



The Pastoral Origin of Semiotically Functional Tonal Organization of Music

Aleksey Nikolsky*

Independent Researcher, Austin, TX, United States

This paper presents a new line of inquiry into when and how music as a semiotic system was born. Eleven principal expressive aspects of music each contains specific structural patterns whose configuration signifies a certain affective state. This distinguishes the tonal organization of music from the phonetic and prosodic organization of natural languages and animal communication. The question of music's origin can therefore be answered by establishing the point in human history at which all eleven expressive aspects might have been abstracted from the instinct-driven primate calls and used to express human psycho-emotional states. Etic analysis of acoustic parameters is the prime means of cross-examination of the typical patterns of expression of the basic emotions in human music versus animal vocal communication. A new method of such analysis is proposed here. Formation of such expressive aspects as meter, tempo, melodic intervals, and articulation can be explained by the influence of bipedal locomotion, breathing cycle, and heartbeat, long before *Homo sapiens*. However, two aspects, rhythm and melodic contour, most crucial for music as we know it, lack proxies in the Paleolithic lifestyle. The available ethnographic and developmental data leads one to believe that rhythmic and directional patterns of melody became involved in conveying emotion-related information in the process of frequent switching from one call-type to another within the limited repertory of calls. Such calls are usually adopted for the ongoing caretaking of human youngsters and domestic animals. The efficacy of rhythm and pitch contour in affective communication must have been spontaneously discovered in new important cultural activities. The most likely scenario for music to have become fully semiotically functional and to have spread wide enough to avoid extinctions is the formation of cross-specific communication between humans and domesticated animals during the Neolithic demographic explosion and the subsequent cultural revolution. Changes in distance during such communication must have promoted the integration between different expressive aspects and generated the basic musical grammar. The model of such communication can be found in the surviving tradition of Scandinavian pastoral music - kulning. This article discusses the most likely ways in which such music evolved.

OPEN ACCESS

Edited by:

Danielle Sulikowski,
Charles Sturt University, Australia

Reviewed by:

Colwyn Trevarthen,
University of Edinburgh,
United Kingdom
Mark Reybrouck,
KU Leuven, Belgium

*Correspondence:

Aleksey Nikolsky
aleksey@braavo.org

Specialty section:

This article was submitted to
Evolutionary Psychology,
a section of the journal
Frontiers in Psychology

Received: 21 November 2019

Accepted: 22 May 2020

Published: 23 July 2020

Citation:

Nikolsky A (2020) The Pastoral
Origin of Semiotically Functional Tonal
Organization of Music.
Front. Psychol. 11:1358.
doi: 10.3389/fpsyg.2020.01358

Keywords: tonal organization, animal communication, honest signal, semiosis, aspects of expression, domestication, kulning vs. yodel, motherese

Abbreviations: AC, animal call (plural ACs); AE, aspect of expression (plural AEs); FF, fundamental frequency (plural FFs); ky, thousand years; kya, thousand years ago; SFTO, semiotically functional tonal organization; TO, tonal organization.

TONAL ORGANIZATION AND MUSICAL MODE

Since antiquity, scholars have been puzzled by the origins of music. Their quest still remains largely unanswered—impeded by the shortage of available data. The current consensus holds that some kind of musilanguage (Brown, 2000) must have preceded the bifurcation of music and language, marking the emergence of behavioral modernity in humans (Cross, 1999). Pitch orientation is seen as the primary structural marker of music, followed by rhythmometric organization (Brown, 2017)¹. This unnecessarily oversimplified view can and should be expanded, since in reality music is organized not in two but in eleven *aspects of expression* (AEs²), each providing its autonomous information channel (**Table 1**):

- Melodic contour,
- Harmony,
- Texture,
- Form/thematicity,
- Tempo,
- Rhythm,
- Meter,
- Articulation,
- Dynamics,
- Register,
- Timbral quality (instrumentation)³.

The problem is that in investigation of music, cognitive scientists rely on “standards” of Western musical theory, produced by Western civilization and therefore specific to certain historic periods and geographic regions. Although Western music system has proved to be the widest spread and the oldest surviving tradition, with its theoretic foundation rooted in the 3rd millennium BC (Dumbrill, 1998; Mathiesen, 1999; Jorgensen, 2003; Christensen, 2008; Crickmore, 2009; Nikolsky, 2016), nevertheless, there are other civilizations that abide by their own musical theories, explicit or/and implicit, documented or/and

orally transmitted (Nettl, 2005). The need to formulate a “meta-theory” applicable to all varieties of musics has been realized only in the 1890s and dealt with by the discipline of systematic musicology (Bader, 2018). However, this discipline too inherited the framework of Western “classical music,” which is just one of many (Nikolsky, 2015b, 2016, 2020; Nikolsky et al., 2020). Since this framework is tailored to incremental frequency changes, the pitch-related AEs have been prioritized in Western musicology, covered by the dedicated disciplines of harmony, counterpoint, and musical form (Christensen, 2008). The other AEs have only recently received attention, after the traditional discipline of musical form was approached semiotically (Bobrovsky, 1978; Mazel, 1979; Ratner, 1980; Nazaikinsky, 1982, 1988, 2013; Lerdahl and Jackendoff, 1985; Berry, 1987; Ruwet and Everist, 1987; Beliayev, 1990b; Molino, 1990; Nattiez, 1990; Aranovsky, 1991, 1998; Monelle, 1992, 2000, 2006; Narmour, 1992; Tarasti, 1994, 1995, 2012; Kholopova, 2002; Arom, 2004; Bonfeld, 2006; Medushevsky, 2010; Tagg, 2012; Turino, 2014; Benjamin et al., 2015; Yust, 2018). Cross-examination of syntactic, pragmatic, and semantic use of conventional musical idioms has revealed that they break into 11 different AEs (**Table 1**). Nine of them are used in monophonic music (without harmony and texture)⁴. Each AE is distinguished by its unique perceptual substrate and idiomatic expressions.

Interspecific comparison of human music to vocalizations of different animal species along these aspects promises a better understanding of the qualitative leap in the emergence of music. The Moscow school of “integrative analysis”⁵ presents a methodology for such interspecific analyses, which I have adapted to identify those typological patterns in AEs of human music that contrast *animal calls* (ACs). These contrasts should be examined to reveal what exactly in human cultural evolution could be responsible for the emergence of new AE patterns that are unique to humans.

Human music is distinguished by its incremental structure (Bresin and Friberg, 2011)—requiring the ability to discriminate

¹Metric organization of rhythm accompanies and supports pitch organization in music (Jones and Large, 1999), jointly supporting the “musical” manner of interpretation of sounds (Huron, 2006). Tonally important pitch-classes are usually stressed by longer durations and/or dynamic accents. However, comparing to rhythm, pitch organization is much more common in known world’s music cultures (found even in music for only percussive instruments, e.g., African talking drums)—there are many forms of music that are characterized by ametric and arrhythmic free timing, but there are very few non-pitch forms. Therefore, pitch organization is a more reliable marker that distinguishes music from language than rhythmometric organization.

²I will use the abbreviation “AE” when speaking of a single aspect of expression, and “AEs” when speaking of multiple aspects of expression.

³The matter of choosing different timbres for different musical expressions has traditionally been handled by the discipline of instrumentation in Western classical music (Banshchikov, 1997). The term “instrumentation” here is somewhat misleading, because it covers not only the qualia of the timbres of musical instruments and their ensembles (trio, quartet, orchestra, orchestral group) but also various types of voices (soprano, tenor, bass), vocal ensembles (duet, trio, choir) and the rules of combining vocals with instruments (Kreitner et al., 2001). Arabic maqam, Persian dastgah, and Indian raga also observe similar rules in their respective practices.

⁴Technically speaking, monophonic music can still engage some idioms that relate to harmony and texture. A melody solo often features a pronounced “harmonic rhythm” (Swain, 2002)—i.e., periodic changes of implied chords (e.g., the “Blue Danube Waltz” theme by J. Strauss Jr.) that can stay regular (as in a metric pulse), be patterned (as in rhythm), or elaborated by expansion or contraction of a pulse period. Monophonic music can also implement changes in texture by patterning a stream of sounds into familiar textural idioms (e.g., the “Alberti figuration” or tremolo on a single tone) which then carry their specific semantic expression, different from other textural components, such as a melodic theme (Skrebkova-Filatova, 1985). However, overall, harmony and texture play a secondary role in monophonic compositional practices, limited to Western classical music alone.

⁵The “integralist school” of structural analysis of music was founded by the father of systematic musicology in Russia, Viktor Beliayev, in the 1920s, during his tenure in the Moscow Tchaikovsky Conservatory (Beliayev, 1990b). Beliayev’s approach was further developed by two leading Moscow theorists, Leo Mazel and Viktor Tzukkerman (Mazel and Tzukkerman, 1967). They sought to integrate thorough structural analysis of a musical work with the psychological and sociological analyses of the expressive means employed in the analyzed musical work (Khannanov, 2005). It was especially Mazel who was concerned with broadening the framework of analysis to encompass not only the domains of melody and harmony, traditional for Western musicology, but also aspects of rhythm, meter, texture, articulation, dynamics, and timbre. After Yevgeny Nazaikinsky’s death in 2006, the leading Russian “integralists” are Valentina Kholopova and Vyacheslav Medushevsky.

TABLE 1 | AEs of music.

Acoustic domain	Aspects of expression in music	Physical substrate of an aspect	Perceptual substrate of an aspect	The overall range of expression of an aspect	Discrete increments within an aspect's range of expression	Gradual inflections within the continuum of an aspect	
Frequency	1. Melodic pitch (consecutive "linear" i.e., "horizontal")	Changes of the fundamental frequency (FF) between the consecutive tones within the same timbral register and voice/part.	Relation of the intervallic size of a frequency change to the Temporal Coherence Boundary ^l and the segregation of audio streams that result from it ^{LV} .	Minimal change—about 100 cents, maximal change—about 1,200–1,600 cents (leaps over an octave characterize a few specific genres ^{XLVII} (e.g., lamentation) or music systems (e.g., anhemitonic pentatonic).	The ambitus of a melody is divided into "degrees" based on the permanence in tuning (stability of pitch level) of tones of the same register that execute the same or similar melodic function within a phrase—the resulting intervals between the degrees define the "interval classes" in a musical mode ^l .	A degree can fluctuate in its frequency within a certain range of values that is usually equal or smaller than the interval between the adjacent degrees ^{III} —either in a form of a temporary alteration of that degree (syntactic inflection) or the portamento gliding between the adjacent tones (pragmatic inflection that traditionally constitutes the subject of intonation in Western music theory) ^{XLVII} .	
	2. Harmonic pitch (concurrent "vertical")	Consecutive changes in the relations of harmonics between the harmonic series of the concurrently sounding tones.	The number of harmonics that share the same frequency values in the harmonic series of each of the concurrently sounding tones ^{IV} .	Minimal matching of harmonics—when the distance between the FF of the concurrent tones is about or below 100 cents, maximal matching—when 1,200 cents apart.	The extent of matching and mismatching of harmonics varies between the degrees of a musical mode, forming a progressive "scale" of harmonic interval classes afforded by that mode, from the most "consonant" to the most "dissonant" interval ^l .	Each harmonic interval can slightly fluctuate in their exact tuning depending on the context of melodic and harmonic relations between the constituent tones within the same musical mode in a music piece, especially major and minor intervals—as a part of expressive tuning ^{XXX} .	
	3. Musical texture	A specific type of arrangement (vertical and horizontal) of all musical sounds within a musical work	The number and relations of familiar conventional structural components in grouping of tones, themselves forming stereotypes of arrangement specific to certain genres ^{XLIII} ; varying along 3 axes: density, range ^{XLIV} and functions ^{XLV} .	The number and relations of familiar conventional structural components in grouping of tones, themselves forming stereotypes of arrangement specific to certain genres ^{XLIII} ; varying along 3 axes: density, range ^{XLIV} and functions ^{XLV} .	The simplest case of texture is a monodic melody where grouping is restrained to shifts in melodic direction and leaps ^{XLV} without rhythmic contrasts (e.g., Paganini— <i>Moto perpetuo</i>); the most complex is the great number of numerous parts/voices that are diverse in their function (e.g., Debussy— <i>La Mer</i>) ^{XLVI} .	Each texture breaks into a number of "stream segments" at its surface level of perception ^{XLVII} ; forming discrete components—"textural cells" used as bricks in constructing a texture by vertical (chords) and horizontal grouping (motifs) of various complexity, functionality and hierarchic relations—ascribed specific semantic values ^{XII} .	N/A
	4. Musical form	Changes of the thematic material—a complex of musical structures consecutively ordered within a music work ^{XXX} .	Repetition, variation, contrast or recapitulation (i.e., the return of a thematic material after some other material) of a specific thematic material, identified by some salient feature(s) ^{XXXIII} .	Repetition, variation, contrast or recapitulation (i.e., the return of a thematic material after some other material) of a specific thematic material, identified by some salient feature(s) ^{XXXIII} .	The simplest form is an exact reproduction of the same material (AA). The most complex is the "unveiling form" based on the ongoing contrast (ABCD. . .) ^{XXXIX} .	Changes in thematic material break a music work into discrete sections, with hierarchic relations between the changing phrases (A-B) and their changing constituent motifs (a-b), forming different hierarchic levels (e.g., A-B-A = ab-cd-ab).	Each section of a musical form can employ thematic materials of various salience, ranging from highly concentrated (symphony) to highly dispersed (prelude) ^{XL} ; the transition from the concentrated to the dispersed state can occur within the same section gradually ^{XL} .
Time	5. Rhythm	Relative duration of consecutive tones is quantized according to a certain division ratio (2,3. . .).	Grouping of the consecutive tones based on their perceived proportions of duration and position within a group ^{VI} .	Minimum—according to the Western classical music theory, semihemidemisemiquaver, or, 128th note (=6th division of a whole note), maximum—brevis (=2 whole notes)—relative to tempo; in absolute terms, from 20 msec ^{VII} to 1,800 (2,000) ms ^{XXXI} .	Rhythmic proportions are estimated in terms of binary or ternary divisions that produce a set of standard durations—i.e., "time classes" (=rhythmic values) engaged in a composition—usually 3–5 divisions, one or two of which are most frequently used, forming a "metric grid" employed to round up the actual duration of a tone to a valid rhythmic value.	Rhythmic values can fluctuate in their actual duration in the so-called "expressive timing" ^{VII} that exaggerates rhythmic contrasts by prolonging anchored tones while shortening tones in passages, ornaments, or short tones in those rhythmic figures that consist of contrasting durations—e.g., overdotting in the so-called "punctured" rhythm ^{VIII} ; such inflections overtake the normative ratios that govern the rhythmic divisions ^{VII} .	

(Continued)

TABLE 1 | Continued

Acoustic domain	Aspects of expression in music	Physical substrate of an aspect	Perceptual substrate of an aspect	The overall range of expression of an aspect	Discrete increments within an aspect's range of expression	Gradual inflections within the continuum of an aspect
	6. Meter	Number of unstressed beats, grouped together with the stressed beat, before the occurrence of the next stressed beat—perceptually and statistically prevailing in a musical movement.	Grouping of beats based on the perceived periodicity of stresses generated by longer and louder tones as well as changes in melodic direction and harmony ^{VI} .	Minimal size group—1, i.e., “spondaic” pulse (every single beat is stressed), maximal size group—24 (compound ternary pulse, made by 4 divisions in 3-level hierarchy 3:6:12:24) ^{XXXII} ; in absolute terms, periodicities from 100 to 6,000 ms ^{IX} .	Metric grouping is estimated in binary or ternary increments of an entire group that can be isochronous or non-isochronous ^{IX} , so that a metric pulse within a meter can proceed by symmetric or asymmetric increments (e.g., common time can switch between 1/1, 2/2, 4/4, 8/8 and 16/16 pulses, or 3/8 + 3/8 + 2/8, 3/8 + 2/8 + 3/8 etc.) ^L .	Metric stress can fluctuate within a metric group, making the metric increments inside it acquire or lose metric weight: this is achieved by placing the anacrusis on a different position within a metric group (e.g., in 4/4 pulse, placing the anacrusis on the fourth beat shrinks the group to three beats—such shift alters the metric pulse without replacing it by another pulse ^{LV} .
	7. Tempo	The average pace of beat within a span of musical movement, which retains a specific character of motion (e.g., hasty or lazy).	An overall impression of a characteristic movement of a certain type (e.g., walking, jogging, running, hopping) akin to a gait—estimated based on the interaction of the pace of beat and the rhythm which determines the choice of tempo by a performer ^{IX} enabling an “absolute tempo” (i.e., the optimal pace for a given music piece) ^X whose importance is reflected by the invention of the metronome ^{XI} .	Minimal metronome value—usually, 40 bpm, maximal metronome value—208 bpm, traditionally tied to heartbeat and gait rates that characterize a particular form of locomotion and its related affective state ^{XXII} , in absolute terms, within the range of the beat value from 300 to 900 ms ^{VII} —inferred based on the density of pitch changes per metric unit of time ^{XXXIII} .	Each tempo is defined by a specific velocity and a character of musical movement (e.g., <i>presto</i> rushes, while <i>allegro</i> does not), so that every musical culture works out a set of standard tempi—for classical music of the Common practice period it is a 12-tempi system ^{XII} where a tempo, optimal for a given music piece, is defined as a range of bpm values within which the “feel” for that tempo remains the same ^{XIII} —narrow enough to consider a “perfect tempo,” ^{XXXIV}	Velocity of each tempo can be adjusted without changing its character—music practice often generates rules for temporary minor fluctuations, reflected by a set of modifier terms: e.g., for Western classical music these are <i>meno mosso</i> , <i>ritenuto</i> , <i>rallentando</i> , <i>piu mosso</i> , <i>stretto</i> , <i>accelerando</i> ; tempo inflections can also be canceled (<i>tempo giusto</i> = strict time) or added (<i>tempo rubato</i> = constant slowing and speeding within the same phrase) ^{XIII} ; the velocity curves for such inflections seem to be fixed by convention in reference to the cultural standards of locomotion ^{XXXV} .
	8. Articulation	The manner of attaching/detaching of successive tones within the same register and part/voice.	Shortening of the nominal rhythmic value (akin to the plucking sound production of mandolin or xylophone)—or, extending it (akin to echo) by the overlap of the end of one tone with the onset of the following tone (thereby generating a momentary harmonic interval in a monophonic line).	Minimal use of articulation— <i>non-legato</i> , maximal detaching— <i>staccatissimo</i> , maximal attaching— <i>legatissimo</i> ^{XIV} ; these concepts do not seem to follow any absolute criteria.	Performance practice generates styles of consecutive rendition of textural elements (e.g., melody, chords, figurations) that fill a range from the most abbreviated to the most extended articulation in a gradient manner: <i>staccato</i> , <i>marcato</i> , <i>mezzo staccato</i> , <i>non-legato</i> , <i>portato</i> , <i>tenuto</i> , <i>legato</i> ^{XV} —their contrast often generates groups (e.g., a 2-tone legato-tenuto or a 2-tone legato-staccato) ^{XLII} .	Many articulation styles form a continuous range of shortening or extending a rhythmic value, depending on the musical context—which establishes the performing conventions ^{XV} ; a common case of contextual influence is adjusting the exact extent of legato ^{XVI} and staccato ^{XVII} depending on how high or low the register to which an articulated tone belongs is placed in the ambitus—the most common axes in flexing the articulation styles are connectedness, discreteness and compactness ^{XLII} .
Amplitude	9. Dynamics	Changes of amplitude between consecutive or/and concurrent tones within the musical texture.	Relative increase or decrease in intensity of a particular tone, textural element (e.g., melody, bass), component (chord), segment (accompaniment) or the entire musical texture.	Minimal dynamics— <i>pianissimo</i> (<i>ppp</i>), maximal— <i>fortissimo</i> (<i>fff</i>); these concepts do not seem to follow any absolute criteria, yet are present in many if not all music cultures, from Ancient Greece on ^{XXXVI} .	Music practice generates dynamic distinctions that generally correspond to the extent of affective intensity of the music ^{XVIII} and form a “scale” of dynamics increments: <i>fortissimo</i> , <i>mezzo forte</i> , <i>forte</i> , <i>mezzo piano</i> , <i>piano</i> , <i>pianissimo</i> ^{XX} —each featuring a range, narrower for experienced music users, highly variable in reproductions of the same music, but stabler per person ^{XX} .	Dynamics also uses gradual changes, usually to support climaxes and intensity contrasts ^{XXXVII} ; such flexions can be graded: positive— <i>piu forte</i> , <i>poco crescendo</i> , <i>molto crescendo</i> , <i>rinforzando</i> ; and negative— <i>meno forte</i> , <i>poco diminuendo</i> or <i>molto diminuendo</i> , <i>morendo</i> ; although these terms appear in Western tradition only in the 19th century, similar notions seem to exist in implicit non-Western music theories to support flexions of the expressive timing ^{XXI} that is most crucial for phrasing ^{XIV} .

(Continued)

TABLE 1 | Continued

Acoustic domain	Aspects of expression in music	Physical substrate of an aspect	Perceptual substrate of an aspect	The overall range of expression of an aspect	Discrete increments within an aspect's range of expression	Gradual inflections within the continuum of an aspect
Timbre	10. Register	Contrasting changes in tonal quality between timbrally homogenous groups of tones within the ambitus of a music work.	Registral position in music is evaluated similarly to pitch—in terms of gradation in higher/lower placement within the ambitus that is employed in a music work—while accounting for timbral similarity in sound quality between adjacent pitches (e.g., registers that are darker/lighter or thicker/thinner in sound) ^{XXII} .	The lowest register forms one pole in the range of musical tones possible for vocal and instrumental production, while the highest register forms the opposite pole; for vocals and such instruments as flute or clarinet, the highest register is the strongest and most vibrant, whereas the lowest—the weakest and dullest; for brass, in contrary, the lowest register is the strongest and the most vibrant ^L .	Human voices and musical instruments have break-up points in their tessitura, where the tonal quality noticeably changes, breaking in a few registers, each distinguished by its own coloration; thus, clarinet has four registers: somber “chalumeau” (E4-E5), dull “transition” (E5-B5), bright “clarino” (B5-C7), piercing “altissimo” (C7-A7) ^{XXIII} ; 4 (3) registers are typical for most singing voices ^{XXIV} and musical instruments ^{XXV} , but different cultures adopt different attitudes to registers: some smoothen registral contrasts, while others increase these contrasts ^{XXVI} .	Some musical traditions cultivate an overlap between the neighboring registers, extending the span of each register, thereby increasing its continuity—e.g., the countertenor can sing as baritone or bass ^{XXVII} ; additionally, musical instruments usually develop an arsenal of performing devices to diversify their timbre providing “flections” of their “principal” timbre—e.g., <i>pizzicato</i> , <i>col legno</i> , <i>con sordino</i> , <i>sul ponticello</i> and <i>sul tasto</i> on string instruments ^{XXVIII} ; similar devices are used by vocalists (<i>parlando</i> , <i>aspirare</i> , <i>fioco</i>); yet another common source of “flexing” the timbre and register is to stress a particular harmonic in the instrumental sound, thereby recoloring its timbre ^{XXIX} .
	11. Instrumentation	Selection of a type of musical instrument and vocals most suitable for a specific expression	Timbres of individual instruments and vocals can blend into a new timbre (e.g., oboe and clarinet), remain discrete yet complement each other (e.g., flute and oboe), or repel (e.g., harp and horn), depending on similarity and synchrony of spectral centroids and attacks ^{XLX} ; and salience of individual harmonics ^{XXX} .	The simplest instrumentation is sustaining a single timbre per piece of music (solo); the most complex is the combination of orchestra and choir that features multiple foreground and background layers, changing over time, with contrasts between tutti, soli, and orchestral and/or choral groups ^{LII} .	Each type of musical instrument and vocals constitutes a specific tone color in a palette of a music-maker ^L ; certain combinations of instruments (string trio, wind quintet, orchestra) ^{LIII} and vocals (duet, quartet, choir) ^{LIV} form stable settings used to create music of certain semantic content depending on the tonal quality and technical capacities of the instruments ^{LII} .	N/A

This table summarizes the structural characteristics of each of the principal AEs according to the treatises on music theory and relevant psychoacoustic research during the last 70 years. In the footnotes to the table, I have provided the relevant sources for those readers who are interested to find out more. All AEs feature incremental organization. Most AEs are quantifiable: melodic and harmonic intervals, textural elements and components, themes, metric pulses, rhythms, and articulation groups. Of 11 AEs, 9 (melody, harmony, form, rhythm, meter, articulation, dynamics and register) exhibit gradual inflections that are systemically used to increase the expressiveness of music.

^I(Noorden, 1975, 40–67); ^{II}(Nikolsky, 2015b); ^{III}(Garbuzov, 1948); ^{IV}(Benson, 2007); ^V(McDermott et al., 2010); ^{VI}(Jones, 2016); ^{VII}(Clarke, 2007); ^{VIII}(Fabian and Schubert, 2008); ^{IX}(London, 2004); ^X(Levitin, 1994); ^{XI}(Fallows, 2001); ^{XII}(Nazaikinsky, 1972); ^{XIII}(Garbuzov, 1950); ^{XIV}(Keller, 1973); ^{XV}(Jerkert, 2003); ^{XVI}(Repp, 1995); ^{XVII}(Repp, 1998); ^{XVIII}(Dean et al., 2011); ^{XIX}(Berndt and Hähnel, 2010); ^{XX}(Garbuzov, 1955); ^{XXI}(Todd, 1992); ^{XXII}(Drabkin, 2001); ^{XXIII}(Miller, 2014); ^{XXIV}(Titze, 1988); ^{XXV}(Patterson et al., 2010); ^{XXVI}(Yemelyanov, 2000, 46); ^{XXVII}(Ravens, 2014); ^{XXVIII}(Garbuzov, 1956); ^{XXIX}(Nazaikinsky and Rags, 1964); ^{XXX}(Rags, 1980); ^{XXXI}(Fraise, 1982); ^{XXXII}(Harding, 1983); ^{XXXIII}(Madison and Paulin, 2010); ^{XXXIV}(Gabrielsson, 1999); ^{XXXV}(Friberg and Sundberg, 1999); ^{XXXVI}(Chew, 2001); ^{XXXVII}(Thiernel, 2001); ^{XXXVIII}(Réti, 1951); ^{XXXIX}(Mazel, 1979); ^{XL}(Kholopov, 2006); ^{XLI}(Val'kova, 1992); ^{XLII}(Braudo, 1961); ^{XLIII}(Huron, 1989); ^{XLIV}(Benward and Saker, 2009); ^{XLV}(Skrebkova-Filatova, 1985); ^{XLVI}(Berry, 1987); ^{XLVII}(Cambouropoulos, 2010); ^{XLVIII}(Rags, 1999); ^{XLIX}(Sandell, 1995); ^L(Kholopova, 2002); ^{LI}(Meyer, 2009); ^{LII}(Banshchikov, 1997); ^{LIII}(Kendall and Carterette, 1993); ^{LIV}(Kreitner et al., 2001); ^{LV}(Rothstein, 1989); ^{LI}(Cambouropoulos, 2008).

changes in at least 9 AEs (Table 1). Their categorization into “classes” seems to be modeled after pitch. A music-maker breaks the range between the lowest and the highest pitch classes (i.e., ambitus) within a music work into “degrees,” forming a set of pitch classes to construct music. Similarly, other AEs divide the continuum between their marginal values into step-like increments, the assortment of which can structurally characterize a musical work. Pitch-class sets receive their analogs in sets of the following classes, intuitively selected by a music-maker for a particular expression per composition:

- “time-classes” (number of rhythmic values i.e., “divisions”),
- “pulse-classes” (number of periodicities in a metric grid),
- “tempo-classes” (number of musical movements)⁶,
- “articulation-classes” (number of styles of connecting consecutive tones),
- “dynamics-classes” (number of dynamic gradations),
- “register-classes” (number of zones of different tonal coloration),
- “texture-classes” (number of textural components),
- “form-classes” (number of themes).

Such discrete classes coexist with gradual inflections for each class (Table 1). Evidently, music is designed to integrate multiple AEs in a complex admixture of their patterns of expression. Music defaults to the *integration* of concurrent tones in contrast to the *segmentation* tendency of speech (Bregman, 1994)—people can sing together, yet when speaking, they always take turns (Brown, 2007). Here, AC sides with music rather than speech, evident in the widespread animal chorusing. Integrative power of music makes the concept of “musical mode” indispensable for understanding the rise of music. “Mode’s” reduction to “scale,” adopted by some researchers (i.e., Pfordresher and Brown, 2017) constitutes a fundamental error in confusing the purely quantitative and formalistic concept of “scale” with the qualitative and content-oriented concept of “mode” (see Nikolsky, 2015b). Musical mode is more than a mere set of pitch-classes selected to make music—it also encapsulates the rules for their interconnection and the semantic range of suitable expressions (Wulstan, 1971; Alekseyev, 1976; Kholopov, 1976, 2005; Bytchkov, 1987, 1997; Lester, 1989; Belayev, 1990a; Porter et al., 2001; Powers and Wiering, 2001; Straehley and Loebach, 2014; Winnington-Ingram, 2015).

In essence, “mode” constitutes the generalization of a particular melodic typology, characteristic for a given musical genre, which supplies that mode with semantic denotations (Nazaikinsky, 2013). Nothing similar exists in speech. Music is unique in its holistic appreciation of sounds *per se* (Patel, 2010). Hence, the idea of *euphony*—pleasant concordance of sounds in

specific expressions—is quintessential for “mode,” as emphasized by Russian theorists.

The same principles apply to “rhythmic modes,” conceptualized within Western (Roesner, 2001) and some non-Western civilizations (Clayton, 2000). Rhythmic divisions, utilized in a composition, complement one another in expression of musical movement and in combinatory rules. A rhythmic *modus* in Western medieval theory, Arabic *maqam*, Iranian *dastgah*, or Indian *raga* incorporates not only a specific progression of rhythmic values but a specific “ethos”—an abstracted emotional quality projected by music on society at large (Shestakov, 1975). Each rhythmic *modus* in the abovementioned music systems is characterized *semantically* by its affiliation with a certain ethos and *structurally* by certain proportions between the duration values used in a music work. Rhythmic *modus* resembles pitch *modus* by incorporating a set of rules. Just as pitch-classes are allowed to follow or not follow one another, or require an alteration for ascending or descending motion, rhythm-classes are restricted to certain ratios which can be altered in a certain way (e.g., a dotted rhythm can be “over-dotted” in a suitable context).

The idea of concordance and appreciation that underlies the overwhelming majority of known traditional music cultures justifies the conceptualization of each AE as a carrier of its proprietary “mode.” Every musical piece can be defined by identifying its melodic, harmonic, rhythmic, metric, tempo, articulation, textural, and timbral modes.

Together, these modes constitute “tonal organization” (TO) in music. Conceptualized by François-Joseph Fétis (1840), TO is a method of joining musical tones together according to the sensibility of music-users (Fétis, 1994, XXV). Unlike tonemes of tonal languages, musical TO affects *all* tones, generates *complex* functional relations between them, and involves rhythm-metric, dynamic, articulatory, and registral arrangements. Speech might also use similar arrangements (Patel, 2006). But music requires a special analytic attention where changes in the melodic contour are quantized into pitch-classes that are continuously cross-compared—unlike the linguistic “vowel pitch” (Walker, 1997, 322–3). Such syntactic pitch-parsing is as imperative for music as word-parsing is for language. Semantics provides yet another distinction: verbal syntax specializes in conveying referential meaning, whereas music specializes in emotional expression⁷ (Gabrielsson and Lindström, 2001; Juslin, 2001, 2005, 2011, 2013; Cook, 2002; Krumhansl, 2002; Gabrielsson and Juslin, 2003; Dissanayake, 2008; Johnson-Laird and Oatley, 2010; Trainor, 2010; Perlovsky, 2012; Altenmüller et al., 2013b; Eerola and Vuoskoski, 2013; Eerola et al., 2013; Peretz, 2013; Nikolsky, 2015a, 2020; Schiavio et al., 2016). Such distinction has been

⁶The word “movement” here refers to a principal division of a longer music work into sizeable sections, each distinguished by its own metric organization and tempo: e.g., a 4-movement symphony or a 3-movement sonata. The concept of movement emerged in 16th-century Western classical music to reflect on the old practice of switching from one tempo to another within the same piece of music (Sadie, 2001). However, by no means the use of multiple movements within the same work is exclusive to Western civilization. Well known are non-Western genres of music that employ cyclic arrangement, such as Arabo-Andalusian *nubah* (Touma, 1996) or Javan court Gamelan music (Sutton, 1991).

⁷Ontologically, it is necessary to distinguish between “meaning” in a natural language and “meaning” in a cultural system of symbols (such as music)—especially in light of the difference in their acquisition: thus, under experimental conditions non-human primates can acquire some symbolic systems but not a full-fledged human language (Balari et al., 2011). It seems that the verbal combinatorial semiosis of referential meaning is fundamentally different from retrieving imagery, be it emotional or motivational information assigned to cultural symbols. This distinction is crucial for the investigation of origins of human language and music. Here, music, despite its combinatorial nature, occupies a place closer to signal-like semiotic systems, which *makes music more accessible to hominins than language*.

fundamental for the musical practices and theories of most musical traditions before Western classical music was swept away by the 20th century modernistic “revolution.” This distinction became revived after emotion and music attracted intense neuropsychological research in the 1980s.

Music’s social nature—evident in entrainment⁸ (Tarr et al., 2014)—and emotionality—evident in chills (Altenmüller et al., 2013a)—are critical for distinguishing music: neither entrainment nor chills characterize verbal communication. And both are closely related through emotional contagion (Trost et al., 2017). This music/language distinction must have been already present in musilanguage, since in AC referential and motivational information is coded differently (Manser, 2010). However, music differs from ACs by encoding affective information according to the conventional modes of numerous AEs, as we shall see. Hence, the structural definition of music should be:

TO of multiple AEs that *entrains* listeners and performers and *transposes* performers’ intentions to emotionally stir listeners through vocal and/or instrumental performance.

Pitch contour, rhythm/meter, and dynamics (the most salient AEs) together constitute the principal structural criteria of music.

EMIC AND ETIC APPROACHES TO TONAL ORGANIZATION

The proposed definition is instrumental for engaging an additional source of evidence in the quest for the origins of music—the comparative structural analysis of world’s archaic indigenous musics, earliest forms of music-making by human infants, and animal vocalizations. The modern advances in computer science support the acoustic and statistical analyses of vast datasets unavailable before. Such investigation could radically update the evolutionary theory while resolving the current situation in comparative ethnomusicology that is nothing short of a crisis (Savage and Brown, 2013).

Many cognitive scientists remain unaware of the profound ideological shift in Western ethnomusicology that occurred during the last half-century. In essence, the study of “text” became replaced by the study of “people” (Zemtsovsky, 1997)⁹.

⁸Entrainment (from French “*en-*” + “*trainer*”—to drag something along) is the term used in physics to address a wide range of phenomena where two oscillators are coupled, and one of them gradually comes into synchrony with the other, becoming locked in a phase. Entrainment of two pendulum clocks was discovered by Christiaan Huygens in 1666 but was explained only few centuries later. In early 20th century, other manifestations of entrainment were unveiled in acoustics (coupling of the organ pipes) and biology (glimmering fire-flies)—until it was generalized as a universal physical phenomenon (Pikovsky et al., 2001). Its biomusicological manifestations were identified in the 1990s, at first in relation to music therapy, and thereafter as an integral part of perception of rhythm and meter (Large and Kolen, 1994), of great importance to the evolution of music (Fitch, 2012).

⁹Thus, Titon (2015), one of the leading Western ethnomusicologists of today, goes as far as defining the discipline of ethnomusicology as “*the study of people making music*”—rather than “the study of music” as the term “musicology” indicates (the study of human societies is conducted by another discipline—“anthropology,” reflected in the etymology of its name). Paradoxically, modern Western “people’s ethnomusicology” still shuns the Soviet ethnomusicology which shared the same approach, holding music as “belonging” to people and “reflecting” people’s mentality, while remaining totally free of the anti-textual bias (Panteleeva, 2019).

The turning point was marked by Gourlay (1982) at the 1979 Oslo Conference of the IFMC by a call for “humanizing ethnomusicology” to abandon “the pretense of objectivity.” Timothy Rice reflected this departure in his influential article “Remodeling Ethnomusicology” (Rice, 1987). At the heart of this transformation lies the emic/etic antithesis, introduced by Pike (1967) in 1957 to oppose the “insider’s” versus the “outsider’s view” in the researcher’s position toward an object of study. Ever since, this opposition has grown into a schism between Western social and cognitive scientists (Headland, 1990). Harris (1964) adapted Pike’s approach for social sciences, conceptualizing “emic” as a specific culture, *mentally* “native” to an “insider,” whereas “etic”—as cultures, experienced not mentally, but *behaviorally* due to their “foreignness” to an “outsider.” Hence, Harris’ claim that an outsider is capable of only grasping the superficial behavioral patterns through direct observation. Harris’ followers wanted to *abstain* from any “mentalization” of observed facts to avoid their misrepresentation (Harris, 1990). Pike’s followers, in contrary, interconnected mental and behavioral aspects, holding that etics and emics present respectively physical and cultural aspects of analysis, so that an outsider *can* learn to analyze like an insider, and vice versa (Pike, 1990).

