org/10.1093/musqtl/gdp030.

———. 2016. *Yodeling and Meaning in American Music*. Jackson: University of Mississippi Press.

Discography

Arom, Simha; International Institute for Comparative Music Studies and Documentation (IICMSD) Berlin (eds.). 1994. *Aka Pigmy Music*. Audivis.

Fink-Mennel, Evelyn. *Johlar und Juz: Registerwechselnder Gesang im Bregenzerwald*. Graz and Feldkirch: Wolfgang Neugebauer Verlag.

Sallée, Pierre; Harmonia Mundi (eds.). 2001. *Gabon. Music of the Bibayak pygmies. Epic cantors*, Harmonia Mundi.

Walser, Don. 2011. *Just Me and My Old Guitar*. Mark Rubin / Rubinchik Records.

Appendix

Examples	**Task**	**Software, R packages**
1, 2, 3, 4	Transcription	MuseScore
5	Spectral data retrieval	Overtone Analyzer
6	Defining formants	PRAAT
8, Video examples	Video Editing	Windows Video Editor, Audacity
15, 3D Plots	3D Surfaces	R, {plotly}
13	ridgeline plots	R, {ggplot2}, {ggridges}
9, 11, 12	scatterplots (vowel plots; cardinality and height)	R, {ggplot2}
14	Line graphs	R, {graphics}

Table 3. Software and R {packages} used for the examples.