For ethnomusicology, emic/etic problem was discussed at the 32nd ICTM Conference, 1993, Berlin. The consensus recognized that insider and outsider perspectives were inseparable and complementary to each other: emic data was to be fit into etic categories, disregarding whether they were actually recognized by the insiders (Baumann, 1993). However, in the following decade Western ethnomusicology became progressively politicized against a supposed “Western bias”—equated with any form of etic evaluation. Some authorities went as far as viewing cross-cultural scientific investigation of music as “cultural colonialism” (see Agawu, 2003).

The purist emic approach replaces the scientific method of investigation with the insider’s description of a native culture in a social context (Myers, 1993, 222–3). The reason for this is that the scientific method by itself is a product of Western civilization (Messner, 1993). Thus, Gourlay (1984) explicitly defies any objective inquiry about music by means of scientific investigation¹⁰. Becker (1986) declares musical systems as being “incommensurable,” and any scientific study of non-Western music as being “immoral.” She insists that each musical culture should be investigated only in its own native terms and *not* evaluated against another culture—the only way for a researcher to study music is to merge with the indigenous community, learn its language and jargon, and collectively make music. In effect, this utilitarian ethno-unilateral approach to music precludes the study of its origins (Dobzhanskaya, 2012). No wonder, in the West, comparative musicology became abandoned,

¹⁰Gourlay argues that *no musicological study of African music by outsiders is justified*, because “in no African language about which we have information, and in many used by other peoples who have oral rather than written traditions, is there a word corresponding to the English term ‘music.’” So, according to Gourlay, “where the term ‘music’ is unknown to the people in question, one can conclude only that what we are presented with is *the investigating scholar’s* concept of his/her ‘music.’”

musical universals denied, and music history fragmented into a bunch of disconnected “histories” (Savage and Brown, 2013). Unfortunately, despite its severe shortcomings, the “emic bias” has penetrated into psychoacoustics (i.e., see Parncutt and Hair, 2011)¹¹.

Certainly, not all Western ethnomusicologists abstain from the musicological analysis (Arom, 2010) and deny the validity of objective etic approach (Alvarez-Pereyre and Arom, 1993). Nevertheless, the anti-analytical trend¹² has taken its toll, establishing a conviction that any research of structural universals is inevitably ethnocentric and inadmissible for ethnomusicology (Nattiez, 2012). Disregarding musical text in sake of musical behavior is symptomatic of a shift away from comparative musicology to fractured sociomusicology of isolated musical communities (Nettl, 2010, 70–92). Many contemporary American ethnomusicological papers are published without a single example of structural analysis to support the author’s claims, basing their claims on entirely behavioral, and not musicological, data—paradoxically conducting *musicological* research without looking into *music per se* (Zemtsovsky, 2002)¹³. Consequently, cognitive scientists interested in comparative music theory and musicological analysis have no choice but to rely on the old publications in English and new ones in other languages (especially those coming from Eastern Europe and Asia, where the influence of politicization is weaker).

The summary of etic/emic arguments, crucial for investigation of TO, demonstrates that proponents of emic approach strongly overvalue it while writing off its fundamental flaws (Table 2).

TO is identifiable based on the etic information alone, and its few potential shortcomings are easily amendable by emic references (Dasen, 2012). Purely etic approach has been a status quo in organology, where musical instruments are identified according to etic principles, disregarding emic views (Baumann, 1993). And there is no reason why the entire field

¹¹Parncutt and Hair subscribe to Gourlay’s defiance of a scientific investigation for those phenomena that do not find a corresponding term in a native language. They categorically insist that the *research of consonance and dissonance be constrained only to music of such cultures that define the concepts of consonance and dissonance*: “if musicians in that culture do not talk directly or indirectly about C/D [consonance/dissonance], it is considered irrelevant.” By this logic, there is no gravity in those countries whose native people do not have a word translatable in English as “gravity.” Parncutt and Hair see the goal of studying music in “documenting the musical and music-theoretical discourses of the insiders about which tones and rhythms should be played together and why, and considering the political and psychological mechanisms that are allowing Western music to dominate world music”—undoubtedly, a controversial and a politically biased agenda.

¹²To substantiate this criticism that is rarely voiced in modern Western literature, I shall quote one of the biggest authorities in ethnomusicology (the emphasis is added by me): “*Functional analyses of musical structure cannot be detached from structural analyses of its social function: the function of tones in relation to each other cannot be explained adequately as part of a closed system without reference to the structures of the sociocultural system of which the musical system is a part, and to the biological system to which all music makers belong*” (Blacking, 1974, 30–31).

¹³One of the main reasons for the drop in standards of musicological and ethnomusicological analyses is that in the US and UK academic curricula, music theory in general, and music analysis in particular, have been offered as rudimentary undergraduate courses (Agawu, 2004). In contrast, in countries of the former Soviet Union, music analysis has been taught at the highest level of scholarship that requires at least 10 years of study before attaining a level of training where an analyst is expected to capture and interpret the totality of expressive means employed in a music work (Khannanov, 2005).

of ethnomusicology should not be treated in the same way. The etic approach is unique in enabling a “progressive” accumulation of knowledge where the mistake of one researcher can be corrected by another. Etic self-sufficiency is evident in the fields of ethology and developmental psychology. Neither human babies nor animals can provide emic information—which by no means invalidates the acoustic analysis of their communication.

In light of this, studying TO is paramount for establishing the objective ground for interdisciplinary scientific research of the evolution of music across the synchronic and diachronic varieties of music systems. TO’s role for musicology is comparable to the role of phonology in linguistics: TO specifies a set of acoustic attributes and their oppositions to encode and convey information. Together, they form the “surface level” that underlies the musical syntax and semantics, and provide the material base for any music culture (Cambouropoulos, 2010).

TONAL ORGANIZATION DISTINGUISHES HUMAN MUSIC FROM ANIMAL COMMUNICATION

The very ability to enjoy “harmonious” sounds most likely emerged as a byproduct of satisfying the need to bring individual emotions in accordance with the interests of a social group (Panksepp and Bernatzky, 2002). Musical anhedonia in humans is exceedingly rare, indicating that music evolved as a direct auditory pathway toward the emotional reward centers in the brain (Loui et al., 2017). Music is probably a human invention that came-into-being to shape important brain functions through the hedonistic effect of appreciating sounds (Patel, 2010). Patel’s (2008) theory of “transformative technology of the mind” reconciled the adaptionist (Darwinian) and the non-adaptionist (Spencerian) approaches, based on the latest cognitive research, and provided the foundation for the theory of “mixed origins of music” (Altenmüller et al., 2013b) that explains how human affective signaling system has transformed the human brain and created music. Emotive specialization and emergence of “musical emotions” must have followed the formation of human auditory-affective circuitry (Altenmüller et al., 2013a).

Centrality of affective signaling brings animal communication closer to music than to speech (Fitch, 2006). Animal signals usually express affective states according to their innate “vocabulary,” are volitionally produced, and are actually felt (Fitch, 2010, 179–81). TO shares more similarities with animal vocalizations than with phonetics, since consonants, crucial for verbal parsing, are unique to human speech—unlike vowels that are more similar to singing and ACs (Kolinsky et al., 2009). Vowels determine verbal prosody which is the primary means of conveying emotions through speech.

Most likely, the musilanguage’s TO resembled the model of vocal production, common for primates and human infants—a reflex-like vocalization (e.g., pain-shrieking), triggered by specific stimuli, and hard-wired for animals but modifiable for humans (Jürgens, 1995). Humans start developing the repertory of cries by differentiating timbral and contour features just a few months after birth (Wermke and Mende, 2009), whereas for most animals, call structure is not modifiable by acoustic experience

TABLE 2 | Pros and cons (P/C) of purely etic, emic, and combined “etic + emic” approaches to analyzing music structures.

Task	Emic P/C	Etic P/C	Etic + Emic P/C
Categorization	<p>Minimal distortion of the original concepts underlying music-making and adequacy of applying theory to practice</p> <p>Possibility that an emic expert would not recognize an important category that is objectively evident in a given culture - due to its novelty and its contradiction to the traditional ideology or a dominant practice</p> <p>Possibility that an entire culture (e.g., herding vocalizations) or certain genres (spells) that appear as music to outsiders might be viewed as non-music by locals for religious, ideological, or practical reasons</p>	<p>Imposing culturally foreign theoretical concepts onto indigenous music practice, and misunderstanding of its typology</p> <p>Accurate categorization of all audible musical features—even if for some religious, ideological, historic, or ethnic reasons they are not noticed, ignored, misinterpreted, or forgotten in an indigenous music theory</p> <p>All forms of musicking that appear as “music” to an outsider and satisfy the criteria of music are recognized as such, even if local experts for some reason do not consider them as music – thereby making the scope of study wider</p>	<p>Adequate categorization</p> <p>Emic categorization should be discarded</p> <p>Emic categorization should be discarded</p>
Capturing important data	<p>Correct identification of all the relevant data as long as it is adequately reflected in emic music theory and recognized during audition by an emic expert</p> <p>Omission of features of TO that have not been perceived by an emic expert (e.g., due to hearing or performing problems), conceptualized by the local emic theory, rejected as foreign or inappropriate, or misinterpreted in terms of the ideologically dominant tradition</p> <p>Omission of those genres or styles of music that are tabooed within a culture (e.g., funeral, totemic, or ancestral), which can severely distort the perspective on such musical culture</p>	<p>Possibility of omitting information that is crucial for a local music culture, but totally unfamiliar to an outsider</p> <p>Complete and comprehensive collection of all the data that is physically present in local music practice and whose typology has been known in other related or similar musical cultures</p> <p>Acceptance of observable genres or styles tabooed for insiders by the outsider researchers, allowing them to integrate this information with the rest of what is known about this culture</p>	<p>Adequate data collection</p> <p>Emic manner of data collection should be discarded</p> <p>Emic taboos should be discarded</p>
Examination of data	<p>A local expert has no problem understanding and interpreting the familiar musical material</p> <p>Possibility that local music theory can be culturally or ethnically biased or tainted in relation to specific kind of data, and local experts would rigidly discard this data from any consideration in their analysis</p> <p>The data pool for examination available to any single local expert is severely limited by their personal experience, and therefore, is prone to lead to erroneous generalizations</p> <p>An insider expert is prone to errors due to accidental lapses in attention – and these errors are hard to detect and correct</p>	<p>Either an outsider must acquire the basic knowledge of local music material, or an insider has to learn some general theory</p> <p>Uniformity of the analytic approach to the musical material that characterizes a given culture secures equal treatment of all the collected data, making the findings more comparable with other findings</p> <p>The use of standardized procedures enables automated data analysis, greatly expanding the data pool, and increasing the reliability of the findings</p> <p>Standardized procedures secure the use of computer technology and collective work of numerous researchers, preventing and correcting accidental errors</p>	<p>Adequate analysis</p> <p>Emic methodology should be discarded</p> <p>Emic data can be used only if it does not contain inaccurate generalizations</p> <p>It will be safer to discard the emic data unless it is confirmed by another expert</p> <p>Adequate verification</p>
Verification of results	<p>Results of the emic expert’s analysis are verifiable by interviewing the insiders and conducting the statistical analysis of their responses, which can identify a problem but cannot resolve it</p> <p>Insiders might follow a system of irrational beliefs (i.e., spiritual) that preclude any rational verification of the results of emic-based TO analysis</p> <p>There is no reliable way to reconcile any disagreement between two disagreeing insider experts</p>	<p>Results of the etic analysis cannot be directly verified by the indigenous music users, requiring a qualified interpreter and limiting the reliability of statistical analysis of the obtained data</p> <p>Irrational criteria can be separated from the rational, and both can be effectively verified: rational by testing, and irrational by interviewing</p> <p>Disagreements between informants are rationally resolvable by testing more subjects and by conducting additional interviews, as needed</p>	<p>Emic evaluation should be discarded</p> <p>Emic evaluation should be discarded</p> <p>Adequate comparison</p>
Comparative Study	<p>Natural coherence of a local culture as regards to ideological, structural, social, and semantic data facilitates its cross-cultural comparison, if the compared cultures are similar in composition</p> <p>As a rule, cross-cultural comparisons are impossible due to the general mismatch of categories and definitions that vary between different emic music theories, both verbally and conceptually - even between linguistically related cultures: there is no way for emic experts in each of the cultures to secure a common ground</p> <p>To compare a few cultures, a researcher has to learn the conventions of music making, language, and customs of each culture, which is extremely time consuming and still does not guarantee adequate understanding of each culture</p>	<p>Reliable evaluation of a local culture is possible only in those domains that are physically manifested, and match the researcher’s experience – otherwise, a comparison might be erroneous</p> <p>Cross-cultural comparisons are easy and comprehensive if the general music theory is equipped with means to identify and classify all the principal varieties in acoustic attributes of tonal organization that are encountered in each of the local music cultures; then, the generalized category will secure a common ground for cross-cultural comparison</p> <p>To compare numerous cultures, an outsider is not obliged to visit and stay in a given location, learn to make music in emic way, and speak the native language, which spares time and enables learning about a greater number of cultures</p>	<p>Emic comparison should be discarded</p> <p>Emic comparison can be discarded</p>

Pros are colored blue, and cons red. The number of cons for the emic approach (11) doubles the number of cons for the etic approach (5). Emic cons are more detrimental for the outcome of the analysis. The etic approach, even at its worse, still allows the researcher to infer valid principles of TO in a sufficiently large pool of samples of musical styles/genres - which, in the long run, secures correction of mistakes by subsequent researchers. At its worse, the emic approach precludes any comparative study and invalidates the study of TO in an isolated music culture where its members do not regard certain sound production as “music” (i.e., incantations, spells, herding vocalizations). The combined etic/emic approach effectively corrects the shortcomings of a purely etic approach, but in most cases, it fails to correct the shortcomings of a purely emic approach.

(Hauser, 1996, 315). Call-learning occurs in a few songbird species, but for most birds, songs are innately encoded, and life experience only activates their retrieval (Marler, 1997).

A call serves as the basic unit in animal communication¹⁴ and usually conveys specific affective information (Hauser, 2000). Different calls are combinable in “mixed bouts” that are different from “pure bouts” (single call) by triggering a sequence of emotion-based behavioral responses in other animals. Each call’s significance is hard-bound to its acoustic structure. Despite their superficial similarity with music, “mixed bouts” lack transposability of intentions: each call comes only in response to the actual stimulus present in the environment (Zuberbühler, 2017). Transposability is the landmark of music—the same structural pattern is intended to express the same idea across different instances of use, without which musical genres would be impossible: e.g., most lullabies are recognized cross-culturally by their set of structural features (Trehub et al., 1993). Genres are based on reproduction and transposability, and usually form genre systems to support important social practices (Samson, 2001), which enables music to reflect perceptual reality. Animal-learned vocalizations miss such comprehensiveness and generalization. They are limited to:

- display of fitness (Naguib and Riebel, 2014),
- a single season and gender (Slater, 2011),
- mating or defending situations (Slater, 2001).

Syntactically, AC overall lacks a combinatorial organization¹⁵. It resembles the one-word holophrastic communication of human infants by depending on a directly observable context and on an “analog” signal-emotion correspondence (Johansson, 2005). The same applies to animal “*phonocoding*”¹⁶ (Marler, 2001): it excludes categorical perception, rhythm, hierarchical structure, and adjacent transitional probabilities (Yip, 2006).

Indispensable for speech and music, compositionality completely eludes ACs—along with listener’s capacity to

¹⁴In some songbirds, the innate encoding consists of smaller elements, resembling syllables, and following simple rules for how to order them, so that a bird actually learns to “assemble” its song. However, the assortment of such elements is very limited, making songs signal-like, restrained to a single species. Playback of isolated syllables of such songs either does not elicit response or produces a weak reaction in other conspecific birds (Searcy, 1992). Perhaps the rearrangement of elements constitutes not a pragmatic, but a “syntactic” production unit—thus, zebra finches were found to stop at syllabic breaks in a song, when detracted (Cynx, 1990). Rearrangement of “syllables” is also used by a few primate species (gibbon) to disclose the identity of a caller for conspecific animals (Marler and Mitani, 2008).

¹⁵Although it is not uncommon for ACs to form a sequence according to a rule-based structure, noticeable by conspecific animals (Fitch, 2010, 182), changes in such structures apparently do not result in the changes of meaning of the entire song (Hauser, 2000). The most syntactically elaborated bird and whale songs use combinatorial features, albeit minimal. However, despite having a componential structure, such animal song in its entirety presents a *single* piece of information learned from the animal’s parent holistically rather than incrementally, element by element (in contrast to how humans learn), and is therefore highly stereotypical in form (Hurford, 2012, 3–99).

¹⁶The concept of phonocoding (i.e., “phonological coding”) was introduced to oppose “lexicoding” of human speech (Marler, 2001). Phonocoding refers to the capacity to generate new sound patterns by recombining the constituent elements and components of known conventional signals. This capacity is minimal in non-human primates, but common in learned vocalizations of songbirds and whales, which, however, remain primarily non-symbolic and affective.

continually (re)-organize behavior as the song unveils. Non-human communication, as a rule, employs a “one-ended” system: a signaling animal emits a signal unconsciously, not for any specific receiver but as a physiological reflex conditioned to a particular type of stimuli (Hauser, 2000). Such intention-free transmission precludes semiosis¹⁷—since sender and receiver must share signs and codes to actually transmit information.

A cumulative “two-ended” semiosis, where the receiver signals in response to the sender and vice versa, is unique to humans, and emerges as a result of technological complexity of human life. Dennett (1983) called this “second-order intentionality”—i.e., the receiver’s beliefs and desires about the sender’s beliefs and desires—in distinction from the “first-order intentionality” that is limited to the receiver alone.

- *First-order* intentionality is characterized by a one-ended conscious processing of unconsciously emitted signal—here, the unintended signaling receives an intentional interpretation.
- *Second-order intentionality* requires a two-ended premeditation of a signal: the signaler has to consider the receiver’s competence, and the receiver must be looking for information while considering the signaler’s circumstances.

Subsequently, the state of knowledge is changed on both ends of such communication, which, so far, has not been found in any non-human animal. Most common for ACs is *zero-order intentionality*—the signaler does not consciously intend to convey a piece of information, but *instinctively* engages a specific signal structure, triggering a similarly automatic response of the receiver.

Two-ended communication generates an unlimited diversity of structure due to infinite recombinations of a finite set of discrete elements that do not carry meaning on their own—what Abler (1989) calls “particulate principle.” It is peculiar to human language and music, finding only embryonal equivalents in a few animal species (Hauser, 2000). Complexity, comparable to human, is evident in some birdsongs, but serves to impress mates and intimidate competitors rather than conveying a specific message (Marler and Slabbekoorn, 2004)—likely forming a parallel (not prototype) to human evolution (Fitch, 2010, 184).

¹⁷The term “semiosis” here refers to the Peircean concept of conveying information by encoding it into signs by one party and decoding it by another party—a “two-ended” system. A “one-ended” call can be somehow interpreted in relation to the situational context by the listening animal, but this interpretation can radically differ from the actual state of the sender: e.g., bird’s mating call might be interpreted by a nearby cat not as a signal of readiness for mating but as a signal for hunting. Then, the integrity of the information passed from sender to receiver is not preserved. Within this context, the use of the term “meaning” in regard to an AC, adopted in biosemiotics (Sebeok, 1994, 111), is confusing, since “meaning” implies that someone “means” something by displaying a specific sign. More accurate here would be to employ the term “*significance*” (as in “to signify”) instead of “meaning.”

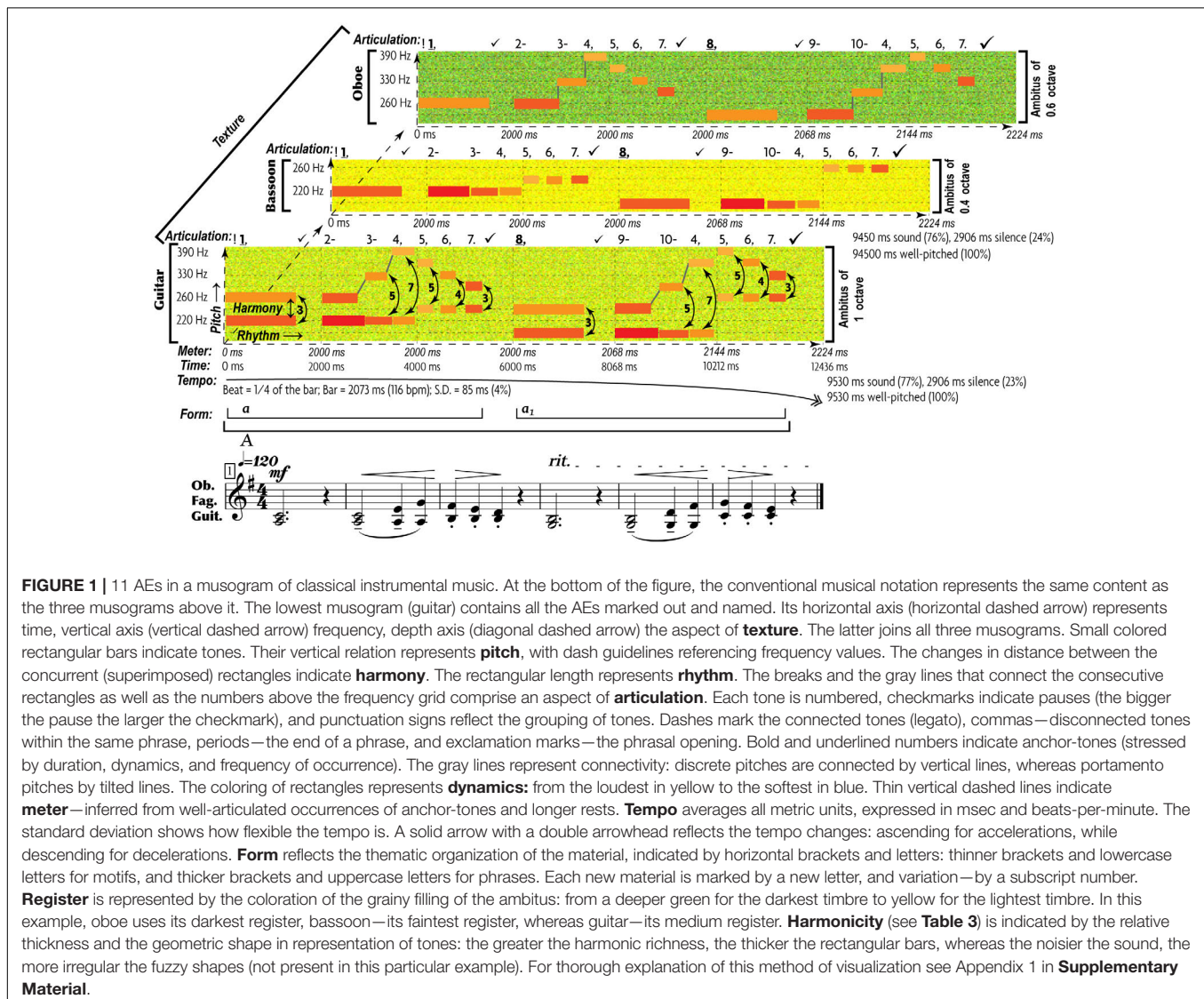


FIGURE 1 | 11 AEs in a musogram of classical instrumental music. At the bottom of the figure, the conventional musical notation represents the same content as the three musograms above it. The lowest musogram (guitar) contains all the AEs marked out and named. Its horizontal axis (horizontal dashed arrow) represents time, vertical axis (vertical dashed arrow) frequency, depth axis (diagonal dashed arrow) the aspect of **texture**. The latter joins all three musograms. Small colored rectangular bars indicate tones. Their vertical relation represents **pitch**, with dash guidelines referencing frequency values. The changes in distance between the concurrent (superimposed) rectangles indicate **harmony**. The rectangular length represents **rhythm**. The breaks and the gray lines that connect the consecutive rectangles as well as the numbers above the frequency grid comprise an aspect of **articulation**. Each tone is numbered, checkmarks indicate pauses (the bigger the pause the larger the checkmark), and punctuation signs reflect the grouping of tones. Dashes mark the connected tones (legato), commas—disconnected tones within the same phrase, periods—the end of a phrase, and exclamation marks—the phrasal opening. Bold and underlined numbers indicate anchor-tones (stressed by duration, dynamics, and frequency of occurrence). The gray lines represent connectivity: discrete pitches are connected by vertical lines, whereas portamento pitches by tilted lines. The coloring of rectangles represents **dynamics**: from the loudest in yellow to the softest in blue. Thin vertical dashed lines indicate **meter**—inferred from well-articulated occurrences of anchor-tones and longer rests. **Tempo** averages all metric units, expressed in msec and beats-per-minute. The standard deviation shows how flexible the tempo is. A solid arrow with a double arrowhead reflects the tempo changes: ascending for accelerations, while descending for decelerations. **Form** reflects the thematic organization of the material, indicated by horizontal brackets and letters: thinner brackets and lowercase letters for motifs, and thicker brackets and uppercase letters for phrases. Each new material is marked by a new letter, and variation—by a subscript number. **Register** is represented by the coloration of the grainy filling of the ambitus: from a deeper green for the darkest timbre to yellow for the lightest timbre. In this example, oboe uses its darkest register, bassoon—its faintest register, whereas guitar—its medium register. **Harmonicity** (see Table 3) is indicated by the relative thickness and the geometric shape in representation of tones: the greater the harmonic richness, the thicker the rectangular bars, whereas the noisier the sound, the more irregular the fuzzy shapes (not present in this particular example). For thorough explanation of this method of visualization see Appendix 1 in **Supplementary Material**.

The structural criterion for emergence of the *Semiotically Functional TO* (SFTO)¹⁸ in music is therefore manifested in the introduction of particulate organization in *phonocoding*.

THE TIMEFRAME OF TONAL ORGANIZATION OBTAINING FULL SEMIOTICALLY FUNCTIONAL CAPACITY

The current consensus holds that music was gradually formed since the appearance of *Homo heidelbergensis* about 600,000 BP, leading to an artistic “explosion” circa 40,000, when the earliest

¹⁸By “semiotically functional,” I mean that a music-maker selects the elements and components of tonal organization for each of the aspects of expression in music based on their efficacy in conveying specific affective information (“musical emotion”) to his/her listeners and/or partners in performance. In this sense, the AC can be considered “semiotically dysfunctional”—not supporting a successful two-ended communication (delivery of the intended message) between the sender and the receiver.

bone “flutes”¹⁹ were produced “en masse” (Morley, 2013, 219–25). Although flutes prove the existence of TO in the Aurignacian culture, this tells nothing of whether their sounds served a one- or two-ended communication. In all likelihood, TO did

¹⁹The word “flute” here is used informally: there is not enough archeological evidence to conclude if the earliest instruments were flutes or clarinets. The oldest artifact is a bone fragment from Hauer Fteah, Libya, with a single hole, dated 90–110,000 years ago (Blench, 2013). Most archeologists do not recognize it as man-made. Next in line is the 47,000 years old 3-hole artifact from Divje Babe, Slovenia, uncovered in 1995. It was interpreted as a bone bitten by a carnivore (D’Errico et al., 1998). However, experimental testing has demonstrated that none of the cave bear, wolf or hyena dentition could punch two holes without cracking and splitting the bone (Turk et al., 2001). Nevertheless this argument was not accepted by the supporters of non-human origin of the Divje Babe artifact (Morley, 2006). Subsequent tomographic analysis has concluded that the Divje Babe artifact was man-made (Tuniz et al., 2012). Slovenian researchers have presented additional reasons for its man-made origin (Turk, 2014). In spite of this, another recent British study has restated the bite origin hypothesis (Diedrich, 2015)—though, without addressing the 2012 and 2014 studies’ arguments. The third in timeline and unequivocal in its provenance, is the 5-hole Hohle Fels-1 flute, 35,000 years old (Conard et al., 2009).

not *communicate* musical emotions but merely *accompanied* the behavioral display of actual real-life emotions—as it happens in reflex-driven animal vocalizations (Seyfarth and Cheney, 2017). Their acoustic form is shaped by the physiological impact of emotion on the vocal organs plus Pavlovian-style priming.

Semiosis originates in an ongoing interaction between signalers and receivers within the reference-framework of the *same* environment—forging communication rules through the dialectics of ritualization and devaluation (Wiley, 1983). Ritualized signals establish conventions via encoding/decoding interaction between the acquainted individuals. Once established, convention becomes “devalued”—abused by “bluffing calls” of the unacquainted signalers trying to take advantage of the established reactions of the receivers. Increase of dishonest signaling causes the signaler to substitute the signal or modulate it along a single acoustic dimension until an “evolutionary stable strategy” is formed, marking a stationary equilibrium within the population—which ultimately fixes the convention (Maynard-Smith, 1976). Here, “signaling efficacy” obtains its formative power: as natural selection optimizes a signal to support the signaler’s visual display, successful decoding starts relying on whatever the receiver finds most comfortable to detect, discriminate, and remember (Guilford and Dawkins, 1991). Together, strategic design and efficacy determine the ultimate structure of a signal.

The road from animal call to musical phrase goes through the ritualization of innate physiological and behavioral cues that animals use to exchange information (Maynard-Smith and Harper, 2003)²⁰. Ritualized signals differ from cues by being more conspicuous, redundant, stereotypical, and containing alerting components (p. 72). Nevertheless, they remain “concrete” (bound to a single context) like cues (Fitch, 2010, 184) and unlike “transposable” music. For ritualized signal to evolve into musical phrase, its meaningful features must be abstracted to become non-signal-specific and form an AE of TO—a conventional dimension of gradient change along some axis.

The end result of such abstraction is the multifactorial nature of music communication (**Figure 1**): each emotional/motivational state is represented not by a *dedicated signal* but by the configuration of *numerous AEs* (Juslin, 2005). Conventional musical notation is poorly suited for incremental representation of AEs other than rough indications for melody/harmony, rhythm/meter, and form. Waveforms display rhythm and dynamics in finer detail, but miss other AEs. Spectrograms decently represent melody, rhythm, articulation, register, harmonicity, and dynamics, but miss harmony, tempo, meter, and texture. This necessitates the use of a special notation—such as *prosogram*, developed by Mertens (2004) for analyzing speech. Although applicable to monophonic vocal music in visualizing pitch, rhythm, articulation, dynamics,

²⁰Maynard-Smith and Harper give an example of such ritualized physiological cues as thermoregulation that causes animals to raise their feathers/hair to reduce body temperature, heightened in social interaction—which makes an animal appear larger and promotes dishonest signaling of increased body size in instances of confrontation (p. 68). Other physiological cues are respiration, urination/defecation, pupil dilation, and yawning (p. 69). The ritualized behavioral cues include “intention to move” which signals the beginning of a significant action (a bird taking a few false starts before flying), “protective movement,” and “displacement behavior” (p. 70).

harmonicity, and register, prosogram ignores harmony, tempo, meter, texture, and form. To overcome these limitations, I propose a similar approach to music—“*musogram*”²¹. Its advantages over conventional notation in capturing 11 AEs are demonstrated in the simplest case of classical music (**Figure 1**). It introduces the conventions, necessary to read the upcoming figures.

Multifactorial visualization reveals the expressive contribution of all AEs. Each AE features structural patterns representing specific emotional states across cultures, genres, and styles—at least for basic emotions (**Table 3**)²². Configuration of such patterns distinguishes one emotional expression from another. If multiple expressions share the same pattern of AE (e.g., legato characterizes both sadness and tenderness), the combination of a few aspects (e.g., “articulation + meter”) differentiates them.

Multifactorial particulate semiosis shapes musical signs—each AE features SFTO, which enables “natural selection” for the most effectively communicated expressions. AC can be multifactorial but lacks particulate semiosis. Verbal semiosis is particulate but mostly unifactorial: phonetic organization is its primary source²³.

Basic emotions can be recognized across musical cultures (Mohn et al., 2010) and can be acoustically described (Eerola and Vuoskoski, 2013). Therefore, at least some of their musical markers share biological roots with mammalian ACs (Zimmermann et al., 2013). The birth of SFTO is trackable by comparing the multi-cultural markers of typical musical expressions of basic emotions to equivalent AC expressions and by inferring their differences and commonalities (**Table 4**). Common traits indicate music’s inheritance from ACs, whereas contrasting traits—innovations brought about by cultural evolution.

Music and ACs have in common only regularity/irregularity and articulation. They both find a perfect match between human music and AC (5 out of 5 emotional states). The next closest match (4 out of 5) is “harmonicity.” That is why these two aspects of TO (articulation and harmonicity) must be the most ancient, possibly retained from the pre-human times. In contrary, “register” shows a nearly perfect

²¹For thorough explanation of the visual representation of the multifactorial organization of music, a way of its quantification, and its difference from the prosogram approach by Mertens, see Appendix 1 “A New Method of Modal Multifactorial Analysis of Tonal Organization in Music” in **Supplementary Material**.

²²Musicological literature identifies many more structural patterns of different AEs than the patterns listed in **Table 3**—and their semantic references include many more affective states than merely five basic emotions. Much of this information is dispersed in the treatises on music theory, some of which are cited in the beginning of this paper. There are very few books that list such structural patterns in a manner of the 18th century treatises of “musical lexicon” (Cooke, 1959; Mattheson and Harriss, 1981; Bartel, 1997; McCreless, 2002; Vashkevich, 2006). However, only isolated patches of such literature have attracted attention of psychoacousticians and received experimental trial (Kaminska and Woolf, 2000). For this reason, the metareviews on research in “musical emotions” tend to focus exclusively on 5 basic emotions.

²³Although tempo, rhythm, prosodic contours, and registers contribute meaningful motivational and attitudinal information to verbal communication, by no means can they be regarded as its primary semiotic aspects. Without knowing the lexic meaning of words of a particulate language, inferred from phonetic structures of auditioned speech, no adequate understanding of that speech is possible. This is in polar opposition to musical semiosis, where tempo, rhythm, melodic contour, and register directly convey the most important information, whereas keeping the referential meaning optional.

mismatch, testifying that humans cardinally reorganized the use of registers in music. The rest of the AEs display mixed results. If to generalize by emotional states rather than by expressive aspects, then none of the emotions display a full match or a full mismatch. Evidently, coding of emotions in human music has developed its own proprietary acoustic attributes. This confirms that ACs are mostly conspecific. Heterospecific²⁴ generalities support only a rough distinction between “positive” versus “negative” emotions (Snowdon et al., 2015). Human communication inherits from ACs just 2 general semiotic oppositions: (1) positive/negative affectation and (2) low/high intensity of an affective state (Brudzynski, 2013). High-intensity “strong emotions” (Grewe et al., 2005) have evolved into chill-like experiences of music—in contradistinction to the “mundane” use of language (Silvia and Nusbaum, 2011). However, “strong emotions” *per se* could not support musical semiosis because the stimulus-response relationship between chill and music structure has not been experimentally reproducible—music chills seem to occur intermittently (Altenmüller et al., 2013a).

Both incremental and gradual changes in multiple AEs (Table 1) are peculiar to human music, whereas holistic tempo, dynamics, rhythm, and melodic contours are mutual for music and ACs. Musical meter, articulation, and harmony are also traceable to, respectively, ACs’ regularity/irregularity, pausing/continuing, and periodicity/harshness.

However, the cross-examination of TO in expression of 5 basic emotions in music versus ACs reveals that many AE’s patterns are unique to music (Table 5). Moreover, humans completely invert the acoustic characteristics of animal’s affective states:

- Ascending/descending pitch (anger-tenderness),
- Fast/slow tempo (happiness-tenderness),
- Soft/loud dynamics (happiness-fear),
- High/low register (happiness/sadness-anger/fear),
- Harmonicity/inharmonicity (tenderness-anger).

This indicates massive remapping of the instinctive vocal encoding of affective states, achieved throughout the cultural evolution of Homo.

What could have caused such changes?

For many AEs, their cultural origin is obvious: metric pulses usually break into a default binary pulse (Potter et al., 2009), following the left/right paradigm instituted by bipedalism (London, 2004). Rubato patterns (*ritenuto/accelerando*) also relate to bipedal locomotion (Honing, 2003), so as tempo which is synchronizable to gait or heartbeat (Fraisse, 1982). Melodic intervals follow another locomotive paradigm of stepping/leaping (Nikolsky, 2015b)—each successive tone either “stands” (unison), “steps” (2nds and fast 3rds), or “leaps” (>3rd)—unlike harmonic intervals that are factored

by consonance/dissonance relations (a much later historic semiotic development). Articulation grouping relies on yet another biological factor—the breathing cycle (Alekseyev, 1976, 130). Taking a breath terminates a phrase, imposing a “clausal structure” on the melody (Fenk-Oczlon and Fenk, 2009b). The “breath group” prototypes the “articulation group” via a “breathing pulsation” (Etzel et al., 2006). Noteworthy, breathing pulse takes over metric control in ametric forms of music-making (Wallin, 1983). Locomotive and respiratory AEs must have formed long before Homo.

The rhythmic aspect of music possibly emerged from the quantification of verbal rhyming, following the language development (Kharlap, 1972)²⁵. Melodic contours also relate to verbal prosody. The timeline of language formation remains controversial: the “saltational” scenario regards language as a sudden mutation 50–100 kya, whereas the “gradual” scenario qualifies it as part of evolution throughout millions of years (Hillert, 2015). Paleoneurology points to the Middle Pleistocene as a birthtime of language (Quam et al., 2017). Since musical rhythm and melodic contours rely on fine vocal control, their addition to TO must have followed the accumulation of extensive lexic vocabulary within a phonological organization of language (Tallerman, 2013). This ties the emergence of multifactorial TO (which is hardly possible without engaging melodic contour and rhythm) to Homo sapiens and the Upper Paleolithic, as indicated by the proliferation of bone “flutes.” During 1995–2009, over 120 bone pipes were recovered across Europe, dated 36–30 kya and concentrated up to 3 “flutes” per cave (Conard et al., 2009). Evidently, melodic music suddenly became popular in the Aurignacian.

Discreteness of pitch is evident in the construction of Paleolithic “flutes”: holes are drilled in particular spots in order to generate sound of a particular pitch, and there is evidence of common patterns in the intervallic distances between the placement of the holes, suggestive of the commonality of certain melodic intervals in Aurignacian music-making (Nikolsky, 2015b, Appendix II). Discreteness of pitch was very likely to have been accompanied with the discreteness of rhythm, since stressing a pitch as a rule relies on extending its time-value relative to other pitches. Pitch hierarchy is supported by rhythmic contrasts between shorter timing of modally insignificant pitch-classes as well as longer timing of modally important pitch-classes (Krumhansl, 1990).

However, Aurignacian music most certainly lacked SFTO—semiotization of rhythm and directionality requires an extensive period of exploration. This is obvious in the acquisition of musical skills throughout infancy: infants babble—engage in meaningless play with melodic contours—before learning to compose musically *expressive* vocalizations (Moog, 1976; Dowling, 1984; Swanwick et al., 1986; Holahan, 1987; Hargreaves,

²⁴The opposition of conspecific and heterospecific distribution of acoustic features that characterize the vocal expression of a particular affective state in AC allows a researcher to identify those patterns of AEs that match cross-cultural features of corresponding affective states in “musical emotions” of human music. The patterns of expression that are present across multiple animal species are more likely to form the equivalents of “universal” traits of human “musical emotions” than those patterns that are found only within the very same animal species.

²⁵However, the idea of rhyming seems to have a precursor in ACs. Thus, humpback whales match the constituent syllables in some of their songs (Payne, 2001). A similar organization was noticed in mockingbird songs (Thompson et al., 2000). Its underlying cause is perhaps simplification of memorizing a complex song. Yet another cause could be the employment of repetition of a particular syllable in a song for a certain number of times as a conspecific marker for certain bird species (Fitch, 2010, 183). Hearing such birdsongs might have prompted humans to invent rhyming.

TABLE 3 | The configuration of structural patterns for each AE, typically used to express five basic emotions.

Acoustic domains	Aspects of expression in music	Range of the expression of an aspect	Happiness' acoustic markers	Sadness' acoustic markers	Anger's (aggression) acoustic markers	Fear's (anxiety) acoustic markers	Tenderness' (love) acoustic markers
1. Frequency	1. Melodic pitch (consecutive "linear")	High/low relation of tones	Prevalence of ascending contour within a diverse set of contours, wide ambitus, leaps, sharp zigzags, and sharpened intonation	Prevalence of descending and wave-like contours, narrow ambitus, mainly stepwise motion, flat and sliding down intonation	Prevalence of ascending contour, with little diversity of other contours, frequent leaps with sharpened melodic contours and tendency to short motifs	Prevalence of ascending contour, of little variation, wide ambitus, many interruptions, frequent leaps, including extreme, use of angular and wave-like intonations	Fairly narrow ambitus, prevalent steps with occasional leaps, rising intonation, wave-like shapes
	2. Harmonic pitch (concurrent "vertical")	Concordant/discordant combination of tones	Major, diatonic, prevalence of medium size perfect intervals of fourth and fifth	Minor, chromatic, dissonance, prevalence of small intervals	Minor, chromatic, strong dissonance (up to atonal), large intervals, esp. major seventh, augmented fourth	Minor, strong dissonance, diverse intervals	Major, diatonic, general consonance
	3. Form (thematicism)	Sameness/diversity	Relative simplicity	High complexity	High complexity	Relative complexity	Quite low complexity
2. Time	4. Tempo	Fast/slow metric pulse	Mostly fast, with very restricted use of rubato	Slow, with strong rubato and prevalence of ritenuto	Fast, with minimal rubato and general tendency to use accelerando	Fast, with strong rubato and many abrupt changes	Slow and moderate, strong rubato but no abrupt changes
	5. Rhythm	Short/long relation of tones	Sharp contrasts of tones, yet smooth succession of groups of tones	Smoothed contrasts of tones, yet firm patterns, many long tones	Very sharp contrasts of tones, complex patterns with sudden changes, many short tones	Abrupt changes of tones and rhythmic groups, with overall diversity in rhythm, many short tones	Smoothed contrasts of tones, yet with rhythmic diversity, many long tones
	6. Meter	Short/long periodicity of stressed beats	Strong regularity with minimal deviations	Tendency to irregularity	Tendency to syncopation and irregularity	Pronounced irregularity and variability	Strong regularity with moderate variability
	7. Articulation	Styles of attaching/detaching of successive tones	Prevalence of staccato, with overall great diversity of styles	Prevalence of legato, with little diversity of styles, many pauses	Prevalence of staccato, with moderate diversity of styles, occasional legato	Prevalence of staccato (stressed), with great diversity of styles, many pauses	Prevalence of legato, with little diversity of styles, many pauses
3. Amplitude	8. Dynamics	Loud/soft relation of tones	Prevalence of loud and medium loud, with limited crescendo and diminuendo	Prevalence of soft and medium soft, with medium crescendo and diminuendo	Prevalence of very loud, with very little dynamic change, accents tend to fall on unstable tones	Prevalence of soft and medium soft, yet with diverse dynamic changes, mostly abrupt	Prevalence of medium soft, few dynamic changes, accents tend to fall on stable tones
4. Timbre	9. Register	Relation of homogenous groups of tones in their tonal quality	Prevalence of bright register, raised singing formant, brightness	Prevalence of bright register, low singing formant, dullness	Bright register with little changes, raised singing formant, harshness	Prevalence of bright register with abrupt registral changes, general mellowness	Prevalence of dark register, lowered singing formant, general mellowness
	10. Harmonicity, attack and vibrato	Periodic/non-periodic spectral content	Harmonic richness, fast attack, medium vibrato with mid-fast rate	Harmonic scarcity, slow attack, small vibrato range with slow rate	Harmonic richness, much spectral noise, fast attack and decay, large vibrato range, mid-fast rate	Contrasts of harmonic richness and scarcity; gentle attacks; small irregular vibrato with fast rate	Harmonically moderate, slow attacks, small vibrato range with mid-fast rate

This table is compiled based on a number of meta-reviews of experimental research on emotional responses to listening to music (Gabrielsson and Lindström, 2001; Gabrielsson and Juslin, 2003; Juslin and Laukka, 2003; Juslin, 2005). The data is categorized according to the musicological nomenclature: all acoustic attributes are broken into 10 AEs across 4 acoustic domains. The aspect of texture is missing, because it was not controlled for in the experimental studies of the acoustic structural patterns that characterize "musical emotions". The aspect of harmonicity constitutes an organic part of the aspect of instrumentation, listed in the beginning of this paper. This potentially confusing mismatch occurs as a result of the discrepancy in musicological and psychoacoustic scholarships: as a rule, musicians are ignorant of harmonicity, while psychoacousticians are ignorant of instrumentation. Harmonicity can be defined as the extent to which the spectrum of a complex tone is made of its component frequencies that are integer multiples of its fundamental frequency (FF). This is usually measured as the ratio of harmonics to noise. Slow attack and great vibrato generally tend to reduce harmonicity in a monophonic tone.

TABLE 4 | Acoustic attributes of typical animal vocalizations used by different species to display their affective state, grouped according to AEs of human music.

Acoustic domain	AEs of calls	Range of an AE	Pleasure features	Dissatisfaction features	Aggression features	Alarm/disturbance features	Appeasing features
1. Frequency	1. Pitch (melodic contour)	High/low relation of tones in a call	Great variability of frequency modulations	Descending contour accompanies the increase in emotional intensity	Prevalence of descending contour, wide ambitus; yet use of rising pitch, few modulations; leaps accompany the rise in emotional intensity	Prevalence of ascending contour, wave-like and flat melodic shapes, great modulations; leaps accompany the rise in emotional intensity	Prevalence of ascending contour, narrow ambitus, great variability of modulations
	2. Harmony	n/a	n/a	n/a	n/a	n/a	n/a
	3. Form (complexity)	Compact/expanded call structure	n/a	Long and short calls, longer as intensity increases	Long calls, but generally shorter than alarm calls	Longer calls	Longest calls
2. Time	4. Tempo	Fast/slow pace of calling	Slow	The slowest, but speeding up as emotional intensity increases	Fast, speeding up more as emotional intensity increases	Fast, speeding up more as emotional intensity increases	Fast
	5. Rhythm	Short/long duration of tones in a call	Generally short tones	Generally long tones, often repeated	Prevalence of long tones, longer as emotional intensity increases	Prevalence of short tones, longer as emotional intensity increases	Short tones, very repetitive
	6. Meter (regularity)	Periodicity of calls	Regularity	Irregularity	Irregularity	Irregularity	Regularity
3. Amplitude	7. Articulation	Attaching/detaching of tones in a call	n/a	Distinct pausing	Abrupt ending of a call, staccato, shortened pauses	Smoother ending of a call, staccato, shortened pauses	Smoother ending of a call, legato, distinct pausing
	8. Dynamics	Loud/soft relation of tones in a call	Generally soft	Medium soft, the increase in amplitude accompanying the increase in emotional intensity	Very loud, with abrupt onset and offset, the increase in amplitude accompanying the increase in emotional intensity	Medium loud, without abrupt onset and offset, greater increase in amplitude as emotional intensity increases	Soft, rhythmic sequences of amplitude modulation
4. Timbre	9. Register	Bright/dark relation of homogenous groups of tones in a call	Dark register (e.g., purrs and growls)	Dark register (e.g., meows), no change as emotional intensity increases	Very dark register, but increase in intensity engages brighter register (more so than in alarm calls), wide range	Mostly dark register, sometimes very dark, but increase in intensity engages brighter register	More of dark register (in some species bright), no change as emotional intensity increases
	10. Harmonicity	Periodic/non-periodic spectral content of tones in a call	n/a	Tendency to use pure tones (whistle-like or flute-like)	Harsh, non-harmonious, noisy, broadband (growl, bark), harsher as emotional intensity increases	Mostly pure tone (whistle-like), can be noisy and broadband; both, harsher and/or richer in harmonics as intensity increases	Harmonic richness (grunt), mostly noisy, but much less harmonious as intensity increases (coo)

The data for this table is compiled from numerous meta-reviews (Morton, 1977; Peters, 1984; August and Anderson, 1987; Snowdon, 2003; Briefer, 2012; Altenmüller et al., 2013; Zimmermann et al., 2013; Snowdon et al., 2015). According to the classification scheme of Brudzynski (2013), human and animal affective states are equated in the following ways: human "happiness" is equated to animal "pleasure" (satisfaction), human "sadness"—to animal "dissatisfaction" (social isolation from a bonded party), human "anger"—to animal "aggression" (agonistic behavior, conflict with display of threat or combat), human "fear"—to animal "alarm/disturbance" (anxiety at the presence of threat or intimidation by a novel environment), human "tenderness/love"—to animal "appeasing" (affiliation—physical contact without agonistic behavior, e.g., grooming, and play). Those acoustic features that agree between human and animal expressions of the same affective state are marked blue, whereas the disagreeing features—red. Features that are not covered in research literature are marked "n/a." The aspect of "harmony" is clearly not applicable to animal vocalization. The aspect of "form" bears only distant relation to "musical form": AC's compactness loosely corresponds to simplicity of structure, whereas lengthiness—to complexity. Aspect of "meter" also finds only partial correspondence in regularity or irregularity of call units in the AC bouts. The timbral coloration is reflected by the aspect of "harmonicity" rather than "instrumentation" that manages timbre in human music.

TABLE 5 | The acoustic attributes of typical expression of 5 basic emotions in human music that find no correspondences in animal communication (based on **Tables 3, 4**).

Acoustic domains	AEs of music	Range of an aspect	Happiness features	Sadness features	Anger (aggression) features	Fear (anxiety) features	Tenderness (love) features
1. Frequency	1. Melodic pitch (consecutive "linear")	high/low relation of tones	prevalence of ascension in the overall diverse contours, leaps, zigzags, sharpened intonation, wide ambitus	smooth contours, mainly steps, wave-like shapes, flat and falling intonation, narrow ambitus	prevalence of ascending contour, with sharpened shape of melodic contours	little variation in contours, interruptions, use of angular shapes, wide ambitus	prevalence of descending contour, mostly steps with occasional leaps, rising intonation, wave-like shapes
	2. Harmonic pitch (concurrent "vertical")	concordant/discordant combination of tones	major, diatonic, prevalence of perfect 4th and 5th	minor, chromatic, dissonance, mostly small intervals	minor, chromatic, strong dissonance, wider intervals (major 7th, aug. 4th)	minor, strong dissonance	major, diatonic, general consonance
	3. Form (complexity)	sameness (simplicity)/diversity (complexity)	relative simplicity	high complexity	n/a	n/a	low complexity
2. Time	4. Tempo	fast/slow metric pulse	fast, with very restricted rubato	strong rubato and prevalence of ritenuto	minimal use of rubato	strong rubato and many tempo changes	slow and moderate, with moderate rubato
	5. Rhythm	short/long relation of tones	sharp contrasts of tones, yet smooth succession of groups	smoothened contrasts of tones due to frequent use of rubato	very sharp contrasts of tones, complex patterns with sudden changes	abrupt changes of tones and groups, general prevalence of diversity	smoothened contrasts of tones due to rubato use
	6. Meter	periodicity of beat grouping	minimal deviations	n/a	tendency to syncopation	pronounced variability	moderate variability
3. Amplitude	7. Articulation	styles of attaching/detaching of successive tones	prevalence of staccato, great diversity of styles	prevalence of legato, with little diversity of styles	moderate diversity of articulation, occasional legato	stressed staccato (marcato), great stylistic diversity, many pauses	little diversity of styles
	8. Dynamics	loud/soft relation of tones	prevalence of loud and medium loud, with limited crescendo and diminuendo	prevalence of soft, with medium crescendo and diminuendo	mostly loud, with very little dynamic changes, accents tend to fall on unstable tones	prevalence of soft and medium soft, yet with diverse dynamic changes, mostly abrupt	few dynamic changes, accents tend to fall on stable tones
4. Timbre	9. Register	bright/dark relation of homogenous groups of tones	prevalence of bright register, raised singing formant, brightness	prevalence of bright register, lowered singing formant, dullness	prevalence of bright register, with little change, raised singing formant, harshness	prevalence of bright register with abrupt registral changes, mellowness	prevalence of dark register, lowered singing formant, mellowness
	10. Harmonicity, attack and vibrato	periodic/non-periodic spectral content of tones	harmonic richness, fast attack, medium vibrato with mid-fast rate	slow attack, little vibrato with a slow rate	harmonic richness, fast attack and decay, large vibrato with a mid-fast rate	gentle attacks; little and irregular vibrato with a fast rate	slow attacks, little vibrato with a mid-fast rate

These attributes constitute a stock of TO features developed in the process of evolution of human music from hominin musilanguage. This includes changes in vertical harmony, in metric pulse, and in complexity of musical form; contrasts in melodic contour, in directionality of melodic intervals (sharpening for ascending, flattening for descending dyads), and in thematic material; diversity of rhythm, articulation and tempo; and ambitus size. Animal vocalizations do not seem to engage these categories in meaningful differentiation of calls.

1996). Most children pass through a music-babbling stage when 12–18 months old (Gembris, 2006). Universality of babbling suggests the universality of prolonged sensorimotor trials in music-making before semiotic rules are formed. Babbling abstracts melodic directions and intervals, allowing an infant to master particulate semiosis. Similarly, early humans had to long experiment with meaningless melodic play for the SFTO conventions to emerge.

CROSS-CULTURAL “SCRIPTS” IN THE FORMATION OF SEMIOTICALLY FUNCTIONAL TONAL ORGANIZATION

Tool-making technologies (Ambrose, 2001) and “social scripts”—i.e., fixed generalized patterns of social behavior (Aiello, 1998)—most likely served as syntax precursors by providing explicit models for combining numerous elements into a structured sequence (Wildgen, 2004). Paleolithic proxies for syntactical language include composite tools (Ambrose, 2010), fire (Brown et al., 2009), knot-making (Camps and Uriagereka, 2006), cooperative hunting (Chase, 2006, 52), symbolic behaviors (McBrearty and Brooks, 2000), and burials (Mellars, 2004). The same proxies apply to syntax-related features of musical TO. All the AEs of music listed above (perhaps, except harmonicity) are engaged in the syntactic organization of music. Phrasal ends are usually marked by descending pitch, lower register, more concordant harmony, slowing of tempo, longer rhythmic value(s) placed on metrically strong time, reduction in loudness, and clear caesuras in articulation which separate the end of one formal unit (phrase, sentence) from the beginning of the following unit. In addition, there is evidence of a link between structures of tonal and social organization in indigenous societies (Blacking, 1967; Davidson, 1970; Lomax, 1977; Berliner, 1993; Arom and Voisin, 1997; Kubik, 1999)—which indicates that social structures might have also served as proxies for music syntax.

Making bone “flutes” was extremely tedious, demanding skills and expertise (Münzel and Conard, 2009). Why to invest into a “pitch toy” rather than to merely vocalize?

Cave-inhabitants must have supported flute-makers in the same way as they supported cave-artists—their exquisite labor required narrow specialization, precluding participation in hunting/gathering. In animistic ideology, depictions linked hunters to prey, providing means to benefit the outcome of hunting (Hauser, 1999, 1–4). Magic—not aesthetics—governed rock art, turning depiction into a shamanic occupation²⁶. Shamanic music resembles shamanic depiction by cross-linking the signified to the signifier (Hubbard, 2003). In northern shamanic traditions, both melodic and pictorial contours are believed to affect the corresponding real objects (Novik, 2004, 67–85). Archeological evidence also links most resonant locations in caves with rock art in Paleolithic sites, suggesting the

combined ritualistic use of images and music (Reznikoff, 2008; Morley, 2013; Mills, 2016). Hence, a Paleolithic “flute” was most likely a talisman used in rituals (Marshack, 1990). Its manufacturing from the bone of a particular animal (Wyatt, 2016) must have carried more significance for Aurignacians than the pitches it produced.

For melodic semiosis to occur, rhythm and directionality must first be abstracted into AEs. Abstraction of directionality probably followed rhythm: *salience of the melodic direction depends on rhythmic values, but not vice versa*. Tracking the melodic contour within the tonal “grid” constitutes the backbone of melodic organization (Deutsch, 2013), just like tracking the rhythmic grouping within the metric grid supports the temporal organization (Large, 2008). Reference to tonal hierarchy interferes with rhythm-metric perception by biasing the attention toward pitch (Prince et al., 2009). Their conflict indicates that users of non-Western music discriminate rhythm-meter better than users of Western tonality (which agrees with the observations of ethnomusicologists). This suggests that *frequency reference-frame* emerged later than *rhythm-metric*.

Developmentally, acquisition of rhythmic hearing usually precedes melodic hearing (Shatkovsky, 1986). Infants seem to acquire rhythm-discrimination skills earlier than pitch-discrimination (Trehub and Hannon, 2006)²⁷. The perceptual foundations of rhythm/meter are manifested just a few days after birth, as a part of developmentally crucial rhythmic interaction between infants and caregivers, occurring spontaneously and requiring little experience—reflecting its evolutionary importance for bonding (Trainor and Hannon, 2013). In verbal acquisition, rhythm too obtains semantic functionality earlier than prosodic contour (Shvachkin, 1948). According to the vast data collected through administration of early musical education in USSR, rhythmic hearing lays the foundation for vocal musical skills—followed by learning to reproduce melodic contours (Kirnarskaya et al., 2003, 168–170). Impressions that not only rhythm can influence melodic perception by directing the attention to longer tones, but that melodic features carry the reverse influence onto rhythm, are based on the misnomer between rhythm and meter (McAuley, 2010). Melodic intervals, contours, and “tonal accents” help to infer meter, but play no major role in identification of rhythmic values. On the contrary, judgments of melodic similarities are significantly affected by rhythm, especially in folk music (Eerola et al., 2001)²⁸. Even for experienced Western musicians the distinction between

²⁷The earliest age when infants show the ability to recognize changes in pitch contour is 5 months (Chang and Trehub, 1977). Majority of studies demonstrate such capacities in older children, 6 months and up (Trainor and Hannon, 2013). The ability to recognize changes in rhythmic values of a familiar music seems to emerge quite earlier—at 2 months of age (Demany et al., 1977).

²⁸Metricality, along with tonality, influence primarily the Western musicians: non-musicians process melodic contours mostly according to the distribution of longer rhythmic values (Monahan et al., 1987). Non-trained listeners simply cannot ignore rhythm—as it governs their melodic recognition (Jones and Ralston, 1991). Majority of young and inexperienced listeners at first parse melody by rhythm and only then by pitch contour and mode (Halpern et al., 1998). Tempo/rhythm descriptors are much more prevalent in listeners’ judgments of thematic similarity than of pitch contour (Addessi and Caterina, 2000; McAdams, 2004).

²⁶Thus, newer paintings often covered the older ones: hiding the underlying image did not matter—once painted, an image was “brought to life,” and stayed “alive,” even if masked—just as a person who disappears from our sight does not die (Uspensky, 1995, 173–181).

rhythms is more salient than the distinction between pitches (Monahan and Carterette, 1985)²⁹.

Important Upper Paleolithic cultural proxies promote the abstraction of rhythm—not of melodic contour. Metric pulse is transposable from bipedal gait into such a common Paleolithic activity as stone-knapping. Each knapper prefers his own tempo and rhythm (Whittaker, 1994, 81)—quite similar to individual gait preferences (Whittle, 2007). Knappers' heartbeat provides a metric reference (Zubrow and Blake, 2006). Two knappers might have accidentally discovered the expressive capacity of rhythm through their entrainment, thereby forming the world's first musical instrument (Montagu, 2004). Group "musical" knapping was observed amongst Aboriginal women in Queensland (Duncan-Kemp, 1952, 27). Rock slides and gongs are drummed across the globe in rituals related to fertility cults (Fagg, 1997, 38). The ritualistic context provides feeling of contentment or awe, abstractable into a semantic value for the knapping/grinding sound, turning its rhythm into a sign—and the archeological evidence for collective stone-knapping is present in Neolithic sites at Sanganakallu-Kupgal, India (Boivin et al., 2007). Even earlier, stationary lithophones were drummed in Solutrean-Magdalenian caves (pecked rock surfaces were found in Africa)—suggestive of the existence of portable lithophones (Blake, 2011). The weird-sounding cave echo might have prompted specific affective connotations (Cross and Watson, 2006).

Unlike rhythm, *pitch directionality* finds no proxies in the Paleolithic³⁰. A set of meaningful pitch contours could have originated in verbal prosody, but paleolinguists connect the development of the fully phonemicized semantic languages to population growth *after* the Last Glacial Maximum (Robb, 1993). Deeply social, language is imperative for accumulation of knowledge, which depends on population density to avoid "bottlenecks" due to climate changes and extinctions. Cultural evolution stabilized only after 50 kya—most certainly, because of the advancement of language (Klein, 2009). In all the prehistory, the transition to Holocene stands out as the grand leap in innovation, called to subsist an ever-growing population (Richerson et al., 2009). Powell et al. (2009) developed a demic model to estimate the critical population density capable of sustaining the innovation growth to offset the innovation loss: for Europe it was 45 kya. Prior to 20 kya, prehistory consisted of a chain of major discontinuities in cultural transmission (d'Errico and Stringer, 2011). Technically, the archeological concept of "culture" applies only starting from the Neolithic (Probst, 1991, 227).

²⁹Of course, the influence of rhythmic features on the judgment of melodic similarity is far from being simple and direct. Other factors, such as tempo and harmonization, can affect the extent of autonomy of temporal and frequency-related aspects of music (Prince, 2014).

³⁰There are accounts of "tone-painting" where the contour of the hills is represented through the melodic contour in songs of indigenous hunter/gatherers of Northern hemisphere (Krushanov, 1987, 234) whose life style is comparable to that of Aurignacians. However, the idea of such representation most probably was inspired by the need in mnemonic aid in long-distance navigation during migrations with reindeer herds, which doubtfully existed earlier than a few thousand years ago (see the last chapter). Such tradition had chances to survive the ongoing extinctions in harsh climate only as a part of a reliable subsistence strategy for a fairly large population.

The first archeological symbolic "culture" of pan-European scale is the Gravettian, whose common trans-European traits are both socio-economic and spiritual, with regional differences confined to the material techno-complex (Kozłowski, 2015). The continent-wide cultural unity is evident in the omnipresence of "Gravettian Venuses" over most of Europe (Soffer et al., 2000)³¹. Denser population turns language from means of inter-group cooperation that compensates for local ecological deficits into a life-long ethnic marker, akin to the cranial configuration (Robb, 1993). Personal ornaments in Gravettian burials manifest similar function of the "ethnic badge," differentiating age classes across the puberty threshold (Zilhão, 2014).

Social restructuring by ethnos and age hardly occurred without the involvement of music, closely affiliated with funeral and puberty rites. The Gravettian funerary practice strongly suggests the existence of burial rituals regulating the emotive interaction between the group's members, the dead, and the landscape as part of a greater ritual system, underpinned by cosmological beliefs (Pettitt, 2010). The remnant of such socio-eco-cosmological interconnection with TO, providing its semantic foundation, is the ancient doctrine of **ethos**³²—renowned in Hellenic civilization (Mathiesen, 1984), but certainly much older (Farmer, 1965) and geographically wider (Manuel and Blum, 2011). The roots of ethos must lie in the Gravettian trans-European spiritual unity.

CONTRIBUTION OF MULTI-DIMENSIONAL AND MULTI-EMOTIVE SEMIOSIS TO THE EVOLUTION OF MUSIC

Human melodic universals remap animals' universals. Animal anger is characterized by descending contour, whereas animal appeasing—by ascending contour. Music reverts the registers for happiness, sadness, fear, and anger from low to high. Why?

Music contributes to the conservation of knowledge by bonding social groups and incentivizing linguistic communication. This capacity came in play after the Younger Dryas (11 kya), when global warming enabled colonization of Eurasia. Widely dispersed populations created a few flexibly bounded

³¹Broad-scale technological clustering originated in the earlier Aurignacian tradition—attributed to the long-term influence of the ethnolinguistic variation. Forming of a continental culture during the Gravettian indicates the increased language contacts between different "clusters," establishing pan-European networks of informational exchange (Zilhão, 2014).

³²The term "ethos" was coined in Archaic Greece, where it originally meant "custom," but by Classic times it obtained the meaning of a certain affective "character," associated with a particular musical melodic mode. "Ethos" embodied the consensus within a community as to which affective states would be generally "good" or "bad" for that community. The doctrine of "ethos" is closely related to the concept of "harmony of spheres," attributed by Hellenic sources to Pythagoras, who presumably learned it from Babylonians. The discussion of ethical value of this or that musical emotion and its suitability for astrological dispositions constituted an important part of public discourse in Ancient civilizations of Near and Far East, as well as Central Asia.

“social territories³³,” developing the dialect continuums by linkages among groups due to intermarriages during population shortfalls (Robb, 1993). Population growth and sedentism accompanied rapid neolithization, promoting ethnogenesis and thereafter fissioning language into language families as regional cultural differences cumulated (Robb, 1991). Such line of development benefited from the social bonds established by music.

The absence of music-like particulate emotional communication must be one of the reasons why chimpanzees do not accumulate cultural traditions. Some chimpanzees acquire a culture of tools but due to the lack of transposability and abstraction cannot transmit it (Whiten, 2011). However, it is music, not language, that engages reproduction, transposability, and abstraction of idiomatic patterns of each of its AEs.

Human remapping of pitch encoding most probably originates from the continuous practice of:

- **Frequent rotation between aesthetic emotions:** ACs prioritize negative emotions due to greater urgency of their triggers (August and Anderson, 1987). Human music is balanced between negative and positive expressions because of the mentalization of aesthetic emotions (Juslin, 2013). Expression of negative emotions can be pleasurable whenever it occurs in a non-threatening situation, is aesthetically appealing, and seems somehow useful or appropriate (Sachs et al., 2015). Thus, abstraction of emotions enables older children to learn to appreciate sad music (Schubert and McPherson, 2015), whereas at 5–7 months, infants overwhelmingly prefer happy to sad music (Nawrot, 2003). By 4 years, children start intentionally expressing positive and negative emotions in singing (Welch, 2006), distinguishing happy/sad and angry/fearful musics (Eerola and Vuoskoski, 2013). This line of development is also applicable to cultural evolution. In both cases, changes of musical emotions sharpen *contrasts in patterns of their musical expression*—resembling *phonemic oppositions in phonology*.
- **Multifactorial musical semiosis:** Zero- and first-order intentionality separates animal signals from second-order intentionality of humans (Seyfarth and Cheney, 2017). Although non-human primates can coordinate the produced signal with the listeners’ response, modulating the acoustic features of their calls accordingly, modulation usually engages a single parameter—falling short of the complex multidimensional nature of emotional communication in verbal prosody and music (Filippi, 2016). Simultaneous interactive control over *multiple AEs* is peculiar to music alone. Thus, in expression of anger, prevalence of ascending contour and high register conveys physical strain, while the side-effect of their monotony is compensated by a diverse contrasting rhythm and spectral

content, projecting agitation (Table 2). AC’s anger does not engage such interaction. It conserves a *unifactorial* timbral quality³⁴ (Table 3).

All AEs differ in musical expression of love (Figure 2) and anger (Figure 3), as evident in musograms³⁵ of indigenous Siberian songs that Russian theorists believe to represent the earliest forms of TO (Alekseyev, 1976, 1986; Brodsky, 1976; Zemtsovsky, 1983; Mazepus, 1993; Mazepus and Galitskaya, 1997; Novik, 1999; Zabolotskaya, 2009; Dobzhanskaya, 2011, 2016; Nikolsky, 2015b; Sheikin, 2017, 2002).

Unlike the expression of love, anger engages a wider ambitus, greater leaps, contrasting registers, harsh timbres, loudness, shorter and richer rhythms, reduced regularity and tonal stability, increased tempo fluctuations, staccato articulation, and thematic complexity (Figures 2, 3). However, gorillas express anger differently: “call-motifs” remain always isolated and slow-paced, featuring neither a clear melodic contour (due to its enormous bandwidth) nor rhythm (Figure 4).

If humans *consciously* manipulate numerous *learned* expressive parameters in music, animals *instinctively* “center” on a single biologically “hard-wired” parameter to reflect their emotional intensity. Human infants start their development at the same level where animal cubs start theirs, but quickly advance. Newborns employ just 2 vocalization types: negative and positive (Loewy, 1995). Cries of hunger, cold, distress come first as biological reflexes (Zeskind, 1985). However, the similarity of an infant’s supralaryngeal vocal tract to that of the primate cub’s does not stop the infants from trying to imitate his/her caretaker’s vocalizations (Lieberman, 1985)³⁶. Infant cries start varying in temporal and frequency characteristics as the infant ages (Papoušek and Papoušek, 1995). Loudness, timbre, register, attack speed, FM range, and harmonicity are progressively mastered as markers of different cry-types (Golub and Corwin, 1985). An infant builds a repertory of melodic contours assigned to specific situations and used as building blocks to inform the caretaker about his/her state and to receive a desired treatment (Wermke and Mende, 2009). Such ongoing two-ended communication lies at the heart of musicality (Trevarthen, 2019).

Call/cry-repertory building appears to be universal in human development (Wermke et al., 2007), very likely paralleling the phylogenetic evolution of music (Foster, 1994). Similarities between the structure and function of human and non-human vocalizations were discovered in crying, motherese, and babbling (Snowdon, 2003). Fluent switching from one cry-type to another, corroborated by the caretaker’s response,

³⁴It could be said that an animal “centers” (i.e., focuses) on a single aspect of vocal expression, conserving the extent of increase or decrease in intensity of the psychophysiological state that is associated with that vocal expression. This is yet another parallel between AC and the vocalization of a sensorimotor human infant. This is in contrast to the ability of an adult human to simultaneously conserve multiple dimensions of changes in multiple AEs in music.

³⁵For a comprehensive analysis of those musical examples that were selected for musograms in Figures 2, 3, 4, and 7, see Appendix 2 “A Comparative Structural Analysis of Musograms” in **Supplementary Material**.

³⁶Similarity of the anatomy of the supralaryngeal vocal tract of the human baby and the ancestors of *Homo sapiens* provides yet another justification for seeking the TO model of hypothetical Paleolithic music in the musical babbling of 1–2-year-old infants.

³³Thus, Peter Bogucki counts as few as 14 Mesolithic “social territories”—i.e., regions differentiated by the material culture as manifested by archaeological evidence—spread out over the entirety of Western Europe during its transition from the Boreal to the Atlantic periods, c. 7500 kya (Bogucki, 1988, 41–46).

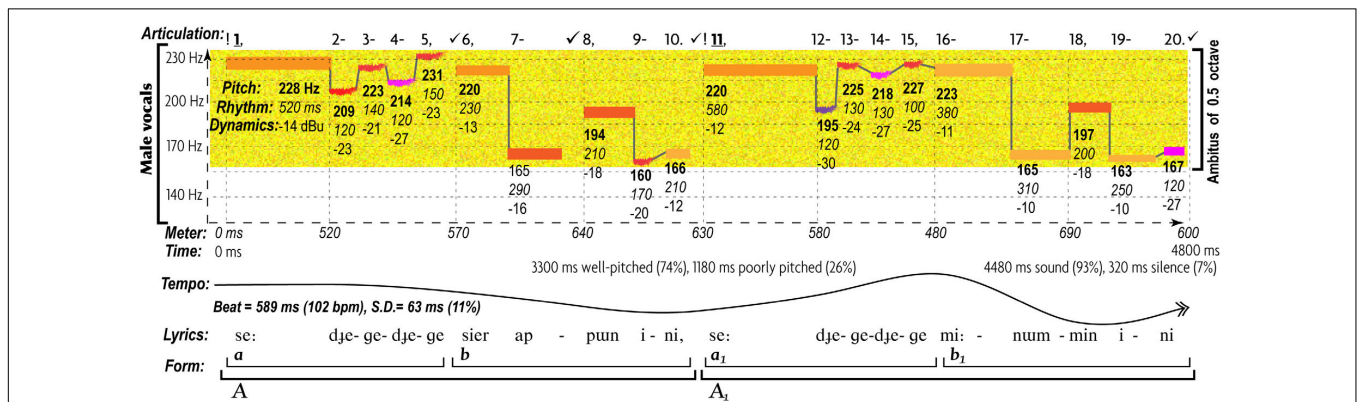


FIGURE 2 | Characteristic patterns of AEs in expression of love in a Yakut traditional lyrical song “Sae Dyige” (may be auditioned at <http://chirb.it/sNcgG1>). By Juslin’s (2005) classification this song fits the “love” music category—in agreement with its lyrics, describing how a woman is anticipating visits of her multiple lovers (Aleksyeyev and Nikolayeva, 1981, 86). The musogram follows the same conventions as **Figure 1**, with minor additions due to the less definite use of pitch in the purely vocal music. Tones of low spectral periodicity (noisy or spoken-like) are represented by fuzzy strips in contrast to high periodicity, represented by rectangular bars. The number under each pitch displays its frequency value in bold, its duration in italic, and its maximal amplitude (the highest value of any of its spectral constituents) in regular font. The lyrics are given in the phonetic transcription. There are two contrasting motifs: “a”—a sustained low anchor tone (tonic function), followed by rapid alternation of steps with rising intonation; and “b”—two descending intonations, the first of which leaps to the alternative anchor (dominant function to mark a cadence), while the second steps down and then gently rises. These two motifs make up a call-like phrase that is regularly repeated. Song is characterized by a narrow ambitus (half-octave), mid-low register, high harmonicity, low complexity, moderate tempo (102 bpm) with little rubato (11%), diverse rhythm (usage of four rhythmic values), regular meter, overwhelming legato (97%), and scarce dynamic changes. For more detailed discussion, see Appendix 2 “A Comparative Structural Analysis of Musograms.”

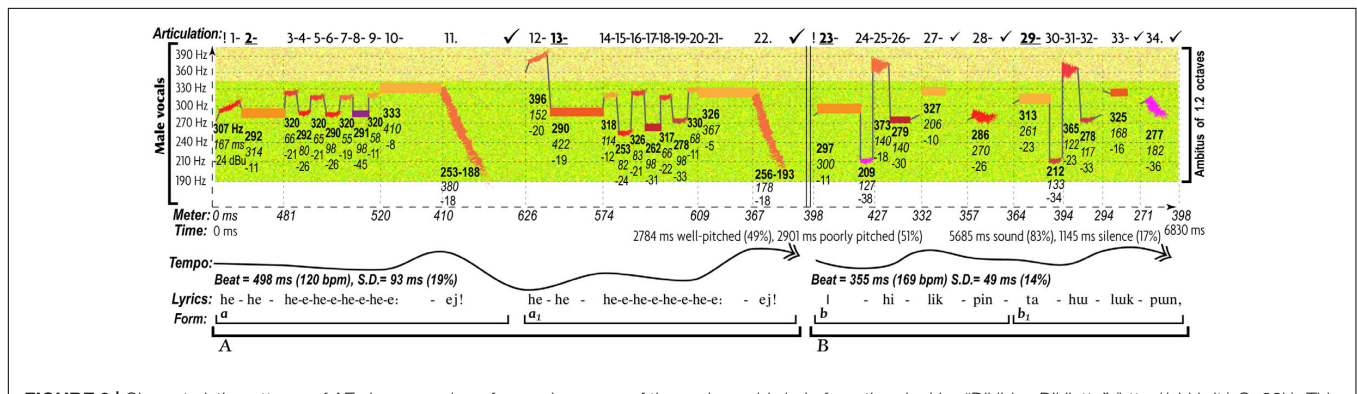


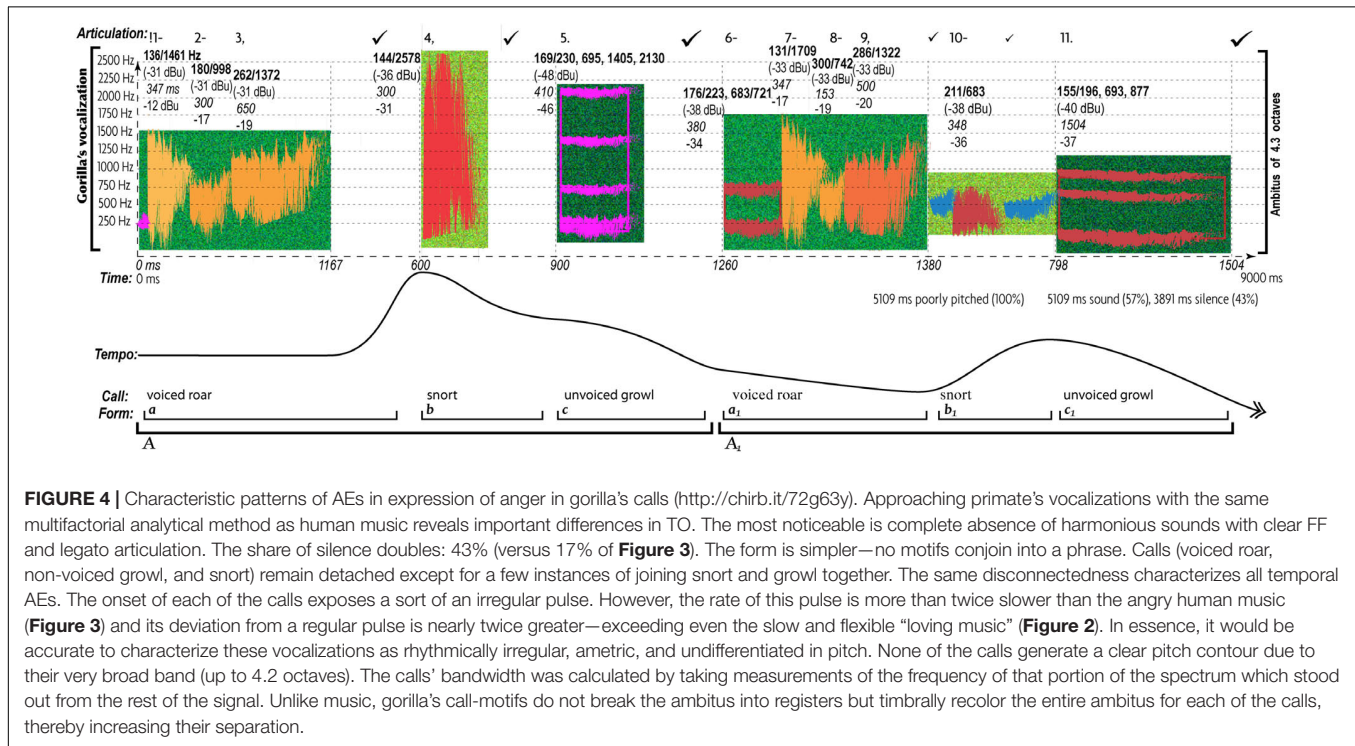
FIGURE 3 | Characteristic patterns of AEs in expression of anger in a song of the underworld virgin from the olonkho “Djiribina Djirilatta” (<http://chirb.it/sCq02k>). This excerpt from the traditional Yakut epic expresses anger of the evil sorcerer toward the heroine, challenging her to a fight (Aleksyeyev and Nikolayeva, 1981, 35). Structural descriptors of most aspects of this song fall in the category of “angry” music (Juslin, 2005). The acoustic markers of all AEs contrast those in **Figure 2**. The ambitus is over twice wider. There are two registers instead of one: low singing and high “shouting”, both are higher than **Figure 2**. The share of well-pitched sounds in the overall duration of music is reduced by 34%. The share of staccato articulation is increased (by 142% in the duration of silence and 40% in the number of pauses). Tones are overall shorter and 50% more diverse in time values, with contrasts between rhythmic groups. The tempo contains abrupt switches, the fastest of which is 66% faster and 73% more variable (rubato) than **Figure 2**. Intonations feature wide leaps, on average 70% wider than **Figure 2**. Thematically, the music is more diverse and complex, using two contrasting materials, “A” and “B” (**Figure 2** had only one). Timbre is harsh (a heightened larynx and intensified pressure).

prompts the cross-examination of the cries’ acoustic parameters. The intensity of temporal expression usually matches pitch expression (frequent leaps require faster tempo to convey excitement and emergency—otherwise the caretaker is not “convinced” to respond urgently enough). Together, the projection of feedback and memorization/cross-relation of cry-types establish the acoustic oppositions between AEs of common musical emotions.

What diverts music from AC is the radical change in communication framework. Animals communicate “face-to-face” in situations that demand immediate action, which

selects signals effective in expressing rapidly changing motivational states, with clear gradations in their intensity (Morton, 1977). Such signaling prioritizes ease of detection, speed of interpretation, signal’s brevity, and a single salient gradient AE (Maynard-Smith, 1976). High redundancy and stereotypicality of selected signals often “fix” them (Simpson, 1997). This precludes combinability of AEs and calls, enabling “dishonest” calling.

Unlike animal calls, traditional indigenous music normally never “lies” (Nikolsky, 2016, Appendix III). A performer, as a rule, expresses emotions he actually feels—even when



impersonating an epic protagonist or a spirit, the singer becomes temporarily “possessed” by them (Novik, 2004, 272). “Putting on an act” is a prerogative of post-Renaissance Western classical performance tradition, and even there the performance canon demands “method-acting” to convince the audience in the realism of musical emotions (Nikolsky, 2015a)³⁷. A non-western traditional song usually appears “westernized” to the indigenous audience when “acted out” formally (Zemtsovsky, 1983). Folk “cover-songs” necessarily engage the performer’s “direct”—rather than “indirect” or “scripted” speech (Zemtsovsky, 1979)³⁸.

Insincerity and falsehood in musical expression did not present a critical issue prior to the 1760s (Charlton, 2009). They both attracted public discourse as a systemic aberration peculiar to a specific class of music (rather than a “defective” sample) only after the entertainment industry became institutionalized (Dahlhaus, 1989, 314). Rise of mass production made “emotional faking” a norm for commercial

popular music—explicitly codified in Irving Berlin’s composition standards (Suisman, 2009)³⁹. So, music started as a decidedly “honest signal” (Levitin, 2009, 141–6) and only recently adopted “acting”—albeit, hardly enough to declare music fundamentally “dishonest”⁴⁰.

Jointly, multi-dimensionality of music and emotional contagion make lying difficult. Music always integrates listeners and performers, and this togetherness promotes sincerity. The particulate structure of musical semiosis effectively reveals dishonesty: at least some of AEs’ insincere expressions are bound to contradict each other, prompting a resolving interpretation. But what in the cultural evolution could have spurred the inclination for aspect-matching?

³⁷Demonstration of musical “method-acting” can be found in the video clip of Andrei Gavrilov performing Rachmaninov’s Prelude in g minor, op.23, No.5 <https://www.youtube.com/watch?v=T3AEfMMYH6A>. Especially telling is the pianist’s facial expressions, as he is getting up from the piano bench after completing his performance—he continues to remain in his “role.”

³⁸Some indigenous traditions have developed professional forms of musical art which require aesthetic evaluation (e.g., Tatar, Kazakh, Mongolian). However, they still fundamentally differ from Western classical music by not taking a musical work as a “script” created by the composer for the performer to adhere to (Zemtsovsky and Kunanbayeva, 2011). Only the Western musician is trained as part of his occupation to accurately “execute” the composer’s script while being aware of the fictiveness of its emotional content. However, application of such treatment to a folk cover song is most likely to come across as fundamentally “inauthentic” and detrimental to the song (Moore, 2011).

³⁹Berlin’s rules Nos. 3, 6, and 9 call for the composer to please the consumer at the cost of insincerity: “the ideas and lyrics must suit *either a male or a female*, so both sexes want to *buy a song*,” “music and lyrics *must* have to do with *things common to everyone*,” and, most explicitly—“songwriter must look upon the song as a mere business, *not take music to heart*.” Berlin’s rules break away from the Western composer’s “canon,” established since the introduction of “*musica reservata*” in the 16th century (Meier and Dittmer, 1956). For this reason, Berlin’s approach provoked criticism of the American popular music in toto, seen by connoisseurs of art music as a “sweet lie” sold (for profit) to the mass audience to replace music that is “truthful” yet unpleasant in revealing “social truth” (Adorno, 1942).

⁴⁰As far as I know, Trehub (2008) remains the only scholar who believes that music, in general, operates by having the performer emotionally deceive the audience. Other scholars who point out that a professional performer can evoke emotions that he/she does not actually feel, realize that this discrepancy is possible only in music that segregates the listener, the performer, and the composer. *This solely happens in Western classical music*. And even within this tradition “deceiving” the audience is still regarded as a fault to be avoided. Noteworthy, Trehub did not respond to Juslin and Västfjäll’s (2008) objection to her criticism.

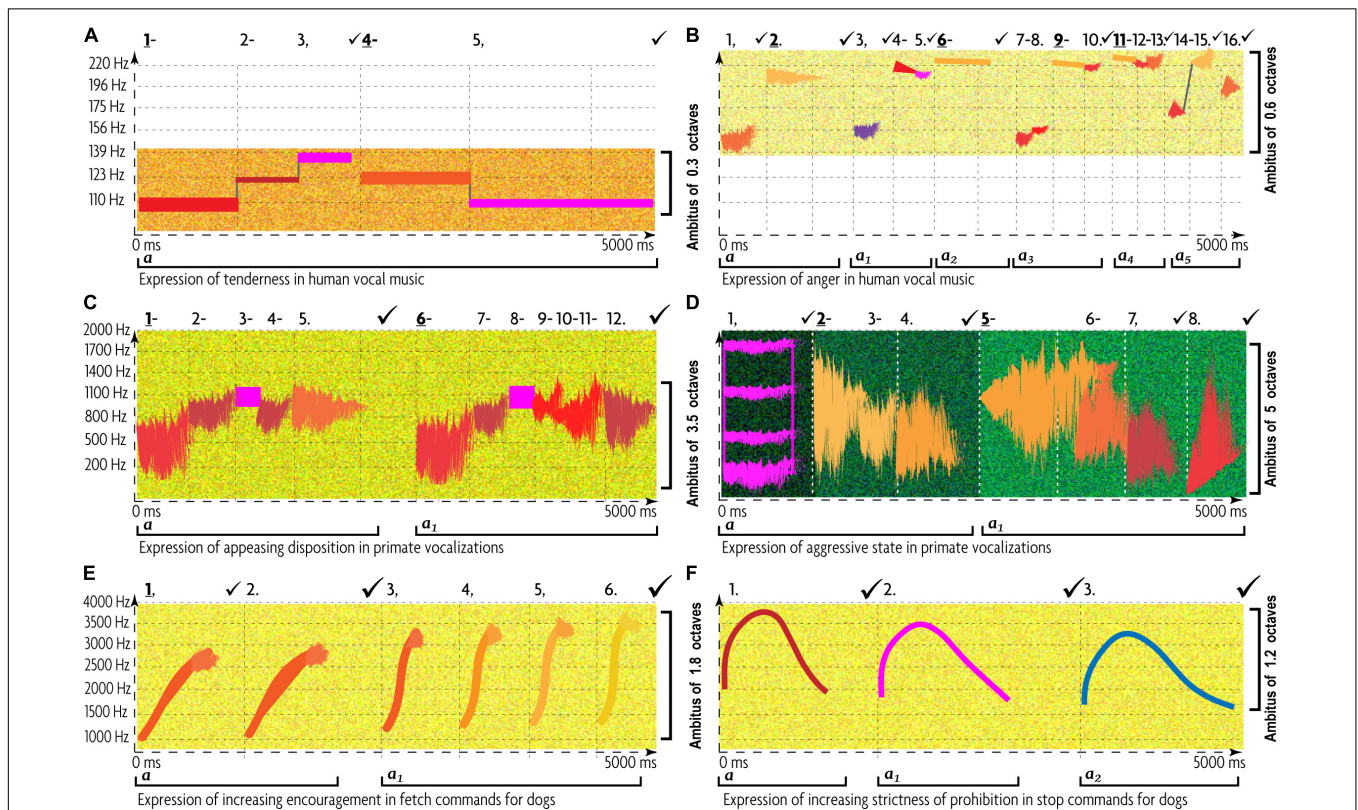


FIGURE 5 | Hybridization of characteristic patterns of ACs and human music in encouraging and prohibiting commands by human trainers to their dogs (McConnell and Baylis, 1985; McConnell, 1990, 1991, 2002; Miklosi, 2015). **(A) Typical expression of tenderness in human music.** This diagram extracts the key features of **Table 2** and **Figure 2**: very few pitch-classes with a low rate of change within a narrow ambitus, wave-like melodic contours filled by stepwise motion in the low register, slow tempo, with long tones and tendency to decelerate, and regular meter yet rhythmic diversity. Articulation is mostly legato, with occasional pauses. Dynamics is soft, stressing the anchor tones. **(B) Typical expression of anger in music** (according to **Table 2** and **Figure 3**): many pitch-classes with high rate of change and wide ambitus, ascending contours, and leaping zigzagging motion in high register. The tempo is fast, with short tones, often accelerating, with irregular pulse, and strong rhythmic contrasts. Dynamics is mostly loud, and accents fall on metrically weak tones. **(C) Typical expression of appeasing disposition in primate vocalizations** (**Table 3**). Many pitch-levels have a high rate of change, following a gradually ascending melodic contour within a relatively narrow ambitus. Tempo is fast, with short tones and long groupings. These features strongly contrast **(A)**, whereas metric regularity, legato articulation, low registration, and soft dynamics resemble **(A)**. **(D) Typical expression of aggressive disposition in primate vocalizations** (**Table 3** and **Figure 4**). There are relatively few pitch changes due to an extremely broad bandwidth, precluding frequent leaping. Long tones are embedded in fast motion with a descending contour in low register. These features oppose **(B)**, whereas meter, articulation, dynamics, and harmonicity resemble **(B)**. **(E) Typical expression of growing encouragement in fetch-whistles for dogs.** This expression combines a tender disposition of a human **(A)** with the appeasing disposition of a dog **(C)**. Therefore, fetch-command has to reconcile the contradictions between AEs' expressions of **(A)** and **(C)**. To accomplish this, the ascending contour becomes steeper, each signal and the time interval between signals become shorter, the ambitus of each signal grows and reaches higher register, and the groupings grow in size (from 2 to 4). Temporal and pitch AEs are co-adjusted, merging traits from **(A)** and **(C)**. **(F) Typical expression of growing prohibition in stop-whistles for dogs.** This expression combines the display of human displeasure, like **(B)**, with the appeasing disposition of the dog **(C)**, while structurally and semantically opposing **(E)**. **(F)** subverts a single long tone to the contrasting gradual fleccions in pitch, where the descending portion receives the greatest significance. The increase in intensity of prohibition is signified by extending the time values and reducing the steepness of the descending curve—in contrast to **(E)**. Dynamics provides yet another axis of opposition: loud for **(E)** versus soft for **(F)**. Most importantly, the **(E,F)** opposition involves a compensatory interaction of the temporal, dynamic, and pitch patterns of AEs. Thus, whenever **(F)** is used in isolation, its softness, slowness, and ametricity might project the impression of passiveness—contrary to the categorical nature of a “stop” command. To avoid this, **(F)**'s melodic curve combines ascending and descending curves whose conflicting relation generates extra tension.

DOMESTICATION OF ANIMALS SETS THE NEED TO MAKE TONAL ORGANIZATION SEMIOTICALLY FUNCTIONAL

The need to command domestic animals underlaid the population explosion of both humans and livestock during the Neolithic Revolution. Animals benefited from human support,

while humans benefited from animal produce. They both had to establish common patterns in their existing codes of vocal communication and adopt new patterns wherever the old patterns were deficient. Aspect-matching of pitch and rhythm was part of “bi-specific translation” of human commands (**Figure 5**). Rhythm reflects the “motion” pattern characteristic for a given “emotion” (Amaya et al., 1996), while pitch—the exertion/effort required by such motion—jointly defining a “sound gesture” (de Götzen, 2004). Perception of pitch and

rhythm relies on the biological components mutual for mammals, thereby supporting heterospecific communication. There is fMRI evidence of shared emotional vocalization systems across species (Belin et al., 2008).

An account of pitch-rhythm interaction comes from dog-training. Long continuous low/descending pitch is universally used to stop a dog, whereas repetitions of short rhythmic high tones—to encourage it—which might comprise a mammalian generality (McConnell and Baylis, 1985). Dog trainers identify pitch contour, rhythm, repetition rate, and amplitude as AEs effective in dog's commands.

Stop/fetch opposition reflects a multi-dimensional compensatory interaction of pitch, rhythm, and dynamics, mutual for both humans and canines. Some of the animal acoustic “universals” became appropriated into this bispecific communication, while others were overruled. Thus, across mammals, greater amplitude generally corresponds to a higher level of arousal (Briefer, 2012). However, it is only the fetch-command that follows this rule, whereas the stop-command, in contrary, adopts soft dynamics to subdue a dog (McConnell, 2002, 49–63). This overriding of the natural association between dominance and loudness highlights the fundamental difference between human and animal communications (Owren and Rendall, 2001):

- Human communication is “*receiver-centered*”—TO caters to information requirements of the listener;
- Animal communication is “*sender-centered*”—TO reflects the psycho-physiological state of the signaler, disregarding the listener.

Human-to-animal communication integrates both strategies:

- *Humans address animals*, treating them like humans, but *perfect the encoding* to secure the desired response. Thus, “doggerel” (Hirsh-Pasek and Treiman, 1982) constitutes dog-directed adaptation of human motherese (Mitchell R. W., 2001).

Pitch contour is a primary AE for most human cultures. Melody is the only aspect that differentiates between the basic musical emotions completely on its own (Table 2)⁴¹. In ACs, pitch does not provide such differentiation (Table 3). Pitch's importance for music pushes human melodies *higher* in register. This is because the low frequencies appear softer (Oxenham, 2013)—making the low contours less salient than the high contours. The same applies to primate hearing and, possibly, other mammals (Stebbins and Moody, 2011). Domestic animals too should follow suit. This incentivizes humans to raise contours characteristic for basic emotions above 1 kHz, where pitch changes are more salient. The only exception is the affection/love

⁴¹Prevalence of ascending contour characterizes happiness, anger, and fear. Happiness differs from anger and fear by employing variety of melodic contours called to diversify an ascending contour. Anger differs from fear by using sharp rather than wave-like contours and by dominance of staccato articulation in pitch changes (fear mixes staccato and legato articulations). Prevalence of descending contour characterizes both, sadness and love. They can be distinguished solely by intonation: flattened with stepwise falling contours for sadness, and sharpened with occasional ascending leaps for love.

signals. Intimacy requires close-distance communication where the softness of low-frequency poses no problems.

Social animals share affective signaling system with humans (Snowdon et al., 2015). This enables effective musical communication between humans and domestic animals—all of whom are “social” (Stricklin, 2001). SFTO in all likelihood evolved gradually, following the schemata of human-to-dog communication. The earliest archeological evidence of domesticated dogs dates back to 15 kya (Larson et al., 2012), but signs of domestication were found in a Gravettian site, at Předmostí (Germonpré et al., 2012). The DNA analysis indicates that a dog-like 33 kya old fossil from Altai is closer to modern dogs than to wolves (Druzhkova et al., 2013). Dog domestication must have been slow, preceded by feeding dogs with leftovers in exchange that they would follow humans and alert them of approaching predators. Dogs are genetically adapted to digest starch, which constituted part of human diet (Axelsson et al., 2013). Similar adaptation occurred in dog's communication system. *It adopted traits of human TO*. Compared to wolves, dogs use more vocal signals, especially bark-based—and barks feature co-modulation of two expressive aspects, amplitude and rhythm (Simpson, 1997). Alerting and territorial barking, both vary in intensity and rate depending on the distance of the dog from the conspecific or heterospecific intruder and the extent of the dog's arousal. At near distances barks become louder and more rapid. Such signaling and the manner of its modification most likely evolved in response to human's selective pressure on dogs to bark territorially at strangers (Simpson, 1997).

Human-to-dog communication most likely prototyped communication to later domesticates: cows, sheep, and goats. The surviving Nordic tradition of *kulning* provides the gist of the Neolithic pastoral music-making.

THE SCANDINAVIAN TRADITION OF KULNING AS A MODEL OF NEOLITHIC MUSICAL SEMIOSIS

Animal husbandry in Scandinavia started ≈1800 BC and reached its “golden age” by 1200 BC. This is when owning larger stocks became prestigious while climate warming enabled outdoor animal maintenance almost year-long (Tesch, 1992). However, winter grazing was hard on bushes and trees, depleting local resources. This, along with subsequent climate cooling, brought about a new housing style, designed to shelter animals together with humans for winter—which characterized Scandinavian pastoralism (Armstrong Oma, 2013). Sharing the house with animals led to acceptance of animals as household members, equal to humans, and categorically as “clean”—even animal dung was used to make wattle and daub walls. Sharing is known to increase bonding. Human dependence on milk products, and animals'—on humans' “room and board” promoted mutual trust and attraction (Armstrong Oma, 2010). From being “products,” animals turned into “producers” of dairy. This brought about *psychological “revolution”* in human-animal relationships, where music acquired the leading role.

Milking required concordance. An irritated animal or milkmaid reduced milk-yield, reducing human nutrition. Humans had to maintain mutual affection toward animals—evident in taboos on swearing/screaming at cattle, widespread across Eurasia (Plotnikova, 1999b). Music ritualized and fortified this union across different cultures (Shevtsov, 1988; Wallin, 1991; Alekseyev, 1995; Ivarsdotter, 1995, 2004; Novik, 1999; Dorina, 2004; Dissanayake, 2005; Kolltveit, 2008; Cheng, 2009; Yoon, 2018), especially evident in surviving traditions of milking songs (Nielsen, 1997; Pegg, 2001; Gioia, 2006b), animal lullabies (Kondratyeva, 1989; Kyrgys, 2002; Tchotchkina, 2003; Kan-ool, 2012), and spells (Kondratyeva, 1996; Kyrgys, 2002; Bordzhanova, 2007; Sodgerel, 2012, 2016; Tiukhteneva, 2017)—which all share the union of musicality and love/care that characterizes human motherese (Trevvarthen, 2019).

Principal traits of such music can be extracted from the current practice of Scandinavian herder's music-making. Its chief task is to control the behavior of the grazing livestock during the warm seasons at distant pastures (Ivarsdotter, 2004). The herder aims at influencing the animal's emotional state over a range of distances, up to a few kilometers. Long-distance transmission requires a special vocal technique and musical instruments. The same musical signals convey different information to livestock and humans: commanding animals while informing animal-owners at the farmstead of their animal's wellbeing. This dual communication has been faceted through a transhumance system known as *shieling* in England (Cheape, 1996), and *fåbod* in Scandinavia (Svensson, 2015)—emerging during the late Bronze Age in response to the scarcity of local winter fodder (Tesch, 1992). In Sweden, the shieling standard was set in Dalarna, and the alternative local traditions are considered its variations (Svensson, 2015). Traces of shieling are spotted across Europe, from the Hebrides to the Carpathians, becoming widespread by the Iron Age (Cheape, 1996). In Norway, the earliest fossil fields of lynchets show signs of cultivation during the late Bronze Age (Skrede, 2005), confirmed by palaeobotanic and archeological dating (Kvamme, 1988).

Shieling is characterized by seasonal migration to a summer station where herders spend their daytime supervising animals, preparing fodder for the coming winter, and produce dairy during evenings (Cabouret, 1984). Since milking, butter- and cheese-making traditionally constituted the women's job, shieling and its music became female prerogatives in Scandinavia. There, milking could dishonor a man, and shieling was managed exclusively by young women (Svensson, 2015). In Ireland, shieling was a family business, whereas in Spain, France, and Switzerland dairy-work and herding were conducted by men.

The gender difference, undoubtedly, played a role in shaping the European pastoral musical traditions. Scandinavian, Icelandic, Alpine, Jurassic, Pyrenean, Apennine, Sardinian, Balkan, Turkish, and Caucasian mountains have sheltered singing styles that originated in the herding culture, and shared a peculiar singing technique based on a forceful high-laryngeal falsetto-like sound production (Wallin, 1991, 510). Wallin (pp. 511–23) summarizes the archeological, anthropometric, and genetic research to support the ethnographic findings of Carl-Allan Moberg (1971). Moberg outlines the core traits of the

archaic *Fåbodväsendet* music: “head-voice” vocal technique, utilitarian function of long-distance signaling, and ideological roots in pagan magic.

The centerpiece of *Fåbodväsendet* tradition is its “maximal-distance” style—“*kula*”—that I distinguish from “*kulning*”—an umbrella-term for the entire *Fåbodväsendet*⁴². Local names for *kulning* (e.g., *lockrop*) imply the alluring of animals by magic properties of sound to suggest certain behavior to the herd, avert evil trolls and predator-animals—following shamanic tradition of maiden singing (Mitchell R. W., 2001). In Swedish mythology, forest spirits possessed their own cattle, and herdswomen (*kulerska*) learned *kulning* from *skogsrå*, “sirens of the woods” (Johnson, 1990). Suggestive power of *kulning* was deemed so high that women lived in *fåbods* alone without any weapons. Folk beliefs attributed this power to beauty. Indeed, well-ornamented high “warbling” register of distant female voice made men and women pause their work and enjoy the sounds (Ivarsdotter, 1986). For humans, *kula* clearly presented an aesthetic object despite bearing utilitarian status of “non-music” (Frödin, 1929)⁴³. For animals, *kula* constituted a “safety call.” Both attitudes focus on *positive* rather than negative emotions—not only to keep the cattle under human control, preventing panic, but also to boost the *kulerska*'s confidence and alertness (Wallin, 1991, 420)⁴⁴. SFTO must have emerged as a set of sonic attributes, perception of which was directly “wired” to reward circuits in brains of humans and domestic animals.

Wallin (1991, 420) rightfully stresses that matriarchy influenced early pastoralism: “the maternal instinct and care” instilled the social holding of attachment to stabilize and reinforce the animal-human affiliation. Distinctively female, *Fåbod* tradition must have prehistoric roots (Johnson, 1990). Motherese undoubtedly prototyped a close-range *kulning*. Animal-directed vocalizations acoustically and functionally resemble lullabies by commanding calmness/happiness—not just in Sweden (Wallin, 1991, 392) but also on the other side of Eurasia, in Altai (Kondratyeva, 1996). Common traits include

⁴²There is a wealth of terms used in Scandinavian countries to refer to herding vocalizations (Rosenberg, 2003, 8). Although the term “*kulning*” (*kolning*) is most commonly used in English in relation to the special technique of the long-distance vocal calling, I follow Wallin (1991, 387) in reserving the term “*kula*” (he uses the alternative spelling “*kola*”) which in Swedish means “to make a distant call” exclusively for long-distance communication. This is necessary, because long-distance “*kula*” calls are routinely inserted in mid-distance and close-distance vocalizations, while it is the long-distance “*kula*” style that distinguishes shieling vocalizations from other forms of traditional Scandinavian music.

⁴³It should be noted that the peculiar status of pastoral music as a form of *heterospecific* communication is responsible for the emic views on *kulning* as **non-music**. This is yet another confirmation of the need in the etic approach. Across Eurasia, herder-made music is distinguished from “normal” music as a form of “magic.” The profession of the herder is traditionally associated with sorcery: herders are believed to sign a contract with the evil forest spirits, receiving magic power for vocal and instrumental music-making in exchange for not using their gifts publicly, under the threat of death (Plotnikova, 1999b). At the eastern end of Eurasia, in Altai, supernatural beliefs are even stronger, reserved not only for professional herders (*chabans*) but for all livestock-owners who use pastoral spells (Kondratyeva, 1996). All vocalizations of this type are considered non-music—to the extent that informants perceive any request to “sing” a spell as being ridiculous.

⁴⁴Noteworthy, despite a 16-hour-long workday and insecurity of living alone without any weapons, shieling jobs were always highly sought after, since women remained in charge of their summer life and enjoyed freedom unavailable to them at the farmstead (Rosenberg, 2014).

prolonged singing, formulaic regularity, vocables, smooth contours, motherese-talking, and caressing (Tiukhteneva, 2017). In animistic societies, both infant-lulling (Kondratyeva, 1989; Farber, 1990; Tchotchkina, 2003; Gioia, 2006a; Milne, 2017; Garroway, 2019) and domestication rites for newborn cattle (Aksyonov, 1964; Johnson, 1990; Kondratyeva, 1996; Plotnikova, 1999b; Kan-ool, 2012; Tiukhteneva, 2017) are associated with magic, achievable by female “charms.”

Similar to lullabies are milking songs (Nielsen, 1997)—used across Eurasia, from Scotland to Mongolia (Gioia, 2006b, 71). Remarkably, when milking, Mongolian herdsman switch to motherese-like “musical talk,” based on animal onomatopoeia (Yoon, 2018). Known cases of male pastoral calling engage falsetto to imitate the female model (Uttman, 2002). Similarly, in surviving pastoral traditions of Altai, lulling is reserved for women, and require throat-singing if sung by men (Tiukhteneva, 2017). Pastoral spells in Altaic tradition constitute female prerogative, but are occasionally performed by men (Kondratyeva and Kopytov, 2017), engaging throat-singing (Kyrgys, 2002, 64). Like falsetto, throat-singing emphasizes harmonics that make melodies appear registrally higher—closer to the female range—and, like female kula, resembling pure tones.

The same applies to whistling signals, used across Eurasia by herdsman to stimulate and/or safe-guard animals (Levin and Suzukei, 2006, 134–40). Just like kulning, in pastoral societies whistling is associated with sorcery (Plotnikova, 1999a) and is thoroughly regulated by taboos (Dzenzelevskii, 1984). Acoustically, whistling comes closest to “kula” in distance-range, loudness, and tonal quality (Eklund and Mcallister, 2015). To command their animals, Altaic herdsman produce whistles audible over 4–5 km, and throat-singing—3 km (Pegg, 2001, 236). Curiously, female “head voice,” required by kula, is called “whistle register” (Sundberg, 1987, 50). And *xöömii* (throat-singing) is considered a form of whistling in Mongolia (Pegg, 1992).

Wallin (1991, 523) sees shieling music as part of the prehistoric expansion of a novel herding culture northwest of Anatolia/Balkan/Caucasus toward Iceland, with its base in Jamtland (Figure 6). Jamtland’s “forest barrow” marked the end of tundra after the glaciers’ retreat, attracting hunters and supporting a mixed pastoral economy that survived at the coldest outskirts of Europe practically unchanged until the late Middle Ages. Geographic and chronological distribution of cattle-herding across Europe, quite well-studied, provides timing references for Wallin’s model. The outcome of this geomusicological⁴⁵ correlation is presented in Figure 6.

Domesticated cattle spread East-to-West along the Mediterranean coastline, encapsulating most of “yodeling” territories ≈6000 BC. The South-to-North expansion took much longer—Central Sweden became pastoralized in the 2nd millennium BC. Dissemination of cattle and Indo-European languages went hand by hand. The Indo-European language family covers most of Europe—except for Finno-Ugric languages of Fennoscandia and Russia. Another notable exception is

Turkey whose Indo-European languages (Hittite, Luwian, Palaic, Lydian) died out during Antiquity. Formation of each new Indo-European language seems to have followed the adoption of husbandry. The yodeling areas correspond to the earlier stages in expansion of the Indo-European languages, conserved by the mountain systems: Taurus, Pontic, and Armenian Highland in Turkey, the neighboring Caucasus, Balkan, and more remote Carpathian, Alps, Jura, Apennine, Sardinian, Corsican, and Pyrenean. The dissemination routes either curve around the mountains or cross them by riverbeds. The oldest routs ran by the Mediterranean coastline along the 40N latitude, supporting the conclusion of Diamond and Bellwood (2003) that the domesticates and languages spread faster to East-West than to South-North. This explains the divergence of pastoral music tradition into two types: Southern yodeling versus Nordic kulning and kulning-likes⁴⁶, distinguished by different bovine genomes. Studies of Y-chromosomal variation have identified two primary taurine haplogroups in Europe, split in two homogenous regions alongside cultural, historic, religious, and linguistic boundaries between the pied or red cows of the Nordic and Baltic/Slavic lands, on the one hand, and the spotted yellow or brown breeds of Switzerland and southern territories, on the other hand (Edwards et al., 2011).

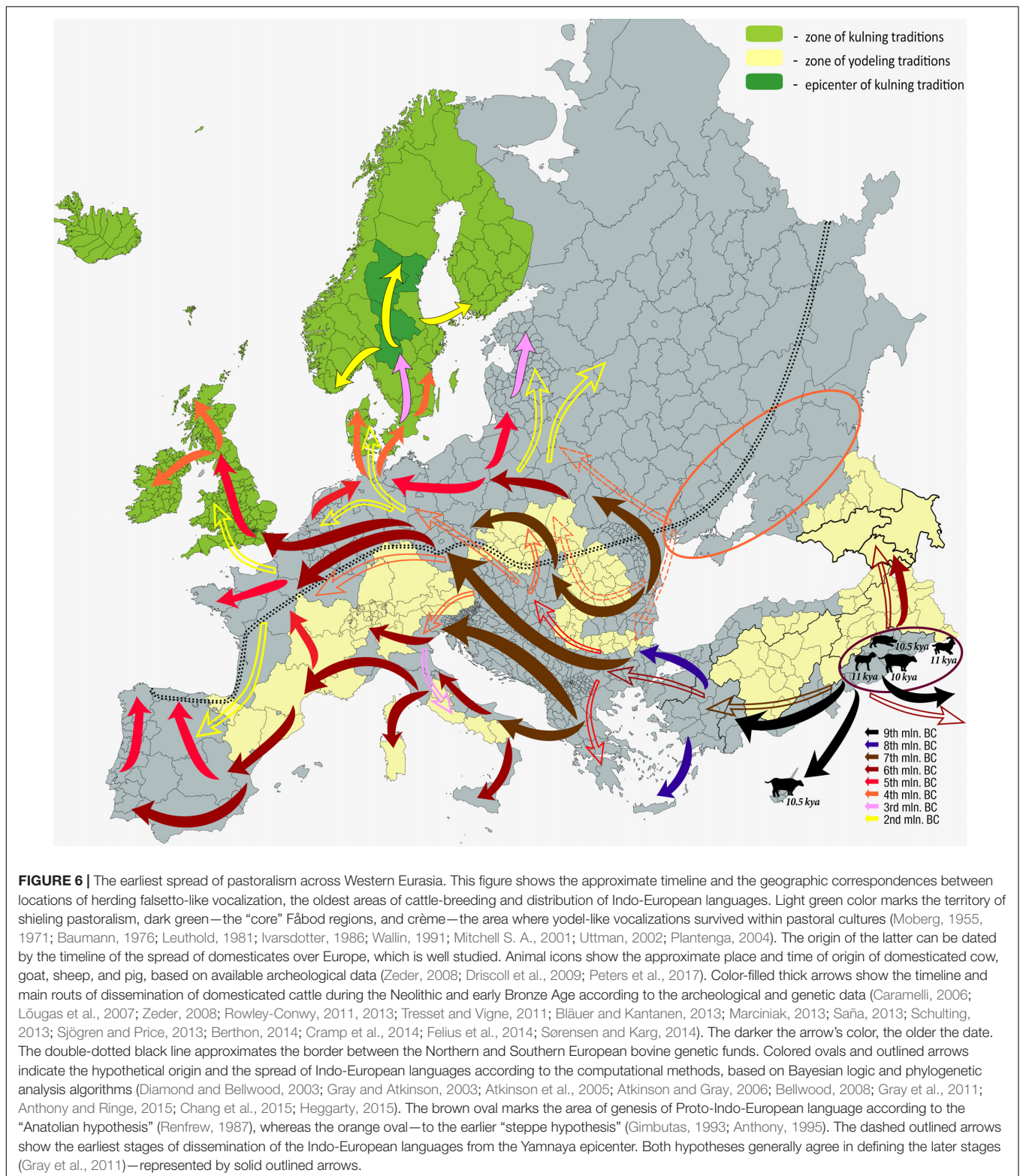
Kulning and yodel form respectively Northern and Southern “dialects” of a cattle-directed “language”—a satellite of the proto-Indo-European. The main role in the Indo-European “domestication package” belonged to cattle—the largest meat- and milk-source of all domesticates. The emergence of cattle-related mythology reflects the importance of cattle and explains the sudden proliferation of cattle burials across Northern Europe ≈3000 BC (Sjögren and Price, 2013)⁴⁷. Symbolic elevation of cattle could characterize the entire Neolithic “revolution” in Eurasia, more noticeable in Scandinavia, where ox symbolism replaced red-deer symbolism after ox overtook deer as the most important food source (Tilley, 1996, 183–4). If wild deer opposed the human sphere as a *utilitarian* object of desire, domesticated ox was included into the human sphere as the *emotional* object of desire. And music is indispensable in supporting emotionality.

Divinization of music (Franklin, 2006) and ox (Campbell, 2017), so prominent in Indo-European tradition, could have a single origin in Indo-Iranian lands—bound to the concept of non-violence (Tull, 1996). Cattle sacrifice is depicted in prehistoric Sujapur petroglyphs (Brooks and Wakankar, 1976). The ritual use of burnt cow dung is still common in Hinduism,

⁴⁶Thus, Finnish “ringing calls” present a form of vocalization that acoustically and culturally resembles Swedish kulning while featuring a few unique traits (Uttman, 2002). Occasionally, ringing calls are performed by men (falsetto), utilize a peculiar lip technique (generating the “phui”-like tonal quality), and engage “darker” vowels.

⁴⁷Cattle definitely carried special symbolic significance in Neolithic England (Ray and Thomas, 2003). Cattle received the same funeral treatment as humans in Danube winter burials as part of the Sun cult which thrived throughout the 4th millennium BC, probably because of drastic swings in solar activity (Horvaith, 2012). The second millennium BC Linear-B tablets from Knossos testify that, unlike sheep/goats, cattle was given names, bestowed with individuality—and was associated with royalty and sacrificial rites (McInerney, 2010, 50–53).

⁴⁵The scope and the method of geomusicology were introduced by George Carney (Nash and Carney, 1996). Izaly Zemtsovsky formulated an analogous approach in his proposal to establish a new discipline of ethnomusicology (Zemtsovsky, 2005).



traceable to the 3000 BC Ashmounds (Boivin, 2004). The Shiva-bull affiliation is evident in the Bronze Age Harappan “Proto-Shiva” (Hiltebeitel, 2011). Harappan symbolism clearly elevates the cattle over other domesticates, evident in the buffalo figurine

amulets and seals that are likely to assimilate the west-bound Indo-Iranian cult of Mother Goddess, eventually forming the “Sacred Cow” concept (Lodrick, 2005). This corresponds to veal and cow-milk becoming primary foods during Rigvedic and

Vedic times—there were people at that time who lived on milk alone (Prakash, 1961, 12). Milk products were used in rituals and offerings to gods, certainly accompanied by music, promoting the transformation of cow into the symbol of femininity and fecundity in Vedic literature (Brown, 1964). Consecration of cow gave it purity: even its urine and dung were used for healing and cleansing (Korom, 2000).

The cultural context of kulning and the tradition of home-sharing with cattle strongly resembles the Vedic cultural blend of non-violent femininity, cow-worship, and magic. It is not accidental that kula finds a nearly perfect match in Tibetan traditional pastoral songs with long rhythmically free undulating phrases, extremely tense timbre of high quasi-falsetto voice, generous ornamentation, and an ongoing variation (Stuart, 2008, XXIV). This is the most ancient of the three major forms of Tibetan music, peculiar to a nomadic pastoral culture, and originating from cattle calls (Crossley-Holland, 1967). Like kulning, it incorporates parlando and recitative for close-distance vocalization to animals, and also includes milking songs (Plantinga, 2004, 113).

Introduction of milk revolutionized the Neolithic lifestyle, supporting the psychological revolution in human-animal relations and bi-specific musical communication—especially in Northern Europe, where milk quickly replaced fish as the main food—manifested by the widespread adoption of milk-storing pottery (Cramp et al., 2014). The archeological evidence agrees with the genetic evidence of the time of emergence of lactase persistence⁴⁸. Lactase persistence reflects the adaptation to diet (Hancock et al., 2010)—without which adults have lactose intolerance and nutritional loss (Campbell et al., 2005). Ill effects of malnutrition coexisted with milk-bound diseases during the adoption of the milk-based diet. Mycobacterium tuberculosis existed 40,000 years ago, but became pathological for humans only from 6200–5500 BC onward (Hershkovitz et al., 2015) - by the time when the spread of husbandry reached Central Europe. Seemingly “the same” milk could either kill or nurture life—which must have promoted new supernatural beliefs and rituals to “exorcize” milk-production in replacement of the earlier hunter/gatherer rituals. Music, so common for religious applications, most certainly supported this reform.

For Europe, geographic distribution of Indo-European languages⁴⁹ (Heggarty, 2015) goes hand in hand with the distribution of taurine mtDNA that descends from the Fertile Crescent (Caramelli, 2006). And subdivision of the bovine European genetic pool into Northern/Southern genotypes

(Edwards et al., 2011) matches the distribution pattern of lactase persistence: 40% of adults in Greece versus 90% in Scandinavia/England (Curry, 2013). Those populations that consumed more dairy have higher occurrence of lactase persistence (Bersaglieri et al., 2004). Evidently, milk dependence was more than twice higher in the North. The Indo-European expansion occurred through the farmers’ immigration and interaction with local foragers rather than by technological import alone (Rowley-Conwy, 2011). Greater lactase persistence in the North reflects the dairy’s effectiveness in providing nutrients, the convenience of its storage in cold climate, the insurance against bad harvests (Gerbault et al., 2013), and health benefits of increased vitamin D consumption in low-sunlight conditions (Flatz and Rotthauwe, 1973).

Kulning emerged to nourish the symbiotic co-dependence of humans and cattle in harsh Nordic conditions that demanded stronger bonding than those of more diverse pastoral economies of Southern yodel territories, therefore employing a **female pastoral model**.

The biggest contender for the Indo-European language family in Northern Europe—the Uralic family (Diamond and Bellwood, 2003)—relates to another domesticate: the reindeer. Reindeer hunting was essential for colonization of Eurasian Arctic/Subarctic (Gordon, 2003). However, reindeer domestication still remains in its early phase (Reimers and Colman, 2009). The distinction between reindeer-hunting and reindeer-herding remains vague—even reindeer owners often do not know if a particular reindeer is “wild” or “domestic” (Ventsel, 2006)⁵⁰. Leading fences and corrals have been used for hunting wild reindeers and only recently have they become “domestic” accessories (Aronsson, 1991). Reindeer pastoralism emerged gradually from taming individual reindeers for transportation and decoy-hunting—compensating for the depletion of wild reindeer population (Vorren, 1973) that occurred during the 13–16th centuries (Hansen and Olsen, 2014, 175)⁵¹. Reindeer domestication must have started in parallel with cattle domestication in Norway/Sweden but lingered into the Middle Ages—absorbing cultural traits of human-to-cattle communication.

The principal psychological trait of kulning is the “humanization” and child-like patronizing of cattle. Similar attitude characterizes reindeer pastoralism: animal is treated like a family member whose life is valued and its attitudes are respected (Ingold, 1986). Kulning, yodel, and reindeer-communication should all be regarded as various “**languages of domestication**,” generated by borrowing “acoustic traps and snares”—i.e., onomatopoeic decoy calls—from hunters

⁴⁸Lactase persistence was completely absent in early Neolithic population 5500 BC (Burger et al., 2007), making its first appearance in Scandinavia in 3400 BC (Malmström et al., 2010), by 3000 BC in Iberia (Plantinga et al., 2012) and taking over Europe thereafter (Marciniak and Perry, 2017). This timeframe agrees with the scenario represented in Figure 6.

⁴⁹The Indo-European family contains 144 languages divided amongst 11 distinct branches—with even more languages most certainly having existed in the past but gone extinct (Diamond and Bellwood, 2003). In Europe, non-Indo-European languages are limited to merely 11 documented languages (only 8% of the total number of languages): Etruscan, Basque, Iberian, Tartessian, Estonian, Finnish, Urartian, Sumerian, Hurrian, Hattic, and Mitannian—plus 3 undocumented languages: Pictish, Lepontic, and Ligurian (Robb, 1993).

⁵⁰Herders routinely let their reindeers graze unsupervised for a rather extensive length of time. Inevitably, many animals become lost, turn wild, and can then be hunted (Stépanoff et al., 2017). Also, the herder’s strategy of searching for his lost animals strikingly resembles that of hunting.

⁵¹This caused the import of non-native reindeers via the emerging Russo-Finno-Scandinavian markets and transition to pastoralism (Roed et al., 2018). Genetic evidence points to 3 epicenters of reindeer domestication: Fennoscandia, Western and Eastern Russia (Roed et al., 2008). Reindeer domestication took about 6000 years. Its earliest evidence comes from the 4000 BC petroglyphs (Helskog, 2012), a 1510–1130 BC burial (Murashkin et al., 2016), and the paleolinguistic tracking of words for reindeer that date back to 1500–1000 BC (Aikio, 2006).

and syntactically reorganizing them into “animal-directed” words to control the herd, its leader, and the individual animals (Alekseyev, 1995).

Kulning and yodel are *Indo-European* musical “cow-languages,” later adapted for goats/sheep as they became personalized like cows⁵², whereas reindeer-vocalizations make a *Finno-Ugric* “reindeer-language.”

Kulning’s SFTO was forged by long-distance delivery of the desired subharmonic structure. Kula is characterized by dynamic maximization (80–100 dB SPL at 50 cm)⁵³ while fixing 4 formants at FF, 1700, 3,000 and 4,000 Hz throughout all frequency changes, restraining vibrato, and raising the larynx above the resting position (Johnson et al., 1982). Elevating laryngeal position up to 4 cm increases the sub-glottal pressure tenfold as compared to talking (Ivarsdotter, 1986). Somehow, this causes no distortions, and kula’s “harmonic signature” remains virtually unchanged at close- and mid-distances (1–11 m)—contrasting the “classic” falsetto (Eklund and Mcallister, 2015). Harmonic conservation is still observable at 22 m in kulning, albeit varying between different performers (Eklund et al., 2019). Evidently, kula is designed to transmit *kulerska*’s harmonic and melodic “signatures” to the herd at distances common in herding (Rosenberg, 2014).

Long-distance spectral optimization is known in intergroup communication of some primates (Waser and Waser, 1977). However, optimization to preserve subharmonic structures is unique to kula.

Kula’s sounds are supposed to stand out in the environmental soundstage by featuring unnaturally hyper-periodic noise-free spectrum. Kula’s harmonicity aligns with “pleasantness”—following the cow-bell paradigm. Animal-bells were used in Scandinavia at least from 1–4th centuries (possibly, from the beginning of the Bronze Age) to repel evil spirits, mark a human-controlled territory, and decorate the herd’s leading animal (Kolltveit, 2008). For cattle, the bell signified human control, herd-leader’s authority, and a safety signal. Humans associated bells with nature, peacefulness, goodness, and protection, employing bells to “borrow” the land from the forest spirits (Emsheimer, 1991, 43). Therefore, overall harmonicity signifies strongly positive values—in line with kulning’s perceived beauty and safety/care. Across the animal world, too, harmonicity

(pure-tonedness) and inharmonicity are meaningful along the friendliness/fear opposition (Morton, 1977).

Long-distance transmission requires high intensity and register. For 1 km, the most effective transmission occurs at ≈ 2 kHz (= C7) (Graf, 1980; Gray and Atkinson, 2003)—the range of a piccolo flute. Perhaps, whistling prototyped kula. Whistles are common in communication with dogs and the herd. And whistles exceed calling and yodeling in long-distance intelligibility: correct identification of whistles at 170 m distance is 95% versus 58% for yodeling and 70% for calling (Titze et al., 2018). *Bi-factorial* changes of rhythm/pitch-contour in whistling signals would pave the road for *tri-factorial* changes of rhythm/pitch/phrase-length in kula.

Long-distance communication eliminates mimics and gestures from semiosis, making it rely exclusively on acoustic attributes and demanding long-term memory (Wallin, 1991, 390). Exclusion of visual cues promotes the prolongation of a musical expression to facilitate its recognition and memorization. Therefore, phrase length reflects the distance: longer distances require longer phrases (p. 391). Changes in distance generate musical syntax (Figure 7).

Close distance promotes short phrases of multi-registral motherese-like recitative where only the “reciting tones” are pitched, and exaggerated leaps employ legato and portamento (Figure 7B). Pitches have tendency to monotony in low register at phrasal ends, which generates tonicity. Vocalizations are mostly stimulating and diverse in their referential/propositional content.

Middle distance makes motherese inaudible, instead requiring a different approach. Vocalizations become euphonized: engaging “parlando” rather than recitative⁵⁴, “smoothing” the leaps, increasing the share of pitched tones, and stressing rhythmic patterning and ordering. The calming effect of these adjustments, inappropriate for stimulating applications that are mostly common for mid-distance communication, is compensated by intensifying dynamics, structural contrasts, and staccato articulation (Figure 7A). Notwithstanding diversification, the highest-register “peak-tones” at motivic beginnings are often monotonous, prototyping the musical “leading-tone” by requiring some sort of continuation (as in a melodic resolution).

Longer distance further increases the share of musicality and pleasantness in herding vocalizations. They prioritize audition over visualization by engaging “call-phrases,” made of exclamatory imperatives and summoning, free from referential/propositional context (Wallin, 1991, 417). Verbalized vocalization is replaced by a wordless kula (p. 410). Simple phrase-sentences consist of motif chains akin to incipits, climaxes, and cadences of Gregorian tunes (Helmer,

⁵²Ivarsdotter describes how goat-calling follows the model of cow-calling, adapting it to the livelier nature of goats, notorious for their proneness to naughtiness (Ivarsdotter, 2004). Similarity of cow-calling (*Kolock*), goat-calling (*Getlock*), and sheep-calling (*Fårlock*) is obvious from listening to their archive recordings published by Swedish radio (Ivarsdotter, 1995). The same similarity is retained in pastoral incantations and spells that survive in Altai region—all three types of calling differ primarily in the prevalence of different phonemes for each of these three animals (Tiukhteneva, 2017). The musical characteristics of all three types of calling closely resemble one another (Kondratyeva and Mazepus, 1999). This suggests that similarity between cow, goat, and sheep pastoral communication is a wide-spread Eurasian phenomenon.

⁵³The highest SPL level is reached at a 30 cm distance from the sound source (125 dB) which exceeds the ear’s pain threshold at 120 dB (Rosenberg, 2014). The average SPL of kula at 1000 Hz is 113 dB. This is dynamically comparable to an operatic soprano singing fortissimo, except that the soprano’s technique requires maintaining a fixed larynx configuration at a low position. However, the maximal SPL of the soprano does not exceed 90 dB near the lips and does not change much from modulating the pitch (Johnson, 1984).

⁵⁴The term “parlando” was adopted by Anna Johnson in her report (Johnson, 1979) despite the traditional use of this term to refer to Western operatic singing that imitates speech and engages speaking “voice registers” (Sicoli, 2015) despite the absence of such intention for *kulerska*. Sung out words of close-distance kulning surprisingly resemble the operatic “parlando” sound. The kulning parlando contrasts the recitative kulning that minimizes voicing and remains much closer to talking than to singing, especially in its dynamics. The opposition of kulning parlando and kulning recitative resembles the opposition of operatic parlando and secco recitative, on the one hand, and the genre of melodrama that became popular in Western classical music in the 19th century, on the other hand.

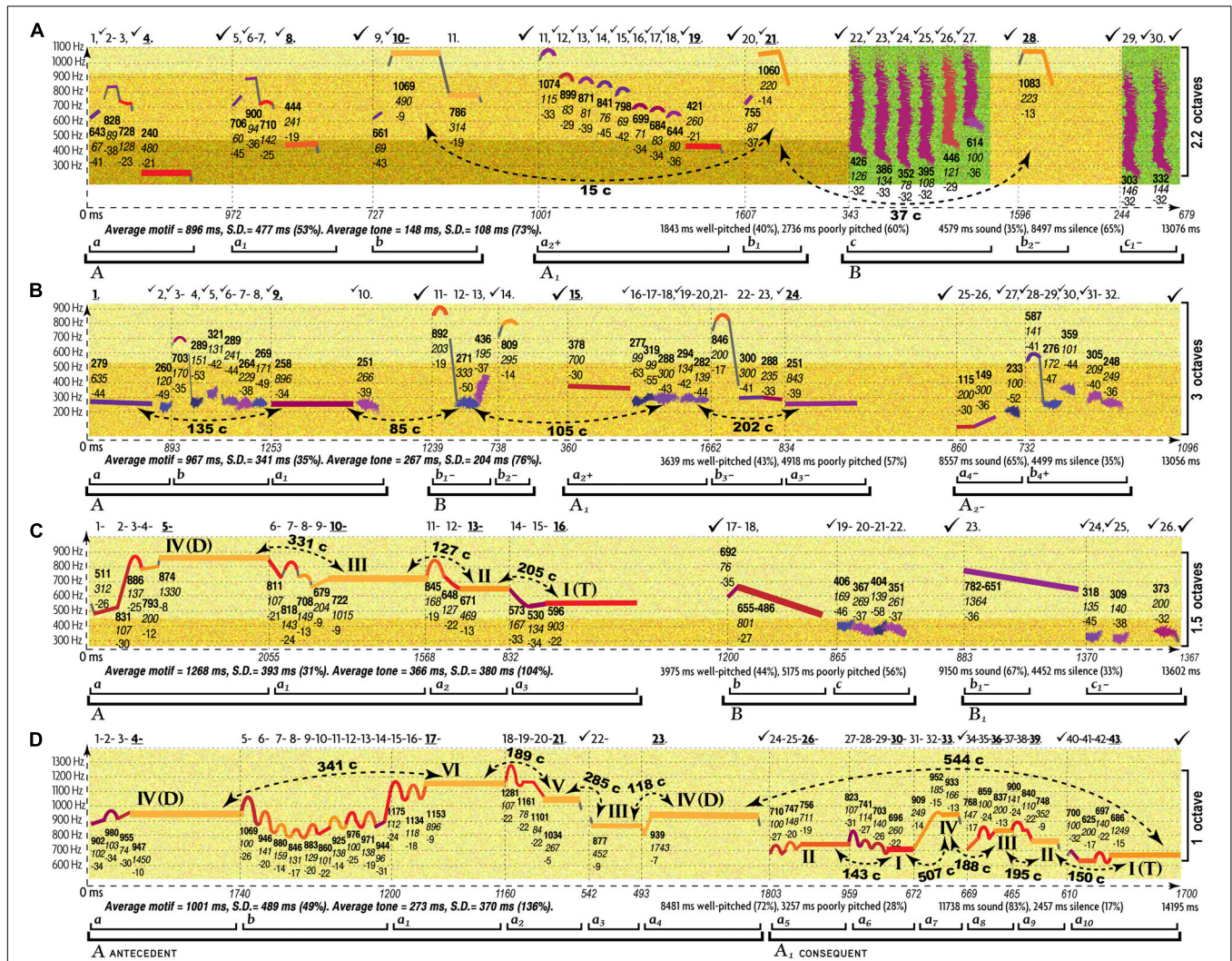


FIGURE 7 | Patterns of TO in four main types of vocalization in the vocal tradition of kulning. Since kulning is essentially ametric and a verbal (except for the closest range recitative), its analytic charts omit lyrics. Unlike the previous figures, the vertical dash lines indicate the onset of motifs. The colored arc-line symbol represents an ornamental melismatic shake. **(A) Stimulative medium-distance kulning: parlando (a), exclamation (b), and onomatopoeia (c) motifs** (<http://chirb.it/ntlxrM>). This style is designed to compel the entire herd to move in the desired direction and, most probably, sets a model of interaction with animals for the other three styles. The three motifs achieve stimulation, each in a different way, contrasting one another in register, harmonicity, rhythm, and articulation. Motif “a” alerts by its staccato zigzag leaping between two registers. Motif “b” combines stimulation (staccato leap up to the “shrieking” register) with relaxation (legato leap down to the long tone). The “shrieking” peak-tones maintain the same pitch level (melodic regularity)—reflected by the dotted double-arrows (numbers indicate the frequency discrepancy in cents). Motif “c” teases the cattle by imitating dog’s barking. The stimulative specialization of **(A)** is manifested in its prevalence of staccato, loud dynamics, three registers within a wide ambitus, exuberance of leaps, and brevity of motifs and tones. Noteworthy, the motifs “a₂” and “c” resemble the “fetch-command” archetype (**Figure 5E**). **(B) Stimulative close-distance kulning: recitative (a) and motherese (b) motifs** (<http://chirb.it/8K3Lqg>). **(B)**, like **(A)**, is stimulative but dynamically gentler due to closer distance (~9 dB softer). This allows for diverse motherese-like prosodic exaggerations in motivating individual animals. Motif “a” expresses love/care by greatly prolonging the “recitative tone,” sustaining its pitch and harmonicity. Motif “b” stimulates animals by briefly stressing the upper “head-voice” register with a shake-like embellishment, then sliding it all the way down to the low talking voice. Compared to **(A)**, **(B)** is smoother: fewer registers, less staccato, and longer motifs and tones. **(B)** tends to support a monotone (a predecessor of tonicity), most noticeable at phrasal ends. **(C) Inhibitive longer-distance kulning: simple kula (a), exclamation (b), and parlando (c) motifs** (<http://chirb.it/n6f0sv>). This style functionally opposes **(B)** by commanding the herd to stop grazing and to go home, implying that it is no longer safe to stay out. The chief function of “a” (kula) is to instill confidence in the herder’s control over the animals. “Kula” typically consists of a chain of motifs stitched together to form a characteristic shape of steep ascension to the crest point and thereafter a gradual fall-off. However, motifs might differ according to their phrasal functions: initiation, climax, decay, and cadence. The resulting kula receives a basic modal TO: anchor tones constitute “degrees” of the mode, forming a fifth between the marginal degrees and dividing it in wider upper and narrower lower parts. The Roman numerals indicate degrees (I = stable is marked as T = “tonic”). The “b” motif presents “exclamation”: a gradual sliding down (~4th), softer than in **(A)**, and shaped like the “stop-command” (**Figure 5E**). Similarly shaped is the parlando “c” motif, much smoother than **(B)** due to its prevailing legato, freer rhythm, more homogenous registers, and longer motifs and tones. **(D) Tropotrophic maximal-distance kulning: exclusive use of complex kula sentences** (<http://chirb.it/gpyC7t>). Delivering signals over a kilometer requires taking multiple short caesuras throughout the span of the kula’s descending formula, which distinguishes **(D)** from **(C)** by making kula complex. Motifs make up phrases, and phrases—sentences, all of which create modal complexity: anchor-tones form intervallic relations that define degrees within a mode (usually, 5–7 degrees). Upper degrees open kula, forming an antecedent cadence (marked by letter “D”—“dominant” function). Lower degrees end kula with the consequent cadence (marked by “T”—“tonic”), providing resolution.

(Continued)

FIGURE 7 | Continued

Compared to (C), sentences in (D) are longer, rhythmically freer, more homogenous (by maintaining legato, a single register, the narrowest of ambitus for all kulning styles, and no leaps). Relaxation, secured by modal resolution, is supported by beautification: exclusive use of legato in smoothly shaped phrases and exquisite ornamentation (shakes, trills). (D) differs from (C) by sacrificing dynamic shaping on a phrasal level and, instead, reproduces the same dynamic contour on a motivic level—the final long tone is almost always the loudest in a motif (i.e., stable). Increased homogeneity and melodic consonance (i.e., absence of leaps) are called to motivate the herd not to depart any further beyond the range of hearing kula.

1975). Each phrase is distinguished by a wavelike melodic-dynamic “envelop” with an abrupt quick rise and a gradual prolonged fall. Kula pushes vocalizations higher, squeezing their ambitus, homogenizing timbres and legato articulation, while loosening the rhythm (Figure 7C). This triggers the modal genesis: kula’s anchor-tones turn into degrees, with more-or-less sustained pitch values. The lowest degree becomes “tonic,” in contrast to the unstable upper degrees, thereby forming tetrachord-based modes.

Maximum-range communication complicates kula by introducing hierarchic structuring (motifs-phrases-sentences) and by engaging the contrasting phrasal functions (initiation/climax/interruption/termination). The stimulating effect of the increased syntactic contrasts, undesirable for maximum-range communication that focuses on keeping the animals calm, is compensated by greater melodic homogeneity: maximizing legato, sentence-length, and dynamics, while minimizing melodic-intervallic, rhythmic, and registral diversity (Figure 7D). Longer span necessitates inter-phrasal caesuras, marking multiple phrases within long sentences, joined by stereotypical declining inter-phrasal melodic and dynamic “envelops.” Melody relies on pentachordal skeleton, divided in upper major and lower minor 3rds, often supported with quartal/quintal infix (Johnson, 1979). Kula breaks in a series of antecedent-consequent sentences that engage different pentachord/tetrachord(s)—usually conjunct. This produces heptatonic modes (Figure 8).

The ongoing unveiling of musical structures makes kulning particulate by stacking up certain phrasal types while avoiding certain other combinations. This establishes syntactic rules and implicit music theory of TO for herders and herds. Herders perceive kulning as improvised “musical work in progress” (akin jazz improvisation) that elaborates a specific “theme” selected by the *kulerska* (Rosenberg, 2014). Herded animals probably perceive kulning as a series of programmed Pavlovian-conditioned routines. In both cases, compositionality promotes particulate semiosis: the meaning of a streak of phrases consists of the sum of the meanings of each of the constituent phrases. In effect, kulning tells a “continuing story” of the day, going through an elaboration of a musical theme (Rosenberg, 2014).

The herd’s daily movement generates SFTO by stitching/restitching phrases of 5 syntactic-semantic types (Table 6):

- Kula (tropotrophic),
- Exclamations (phatic),
- Onomatopoeia (ludic),
- Parlando (imperative),
- Motherese/recitative (endearing/motivating).

Genesis of SFTO follows the path of human-to-dog whistling communication. Noteworthy, kulning’s exclamations and onomatopoeic calls engage stop- and fetch-whistle features (see Figures 5E,F).

The proof for SFTO’s pragmatic efficacy is in the herd’s fulfilling of the shepherd’s commands (Wallin, 1991, 410).

Yet another source of semiosis for kulning was phonemic symbolism. Complete absence of words in kula and minimal wording of motherese suggest the *prelingual* existence of kulning. Wallin (1991, 410–413) rightfully emphasizes that there is no reason to label kula’s sounds as “*phonemes*”—they are mere *homologues* to vowels and consonants, shaped by the anatomic-physiological conditions of breathing and acting while uttering. The same applies to traditional Alpine yodel (Fenk-Oczlon and Fenk, 2009a). Yodel and kulning vocables are formed not by phonological oppositions of local languages but by the communication distance and the extent of the desired stimulation/inhibition for a given call. Thus, the highest larynx and intensity at the onset of long-distance inhibitive kula-phrases generate a semantically “negative” [i], whereas a relaxed post-climactic position in the mid-distance tropotrophic kula generates a “positive” [å]. Similarly, glottal stops at phrasal beginnings and endings range from a gentle [h] to a harsh [tj], depending on the needed attack and tenuto decay (Rosenberg, 2014). The choice of the most common kulning syllables (Ahlbäck, 2007) can be explained by human/animal’s natural selection for effective distant communication (Wallin, 1991, 390).

Monodization of kulning was imperative in genesis of SFTO.

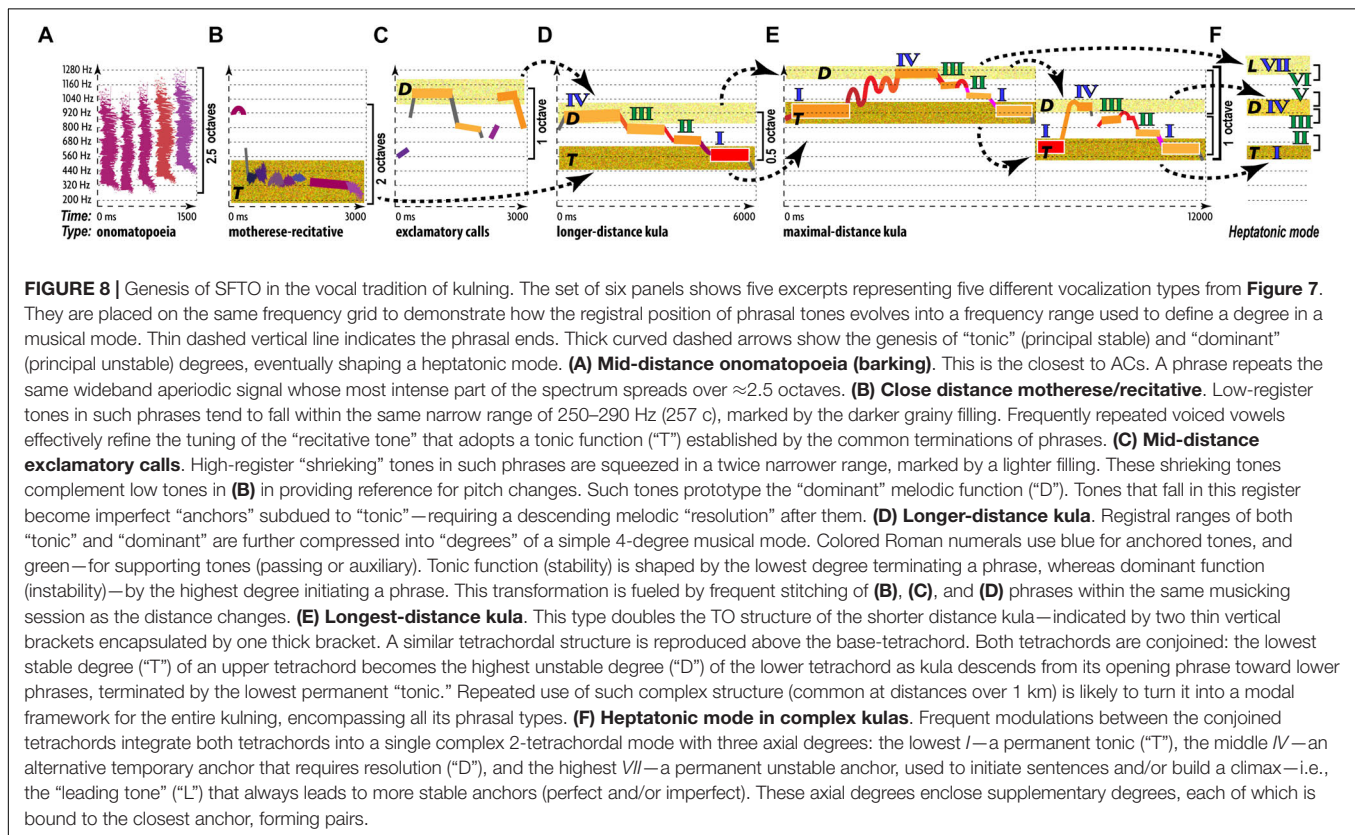
Animal communication usually employs male “chorus,” male-female “duetting,” or “antiphonal” formats (Yoshida and Okanoya, 2005). Musicologically, this corresponds to a special type of texture—“isophony”: the ongoing out-of-sync multi-part reproduction of the same thematic material (Nikolsky, 2018). Isophonic jumble precludes SFTO. For multifactorial patterning to emerge, each vocalizer must clearly hear his/her voice in order to track spectral changes without any contamination by a partner. This is how infants learn to make their own songs and how children acquire “musical ear” (Nikolsky, 2020). Even in non-European traditions that are exclusively *polyphonic*, such as Aka Pygmy, motherese and children-made music remain *monophonic* (Rouget, 2011). This is because an auditory stimulus must be objectified to become accessible for reproduction: a relation of 2 tones in certain aspect must be realized as an auditory constant to lay the foundation for construction of a musical mode (Nazaikinsky, 1973). For perception, the listener must discover permanence of the foreground “sound-object” against the background of a sound-stage, and memorize it in order to relate to it all of the subsequent changes in the thematic material.

Just as one cannot learn prosody of a language by listening to the crowd, one cannot learn SFTO by listening to isophony. And

TABLE 6 | Acoustic traits of main motif types and their semantic values in kulning.

Acoustic domains	Aspects of music	Kula singing phrases and sentences	Exclamation calls	Onomatopoeic imitations	Parlando singing and calling	Motherese recitative
1. Frequency	1. Melodic pitch (consecutive)	Prevalence of long descending stepwise motion; narrow ambitus of a 5th breaks in two portions: upper major. 3rd and lower minor. 3rd (with possible infrafix). Well-defined pitches, many melismas. The strongest melodic coherence. "Safety call" function—calming yet keeping alert, under control. Subject to aesthetic evaluation.	Prevalence of a zigzag pitch contour in a brief group of tones with a very short start, poorly defined pitches, medium long leaps, large ambitus of 0.5–1.6 octave. Optional but frequently employed shakes and melismas distinguish the stimulation use from inhibition, both engaged exclusively in the phatic function (no aesthetic value).	Short repetitions of the same imitation of an animal call: poorly pitched, very broad bandwidth, no melismas, melodic coherence is absent, stimulating function (to make the herd move in the necessary direction). Most likely, this is a derivative of a "fetch whistle" command. Utilitarian application (no aesthetic value).	Prevalence of a zigzag shape, like exclamation, but longer and stressing a huge descending leap (up to 1.6 octave), greater than an ascending leap (<7th). Both leaps relax toward the motif's end. They stimulate or inhibit, based on the extent of such relaxation. Utilitarian application with aesthetic value.	Prevalence of drastic contrasts in pitch between flat monotonous pitched talking and extreme zigzag leaping (about 1.7 octave) that is usually embellished with melismas. Stimulating and motivating functions, supported with gestures and mimics. Only a few pitches are clear. Mostly utilitarian application.
	2. Harmony	n/a	n/a	n/a	n/a	n/a
	3. Form (complexity)	Greatest complexity	Simplicity	Greatest simplicity	Low complexity	Medium complexity
2. Time	4. Tempo	Slow, frequent ritenuto toward the phrasal end.	Moderate and/or slow tempos, possible ritenuto.	Moderate, possible accelerando.	Moderate or slow, frequent rubato.	Moderate and lively, with moderate rubato.
	5. Rhythm	Sharp contrasts of melismatic and anchor tones, the greatest rhythmic diversity, no increments.	Contrast of short-long pattern in grouping and totally arrhythmic "breath-groups."	Prevalence of the same relatively short rhythm, grouped by pauses, clear increments.	Contrast of short upper versus long lower tones, and of groups of equal versus patterned rhythm.	Contrast of long initial versus short last tones, and of free "verbal" versus patterned "musical" rhythm.
	6. Meter	Most irregular, often totally "ametric," a dragging feel.	Can contain regular fragments, usually iambic.	Mostly regular (spondaic).	Usually irregular (loose, free).	Irregular, as if always changing iambic-trochaic.
	7. Articulation	Absolute dominance of legato. Frequent and long caesuras between phrases that usually end on long tenuto tone and a descending glide.	Contrasted groups of staccato on ascending leaps and legato on descending steps. Frequent ending on a tenuto glide.	Prevalent non-legato provides ease and clarity of recognition for each of the imitations.	Contrasted groups of staccato for ascending leaps and legato for steps as well as descending leaps. Sometimes tenuto endings.	Syllables within a word usually are legato, while words or vocables are separated by pauses.
3. Amplitude	8. Dynamics	The loudest type. Mid-distance kula phrases are shaped diminuendo, while each long-distance kula sentence is shaped wavelike.	Rather intense dynamics with contrasts between the wavelike shape and the opposition of loud high versus soft low tones.	Only minor dynamic changes within mostly loud levels, copying the typical dynamic envelop of a typical animal call.	Moderate contrast of softer high tones and louder lower tones, in overall soft dynamics.	The softest type, yet with sudden accents, falling on a single syllable in those words that are marked by a zigzag leap.
4. Timbre	9. Register	Single, very high register—the longer the distance, the brighter the tonal quality (piercing or shrilling). Fixed high larynx position, constant brightness.	Multiple registers: shrieking for the highest tones, shrilling (kula-like) for high, "casual" for low tones. Variable larynx position. Recoloring of the tones within a motif.	Single register for each onomatopoeic imitation, usually broadband—in contrast to the narrow-band kula. Usually high larynx. Overall dullness.	Contrast of 2 registers: head-voice (kula-like) for high tones and throat or chest singing voice for low tones. Variable larynx position. Recoloring of the tones within a calling motif.	Contrast of 2 registers: head-voice (kula-like) for one syllable and normal speaking voice for the rest. Variable larynx position.
	10. Harmonicity, attack and vibrato	Clear harmonics (akin to pure tones), ascending portamento attack and descending termination (in longer kulas), minimal vibrato (only to embellish a tone such as a trill).	Clear harmonics (akin to kula) for the highest and longest tones only, ascending and descending portamento, no vibrato.	Prevalence of non-periodic spectrum, harsh, noisy sound, little voicing (only if present in the imitated model), no portamento or vibrato.	Clear and rich harmonics for the lowest and longest tones only (with some vibrato), ascending and descending portamento for leaps only.	Prevalence of non-periodic spectrum, as in speech, with pronounced frequency modulation and noise. No vibrato. Leaps engage portamento.

Ten AEs (in rows) are used in five types of phrases (in columns), each characterized by a unique combination of AE patterns, the most distinctive of which are pitch, rhythm, articulation, dynamics, and register. Each type also is distinguished by its semantic specialization: kula—safety signal for grazing, exclamation calls—social "grooming talk," onomatopoeia—playful teasing, parlando—commanding and convincing, and motherese—endearing and trusting. Except for the long-distance kula (whose sentences can reach up to 15 s), all other types are quite brief (usually, 0.3–2 s) and are intermixed with the same-distance or shorter-distance motif types (i.e., kula phrases can be included in motherese recitative, but motherese cannot be included in kula). The maximal distance squeezes the ambitus into an octave confined to a single highest register. This compresses degrees into steps of a smaller or a longer intervallic size, depending on their phrasal position. Climactic step tends to constitute the interval of major 2nd, whereas cadential step—of minor 2nd, to emphasize and facilitate resolution of tension ("major" for a peak in tension, "minor" for relaxation). The framework of a breath cycle sets the basis for traditional association of major with happiness (climax = inspiration = maximal power) and of minor with sadness (cadence = expiration = collapse). Mid-distance kula transposes heptatonic structures to lower registers, fitting them into a tetrachord (pentachord, if a climactic motif is added). Octave equivalence secures heptatony. Closer distances enable alterations, flattening or sharpening of the unstable degrees, and timbral recoloring. Stacking phrases of contrasting TO and semantic values, learnable by humans and domesticated animals, generates the SFTO. See a fuller version of this table in Appendix 2 "A Comparative Structural Analysis of Musograms," in **Supplementary Material**.



herding music promotes monodic application: herding demands hours of solitary interaction with animals, ideal for testing their response to music-making.

CONCLUSION

Homo heidelbergensis was already anatomically capable of practicing proto-music which was most probably isophonic, lacking the combined coding of pitch/rhythm—without which conventionalization of the semiotically functional melody-making was hardly possible. Isophony supports only group communication of zero- and first-order intentionality, limited and conditioned by the genetically embedded instinctive responses to isophonic formula. Learning of multi-factorial particulate expression and second-order intentionality requires monophonic production. AE’s pattern becomes fully semiotic only when many senders/receivers remember it as the bearer of the same semantic value that connotes a certain affective state—“binding hearer to speaker” through “tying of some social sentiment” (Wallin, 1991, 420).

Emotional contagion is possible in isophonic signals, but it is primed to a single most salient AE—provided all communicators share the necessary neuro-anatomical substrates (Snowdon et al., 2015). Harmony, meter, texture, and form are not supported by non-human brains; neither is a premediated “construction” of an intended message. Animal interpretation of auditory signals is inherently circumstantial—determined by

the signaling context (Zuberbühler, 2017). Therefore, human music is often “misunderstood” by animals, requiring music’s “translation” into animal’s “sonic templates of recognition” (Snowdon and Teie, 2013).

For ACs to evolve into music, a repertoire of patterns of AEs had to be extracted from proto-musicking practice and abstracted into elemental signs to continuously inform someone(s) of the communicator’s affective state, intentions, and needs. Such use emerged in communication with domesticated dogs, thereafter, adapted for herding. Hunting/gathering does not demand such communication. Instead, it prioritizes *collective* collaboration: bringing participants emotionally “in-tune,” binding them into a group to increase one’s powers. Such use makes sense in situations of using loud complex sounds while hunting large prey and repelling human predators in open savannah space (Jordania, 2011). Large groups of big-game foragers tend to prioritize *collective* music-making over *personal*, confining the latter to pre-pubertal age, like Aka Pygmies (Rouget, 2011). *Homo* probably exported isophonic proto-music from Africa to Europe.

The last Glacial Maximum greatly reduced the European population by the Gravettian—until the Magdalenian repopulation (Maier, 2017) enabled the rise of symbolic cultures (Kozłowski, 2015) and ethnolinguistic genesis (Zilhão, 2014). Low-density foraging groups usually form alliances, cemented by linguistic commonalities and intermarriage (Marlowe, 2005). Music surpasses language in its bonding capacities (Nakata and Trehub, 2004). Gravettian proto-music must have adjusted isophony for new cultural applications,

especially religious. Smaller groups generate a smaller sonic “jumble,” facilitating the recognition of specific musical elements. Smaller groups also promote honesty in communication (Richerson and Boyd, 2005). Honest musical expression enables and validates the person-to-person musical communication. This opens doors to the cultural development of a motherese communicative model. Small groups are likely to promote motherese-like duetic and babbling-like solitary music-making. Thus, collective music-making is exceedingly rare in Northern Siberia (Alekseyev, 1967) which has always remained underpopulated (Sikora et al., 2019)—closely resembling life in glaciated Europe.

Motherese talk, lullabies, onomatopoeia, and instinctive utterances supplied the initial material for the formation of bi-specific SFTO. Changes in distance while continuously communicating with the herd put in place the musical modes. The closest distance promotes low-register monotony, middle distance—high-register monotony, long distance—tetrachord-based tonicity, and maximal distance—conjunct pentachord/tetrachord octave-equivalent modes with dominant-tonic functionality. Monotony increases the tuning accuracy of anchor-tones, firstly defining principal degrees (tonic, supertonic, dominant), and then additional unstable degrees (Alekseyev, 1976). Characteristic modal intonations of different phrasal styles and varying position within a breathing cycle charge modal degrees with specific functionality, which directs the formation of semantic values for each of the common modal intonations. This triggers the process of modal evolution as outlined by Beliaev (1963) and elaborated by Nikolsky (2015a, 2016).

Nordic kulning is probably a vestige of an archaic cattle-oriented “domestication language” which descended from yodel—accompanying the northerly spread of Indo-European languages throughout Europe. Other Eurasian domestication languages accompanied the spread of the Uralic and Turkic language families, and were optimized, respectively, for reindeer and horse. Rémy Dor cross-analyzed vocalizations/whistles of herders speaking 20 Turkic languages, from Anatolia to Yakutia, and inferred their syntactic organization (Dor, 2005), identifying their common utterances (Dor, 1993). Like Wallin and Alekseyev, Dor too found continuity between vocalizations of hunters and herders: “somatotropic” vocalizations, designed to make the prey come closer, evolved into “fetch” or “home-return” calls, while “somatofugal” vocalizations evolved into “stop” calls to repel predators. The new class of “somatoneutral” vocalizations emerged in order to keep an animal at a constant distance (like safety-call kula). Strong biological foundation of this distance-governed communication made it well-conserved—practically indestructible—unlike languages or music systems (Dor, 2008).

Domestication languages could underlie modern languages and musics, as traditional beliefs suggest. Swedish rural informants considered kulning an ancient “language” (Moberg, 1971, 145). And on the opposite end of Eurasia, Mongolian herders believe that their music-making is derivative of the “large language,” superior to human language and designed to communicate with animals, nature, and spirits (Pegg, 2001, 235). Altaic *xöömii* most likely constitutes yet another “domestication language.”

Capacity to simultaneously control numerous AEs and second-order intentionality enabled humans to create a heterospecific semiotic system of communicating desirable affective states, which gave humans control over domestic animals, resolved human sustenance needs, and put in place music as we know it. The semiotically functional tonal organization that distinguishes music from speech might have emerged no earlier than during the Neolithic “revolution” as a result of forging new conventions of human-to-animal vocal communication.

DIRECTIONS FOR FUTURE RESEARCH

Comparative examination of human-to-animal signaling for different domesticated animals across different geographic regions can confirm whether the paradigm of “musical domestication language,” divisible in “dialects” and integrable into “language families,” is applicable here.

Collecting a database of patterns of human-to-animal communication would be analogous to building a lexicon of a newly discovered natural human language or to establishing a stock of typical idioms in the musical communication within a novel musical culture. Once established, such database can be statistically analyzed and cross-examined in relation to other databases, e.g., of emotional expressions in music. This could substantiate or invalidate my conclusions.

The perception of specific elements and patterns of human-to-animal communication by humans and animals can be experimentally tested. This could identify syntactic and pragmatic rules that cannot be assessed by acoustic analysis alone. Together, both approaches can evaluate semiotic efficacy of TO in pastoral signaling. This, in turn, can establish whether introduction of herding communication during the Neolithic Revolution was capable of generating SFTO in music to make it break away from the basics of animal communication.

Experimental archaeo-ethnomusicology could provide yet another way of verifying this hypothesis. Members of isolated tribes that maintain a hunter/gatherer lifestyle and use no domestic animals can be introduced to domestic animals and “taught” to use music-like signals to command them. Their progress can be analyzed and compared to patterns of conspecific acquisition of music skills by human infants, as well as to the available archaeological, genetic, and paleo-physiological data.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I am grateful to CT and MR for reviewing the manuscript for this paper, and to Sheila Bazleh for copy-editing it. My special thanks to Leonid Perlovsky, Steven Brown, Piotr Podlipniak, Leon Crickmore, Theodor Levin, Margarita Mazo, and Philipp

Tagg for their critical input in relation to matters of semiotics of music, and to Isaly Zemtsovsky, Eduard Alekseyev, and Frank Scherbaum for reviewing my approach to modal analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01358/full#supplementary-material>

REFERENCES

- Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *J. Soc. Biol. Struct.* 12, 1–13. doi: 10.1016/0140-1750(89)90015-8
- Addressi, A. R., and Caterina, R. (2000). Perceptual musical analysis: segmentation and perception of tension. *Music. Sci.* 4, 31–54. doi: 10.1177/102986490000400102
- Adorno, T. W. (1942). *On Popular Music*. Frankfurt am Main: Institute of Social Research.
- Agawu, V. K. (2003). *Representing African Music: Postcolonial Notes, Queries, Positions*. London: Routledge.
- Agawu, K. (2004). How we got out of analysis, and how to get back in again. *Music Analysis* 23, 267–286. doi: 10.1111/j.0262-5245.2004.00204.x
- Ahlbäck, S. (2007). *The Tonality of Older Swedish Folk Music*. Stockholm: Udda Toner.
- Aiello, L. C. (1998). “The foundations of human language,” in *The Origin and Diversification of Language*, eds N. G. Jablonski and L. C. Aiello (San Francisco, CA: California Academy of Sciences), 21–34.
- Aikio, A. (2006). New and old samoyed etymologies. part-2. *Finnisch Ugrische Forschungen* 59, 9–34.
- Aksyonov, A. N. (1964). *Tuvan Folk Music: Materials and Studies [Тувинская народная музыка: материалы и исследования]*. Moscow: Muzyka.
- Alekseyev, E. Y. (1967). Is there polyphony among the Yakuts? [Есть ли у якутов многоголосие]. *Sovetskaya Muzyka* 5, 97–105.
- Alekseyev, E. Y. (1976). *Problems in the Genesis of Musical Mode (on the example of Yakut folksong): analysis [Проблемы формирования лада (на материале якутской народной песни): исследование]*. Moscow: Muzyka.
- Alekseyev, E. Y. (1986). *Musical Intonation in the Earliest Forms of Folklore. The Aspect of Pitch [Раннефольклорное интонирование: звуковысотный аспект]*. Moscow: Soviet Composer.
- Alekseyev, E. Y. (1995). “The pragmatics, hystronics and rituals in sound imitations of peoples of Northern Siberia [Прагматическое, игровое, ритуальное в звукоподражаниях народов Севера Сибири],” in *Voice and Ritual. [Голос и ритуал]*, eds E. Dorokhova, N. Zhulanova, and O. Pashina (Moscow: The State Institute of Art Studies at the Ministry of Culture of the Russian Federation), 33–35.
- Alekseyev, E. Y., and Nikolayeva, N. (1981). *Models of Yakut vocal folklore [Образцы якутского песенного фольклора]*. Yakutsk: Academy of Sciences of the USSR.
- Altenmüller, E., Kopiez, R., and Grewe, O. (2013a). “A contribution to the evolutionary basis of music: lessons from the chill response,” in *Evolution of Emotional Communication*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford: Oxford University Press), 313–336. doi: 10.1093/acprof:oso/9780199583560.003.0019
- Altenmüller, E., Kopiez, R., and Grewe, O. (2013b). “Strong emotions in music: are they an evolutionary adaptation?,” in *Sound - Perception - Performance*, ed. R. Bader (Heidelberg: Springer), 131–156. doi: 10.1007/978-3-319-00107-4_5
- Altenmüller, E., Schmidt, S., and Zimmermann, E. (eds). (2013). “A cross-taxa concept of emotion in acoustic communication: an ethological perspective,” in *Evolution of Emotional Communication*, (Oxford, UK: Oxford University Press), 339–356. doi: 10.1093/acprof:oso/9780199583560.003.0020
- DATA SHEET S1** | Appendix 1 – A new method of modal multifactorial analysis of tonal organization in music and music-like sounds.
This technical paper contains instructions for identifying the tonal organization in a music work, a music-like vocalization (e.g., infant’s babbling) or music-like animal signals (e.g., bird’s song) – including sounds that are indefinite or modulating in pitch.
- DATA SHEET S2** | Appendix 2 – A comparative structural analysis of musograms used in **Figures 3, 4, 7** of this article.
This document contains a comprehensive analysis of the characteristic traits of tonal organization in the examples of human musical communication, animal vocal communication, and bi-specific communication between domestic animals and their human guardians.
- Alvarez-Pereyre, F., and Arom, S. (1993). Ethnomusicology and the emic/etic issue. *World Music* 35, 7–33.
- Amaya, K., Bruderlin, A., and Calvert, T. (1996). “Emotion from Motion,” in *Proceedings of Graphics Interface '96 Toronto, Ontario, Canada: 22 - 24 May 1996*, eds R. Bartels and W. A. Davis (Toronto: Canadian Human-Computer Communications Society), 222–229.
- Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science* 291, 1748–1753. doi: 10.1126/science.1059487
- Ambrose, S. H. (2010). Coevolution of composite-tool technology, constructive memory, and language. *Curr. Anthropol.* 51, S135–S147. doi: 10.1086/650296
- Anthony, D. W. (1995). Horse, wagon & chariot: indo-European languages and archaeology. *Antiquity* 69, 554–565. doi: 10.1017/S0003598X00081941
- Anthony, D. W., and Ringe, D. (2015). The indo-european homeland from linguistic and archaeological perspectives. *Annu. Rev. Linguist.* 1, 199–219. doi: 10.1146/annurev-linguist-030514-124812
- Aranovsky, M. G. (1991). *Syntactic Structure of a Melody. Research [Синтаксическая структура мелодий. Исследование]*. Moscow: Muzyka.
- Aranovsky, M. G. (1998). *Musical Text. Structure and Properties [Музыкальный текст. Структура и свойства]*. Moscow: Kompozitor.
- Armstrong Oma, K. (2010). Between trust and domination: social contracts between humans and animals. *World Archaeol.* 42, 175–187. doi: 10.1080/00438241003672724
- Armstrong Oma, K. (2013). Human-animal meeting points: use of space in the household arena in past societies. *Soc. Anim.* 21, 162–177. doi: 10.1163/15685306-12341300
- Arom, S. (2004). *African Polyphony and Polyrhythm: Musical Structure and Methodology*. Cambridge, MA: Cambridge University Press.
- Arom, S. (2010). Corroborating external observation by cognitive data in the description and modelling of traditional music. *Music. Sci.* 14, 295–306. doi: 10.1177/10298649100140S216
- Arom, S., and Voisin, F. (1997). “Theory and technology in African music,” in *The Garland encyclopedia of world music: Africa*, ed. R. M. Stone (New York, NY: Garland Publishing), 254–270.
- Aronsson, K. -Å (1991). *Forest Reindeer Herding A.D. 1-1800: An Archaeological and Palaeoecological Study in Northern Sweden*. Umeå: Dept. of Archaeology, University of Umeå.
- Atkinson, Q. D., and Gray, R. D. (2006). “How old is the Indo-European language family? Illumination or more moths to the flame?,” in *Phylogenetic Methods and the Prehistory of Languages*, eds P. Forster and C. Renfrew (Cambridge: McDonald Institute for Archaeological Research), 91–109.
- Atkinson, Q. D., Nicholls, G., Welch, D., and Gray, R. D. (2005). From words to dates: water into wine, mathematic or phylogenetic inference? *Trans. Philol. Soc.* 103, 193–219. doi: 10.1111/j.1467-968X.2005.00151.x
- August, P. V., and Anderson, J. G. T. (1987). Mammal sounds and motivation-structural rules: a test of the hypothesis. *J. Mammal.* 68, 1–9. doi: 10.2307/1381039
- Axelsson, E., Ratnakumar, A., Arendt, M. L., Maqbool, K., Webster, M. T., Perloski, M., et al. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–364. doi: 10.1038/nature11837
- Bader, R. (ed.) (2018). *Springer Handbook of Systematic Musicology*. Berlin: Springer-Verlag.

- Balari, S., Benítez-Burraco, A., Camps, M., Longa, V. M., Lorenzo, G., and Uriagereka, J. (2011). The archaeological record speaks: bridging anthropology and linguistics. *Int. J. Evol. Biol.* 2011:382679. doi: 10.4061/2011/382679
- Banshchikov, G. (1997). *The Rules of Functional Instrumentation [Законы функциональной инструментации]*. Sankt-Petersburg: Kompozitor.
- Bartel, D. (1997). *Musica Poetica: Musical-Rhetorical Figures in German Baroque Music*. Lincoln: University of Nebraska Press.
- Baumann, M. (1993). Listening as an emic/ethnic process in the context of observation and inquiry. *World Music* 35, 34–62.
- Baumann, M. P. (1976). *Musikfolklore und Musikfolklorismus*. Winterthur: Amadeus Press.
- Becker, J. (1986). Is western art music superior? *Musiq. Q.* 72, 341–359. doi: 10.1093/mq/LXXII.3.341
- Beliayev, V. (1963). The Formation of folk modal systems. *J. Int. Folk Music Council* 15, 4–9. doi: 10.2307/836227
- Beliayev, V. M. (1990a). “Modal systems in the traditional music of the USSR [Ладовые системы в музыке народов СССР],” in Viktor Mikhailovich Beliyev [Виктор Михайлович Беляев], ed. I. Travkina (Moscow: Soviet Composer), 223–377.
- Beliayev, V. M. (1990b). Viktor Mikhailovich Beliyev [Виктор Михайлович Беляев]. Moscow: Soviet Composer.
- Belin, P., Fecteau, S., Charest, I., Nicastrò, N., Hauser, M. D., and Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proc. R. Soc. B Biol. Sci.* 275, 473–481. doi: 10.1098/rspb.2007.1460
- Bellwood, P. S. (2008). “Archaeology and the origins of language families,” in *Handbook of Archaeological Theories*, eds R. A. Bentley, H. D. G. Maschner, and C. Chippindale (Lanham, MD: Altamira Press), 225–244.
- Benjamin, T., Horvit, M. M., and Nelson, R. (2015). *Techniques and Materials of MUSIC: From the Common Practice Period through the Twentieth Century*. Stanford, CA: Cengage Learning.
- Benson, D. J. (2007). *Music: A Mathematical Offering*. Cambridge: Cambridge University Press.
- Benward, B., and Saker, M. N. (2009). *Music in Theory and Practice*. New York, NY: McGraw-Hill.
- Berliner, P. F. (1993). *The Soul of the Mbira: Music and Traditions of the Shona People of Zimbabwe*. Chicago, IL: University of Chicago Press.
- Berndt, A., and Hähnel, T. (2010). “Modelling musical dynamics,” in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound, Piteå, Sweden — September 15 - 17, 2010*, ed. K. Brandenburg (Ilmenau: Fraunhofer Institute for Digital Media Technology), 1–8.
- Berry, W. T. (1987). *Structural Functions in Music*. New York, NY: Dover Publications.
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., et al. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74, 1111–1120. doi: 10.1086/421051 doi: 10.1086/421051
- Berthon, R. (2014). Past, current and future contribution of zooarchaeology to the knowledge of the neolithic and chalcolithic cultures in south caucasus. *Stud. Caucasian Archaeol.* 2, 4–30.
- Blacking, J. (1967). *Venda Children's Songs: A Study in Ethnomusicological Analysis*. Chicago: University of Chicago Press.
- Blacking, J. (1974). *How Musical is Man?*. Seattle, WA: University of Washington Press.
- Blake, E. C. (2011). *Stone “Tools” as Portable Sound-Producing Objects in Upper Palaeolithic Contexts: The Application of an Experimental Study*. Dissertation, University of Cambridge, The Department of Archaeology, Cambridge.
- Bläuer, A., and Kantanen, J. (2013). Transition from hunting to animal husbandry in Southern, Western and Eastern Finland: new dated osteological evidence. *J. Archaeol. Sci.* 40, 1646–1666. doi: 10.1016/j.jas.2012.10.033 doi: 10.1016/j.jas.2012.10.033
- Blench, R. (2013). Methods and results in the reconstruction of music history in Africa and a case study of instrumental polyphony. *Azania Archaeol. Res. Afr.* 48, 31–64. doi: 10.1080/0067270X.2013.771016
- Bobrovsky, V. P. (1978). *The Functional Basics of Musical Form [Функциональные основы музыкальной формы]*. Moscow: Muzyka.
- Bogucki, P. I. (1988). *Forest Farmers and Stockherders: Early Agriculture and Its Consequences in North-Central Europe*. Cambridge, MA: Cambridge University Press.
- Boivin, N. (2004). Landscape and cosmology in the South Indian Neolithic: new perspectives on the Deccan Ashmounds. *Camb. Archaeol. J.* 14, 235–257. doi: 10.1017/S0959774304000150
- Boivin, N., Brumm, A., Lewis, H., Robinson, D., and Korisetter, R. (2007). Sensual, material, and technological understanding: exploring prehistoric soundscapes in south India. *J. R. Anthropol. Institute* 13, 267–294. doi: 10.1111/j.1467-9655.2007.00428.x
- Bonfeld, M. S. (2006). *Music - Language, Speech, Thinking: Experience of Systematic Research of Musical Art [Музыка - язык, речь, мышление: опыт системного исследования музыкального искусства]*. Sankt-Petersburg: Kompozitor.
- Bordzhanova, T. G. (2007). *Ritual Poetry of the Kalmyks. The system of genres, poetics. [Обрядовая поэзия калмыков. Система жанров, поэтика]*. Elista: Kalmyk Book Publishers.
- Braudo, I. A. (1961). *Articulation. On pronunciation of melody [Артикуляция. О произношении мелодий]*. Leningrad: State musical publishing.
- Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge MA: MIT Press.
- Bresin, R., and Friberg, A. (2011). Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex* 47, 1068–1081. doi: 10.1016/j.cortex.2011.05.009
- Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *J. Zool.* 288, 1–20. doi: 10.1111/j.1469-7998.2012.00920.x
- Brodsky, I. A. (1976). “On research of music of the ethnicities of North of the Russian Federation [К изучению музыки народов Севера РФСР],” in *Traditional and Contemporary Folk Music Art. Collection of Essays. [Традиции и современное народное музыкальное искусство. Сборник трудов]*, ed. V. Yefimnenkova (Moscow: State Music Pedagogical Institute named after Gnesin), 244–257.
- Brooks, R. R. R., and Wakankar, V. S. (1976). *Stone Age Painting in India*. London: Yale University Press.
- Brown, K. S., Marean, C. W., Hennessey, A. I. R., Jacobs, Z., Tribolo, C., Braun, D., et al. (2009). Fire as an engineering tool of early modern humans. *Science* 325, 859–862. doi: 10.1126/science.1175028
- Brown, N. W. (1964). The sanctity of the cow in hinduism. *Econ. Wkly.* 1, 245–256.
- Brown, S. (2000). “The ‘Musilanguage’ model of language evolution,” in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 271–300.
- Brown, S. (2007). Contagious heterophony: a new theory about the origins of music. *Musiq. Sci.* 11, 3–26. doi: 10.1177/1029864907011010101 doi: 10.1177/102986490701100101
- Brown, S. (2017). A joint prosodic origin of language and music. *Front. Psychol.* 8:1894. doi: 10.3389/fpsyg.2017.01894
- Brudzynski, S. M. (2013). “Vocalizations as indicators of emotional states in rats and cats,” in *Evolution of Emotional Communication*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford, UK: Oxford University Press), 75–91. doi: 10.1093/acprof:oso/9780199583560.003.0005
- Burger, J., Kirchner, M., Bramanti, B., Haak, W., and Thomas, M. G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3736–3741. doi: 10.1073/pnas.0607187104
- Bytchkov, Y. N. (1987). On Systemic Nature of Modal Organization in Music [О системном характере ладовой организации в музыке]. Dissertation, Russian Academy of Music named after Gnessin, The Department of Music theory, Moscow. doi: 10.1073/pnas.0607187104
- Bytchkov, Y. N. (1997). *On Dialectics of Formation and Unfolding of Mode [О диалектике становления и развертывания лада]*. Moscow: Russian Academy of Music named after Gnessin.
- Cabouret, M. (1984). *La vie Pastorale Dans les Montagnes et les Forêts de la Péninsule Scandinave*. Dissertation, Université de Paris-Sorbonne, The Department of Geography, Lille.
- Cambouropoulos, E. (2008). Voice and stream: perceptual and computational modeling of voice separation. *Musiq. Percept.* 26, 75–94. doi: 10.1525/mp.2008.26.1.75
- Cambouropoulos, E. (2010). The musical surface: challenging basic assumptions. *Musiq. Sci.* 14, 131–147. doi: 10.1177/10298649100140S209

- Campbell, A. K., Waud, J. P., and Matthews, S. B. (2005). The molecular basis of lactose intolerance. *Sci. Prog.* 88, 157–202. doi: 10.3184/003685005783238408 doi: 10.3184/003685005783238408
- Campbell, J. (2017). *The Masks of God. Occidental Mythology*. San Anselmo, CA: Joseph Campbell Foundation.
- Camps, M., and Uriagereka, J. (2006). “The Gordian knot of linguistic fossils,” in *The Biolinguistic Turn: Issues on Language and Biology*, eds J. Rosselloi and J. Martiin (Barcelona: University of Barcelona), 34–65.
- Caramelli, D. (2006). The origins of domesticated cattle. *Hum. Evol.* 21, 107–122. doi: 10.1007/s11598-006-9013
- Chang, H.-W., and Trehub, S. E. (1977). Auditory processing of relational information by young infants. *J. Exp. Child Psychol.* 24, 324–331. doi: 10.1016/0022-0965(77)90010-8
- Chang, W., Cathcart, C., Hall, D., and Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91, 194–244. doi: 10.1353/lan.2015.0005
- Charlton, D. (2009). *Opera in The age of Rousseau: Music, Confrontation, Realism*. Cambridge, MA: Cambridge University Press.
- Chase, P. (2006). *The Emergence of Culture: The Evolution of a Uniquely Human Way of Life*. New York, NY: Springer.
- Cheape, H. (1996). Shielings in the highlands and islands of Scotland: prehistory to the present. *Folk Life* 35, 7–24. doi: 10.1179/043087796798254498 doi: 10.1179/043087796798254498
- Cheng, J. (2009). A review of early dynastic iii music: man’s animal call. *J. Near East. Stud.* 68, 163–178. doi: 10.1086/613988
- Chew, G. (2001). *Articulation And Phrasing*. Oxford: Oxford University Press.
- Christensen, T. (ed.) (2008). *The Cambridge History of Western Music Theory*. Cambridge, MA: Cambridge University Press.
- Clarke, E. F. (2007). “Rhythm and timing in music,” in *The Psychology of Music*, (Cambridge, MA: Academic Press), 473–500. doi: 10.1016/b978-0-12213564-4/50014-7
- Clayton, M. R. L. (2000). *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance*. Oxford, UK: Oxford University Press.
- Conard, N. J., Malina, M., and Münzel, S. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 737–740. doi: 10.1038/nature08169
- Cook, N. D. (2002). *Tone of Voice and Mind: The Connections between Intonation, Emotion, Cognition and Consciousness*. Amsterdam: John Benjamins Publishing.
- Cooke, D. (1959). *The Language of Music*. London: Oxford University Press.
- Cramp, L. J. E., Evershed, R. P., Lavento, M., Halinen, P., Mannernaa, K., Oinonen, M., et al. (2014). Neolithic dairy farming at the extreme of agriculture in northern Europe. *Proc. R. Soc. B Biol. Sci.* 281:20140819. doi: 10.1098/rspb.2014.0819
- Crickmore, L. (2009). “The tonal systems of Mesopotamia and ancient Greece: some similarities and differences,” in *The Archaeomusicological Review of the Ancient Near East*, eds R. Dumbrell and M. Marcetteau (London: Academic Press), 1–16.
- Cross, I. (1999). “Is music the most important thing we ever did? music, development and evolution,” in *Music, Mind and Science*, ed. S. W. Yi (Seoul: Seoul National University Press), 10–39.
- Cross, I., and Watson, A. (2006). “Acoustics and the human experience of socially-organized sound,” in *Archaeoacoustics*, eds C. Scarre and G. Lawson (Cambridge, MA: McDonald Institute for Archaeological Research), 107–116.
- Crossley-Holland, P. (1967). “Form and style in Tibetan folksong melody,” in *Jahrbuch für musikalische Volks- und Völkerkunde Band 3*, ed. F. Bose (Berlin: Walter de Gruyter), 9–69. doi: 10.1515/9783111448503-002
- Curry, A. (2013). The milk revolution. *Nature* 500, 20–22. doi: 10.1038/500020a
- Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch (*Taeniopygia guttata*). *J. Comp. Psychol.* 104, 3–10. doi: 10.1037/0735-7036.104.1.3
- Dahlhaus, C. (1989). *Nineteenth-Century Music*. Berkeley, CA: University of California Press.
- Dasen, P. R. (2012). “Emics and Etics in Cross-Cultural Psychology Towards a Convergence in the Study of Cognitive Styles,” in *Proceedings of the 4th Africa Region Conference of the IACCP, University of Buea, Cameroun, Aug. 1-8, 2009*, eds T. M. S. Tchombe and M. Fülöp (Buea: University of Buea), 55–73.
- Davidson, M. (1970). Some music for the Lala Kankobe. *Afr. Music J. Afr. Music Soc.* 4, 103–113. doi: 10.21504/amj.v4i4.1685
- de Götzen, A. (2004). “The sounding gesture: an overview,” in *Proceedings of the 7th International Conference on Digital Audio Effects, DAFX 04: Naples, October 5 - 8, 2004*, ed. G. Evangelista (Naples: Università degli studi di Napoli, Dipartimento di scienze fisiche), 5–10.
- Dean, R. T., Bailes, F., and Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: an experimental investigation. *PLoS One* 6:e18591. doi: 10.1371/journal.pone.0018591
- Demany, L., Mckenzie, B., and Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature* 266, 718–719. doi: 10.1038/266718a0
- Dennett, D. C. (1983). Intentional systems in cognitive ethology: the “Panglossian paradigm” defended. *Behav. Brain Sci.* 6, 343–355. doi: 10.1017/S0140525X00016393
- d’Errico, F., and Stringer, C. B. (2011). Evolution, revolution or saltation scenario for the emergence of modern cultures. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1060–1069. doi: 10.1098/rstb.2010.0340
- D’Errico, F., Villa, P., Llonca, A. C. P., and Idarraga, R. R. (1998). A middle palaeolithic origin of music? Using cave-bear bone accumulations to assess the Divje. *Antiquity* 72:65. doi: 10.1017/s0003598x00086282
- Deutsch, D. (2013). “The processing of pitch combinations,” in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 249–325. doi: 10.1016/b978-0-12-381460-9.00007-9
- Diamond, J., and Bellwood, P. S. (2003). Farmers and their languages: the first expansions. *Science* 300, 597–603. doi: 10.1126/science.1078208 doi: 10.1126/science.1078208
- Diedrich, C. G. (2015). ‘Neanderthal bone flutes’: simply products of Ice Age spotted hyena scavenging activities on cave bear cubs in European cave bear dens. *R. Soc. Open Sci.* 2, 140022–140022. doi: 10.1098/rsos.140022
- Dissanayake, E. (2005). “Ritual and ritualization: musical means of conveying and shaping emotion in humans and other animals,” in *Music and Manipulation: On the Social Uses and Social Control of Music*, eds S. Brown and U. Volgsten (New York, NY: Berghahn Books), 31–56.
- Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Music. Sci.* 12, 169–195. doi: 10.1177/1029864908012001081
- Dobzhanskaya, O. (2011). “Ritual musical folklore of Samoyedic peoples as an object of study: main results and research perspectives [обрядовый музыкальный фольклор самодийских Хаподоькакобъект изучения: основные результаты и перспективы исследования],” in *From Congress to Congress: Proceedings of the Second All-Russian Congress of Folklore Study [Материалы Второго Всероссийского фольклорного съезда]*, ed. A. S. Kargin (Moscow: State Republican Center of Russian Folklore), 300–311.
- Dobzhanskaya, O. E. (2012). “On the concept of ‘music’ in relation to archaic and early folkloric cultures [О понятии «музыка» применительно к архаичным и раннефольклорным культурам],” in *XVI Tzarskoselsk Readings [XVI Царскосельские чтения]*, ed. V. Skvortsov (Sankt-Petersburg: Pushkin Leningrad State University), 256–259.
- Dobzhanskaya, O. E. (2016). The living has sound; the dead is silent. *Anthropol. Archeol. Eurasia* 55, 7–21. doi: 10.1080/10611959.2016.1263485
- Dor, R. (1993). Les Huchements du berger turc: II: du huchement-aux-morts à l’appel des chevaux. *Études Turques Ottomanes* 3, 27–41.
- Dor, R. (2005). Les huchements du berger turc III: interpellatifs adressés au gros bétail. *Turcica* 27, 199–222. doi: 10.2143/turc.27.0.2004362
- Dor, R. (2008). À l’aube du cri: de l’homme à l’animal avant le partage du monde. *Diogenes* 200, 129–139. doi: 10.3917/dio.200.0129
- Dorina, M. V. (2004). *Wind and Percussive Instruments of Sayano-Altaians Turks: An Attempt of Historic-Ethnographic Investigation [Духовые и ударные музыкальные инструменты тюрков Саяно-Алтая (Опыт историко-этнографического исследования)]*. Dissertation, Tomsk University, The Department of Ethnography and History of Science, Tomsk.
- Dowling, W. J. (1984). Cognitive processes in the perception of art. *Adv. Psychol.* 19, 145–163. doi: 10.1016/S0166-4115(08)62350-X
- Drabkin, W. (2001). “Register,” in *The New Grove Dictionary of Music and Musicians*, eds S. Sadie, and J. Tyrrell (Oxford: Oxford University Press).

- Driscoll, C. A., Macdonald, D. W., and O'Brien, S. J. (2009). From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9971–9978. doi: 10.1073/pnas.0901586106
- Druzhkova, A. S., Thalmann, O., Trifonov, V. A., Leonard, J. A., Vorobieva, N. V., Ovodov, N. D., et al. (2013). Ancient DNA analysis affirms the canid from altai as a primitive dog. *PLoS One* 8:e57754. doi: 10.1371/journal.pone.0057754
- Dumbrill, R. (1998). *The Musicology and Organology of the Ancient Near East*. London: Tadmora Press.
- Duncan-Kemp, A. M. (1952). *Where Strange Paths Go Down*. Brisbane: W. R. Smith & Paterson.
- Dzenzelevskii, I. A. (1984). "Taboos in practice of Carpathian sheep-breeders [Запреты в практике карпатских овецодов]," in *Slavic and Balkan folklore: Ethnogenetic commonality and typological parallels [Славянский и балканский фольклор: этногенетическая общность и типологические параллели]*, ed. N. I. Tolstoy (Moscow: Muzyka), 256–277.
- Edwards, C. J., Ginja, C., Kantanen, J., Pérez-Pardal, L., Tresset, A., Stock, F., et al. (2011). Dual origins of dairy cattle farming - Evidence from a comprehensive survey of european Y-chromosomal variation. *PLoS One* 6:e15922. doi: 10.1371/journal.pone.0015922
- Eerola, T., Friberg, A., and Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Front. Psychol.* 4:487. doi: 10.3389/fpsyg.2013.00487
- Eerola, T., Järvinen, T., Louhivuori, J., and Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Percept.* 18, 275–296. doi: 10.1525/mp.2001.18.3.275
- Eerola, T., and Vuoskoski, J. K. (2013). A review of music and emotion studies: approaches, emotion models, and stimuli. *Music Percept.* 30, 307–340. doi: 10.1525/MP.2012.30.3.307
- Eklund, R., and Mcallister, A. (2015). "An Acoustic Analys of 'Kulning' (Cattle Calls) Recorded in an Outdoor Setting on Location in Dalarna (Sweden)," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, eds M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, and J. Stuart-Smith (Glasgow: University of Glasgow), 10–14.
- Eklund, R., Mcallister, A., and Dahlström, K. (2019). "An acoustic analysis of Swedish cattle calls, 'kulning', performed outdoors at three distances," in *Proceedings from FONETIK 2019 Stockholm, June 10–12, 2019*, ed. M. Heldner (Stockholm: Stockholm University), 61–66.
- Emsheimer, E. (ed.) (1991). "Schwedische Schellenmagie (Für Walter Wiora zum 70. Geburtstag)," in *Studia Ethnomusicologica Eurasiatica*, (Stockholm: Kungliga Musikaliska Akademien), 41–52.
- Etzel, J. A., Johnsen, E. L., Dickerson, J., and Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *Int. J. Psychophysiol.* 61, 57–69. doi: 10.1016/j.ijpsycho.2005.10.025
- Fabian, D., and Schubert, E. (2008). Musical character and the performance and perception of dotting, articulation and tempo in 34 recordings of variation 7 from J.S. Bach's Goldberg Variations (BWV 988). *Music. Sci.* 12, 177–206. doi: 10.1177/102986490801200201
- Fagg, M. C. (1997). *Rock Music*. Oxford: University of Oxford.
- Fallows, D. (2001). "Metronome," in *The New Grove Dictionary of Music and Musicians*, ed. G. Grove (Oxford: Oxford University Press)
- Farber, W. (1990). *Magic at the cradle: Babylonian and Assyrian lullabies*. Anthropos: Internationale Zeitschrift für Völker- und Sprachenkunde.
- Farmer, H. G. (1965). The ethos of antiquity. *Islamic Stud.* 4, 25–30.
- Felius, M., Beerling, M.-L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., and Lenstra, J. A. (2014). On the history of cattle genetic resources. *Diversity* 6, 705–750. doi: 10.3390/d6040705
- Fenk-Oczlon, G., and Fenk, A. (2009a). "Musical pitch in nonsense syllables: correlations with the vowel system and evolutionary perspectives," in *Proceedings of the 7th ESCOM 2009*, eds J. Louhivuori, T. Eerola, S. Saarikkallio, T. Himberg, and P.-S. Eerola (Jyväskylä: University of Jyväskylä), 110–113.
- Fenk-Oczlon, G., and Fenk, A. (2009b). Some parallels between language and music from a cognitive and evolutionary perspective. *Music. Sci.* 13, 201–226. doi: 10.1177/1029864909013002101
- Fétis, F.-J. (1994). *Esquisse de L'histoire de L'harmonie*. Hillsdale, NY: Pendragon Press.
- Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front. Psychol.* 7:1393. doi: 10.3389/fpsyg.2016.01393
- Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009
- Fitch, W. T. (2010). *The Evolution of Language*. Cambridge, MA: Cambridge University Press.
- Fitch, W. T. (2012). "The biology and evolution of rhythm: unravelling a paradox," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 73–95. doi: 10.1093/acprof:oso/9780199553426.003.0009
- Flatz, G., and Rothhauwe, H. W. (1973). Lactose nutrition and natural selection. *Lancet* 302, 76–77. doi: 10.1016/S0140-6736(73)93267-4
- Foster, M. L. (1994). "Symbolism: the foundation of culture," in *Companion Encyclopedia of Anthropology*, ed. T. Ingold (New York, NY: Routledge), 366–395.
- Fraisse, P. (1982). "Rhythm and tempo," in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 149–180. doi: 10.1016/b978-0-12-213562-0.50010-3
- Franklin, J. C. (2006). Lyre gods of the bronze age musical koine. *J. Ancient Near East. Relig.* 6, 463–482.
- Friberg, A., and Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *J. Acoust. Soc. Am.* 105, 1469–1484. doi: 10.1121/1.1426687
- Frödin, J. (1929). *Om Fäbodbebyggelsens Utbredning Och Olika Typer i Europa*. Uppsala: Geografiska institutionen, 176–194.
- Gabrielsson, A. (1999). "The performance of music," in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 501–602.
- Gabrielsson, A., and Juslin, P. N. (2003). "Emotional expression in music," in *Handbook of Affective Sciences Series in Affective Science*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (New York, NY: Oxford University Press), 503–534.
- Gabrielsson, A., and Lindström, E. (2001). "The influence of musical structure on emotional expression," in *Music and Emotion. Theory and Research*, eds P. N. Juslin and J. A. Sloboda (Oxford, NY: Oxford University Press), 223–248.
- Garbuzov, N. (1948). *Zonal Nature of Pitch Hearing [Зонная природа звуковысотного слуха]*. Moscow: Russian Academy of Science.
- Garbuzov, N. (1950). *Zonal Nature of Tempo and Rhythm [Зонная природа темпа и ритма]*. Moscow: Academy of Science of the USSR.
- Garbuzov, N. (1955). *Zonal Nature of Hearing of Dynamics [Зонная природа динамики слуха]*. Moscow: Gos Muz Izdat [State Musical Publishing].
- Garbuzov, N. (1956). *Zonal Nature of Hearing of Timbre [Зонная природа тембрового слуха]*. Moscow: Gos Muz Izdat [State Musical Publishing].
- Garroway, K. H. (2019). *Growing Up in Ancient Israel: Children in Material Culture and Biblical Texts*. Atlanta: SBL Press.
- Gembris, H. (2006). "The development of musical abilities" in *MENC Handbook of Musical Cognition and Development*, ed. R. Colwell (New York, NY: Oxford University Press), 124–164. doi: 10.1093/acprof:oso/9780195304565.003.0005 doi: 10.1093/acprof:oso/9780195304565.003.0005
- Gerbault, P., Roffet-Salque, M., Evershed, R. P., and Thomas, M. G. (2013). How long have adult humans been consuming milk? *IUBMB Life* 65, 983–990. doi: 10.1002/iub.1227
- Germonpré, M., Lázníková-Galetová, M., and Sablin, M. V. (2012). Palaeolithic dog skulls at the gravettian poedmosti site, the czech republic. *J. Archaeol. Sci.* 39, 184–202. doi: 10.1016/j.jas.2011.09.022
- Gimbutas, M. (1993). The Indo-Europeanization of Europe: the intrusion of steppe pastoralists from south Russia and the transformation of Old Europe. *WORD* 44, 205–222. doi: 10.1080/00437956.1993.11435900
- Gioia, T. (2006a). *Healing Songs*. Durham, NC: Duke University Press.
- Gioia, T. (2006b). *Work Songs*. Durham, NC: Duke University Press.
- Golub, H. L., and Corwin, M. J. (1985). "A physioacoustic model of the infant cry," in *Infant Crying: Theoretical and research perspectives*, eds C. F. Z. Boukydis and B. M. Lester (Boston, MA: Springer), 59–82. doi: 10.1007/978-1-4613-2381-5_3
- Gordon, B. (2003). Rangifer and man: an ancient relationship. *Rangifer* 23, 15–27. doi: 10.7557/2.23.5.1651
- Gourlay, K. A. (1982). Towards a humanizing ethnomusicology. *Ethnomusicology* 26, 411–420.
- Gourlay, K. A. (1984). The non-universality of music and the Universality of non-music. *World Music* 26, 25–39.

- Graf, W. T. (1980). *Vergleichende Musikwissenschaft: ausgewählte Aufsätze*. Vienna: Föhrenau.
- Gray, R. D., and Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435–439. doi: 10.1038/nature02029
- Gray, R. D., Atkinson, Q. D., and Greenhill, S. J. (2011). Language evolution and human history: what a difference a date makes. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1090–1100. doi: 10.1098/rstb.2010.0378
- Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2005). How does music arouse “chills”? Investigating strong emotions, combining psychological, physiological, and psychoacoustical methods. *Ann. N. Y. Acad. Sci.* 1060, 446–449. doi: 10.1196/annals.1360.041
- Guilford, T., and Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Anim. Behav.* 42, 1–14. doi: 10.1016/S0003-3472(05)80600-1 doi: 10.1016/s0003-3472(05)80600-1
- Halpern, A. R., Bartlett, J. C., and Dowling, W. J. (1998). Perception of mode, rhythm, and contour in unfamiliar melodies: effects of age and experience. *Music Percept.* 15, 335–355. doi: 10.2307/40300862
- Hancock, A. M., Witonsky, D. B., Ehler, E., Alkorta-Aranburu, G., Beall, C., Gebremedhin, A., et al. (2010). Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl.), 8924–8930. doi: 10.1073/pnas.0914625107
- Hansen, L. I., and Olsen, B. J. (2014). *Hunters in Transition: An Outline of Early Sámi History*. Leiden: Brill Publishers.
- Harding, R. E. M. (1983). *The Metronome and Its Precursors*. Oxford: Gresham Books.
- Hargreaves, D. J. (1996). “The development of artistic and musical competence,” in *Musical Beginnings*, eds I. Deliège and J. Sloboda (Oxford: Oxford University Press), 145–170.
- Harris, M. (1964). *The Nature of Cultural Things*. New York, NY: Random House.
- Harris, M. (1990). “Emics and etics revisited,” in *Emics and Etics: The Insider/Outsider Debate*, eds N. Thomas, K. Headland, L. Pike, and M. Harris (Newbury Park: SAGE Publications), 48–61.
- Hauser, A. (1999). *The Social History of Art: From Prehistoric Times to the Middle Ages*. New York: Psychology Press.
- Hauser, M. D. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.
- Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species’ vocalizations. *Cogn. Sci.* 24, 445–475. doi: 10.1016/S0364-0213(00)00026-4
- Headland, T. N. (1990). “Introduction: a dialogue between Kenneth Pike and Marvin Harris on emics and etics,” in *Emics and Etics: The Insider/Outsider Debate*, eds N. Thomas, K. Headland, L. Pike, and M. Harris (Newbury Park: SAGE Publications), 13–27.
- Heggarty, P. (2015). “Europe and western Asia: Indo-European linguistic history,” in *The Global Prehistory of Human Migration*, ed. P. S. Bellwood (Oxford UK: Wiley-Blackwell), 157–167.
- Helmer, R. H. P. (1975). *European Pastoral Calls and Their Possible Influence on Western Liturgical Chant*. New York, NY: Columbia University Press.
- Helskog, K. (2012). Ancient depictions of reindeer enclosures and their environment. *Fennoscandia Archaeol.* 29, 29–54.
- Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., May, H., Lee, O. Y.-C., Feldman, M., et al. (2015). Tuberculosis origin: the Neolithic scenario. *Tuberculosis* 95, S122–S126. doi: 10.1016/j.tube.2015.02.021
- Hillert, D. G. (2015). On the evolving biology of language. *Front. Psychol.* 6:1796. doi: 10.3389/fpsyg.2015.01796
- Hiltebeitel, A. (2011). “The indus valley ‘proto-siva’: reexamined through reflections on the goddess, the buffalo, and the symbolism of vāhanas,” in *When the Goddess was a Woman*, eds V. Adluri and J. Bagchee (Leiden: Brill Academic Publishers), 399–431.
- Hirsh-Pasek, K., and Treiman, R. (1982). Doggerel: motherese in a new context. *J. Child Lang.* 9, 229–237. doi: 10.1017/S0305000900003731
- Holahan, J. M. (1987). “Toward a Theory of Music Syntax: Some Observations of Music Babble in Young Children,” in *Music and Child Development*. New York, NY: Springer, 96–106.
- Honing, H. (2003). The final ritard: on music, motion, and kinematic models. *Comput. Music J.* 27, 66–72. doi: 10.1162/014892603322482538
- Horvath, T. (2012). “Animal deposits in the late copper age settlement of balatonőszöd-temetői dűlő, Hungary,” in *The Ritual Killing and Burial of Animals European Perspectives*, ed. A. Pluskowski (Oxford: Oxbow Books), 115–136.
- Hubbard, T. L. (2003). Further correspondences and similarities of shamanism and cognitive science: mental representation, implicit processing, and cognitive structures. *Anthropol. Consciousness* 14, 40–74. doi: 10.1525/ac.2003.14.1.40
- Hurford, J. R. (2012). *Language in the Light of Evolution: The Origins of Grammar*. Oxford, UK: Oxford University Press.
- Huron, D. (1989). “Characterizing Musical Textures,” in *Proceedings: 1989 International Computer Music Conference, November 2-5*, (Ann Arbor, MI: Michigan Publishing, University of Michigan Library), 131–134.
- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Ingold, T. (1986). Reindeer economies: and the origins of pastoralism. *Anthropol. Today* 2, 5–10. doi: 10.2307/3032710
- Ivarsson, A. (1986). *Sången I Skogen. Studier Kring Den Svenska Fäbodmusiken*. Dissertation, Uppsala University, Department of Musicology, Uppsala.
- Ivarsson, A. (1995). *Lockrop & Vallåtar - Ancient Swedish Pastoral Music*. Stockholm: Musica Sveciae.
- Ivarsson, A. (2004). “And the cattle follow her, for they know her voice ... On communication between women and cattle in Scandinavian pastures,” in *Pecus. Man and Animal in Antiquity*, ed. B. S. Frizell (Rome: The Swedish Institute in Rome), 146–149.
- Jerkert, J. (2003). *Measurements and Models of Musical Articulation*. Available online at: <http://www.speech.kth.se/publications/masterprojects/2004/jerkert.pdf> (accessed July 31, 2019).
- Johansson, S. (2005). *Origins of Language. Constraints on Hypotheses*. Amsterdam: J. Benjamins Pub. Co.
- Johnson, A. (1979). Fäbodmusik i Förvandling: Rapport Från En Exkursion Sommaren 1977. *Svensk Tidskrift Musikforskning* 60, 5–39.
- Johnson, A. (1984). Voice physiology and ethnomusicology: physiological and acoustical studies of the Swedish herding song. *Yearbook Tradit. Music* 16, 42–66. doi: 10.2307/768202
- Johnson, A. (1990). “The sprite in the water and the siren of the woods. On Swedish folk music and gender,” in *Music, Gender, and Culture*, eds M. Herndon and S. Ziegler (Wilhelmshaven: F. Noetzel), 27–40.
- Johnson, A., Sundberg, J., and Wilbrand, J. (1982). Study of phonation and articulation in a type of Swedish herding song. *Dept. Speech Music Hear. Q. Prog. Status Rep.* 23, 99–116.
- Johnson-Laird, P. N., and Oatley, K. (2010). “Emotions, music, and literature,” in *Handbook of Emotions*, eds M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York: The Guilford Press), 102–113.
- Jones, M. R. (2016). “Musical time,” in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford, UK: Oxford University Press), 125–141.
- Jones, M. R., and Large, E. W. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159. doi: 10.1037/0033-295x.106.1.119 doi: 10.1037/0033-295x.106.1.119
- Jones, M. R., and Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Mem. Cogn.* 19, 8–20. doi: 10.3758/BF03198492
- Jordania, J. (2011). *Why Do People Sing: Music in Human Evolution*. Tbilisi: Logos.
- Jørgensen, E. R. (2003). Western classical music and general education. *Philos. Music Educ. Rev.* 11, 130–140. doi: 10.2979/pme.2003.11.2.130
- Jürgens, U. (1995). “Neuronal control of vocal production in non-human and human primates,” in *Current Topics in Primate Vocal Communication*, eds E. Zimmermann, J.D. Newman, and U. Jürgens (Boston, MA: Springer), 199–206. doi: 10.1007/978-1-4757-9930-9_10
- Juslin, P. N. (2001). “Communicating Emotion in Music Performance: A Review and Theoretical Framework,” in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (Oxford, UK: Oxford University Press), 309–337.
- Juslin, P. N. (2005). “From mimesis to catharsis: expression, perception, and induction of emotion in music,” in *Musical Communication*, eds D. Miell, R. MacDonald, and D. J. Hargreaves (Oxford, UK: Oxford University Press), 85–116. doi: 10.1093/acprof:oso/9780198529361.003.0005

- Juslin, P. N. (2011). "Music and emotion: seven questions, seven answers," in *Music and the Mind*, eds I. Deliège and J. Davidson (Oxford, UK: Oxford University Press), 113–138.
- Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008
- Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770
- Juslin, P. N., and Västfjäll, D. (2008). All emotions are not created equal: reaching beyond the traditional disputes. *Behav. Brain Sci.* 31, 600–621. doi: 10.1017/S0140525X08005554
- Kaminska, Z., and Woolf, J. (2000). Melodic line and emotion: cooke's theory revisited. *Psychol. Music* 28, 133–153. doi: 10.1177/0305735600282003
- Kan-ool, A. K. (2012). Pastoral plots in the traditional Tuvian culture of the Erzya kozhuun. [С котоводческиезаговоры в традиционной культуре тувинцев рзинскожууна]. *Bull. East Siberian Acad. Culture Arts* 1, 31–35.
- Keller, H. (1973). *Phrasing and Articulation: A Contribution to a Rhetoric of Music, with 152 Musical Examples*. New York, NY: W.W. Norton.
- Kendall, R. A., and Carterette, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemp. Music Rev.* 9, 51–67. doi: 10.1080/07494469300640341
- Khannanov, I. I. (2005). *Russian Methodology of Musical Form and Analysis*. Dissertation, University of California, Santa Barbara, Department of Musicology, Santa Barbara, CA.
- Kharlap, M. (1972). "Traditional Russian musical system and the problem of origin of music [Народно-русскаямузыкальнаясистема и проб лемапоисхождениямузыки]," in *Early forms of art [ранниеформы искусства]*, ed. Y. Meletinskii (Moscow: Iskusstvo), 246–247.
- Kholopov, Y. (1976). Mode[ад]. *Encyclopedia Musica[музыкальнаяэнциклопедия]* 3, 130–143.
- Kholopov, Y. (2005). "Towards the problem of mode in Russian theoretic musicology [к проблемелада в русскомтеоретическоммузыкальномзнании]," in *Harmony: Problems of Science and Methodology [Гармония:проблемынауки и Методики]*, ed. E. Struchalina (Rostov-na-Donu: RGK (Rostov State Conservatory)), 135–157.
- Kholopov, Y. (2006). *Introduction into Musical Form [теориямузыка: мелодика, ритм, фактура, тематика]*. Moscow: Moscow State Tchaikovsky Conservatory.
- Kholopova, V. (2002). *Theory of Music: Melos, Rhythm, Texture, Thematicism [Теория музыки: мелодика, ритм, фактура, тематика]*. Sankt-Petersburg: Lan.
- Kirnarskaya, D., Kiyashchenko, N., Tarasova, K., and Tzypina, G. (2003). *Psychology of Musical Activities: Theory and Practice [психология музыкальной деятельности: Теория и практика]*. Moscow: Akademiya.
- Klein, R. G. (2009). *The Human Career: Human Biological and Cultural Origins*, 3rd Edn. Chicago, IL: University of Chicago Press.
- Kolinsky, R., Lidji, P., Peretz, I., Besson, M., and Morais, J. (2009). Processing interactions between phonology and melody: vowels sing but consonants speak. *Cognition* 112, 1–20. doi: 10.1016/j.cognition.2009.02.014
- Kolltveit, G. (2008). "Animal bells in early scandinavian soundscapes," in *Studies in Music Archaeology VI. Current Challenges and New Objectives in Music Archaeology*, eds E. Hickmann, J. Orlamünde, and R. Eichmann (Rahden: Marie Leidorf), 147–153.
- Kondratyeva, N. M. (1989). "Lullabies of Teleuts, Tubalars, Kumandins, Chelkans [колыбельныетелеутов, тубаларов, кумандинцев, члканцев]," in *Folkloric Heritage of the Mountain Altai [Фольклорное наследие Горного Алтая]*, ed. V. Edokov (Gorno-Altai: Gorno-Altai state institute of history, language and literature), 20–48.
- Kondratyeva, N. M. (1996). *Pastoral Magic Spells of Telengits [С котоводческиезаговоры теленгитов]*. Dissertation, Novosibirsk State Conservatory named after Glinka, Department of Ethnomusicology, Novosibirsk.
- Kondratyeva, N. M., and Kopytov, R. B. (2017). The experiment in notation of sung pastoral magic spells of Sutkholtuvans [опыт нотирования поющих сказотоводческих заговоров сутхольских тувинцев]. *J. Musical Sci.* 3, 77–83.
- Kondratyeva, N. M., and Mazepus, V. V. (1999). Modal organization of Telengit pastoral spells [Модальная организация теленгитских сказотоводческих заговоров]. *Lang. Indigenous Peoples Siberia* 6, 19–43.
- Korom, F. J. (2000). Holy cow! the apotheosis of zebu, or why the cow is sacred in hinduism. *Asian Folklore Stud.* 59, 181–203. doi: 10.2307/1178915
- Kozłowski, J. K. (2015). The origin of the Gravettian. *Q. Int.* 359–360, 3–18. doi: 10.1016/j.quaint.2014.03.025
- Kreitner, K., Térey-Smith, M., Westrup, J., Holoman, D. K., Hopkins, G. W., Griffiths, P., et al. (2001). "Instrumentation and orchestration" in *The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).
- Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York, NY: Oxford University Press.
- Krumhansl, C. L. (2002). Music: a link between cognition and emotion. *Curr. Direct. Psychol. Sci.* 11, 45–50. doi: 10.1111/1467-8721.00165
- Krushanov, A. I. (1987). *History and Culture of Chukchi: Historic-Ethnographic Sketches [История и культура чукчей: историко-этнографические очерки]*. Leningrad: Nauka.
- Kubik, G. (1999). *Africa and the Blues*. Jackson, MS: University Press of Mississippi.
- Kvamme, M. (1988). "Pollen analytical studies of mountain summer-farming in western Norway," in *The Cultural Landscape, Past, Present and Future*, eds H. H. Birks, H. J. B. Birks, P. E. Kaland, and D. Moe (Cambridge, MA: Cambridge University Press), 349–365.
- Kyrgyz, Z. K. (2002). *Tuvan Throat Singing [Туви́нское горлоное пение]*. Russia: Nauka.
- Large, E. W. (2008). "Resonating to musical rhythm: theory and experiment," in *Psychology of Time*, ed. S. Grondin (Bingley: Emerald Group Publishing Limited), 189–231.
- Large, E. W., and Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connect. Sci.* 6, 177–208. doi: 10.1080/09540099408915723
- Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y. W., Peters, J., et al. (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8878–8883. doi: 10.1073/pnas.1203005109
- Lerdahl, F., and Jackendoff, R. S. (1985). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Lester, J. (1989). *Between Modes and Keys: German Theory, 1592-1802*. New York, NY: Pendragon Press.
- Leuthold, H. (1981). *Der Naturjodel in der Schweiz: Wesen, Entstehung, Charakteristik, Verbreitung: ein Forschungsergebnis über den Naturjodel in der Schweiz*. Altdorf: Robert Fellmann Liederverlag.
- Levin, T. C., and Suzuki, V. (2006). *Where Rivers and Mountains Sing: Sound, Music, and Nomadism in Tuva and Beyond*. Bloomington, IN: Indiana University Press.
- Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Percept. Psychophys.* 56, 414–423. doi: 10.3758/BF03206733
- Levitin, D. J. (2009). *The World in Six Songs: How the Musical Brain Created Human Nature*. New York, NY: Plume Books.
- Lieberman, P. (1985). "The physiology of cry and speech in relation to linguistic behavior," in *Infant Crying: Theoretical and Research Perspectives*, eds C. F. Z. Boukydis and B. Lester (Boston, MA: Springer), 29–57. doi: 10.1007/978-1-4613-2381-5_2
- Lodrick, D. O. (2005). Symbol and sustenance: cattle in South Asian culture. *Dialect. Anthropol.* 29, 61–84. doi: 10.1007/s10624-005-5809-8
- Loewy, J. V. (1995). The musical stages of speech: a developmental model of pre-verbal sound making. *Music Ther.* 13, 47–73. doi: 10.1093/mt/13.1.47
- Lomax, A. (1977). Universals in Song. *World Music* 19, 117–129.
- London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford, New York: Oxford University Press.
- Lóugas, L., Kriiska, A., and Maldre, L. (2007). New dates for the Late Neolithic Corded Ware Culture burials and early husbandry in the East Baltic region. *Archaeofauna* 16, 21–31.
- Loui, P., Patterson, S., Sachs, M. E., Leung, Y., Zeng, T., and Przsinda, E. (2017). White matter correlates of musical Anhedonia: implications for evolution of music. *Front. Psychol.* 8:1664. doi: 10.3389/fpsyg.2017.01664

- Madison, G., and Paulin, J. (2010). Ratings of speed in real music as a function of both original and manipulated beat tempo. *J. Acoust. Soc. Am.* 128, 3032–3040. doi: 10.1121/1.3493462
- Maier, A. (2017). Population and settlement dynamics from the gravettian to the magdalenian. *Mitteilungen Gesellschaft Urgeschichte* 26, 83–102.
- Malmström, H., Linderholm, A., Lidén, K., Storå, J., Molnar, P., Holmlund, G., et al. (2010). High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol. Biol.* 10:89. doi: 10.1186/1471-2148-10-89
- Manser, M. B. (2010). “The generation of functionally referential and motivational vocal signals in mammals,” in *Handbook of Behavioral Neuroscience*, ed. S. M. Brudzynski (New York, NY: Elsevier), 477–486. doi: 10.1016/b978-0-12-374593-4.00043-7
- Manuel, P., and Blum, S. (2011). “Classical aesthetic traditions of India, China, and the Middle East,” in *The Routledge Companion to Philosophy and Music*, eds T. Gracyk and A. Kania (Abingdon: Routledge), 245–256.
- Marciniak, A. (2013). “Origin of stock-keeping and the spread of animal exploitation strategies in the early and middle Neolithic of the North European Plain,” in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, and S. Shennan (Walnut Creek, CA: Left Coast Press), 221–236.
- Marciniak, S., and Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nat. Rev. Genet.* 18, 659–674. doi: 10.1038/nrg.2017.65
- Marler, P. (1997). Three models of song learning: evidence from behavior. *J. Neurobiol.* 33, 501–516. doi: 10.1002/(sici)1097-4695(19971105)33:5(501::aid-neu2)3.0.co;2-8
- Marler, P. (2001). “Origins of music and speech: insights from animals,” in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 31–48.
- Marler, P., and Mitani, J. C. (2008). A phonological analysis of male gibbon singing behavior. *Behaviour* 109, 20–45. doi: 10.1163/156853989x00141
- Marler, P., and Slabbekoorn, H. W. (2004). *Nature's Music: the Science of Birdsong*. Amsterdam: Elsevier.
- Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evol. Anthropol. Issues News Rev.* 14, 54–67. doi: 10.1002/evan.20046
- Marshack, A. (1990). “Early hominid symbol and evolution of the human capacity,” in *The Emergence of Modern Human. An Archaeological Perspective*, ed. P. Mellars (Edinburgh: Edinburgh University Press), 457–498.
- Mathiesen, T. J. (1984). Harmonia and ethos in ancient Greek music. *J. Musicol.* 3, 264–279. doi: 10.1525/jm.1984.3.3.03a00050
- Mathiesen, T. J. (1999). *Apollo's Lyre: Greek Music and Music Theory in Antiquity and the Middle Ages*. Lincoln, NE: University of Nebraska Press.
- Mattheson, J., and Harriss, E. C. (1981). *Johann Mattheson's Der Vollkommene Capellmeister: A Revised Translation With Critical Commentary*. Ann Arbor, Mich.: UMI Research Press.
- Maynard-Smith, J. (1976). Evolution and the theory of games. *Am. Sci.* 64, 41–45.
- Maynard-Smith, J., and Harper, D. (2003). *Animal Signals*. Oxford: Oxford University Press.
- Mazel, L. (1979). *Structuring of Musical Works [Строение музыкальных произведений]*, 2nd Edn. Moscow: Muzyka.
- Mazel, L., and Tzukkerman, V. (1967). *Analysis of Musical Works [Анализы музыкальных произведений]*. Moscow: Muzyka.
- Mazepus, V. V. (1993). *Universal Grammatic Approach to Culturology [Универсально-грамматический подход в культурологии]*. Novosibirsk: State Conservatory named after Glinka.
- Mazepus, V. V., and Galitskaya, S. P. (1997). *Musical Culture of Siberia. Traditional Culture of Indigenous People of Siberia [Музыкальная культура Сибири. Традиционная культура коренных народов Сибири]*. Novosibirsk: Novosibirsk State Conservatory named after Glinka.
- McAdams, S. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Percept. Interdiscip. J.* 22, 207–237. doi: 10.1525/mp.2004.22.2.207
- McAuley, J. D. (2010). “Tempo and rhythm,” in *Springer Handbook of Auditory Research*, eds J. M. Riess, R. Fay, and A. Popper (New York: Springer), 165–199.
- McBrearty, S., and Brooks, A. S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J. Hum. Evol.* 39, 543–563. doi: 10.1006/JHEV.2000.0435
- McConnell, P. B. (1990). Acoustic structure and receiver response in domestic dogs, *Canis familiaris*. *Anim. Behav.* 39, 897–904. doi: 10.1016/S0003-3472(05)80954-6
- McConnell, P. B. (1991). Lessons from animal trainers: the effect of acoustic structure on an animal's response. *Perspect. Ethol.* 9, 165–187.
- McConnell, P. B. (2002). *The Other End of the Leash: Why We Do What We Do Around Dogs*. New York: Ballantine Books.
- McConnell, P. B., and Baylis, J. R. (1985). Interspecific communication in cooperative herding: acoustic and visual signals from human shepherds and herding dogs. *Zeitschrift Tierpsychol.* 67, 302–328. doi: 10.1111/j.1439-0310.1985.tb01396.x
- McCreless, P. (2002). “Music and rhetoric,” in *The Cambridge History of Western Music Theory* (Cambridge: Cambridge University Press), 847–879. doi: 10.1111/j.1439-0310.1985.tb01396.x
- McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Curr. Biol. Biol.* 20, 1035–1041. doi: 10.1016/j.cub.2010.04.019
- McInerney, J. (2010). *The Cattle of the Sun: Cows and Culture in the World of the Ancient Greeks*. Princeton, NJ: Princeton University Press.
- Medushevsky, V. V. (2010). *On Rules and Means of Artistic Influence of Music [О закономерностях и средствах художественного воздействия искусства]*. Moscow: Muzyka.
- Meier, B., and Dittmer, L. A. (1956). The musica reservata of adrianus petit coclico and its relationship to josquin. *Musica Discip.* 10, 67–105.
- Mellars, P. (2004). “Archaeology and the origins of modern humans: European and African perspectives,” in *The Speciation of Modern Homo Sapiens*, ed. T. J. Crow (London: British Academy), 31–47.
- Mertens, P. (2004). “The prosogram: semi-automatic transcription of prosody based on a tonal perception model,” in *Proceedings of the 2nd International Conference on Speech Prosody*, Nara, 549–552.
- Messner, G. F. (1993). Ethnomusicological research, another “Performance” in the international year of indigenous peoples? *World Music* 35, 81–95.
- Meyer, J. (2009). *Acoustics and The Performance of Music Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instruments Makers. 5th (1978)*. Berlin: Springer.
- Miklosi, A. (2015). *Dog Behaviour, Evolution, and Cognition*. Oxford: Oxford University Press.
- Miller, R. J. (2014). *Contemporary Orchestration: A Practical Guide to Instruments, Ensembles, and Musicians*. New York, NY: Routledge.
- Mills, S. (2016). *Auditory Archaeology*. London: Routledge.
- Milne, L. S. (2017). The terrors of the night: charms against the nightmare and the mythology of dreams. *incantatio*. *Int. J. Charms Charm. Charm.* 6, 78–116. doi: 10.7592/incantatio2017.6.milne
- Mitchell, R. W. (2001). Americans' talk to dogs: similarities and differences with talk to infants. *Res. Lang. Soc. Interact.* 34, 183–210. doi: 10.1207/S15327973RLSI34-2.2
- Mitchell, S. A. (2001). Warlocks, valkyries and varlets: a prolegomenon to the Study of North Sea witchcraft terminology. *Cosmos* 17, 59–81.
- Moberg, C.-A. (1955). Om vallåtar. En studie i de svenska fåbodarnas musikaliska organisation [On herding calls: a study of the musical organization of the Swedish summe]. *Svensk Tidskrift r Musikforskning* 37, 1–27.
- Moberg, C.-A. (1971). *Studien zur schwedischen Volksmusik*. Uppsala: Almqvist & Wiksell.
- Mohn, C., Argstatter, H., and Wilker, F.-W. (2010). Perception of six basic emotions in music. *Psychol. Music* 39, 503–517. doi: 10.1177/0305735610378183
- Molino, J. (1990). Musical fact and the semiology of music. *Music Anal.* 9:105. doi: 10.2307/854225
- Monahan, C. B., and Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Percept. Interdiscip. J.* 3, 1–32. doi: 10.2307/40285320 doi: 10.2307/40285320
- Monahan, C. B., Kendall, R. A., and Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Percept. Psychophys.* 41, 576–600. doi: 10.3758/BF03210491

- Monelle, R. (1992). *Linguistics and Semiotics in Music*. Reading: Harwood Academic Publishers.
- Monelle, R. (2000). *The Sense of Music: Semiotic Essays*. Princeton, NJ: Princeton University Press.
- Monelle, R. (2006). *The Musical Topic: Hunt, Military and Pastoral*. Indianapolis, IN: Indiana University Press.
- Montagu, J. (2004). How old is music? *Galpin Soc. J.* 57, 171–182.
- Moog, H. (1976). *The Musical Experience of the Pre-School Child*. London: Schott Music.
- Moore, A. F. (2011). The end of the revival: the folk aesthetic and its ‘mutation.’ *Popular Music History* 4, 289–307. doi: 10.1558/pomh.v4i3.289
- Morley, I. (2006). Mousterian musicianship? The case of the divje babe i bone. *Oxf. J. Archaeol.* 25, 317–333. doi: 10.1111/j.1468-0092.2006.00264.x
- Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford: Oxford University Press.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* 111, 855–869. doi: 10.1086/283219
- Münzel, S., and Conard, N. J. (2009). “Flötenklang aus fernen zeiten die frühesten musikinstrumente,” in *Eiszeit. Kunst und Kultur. Begleitband zur großen Landesausstellung*, ed. E. K. U. Tübingen (Ostfildern: Archäologisches Landesmuseum Baden Württemberg), 317–321.
- Murashkin, A. I., Kolpakov, E. M., Shumkin, V. Y., Khartanovich, V. I., and Moiseyev, V. G. (2016). Kola olenostrovskiy grave field: a unique burial site in the european arctic. *ISKOS* 21, 185–199.
- Myers, H. (1993). *Ethnomusicology: Historical and Regional Studies*. New York, NY: Norton.
- Naguib, M., and Riebel, K. (2014). “Singing in space and time: the biology of birdsong,” in *Biocommunication of Animals*, ed. G. Witzany (Dordrecht: Springer), 233–247. doi: 10.1007/978-94-007-7414-8_13
- Nakata, T., and Trehub, S. E. (2004). Infants’ responsiveness to maternal speech and singing. *Infant Behav. Dev.* 27, 455–464. doi: 10.1016/j.infbeh.2004.03.002 doi: 10.1016/j.infbeh.2004.03.002
- Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago, IL: Chicago Review Press.
- Nash, P. H., and Carney, G. O. (1996). The seven themes of music geography. *The Canadian Geographer/Le Gographe Canadien* 40, 69–74. doi: 10.1111/j.1541-0064.1996.tb00433.x
- Nattiez, J.-J. (1990). *Music and Discourse: Toward a Semiology of Music*. Princeton, NJ: Princeton University Press.
- Nattiez, J.-J. (2012). Is the search for universals incompatible with the study of cultural specificity? *Hum. Soc. Stud.* 1, 67–94. doi: 10.2478/v10317-012-0005-2
- Nawrot, E. S. (2003). The perception of emotional expression in music: evidence from infants, children and adults. *Psychol. Music* 31, 75–92. doi: 10.1177/0305735603031001325
- Nazaikinsky, Y. V. (1972). *On Psychology of Human Musical Perception [О психологии музыкального восприятия]*. Moscow: Muzyka.
- Nazaikinsky, Y. V. (1973). “On constants in perception of music [О Константности в Восприятии Музыки],” in *Musical Art and Science [Музыкальное искусство и наука]*, ed. Y. V. Nazaikinsky (Moscow: Muzyka), 59–98.
- Nazaikinsky, Y. V. (1982). *The Logic of Musical Composition [Логи ка музыкальной композиции]*. Moscow: Muzyka.
- Nazaikinsky, Y. V. (1988). *The Sonic World of Music [Звуковой мир музыки]*. Moscow: Muzyka.
- Nazaikinsky, Y. V. (2013). *Style and Genre in Music [Смль и жанр в музыке]*. Tbilisi: Tbilisi State Conservatoire.
- Nazaikinsky, Y. V., and Rags, Y. N. (1964). “Perception of musical timbres and the significance of the individual harmonics in a sound [восприятии музыкальных тембров и значении отдельных гармоник звука],” in *Application of the acoustic methods in musicology [Применение акустических методов в музыковедении]*, ed. S. S. Skrebkov (Moscow: Muzyka), 79–100.
- Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-one Issues and Concepts*. Champaign, IL: University of Illinois Press.
- Nettl, B. (2010). *Nettl’s Elephant: On the History of Ethnomusicology*. Champaign, IL: University of Illinois Press.
- Nielsen, S. (1997). “Worksongs in Denmark,” in *Historical studies on folk and traditional music: ICTM Study Group on Historical Sources of Folk Music, conference report, Copenhagen, 24-28 April 1995*, eds D. Stockmann and J. H. Koudal (Copenhagen: Danish Folklore Archives), 241–247.
- Nikolsky, A. (2015a). “¿Cómo funciona la emoción musical? [How Emotion Can Be the Meaning of a Music Work],” in *Música y cuerpo: Estudios musicológicos*, ed. T. Cascudo (Baleares: Calanda Ediciones Musicales), 241–262.
- Nikolsky, A. (2015b). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405
- Nikolsky, A. (2016). Evolution of tonal organization in music optimizes neural mechanisms in symbolic encoding of perceptual reality. part-2: ancient to seventeenth century. *Front. Psychol.* 7:211. doi: 10.3389/fpsyg.2016.00211
- Nikolsky, A. (2018). Commentary: the ‘Musilanguage’ model of language evolution. *Front. Psychol.* 9:75. doi: 10.3389/fpsyg.2018.00075
- Nikolsky, A. (2020). “Emergence of the distinction between ‘verbal’ and ‘musical’ in early childhood development,” in *The Origins of Language Revisited: Differentiation from Music and the Emergence of Neurodiversity and Autism*, ed. N. Masataka (Singapore: Springer).
- Nikolsky, A., Alekseyev, E. Y., Alekseev, I. Y., and Dyakonova, V. E. (2020). The overlooked tradition of ‘personal music’ and its place in the evolution of music. *Front. Psychol.* 10:3051. doi: 10.3389/fpsyg.2019.03051
- Noorden, L. (1975). *Temporal Coherence in the Perception of Tone Sequences*. Eindhoven: Institute for Perceptual Research.
- Novik, Y. S. (1999). “Semiotic functions of human voice in folklore and beliefs of Siberian peoples [Семиотические функции голоса в фольклоре и верованиях народов Сибири],” in *Eastern Folklore and Mythology from the Comparative-Typological Aspect*, eds N. Lidova and N. Nikulin (Moscow: Russian Academy of Science), 217–235.
- Novik, Y. S. (2004). *Rite and Folklore in Siberian Shamanism: An Experiment in Correlation of Structures [Обряд и фольклор в си би рском шаманизме: Опыт сопоставления язмрукмур]*. Moscow: Eastern Literature, Russian Academy of Science.
- Owren, M. J., and Rendall, D. (2001). Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evol. Anthropol. Issues News Rev.* 10, 58–71. doi: 10.1002/evan.1014 doi: 10.1002/evan.1014
- Oxenham, A. J. (2013). “The perception of musical tones,” in *The Psychology of Music*, ed. D. Deutsch (San Diego, CA: Academic Press), 1–33. doi: 10.1016/b978-0-12-381460-9.00001-8
- Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Process.* 60, 133–155. doi: 10.1016/S0376-6357(02)00080-3
- Panteleva, O. (2019). How soviet musicology became marxist. *Slavonic East Eur. Rev.* 97:73. doi: 10.5699/slaveasteurorev2.97.1.0073
- Papoušek, H., and Papoušek, M. (1995). “Beginning of Human Musicality,” in *Music and The Mind Machine: The Psychophysiology and Psychopathology of the Sense of Music*, ed. R. Steinberg (Berlin: Springer), 27–34. doi: 10.1007/978-3-642-79327-1_3
- Parncutt, R., and Hair, G. (2011). Consonance and dissonance in music theory and psychology: disentangling dissonant dichotomies. *J. Interdiscip. Music Stud.* 5, 119–168.
- Patel, A. D. (2006). An empirical method for comparing pitch patterns in spoken and musical melodies: a comment on Pearl’s “Eavesdropping with a master: leos Janáček and the music of speech”. *Empiric. Musicol. Rev.* 1, 166–169. doi: 10.18061/1811/24011
- Patel, A. D. (2008). “Music as a transformative technology of the mind,” in *Music: Its Evolution, Cognitive Basis, and Spiritual Dimensions*, ed. M. A. Meyers (Cambridge, MA: John Templeton Foundation).
- Patel, A. D. (2010). “Music, biological evolution, and the brain,” in *Emerging Disciplines: Shaping New Fields of Scholarly Inquiry in and Beyond the Humanities*, eds C. Levander and C. Henry (Houston, TX: Rice University Press), 91–144.
- Patterson, R. D., Gaudrain, E., and Walters, T. C. (2010). “The perception of family and register in musical tones,” in *Springer Handbook of Auditory Research*, eds M. Riess Jones, R. Fay, and A. Popper (New York, NY: Springer), 13–50. doi: 10.1007/978-1-4419-6114-3_2

- Payne, K. (2001). "The progressively changing songs of humpback whales: a window on the creative process in a wild animal," in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 135–150.
- Pegg, C. (1992). Mongolian conceptualizations of overtone singing (xöömii). *Br. J. Ethnomusicol.* 1, 31–54. doi: 10.1080/09681229208567199
- Pegg, C. (2001). *Mongolian Music, Dance and Oral Narrative: Performing Diverse Identities*. Seattle, WA: University of Washington Press.
- Peretz, I. (2013). "Towards a neurobiology of musical emotions," in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: OUP Oxford), 99–126. doi: 10.1093/acprof:oso/9780199230143.003.0005
- Perlovsky, L. (2012). Cognitive function, origin, and evolution of musical emotions. *Music. Sci.* 16, 185–199. doi: 10.1177/1029864912448327
- Peters, G. (1984). On the structure of friendly close range vocalizations in terrestrial carnivores (Mammalia: Carnivora: Fissipedia). *Zeitschrift Säugetierkunde* 49, 157–182.
- Peters, J., Pöllath, N., and Arbuckle, B. (2017). "The emergence of livestock husbandry in Early Neolithic Anatolia," in *The Oxford Handbook of Zooarchaeology*, eds U. Albarella, M. Rizzetto, H. Russ, K. Vickers, and S. Viner-Daniels (Oxford, UK: Oxford University Press), 247–265.
- Pettitt, P. (2010). *The Palaeolithic Origins Of Human Burial*. London: Routledge.
- Pfordresher, P. Q., and Brown, S. (2017). Vocal mistuning reveals the origin of musical scales. *J. Cogn. Psychol.* 29, 35–52. doi: 10.1080/20445911.2015.1132024
- Pike, K. L. (1967). *Language in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague: Mouton de Gruyter.
- Pike, K. L. (1990). "On the emics and etics of Pike and Harris," in *Emics and Etics: The Insider/Outsider Debate*, eds N. Thomas, K. Headland, L. Pike, and M. Harris (Newbury Park: SAGE Publications), 28–47.
- Pikovsky, A., Rosenblum, M., and Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge, MA: Cambridge University Press.
- Plantinga, B. (2004). *Yodel-Ay-Ee-Oooo. The secret History of yodeling around the World*. Abingdon: Routledge.
- Plantinga, T. S., Alonso, S., Izagirre, N., Hervella, M., Fregel, R., van der Meer, J. W., et al. (2012). Low prevalence of lactase persistence in Neolithic South-West Europe. *Eur. J. Hum. Genet.* 20, 778–782. doi: 10.1038/ejhg.2011.254
- Plotnikova, A. A. (1999a). "On the symbolism of whistling [О символизме виста]," in *The World of Sound and Silence: Semiotics of Soound and Speech in Traditional Culture of Slavs [Мирзвучающий и молчащий: Семьюми казэу ка и речив мэри цюнной куьльмуре славян]*, ed. S. M. Tolstaya (Moscow: Indrik), 295–304.
- Plotnikova, A. A. (1999b). "The sonic magic in Slavic pastoral rituals [Магия звука в славянской скотоводческой обрядности]," in *The World of Sound And Silence: Semiotics of Soound and Speech in Traditional Culture of Slavs [Мирзвучающий и Молчащий: Семьюми казэу ка и речи в мэри и ци онной куьльмуре славян]*, ed. S. M. Tolstaya (Moscow: Indrik), 73–84.
- Porter, J., Powers, H. S., Cowdery, J., Widdess, R., Davis, R., Perlman, M., et al. (2001). "Mode," in *Modal scales and traditional music. Middle East and Asia*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).
- Potter, D. D., Fenwick, M., Abecasis, D., and Brochard, R. (2009). Perceiving rhythm where none exists: event-related potential (ERP) correlates of subjective accenting. *Cortex* 45, 103–109. doi: 10.1016/J.CORTEX.2008.01.004
- Powell, A., Shennan, S., and Thomas, M. G. (2009). Late Pleistocene demography and the appearance of modern human behavior. *Science* 324, 1298–1301. doi: 10.1126/science.1170165
- Powers, H. S., and Wiering, F. (2001). *Mode. The Term. Medieval Modal Theory. Modal Theories and Polyphonic Music*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).
- Prakash, O. (1961). *Food and Drinks in Ancient India (From Earliest Times to 200 A. D.)*. Delhi: Munshi Ram Manohar Lal.
- Prince, J. B. (2014). Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 2319–2337. doi: 10.1037/a0038010
- Prince, J. B., Schmuckler, M. A., and Thompson, W. F. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Mem. Cogn.* 37, 368–381. doi: 10.3758/MC.37.3.368
- Probst, E. (1991). *Deutschland in der Steinzeit: Jäger, Fischer und Bauern zwischen Nordseeküste und Alpenraum*. Munich: Carl Bertelsmann.
- Quam, R. M., Martinez, I., Rosa, M., and Arsuaga, J. L. (2017). "Evolution of hearing and language in fossil hominins," in *Primate Hearing and Communication*, eds R. M. Quam, M. Rosa, and J. L. Arsuaga (Berlin: Springer), 201–231. doi: 10.1007/978-3-319-59478-1_8
- Rags, Y. N. (1980). *Garbuzov N.A. - Musician, researcher and pedagogue [Гарбузов Н.А. - Музыкант, исследователь, педагог]*. Moscow: Muzyka.
- Rags, Y. N. (1999). *Aesthetics from the bottom and aesthetics from the top - quantitative ways of approaching [Эстетика снизу и эстетика сверху - квантитативные пути сближения]*. Moscow: Nauchnyi mir.
- Ratner, L. G. (1980). *Classic Music: Expression, form, and Style*. New York, NY: Schirmer Books.
- Ravens, S. (2014). *The Supernatural Voice: A History of High Male Singing*. Woodbridge: The Boydell Press.
- Ray, K., and Thomas, J. (2003). "In the kinship of cows: the social centrality of cattle in the earlier Neolithic of southern Britain," in *Food, Culture and Identity in the Neolithic and Early Bronze Age*, ed. M. Parker-Pearson (Oxford: Hadrian Books), 37–52.
- Reimers, E., and Colman, J. E. (2009). Reindeer and caribou (*Rangifer tarandus*) response towards human activities. *Rangifer* 26, 55–71. doi: 10.7557/2.26.2.188
- Renfrew, C. (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape.
- Repp, B. H. (1995). Acoustics, perception, and production of legato articulation on a digital piano. *J. Acoust. Soc. Am.* 97, 3862–3874. doi: 10.1121/1.413065
- Repp, B. H. (1998). *Perception and Production of Staccato articulation on the Piano*. New Haven, CT: Haskins Laboratories.
- Réti, R. (1951). *The Thematic Process in Music*. New York, NY: Macmillan Publishers.
- Reznikoff, I. (2008). Sound resonance in prehistoric times: a study of Paleolithic painted caves and rocks. *J. Acoust. Soc. Am.* 123, 3603–3603. doi: 10.1121/1.2934773
- Rice, T. (1987). Toward the remodeling of ethnomusicology. *Ethnomusicology* 31, 469–488. doi: 10.2307/851667
- Richerson, P. J., and Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago, IL: University of Chicago Press.
- Richerson, P. J., Boyd, R., and Bettinger, R. L. (2009). Cultural innovations and demographic change. *Hum. Biol.* 81, 211–235. doi: 10.3378/027.081.0306
- Robb, J. (1991). Random causes with directed effects: the Indo-European language spread and the stochastic loss of lineages. *Antiquity* 65, 287–291. doi: 10.1017/S0003598X00079758
- Robb, J. (1993). A social prehistory of European languages. *Antiquity* 67, 747–760. doi: 10.1017/S0003598X00063766
- Røed, K. H., Bjørklund, I., and Olsen, B. J. (2018). From wild to domestic reindeer – Genetic evidence of a non-native origin of reindeer pastoralism in northern Fennoscandia. *J. Archaeol. Sci. Rep.* 19, 279–286. doi: 10.1016/j.jasrep.2018.02.04802.048
- Røed, K. H., Flagstad, Ø, Nieminen, M., Holand, Ø, Dwyer, M. J., Røv, N., et al. (2008). Genetic analyses reveal independent domestication origins of Eurasian reindeer. *Proc. R. Soc. B Biol. Sci.* 275, 1849–1855. doi: 10.1098/rspb.2008.0332
- Roesner, E. H. (2001). *Rhythmic Modes (Modal Rhythm). The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).
- Rosenberg, S. (2003). *Kulning. Musiken och metoden*. Stockholm: Udda Toner.
- Rosenberg, S. (2014). Kulning – an ornamentation of the surrounding emptiness: about the unique Scandinavian herding calls. *Voice Speech Rev.* 8, 100–105. doi: 10.1080/23268263.2013.829712
- Rothstein, W. N. (1989). *Phrase Rhythm in Tonal Music*. New York, NY: Schirmer Books.
- Rouget, G. (2011). Musical efficacy: musicking to survive—the case of the pygmies. *Yearbook Tradit. Music* 43, 89–121. doi: 10.5921/yeartradmusi.43.0089
- Rowley-Conwy, P. (2011). Westward Ho! The spread of agriculture from central Europe to the Atlantic. *Curr. Anthropol.* 52, S431–S451. doi: 10.1086/658368
- Rowley-Conwy, P. (2013). "North of the frontier: early domestic animals in northern Europe," in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, K. Manning, and S. Shennan (Walnut Creek, CA: Left Coast Press), 283–312.

- Ruwet, N., and Everist, M. (1987). Methods of analysis in musicology. *Music Anal.* 6, 11–36. doi: 10.2307/854214
- Sachs, M. E., Damasio, A. R., and Habibi, A. (2015). The pleasures of sad music: a systematic review. *Front. Hum. Neurosci.* 9:404. doi: 10.3389/fnhum.2015.00404
- Sadie, S. (2001). "Movement," in *The New Grove Dictionary of Music and Musicians*, Stanley, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).
- Samson, J. (2001). *Genre*. *The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).
- Saña, M. (2013). "Domestication of animals in the Iberian Peninsula," in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, K. Manning, and S. Shennan (Walnut Creek, CA: Left Coast Press), 195–220.
- Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "Blended" instrument pairings in orchestration. *Music Percept. Interdiscip. J.* 13, 209–246. doi: 10.2307/40285694
- Savage, P. E., and Brown, S. (2013). Toward a new comparative musicology. *Anal. Approach. World Music* 2, 148–197.
- Schiavio, A., van der Schyff, D., Cespedes-Guevara, J., and Reybrouck, M. (2016). Enacting musical emotions, sense-making, dynamic systems, and the embodied mind. *Phenomenol. Cogn. Sci.* 16, 785–809. doi: 10.1007/s11097-016-9477-8
- Schubert, E., and McPherson, G. E. (2015). "Underlying mechanisms and processes in the development of emotion perception in music," in *The Child as Musician: A handbook of musical development*, ed. G. E. McPherson (Oxford: Oxford University Press), 221–243. doi: 10.1093/acprof:oso/9780198744443.003.0012
- Schulting, R. (2013). "On the northwestern fringes: earlier Neolithic subsistence in Britain and Ireland as seen through faunal remains and stable isotopes," in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, K. Manning, and S. Shennan (Walnut Creek, CA: Left Coast Press), 313–338.
- Searcy, W. A. (1992). "Measuring responses of female birds to male song," in *Playback and Studies of Animal Communication*, ed. P. K. McGregor (Boston, MA: Springer), 175–189. doi: 10.1007/978-1-4757-6203-7_12
- Sebeok, T. (1994). *Signs: An Introduction to Semiotics*. Toronto: University of Toronto Press.
- Seyfarth, R. M., and Cheney, D. L. (2017). Precursors to language: social cognition and pragmatic inference in primates. *Psychon. Bull. Rev.* 24, 79–84. doi: 10.3758/s13423-016-1059-9
- Shatkovsky, G. I. (1986). *The Development of Musical Hearing and Skills [Развитие музыкального слуха и навыков]*. Moscow: Russian Ministry of Culture.
- Sheikin, Y. I. (2002). *History of Music Culture of Siberian Ethnicities: A Comparative Historic Investigation [История музыкальной культуры народов Сибири: сравнительно-историческое исследование]*. Moscow: Russian Academy of Science.
- Sheikin, Y. I. (2017). *History of World Musical Culture: From Antiquity to 16th Century AD. [История мировой музыкальной культуры: от древних времен до XVI века н.э.]*. Yakutsk: The Ministry of culture and spiritual development of Sakha Republic.
- Shestakov, V. P. (1975). *From Ethos to Affect. History of musical aesthetics from Antiquity to the 18th century [От этоса к Аффекту. История музыкальной эстетики от античности до XVIII века]*. Moscow: Muzyka.
- Shevtsov, V. N. (1988). "Hunting and pastoral sound imitations and cries among the Khakas [Охотничье-скотоводческие звукоподражания и возгласы у хакасов]," in *Musical Ethnography of Northern Asia [Музыкальная этнография Северной Азии]*, ed. Y. I. Sheikin (Novosibirsk: Novosibirsk State Conservatory named after Glinka), 108–129.
- Shvachkin, N. H. (1948). "Development of phonematic perception in early childhood [Развитие фонематического восприятия речи в раннем возрасте]," in *The Problems of Psychology of Perception and Cognition. Works of the Institute of Psychology [Вопросы Психологии Восприятия и Мышления. Труды Института Психологии]*, ed. V. Teplov (Moscow: Izvestiya of Academy of Pedagogical Sciences of Russia), 101–133.
- Sicoli, M. A. (2015). "Voice registers," in *The Handbook of Discourse Analysis*, eds D. Tannen, H. E. Hamilton, and D. Schiffrin (Chichester, MA: John Wiley & Sons), 105–126. doi: 10.1002/9781118584194.ch5
- Sikora, M., Pitulko, V. V., Sousa, V. C., Allentoft, M. E., Vinner, L., Rasmussen, S., et al. (2019). The population history of northeastern Siberia since the Pleistocene. *Nature* 570, 182–188. doi: 10.1038/s41586-019-1279-z
- Silvia, P. J., and Nusbaum, E. C. (2011). On personality and piloerection: individual differences in aesthetic chills and other unusual aesthetic experiences. *Psychol. Aesth. Creat. Arts* 5, 208–214. doi: 10.1037/a0021914
- Simpson, B. S. (1997). Canine communication. The veterinary clinics of North America. *Small Anim. Pract.* 27, 445–464. doi: 10.1016/S0195-5616(97)50048-9
- Sjögren, K.-G., and Price, T. D. (2013). A complex Neolithic economy: isotope evidence for the circulation of cattle and sheep in the TRB of western Sweden. *J. Archaeol. Sci.* 40, 690–704. doi: 10.1016/j.jas.2012.08.001
- Skrebkova-Filatova, M. S. (1985). *Texture in music [Фактура в музыке]*. Moscow: Muzyka.
- Skrede, A. M. (2005). "Shielings and landscape in western Norway—Research traditions and recent trends," in *"Utmark" - the Outfield as Industry and Ideology in the Iron Age and the Middle Ages*, eds K. A. Bergsvik and A. Engevik (Bergen: Universitetet i Bergen), 31–41.
- Slater, P. (2001). "Birdsong repertoires: their origin and use," in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 49–63.
- Slater, P. (2011). *Bird Song and Language*. Oxford: Oxford University Press, doi: 10.1093/oxfordhb/9780199541119.013.0008
- Snowdon, C. T. (2003). "Expression of emotion in nonhuman animals," in *Handbook of Affective Sciences*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (Oxford, UK: Oxford University Press), 457–480.
- Snowdon, C. T., and Teie, D. (2013). "Emotional communication in monkeys: music to their ears?," in *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford, UK: Oxford University Press), 133–154.
- Snowdon, C. T., Zimmermann, E., and Altenmüller, E. (2015). Music evolution and neuroscience. *Prog. Brain Res.* 217, 17–34. doi: 10.1016/bs.pbr.2014.11.019 doi: 10.1016/bs.pbr.2014.11.019
- Sodgerel, T. (2012). "Herder's flute tsuur in traditional culture of Western Mongolia [Пастушечья флейта цуур в традиционной культуре Западной Монголии]," in *Culture of Mongol-speaking peoples in globalized space. International scientific forum, Elista, 24-27 October, 2012 [Культура монгольязычных народов в глобализирующемся пространстве]*, ed. K. Nadneyeva (Elista: Kalmyk State University), 60–61.
- Sodgerel, T. (2016). Patterns of playing the Mongolian longitudinal flute: features of the repertoire. [Наигрыши на монгольской продольной флейте: особенности репертуара]. *Bull. Saint Petersburg State Instit. Cult.* 3, 126–129.
- Soffer, O., Adovasio, J. M., and Hyland, D. C. (2000). The "Venus" figurines. *Curr. Anthropol.* 41, 511–537. doi: 10.1086/317381
- Sørensen, L., and Karg, S. (2014). The expansion of agrarian societies towards the north - new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. *J. Archaeol. Sci.* 51, 98–114. doi: 10.1016/j.jas.2012.08.042 doi: 10.1016/j.jas.2012.08.042
- Stebbins, W. C., and Moody, D. B. (2011). "How monkeys hear the world: auditory perception in nonhuman primates," in *Comparative Hearing: Mammals. Springer Handbook of Auditory Research*, eds R. R. Fay and A. N. Popper (New York, NY: Springer), 97–133. doi: 10.1007/978-1-4612-2700-7_4
- Stépanoff, C., Marchina, C., Fossier, C., and Bureau, N. (2017). Animal autonomy and intermittent coexistences: north asian modes of herding. *Curr. Anthropol.* 58, 57–81. doi: 10.1086/690120
- Strahley, I. C., and Loebach, J. L. (2014). The influence of mode and musical experience on the attribution of emotions to melodic sequences. *Psychomusicology* 24, 21–34. doi: 10.1037/pmu0000032
- Stricklin, W. R. (2001). "The evolution and domestication of social behavior," in *Social Behavior in Farm Animals*, eds J. L. Keeling and H. W. Gonyou (Wallingford: CABI publishing), 83–110. doi: 10.1079/9780851993973.0083
- Stuart, C. K. (2008). *Life and Marriage in Skya Rgya, a Tibetan Village - Blo Brtan Rdo Rje*. New York, NY: YBK Publishers.
- Suisman, D. (2009). *Selling Sounds: The Commercial Revolution in American Music*. Cambridge, MA: Harvard University Press.

- Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.
- Sutton, R. A. (1991). *Traditions of Gamelan Music in Java: Musical Pluralism and Regional Identity*. Cambridge, MA: Cambridge University Press.
- Svensson, E. (2015). "Upland living. The Scandinavian shielings and their European Sisters," in *Nordic Middle Ages-Artifacts, Landscapes and Society. Essays in Honour of Ingvild Øye on her 70th Birthday*, eds I. Baug, J. Larsen, and S. S. Mygland (Berg: University of Bergen), 289–300.
- Swain, J. P. (2002). *Harmonic Rhythm: Analysis and Interpretation*. Oxford, UK: Oxford University Press.
- Swanwick, K., Tillman, J., and Maccoby, E. E. (1986). The sequence of musical development: a study of children's composition. *Br. J. Music Educ.* 3, 305–339. doi: 10.1017/S0265051700000814
- Tagg, P. (2012). *Music's Meaning: A Modern Musicology for Non-Musos*. Larchmont, NY: Mass Media's Scholar's Press.
- Tallerman, M. (2013). Join the dots: a musical interlude in the evolution of language? *J. Linguist.* 49, 455–487. doi: 10.1017/S002226713000017
- Tarasti, E. (1994). *A Theory of Musical Semiotics*. Indianapolis, IN: Indiana University Press.
- Tarasti, E. (2012). *Semiotics of Classical Music: How Mozart, Brahms and Wagner Talk to Us*. Berlin: Mouton de Gruyter.
- Tarasti, E. (ed.) (1995). *Musical Signification: Essays in the Semiotic Theory and Analysis of Music*. New York, NY: Mouton de Gruyter.
- Tarr, B., Launay, J., and Dunbar, R. I. M. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096
- Tchotchkina, M. P. (2003). *Altai Children Folklore [Алтайский детский фольклор]*. Gorno-Altai: Gorno-Altai state university press.
- Tesch, S. (1992). The long-term development of a settlement region on the coastal plain - the Köpings area. *Acta Archaeol. Lundensia* 4, 283–344.
- Thiemes, M. (2001). Dynamics. *New Grove Diction. Music Music*. doi: 10.1093/gmo/9781561592630.article.08458
- Thompson, N. S., Abbey, E., Wapner, J., Logan, C., Merritt, P. G., and Pooth, A. (2000). Variation in the bout structure of Northern Mockingbird (*Mimus polyglottos*) singing. *Bird Behav.* 13, 93–98.
- Tilley, C. (1996). *An ethnography of the Neolithic: Early Prehistoric Societies in Southern Scandinavia*. Cambridge, MA: Cambridge University Press.
- Titon, J. T. (2015). Ethnomusicology as the study of people making music. *Muzikoloski Zbornik* 51, 175–185. doi: 10.4312/mz.51.2.175-185
- Titze, I. R. (1988). A framework for the study of vocal registers. *J. Voice* 2, 183–194. doi: 10.1016/S0892-1997(88)80075-4
- Titze, I. R., Blake, D., and Wodzak, J. (2018). Intelligibility of long-distance emergency calling. *J. Voice* 32, 1–9. doi: 10.1016/J.JVOICE.2018.08.008
- Tiukhteneva, S. P. (2017). On pastoral magic amongst contemporary Altaians [О скотоводческой магии у современных алтайцев]. *Oriental Stud.* 34, 62–70.
- Todd, N. P. M. (1992). The dynamics of dynamics: a model of musical expression. *J. Acoust. Soc. Am.* 91, 3540–3550. doi: 10.1121/1.402843
- Touma, H. H. (1996). *The Music of the Arabs, Portland*. Oregon: Amadeus Press.
- Trainor, L. J. (2010). The emotional origins of music. *Phys. Life Rev.* 7, 44–45. doi: 10.1016/j.plrev.2010.01.010
- Trainor, L. J., and Hannon, E. E. (2013). "Musical development," in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 423–498.
- Trehub, S. E. (2008). Music as a dishonest signal. *Behav. Brain Sci.* 31, 598–599. doi: 10.1017/S0140525X08005530
- Trehub, S. E., and Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition* 100, 73–99. doi: 10.1016/j.cognition.2005.11.006
- Trehub, S. E., Unyk, A. M., and Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behav. Dev.* 16, 285–295. doi: 10.1016/0163-6383(93)80036-8
- Tresset, A., and Vigne, J.-D. (2011). Last hunter-gatherers and first farmers of Europe. *C. R. Biol.* 334, 182–189. doi: 10.1016/J.CRV.2010.12.010 doi: 10.1016/j.crv.2010.12.010
- Trevarthen, C. (2019). "The psychobiology of the human spirit," in *Early Interaction and Developmental Psychopathology. Infancy*, eds G. Apter, E. Devouche, and M. Gratier (Cham: Springer International Publishing), 1–18. doi: 10.1007/978-3-030-04769-6_1
- Trost, W. J., Labbé, C., and Grandjean, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia* 96, 96–110. doi: 10.1016/J.NEUROPSYCHOLOGIA.2017.01.004
- Tull, H. W. (1996). The killing that is not killing: men, cattle, and the origins of non-violence (ahimsā) in the Vedic sacrifice. *Indo Iranian J.* 39, 223–244. doi: 10.1163/00000096790085150
- Tuniz, C., Bernardini, F., Turk, I., Dimkaroski, L., Mancini, L., and Dreossi, D. (2012). Did Neanderthals play music? X-ray computed micro-tomography of the divje babe 'Flute'. *Archaeometry* 54, 581–590. doi: 10.1111/j.1475-4754.2011.00630.x
- Turino, T. (2014). Peircean thought as core theory for a phenomenological ethnomusicology. *Ethnomusicology* 58, 185–221. doi: 10.5406/ethnomusicology.58.2.0185
- Turk, I. (2014). *DIVJE BABE I. Upper Pleistocene Palaeolithic site in Slovenia. Opera Inst. Ljubljana: Collegium Graphicum*.
- Turk, I., Dirjec, J., Bastiani, G., Pflaum, M., Lauko, T., Cimerman, F., et al. (2001). New analyses of the "flute" from Divje babe I (Slovenia) [Nove analize "piscali" iz Divjih bab I (Slovenija)]. *Arheoloski Vestnik* 52, 25–79.
- Uspensky, B. (1995). *Semiotics of Art [Семiotics искусства]*. Moscow: Languages of Russian Culture.
- Uttman, M. T. (2002). Eine untersuchung der teiltonspektren bei kulning und lockruftechniken anhand von beispielen aus Schweden und Finnland. *Swedish J. Music Res.* 5, 1–18.
- Val'kova, V. B. (1992). *Musical Thematicism, Cognition, Culture [Музыкальный тематизм, мышление, культура]*. Nizhnii Novgorod: State Nizhegorodskii University.
- Vashkevich, N. (2006). *Semantics of Musical Speech. Musical Syntax [Семантика музыкальной речи. Музыкальный синтаксис]*. Tver: Musorgsky Musical College.
- Ventsel, A. (2006). Hunter-herder continuum in anabarski District, NW Sakha, Siberia, Russian Federation. *Nomadic Peoples* 10, 68–86. doi: 10.3167/np.2006.100205
- Vorren, Ø (1973). "Some trends of the transition from hunting to nomadic economy in Finnmark," in *Circumpolar Problems: Habitat, Economy and Social Relations in the Arctic*, ed. G. Berg (Oxford, UK: Pergamon Press), 185–194. doi: 10.1016/b978-0-08-017038-1.50026-1
- Walker, R. (1997). Visual metaphors as music notation for sung vowel spectra in different cultures. *J. New Music Res.* 26, 315–345. doi: 10.1080/09298219708570733
- Wallin, N. L. (1983). Pitch perception as expression for exogene and endogene coordinated oscillations. *World Music* 25, 46–64.
- Wallin, N. L. (1991). *Biomusicology: Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music*. Hillsdale, NY: Pendragon Press.
- Waser, P. M., and Waser, M. S. (1977). Experimental studies of primate vocalization: specializations for long-distance propagation. *Zeitschrift Tierpsychol.* 43, 239–263. doi: 10.1111/j.1439-0310.1977.tb00073.x doi: 10.1111/j.1439-0310.1977.tb00073.x
- Welch, G. F. (2006). "Singing and vocal development," in *The Child as Musician: A Handbook of Musical Development*, ed. G. McPherson (New York, NY: Oxford University Press), 311–330. doi: 10.1093/acprof:oso/9780198530329.003.0016 doi: 10.1093/acprof:oso/9780198530329.003.0016
- Wermke, K., Leising, D., and Stellzig-Eisenhauer, A. (2007). Relation of melody complexity in infants' cries to language outcome in the second year of life: a longitudinal study. *Clin. Linguist. Phonet.* 21, 961–973. doi: 10.1080/02699200701659243
- Wermke, K., and Mende, W. (2009). Musical elements in human infants' cries: in the beginning is the melody. *Music. Sci.* 13, 151–175. doi: 10.1177/1029864909013002081
- Whiten, A. (2011). The scope of culture in chimpanzees, humans and ancestral apes. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 997–1007. doi: 10.1098/rstb.2010.0334
- Whittaker, J. C. (1994). *Flintknapping: Making and Understanding Stone Tools*. Austin, Texas: University of Texas Press.
- Whittle, M. (2007). *Gait Analysis: An Introduction*. Oxford UK: Butterworth-Heinemann.

- Wildgen, W. (2004). *The Evolution of Human Language: Scenarios, Principles, and Cultural Dynamics*. Philadelphia: John Benjamins Publishing.
- Wiley, R. H. (1983). "The evolution of communication: information and manipulation," in *Animal Behavior: Communication*, eds T. R. Halliday and P. J. B. Slater (London: Blackwell Publishers), 156–189.
- Winnington-Ingram, R. P. (2015). *Mode in Ancient Greek Music*. Cambridge, MA: Cambridge University Press.
- Wulstan, D. (1971). The origin of the modes. *Stud. Eastern Chant* 2, 4–20.
- Wyatt, S. (2016). "Musicality and cultural adaptation," in *Studien zur Musikarchäologie. Klang – Objekt – Kultur – Geschichte*, eds R. Eichmann, L.-C. Koch, and J. Fang (Berlin: Verlag Marie Leidorf GmbH), 169–194.
- Yemelyanov, V. (2000). *Voice Development: Coordination and Training [Развитие голоса: Координация и тренинг]*. Saint Petersburg: Lan.
- Yip, M. J. (2006). The search for phonology in other species. *Trends Cogn. Sci.* 10, 442–446. doi: 10.1016/J.TICS.2006.08.001
- Yoon, S. (2018). What's in the Song? Urtyu duu as sonic "Ritual" among mongolian herder-singers. *MUSI Cultures* 45, 92–111.
- Yoshida, S., and Okanoya, K. (2005). Evolution of turn-taking: a bio-cognitive perspective. *Cogn. Stud.* 12, 153–165. doi: 10.11225/jcss.12.153
- Yust, J. (2018). *Organized Time: Rhythm, Tonality, and Form*. Oxford, New York: Oxford University Press.
- Zabolotskaya, P. Y. (2009). *Folkloric Theater of Yakuts: An Attempt of Historic-Hystrionic Research [Фольклорный театрякутов (опыт историко-театроведческоо исследования)]*. Ulan-Ude: East Siberian state academy of culture and arts.
- Zeder, M. A. (2008). Domestication and early agriculture in the mediterranean basin: origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11597–11604. doi: 10.1073/pnas.0801317105
- Zemtsovsky, I. (1979). "On creative nature of folklore [О творческой природе фольклора]," in *Stylistic Trends in Soviet Music 1960-70s [Стилевые тенденции в Советской музыке 1970-70 годов]*, ed. A. N. Kriukov (Leningrad: Lenuprizdat), 137–147.
- Zemtsovsky, I. (1983). "Song as a historic phenomenon [Песня как исторический феномен]," in *Popular Song: Problems of Study [Народная песня. Проблемы изучения]*, ed. V. Gusev (Leningrad: State Academy of Theatric Arts), 22–35.
- Zemtsovsky, I. (1997). An attempt at a synthetic paradigm. *Ethnomusicology* 41, 185–205. doi: 10.2307/852602
- Zemtsovsky, I. (2002). The apology of text [Апология текста]. *Music. Acad.* 4, 100–110.
- Zemtsovsky, I. (2005). Neither East nor West; in between but not a bridge: A riddle for a new discipline, the ethnogeomusicology. *Muzikologija: Časopis Muzikološkog Instituta Srpske Akademije Nauka i Umetnosti* 5, 195–203. doi: 10.2298/MUZ0505195Z
- Zemtsovsky, I., and Kunanbayeva, A. (2011). "Homo Lyricus, or lyrical song in ethnomusicological stratigraphy of 'folk culture' [Homo Lyricus, или лирическая песня в этномузыкальной стратиграфии «фольклорной культуры»]," in *Classic folklore today. Proceedings of the conference "90 years anniversary of B.N.Putilov's birthday*, (Sankt-Petersburg: Dmitri Bulanin), 199–249.
- Zeskind, P. S. (1985). "A developmental perspective of infant crying," in *Infant Crying: Theoretical and research perspectives*, eds C. F. Z. Boukydis and B. M. Lester (Boston, MA: Springer), 159–185. doi: 10.1007/978-1-4613-2381-5_8
- Zilhão, J. (2014). "The upper palaeolithic of Europe," in *The Cambridge World Prehistory*, eds C. Renfrew and P. G. Bahn (Cambridge, MA: Cambridge University Press), 1753–1785.
- Zimmermann, E., Leliveld, L., and Schehka, S. (2013). "Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices," in *Evolution of Emotional Communication*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford, UK: Oxford University Press), 116–132. doi: 10.1093/acprof:oso/9780199583560.003.0008
- Zuberbühler, K. (2017). "The primate roots of human language," in *Primate Hearing and Communication*, eds R. M. Quam, M. Rosa, and J. L. Arsuaga (Berlin: Springer), 175–200. doi: 10.1007/978-3-319-59478-1_7
- Zubrow, E. B. W., and Blake, E. C. (2006). "The origin of music and rhythm," in *Archaeoacoustics*, eds C. Scarre and G. Lawson (Cambridge, MA: McDonald Institute for Archaeological Research), 142.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nikolsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.