# A constraint-based approach to structuring language and music

Gilbers, Dicky; Rebernik, Teja

*Published in:*
How Language Speaks to Music

*DOI:*
10.1515/9783110770186-004

Dicky Gilbers & Teja Rebernik

# A constraint-based approach to structuring language and music: Towards a roadmap for comparing language and music cross-culturally

**Abstract:** We pursue the hypothesis that musical differences between cultures are based on linguistic, especially phonological, properties of the culture's spoken language. To study this hypothesis, we present a general constraint-based framework for describing the structural similarities between music and language. Music and language are structured by the fact that some sounds are more important than others, based on cognitive strategies which we present here as universal well-formedness conditions. However, which sounds are considered to be most salient differs across cultures, as evidenced by the world's many linguistic and musical typologies. The first goal of our research approach is to identify these universal well-formedness conditions (e.g. prominence of strong elements based on the syllable/chord structure and domain marking based on intonation/melody patterns, pauses) for speech and music. The second goal is to assess how cultures differ from each other in terms of the relative salience assigned to these conditions (i.e. how these conditions are "ranked"). The current paper is meant to be an introduction to a new approach with focus on the identification of general well-formedness conditions. We introduce similar conditions for the description of language and music in order to make comparison of the two disciplines more fruitful. The goal of our research approach is to create a theoretical and methodological map to aid more detailed culture-specific comparisons. The ultimate aim is to provide a comprehensive typological overview for which we will start with a selection of culture families following the World Atlas of Language Structures online (Dryer & Haspelmath, 2013) for language and the Global Jukebox (Wood & Arèvalo, 2018) for music.

**Acknowledgements:** We would like to thank Matt Coler, Steven Gilbers and the participants of the Workshop "Prosody from a cross-domain perspective: How language speaks to music (and vice versa)", 41st Annual Conference of the German Linguistic Society, Bremen, March 2019, for comments and input. We would also like to thank two anonymous reviewers who significantly contributed to improvements on the earlier version of the manuscript.

**Dicky Gilbers, Teja Rebernik,** University of Groningen, The Netherlands

https://doi.org/10.1515/9783110770186-004

**Keywords:** linguistic and musical typology, optimality theory, universality and variation

## 1 Introduction

Similarities between music and language can be found at various levels and defined in different ways. In his book *Music, Language, and the Brain* (Patel, 2010), the author proposes that the comparison can be done on six levels: *sound elements* (pitch in music and timbre in language) serve as the organizing force; *rhythm*, as the systematic patterning of sound, shows that both domains group smaller sounds into higher-level units;[1] spoken and musical *melody* can be directly compared in terms of pitch patterning or contour; *syntax* binds both language and music in terms of a hierarchical, logical structure (i.e. words form sentences, tones form chords); *meaning* can be conveyed by both language and music, although musical meaning is a lot more difficult to define; finally, the domains of language and music can be compared from an evolutionary perspective (i.e. to what extent humans evolved their musical and linguistic abilities by natural selection). In this paper, we focus especially on the more structural levels of rhythm, melody, and syntax.

Others, such as Jackendoff (2009), propose a different way of looking at the language-music connection: for processing language and music, individuals must have the memory capacity for storing representations, integrate these representations in different combinations, create expectations, possess fine-scale control of vocal production, express desire to imitate others, invent new items, and join in with others to produce something together. The general idea, linking different views, is that language and music are uniquely human.[2] Both language and music are highly systematic particular sound systems that are innately perceived and occur in all cultures. This observation raises a compelling question: what are the common (universal) cognitive strategies that are used to process the stream of sounds in order to structure language and music?

The idea that there are structural similarities between language and music did not start with Patel's aforementioned book nor did it end there. Fenk-Oczlon and Fenk (2009), for example, discuss parallels between the musical and linguistic

[1] However, temporal periodicity plays a significantly smaller role in speech than it does in musical meter, where it serves "as a mental framework for sound perception" (Patel, 2010).

[2] Comparatively, in holistic sound systems, like those used by many animals, each sound is associated with a particular meaning, but sounds are not recombined to form new meanings.

building blocks (namely, musical intervals and vowels) or the length and size of utterances (clauses in language and phrases in music, respectively). They find that both linguistic and musical utterances typically have a duration of about 2 seconds and 5–10 "pulses". Likewise, sound inventories consist of between 3 and 12 elements (most frequently 5) for both notes and vowels. Other researchers take a more experimental approach, for example by comparing speech rhythm and (classical) music rhythm of different European languages using the nPVI index[3] (see e.g., Patel & Daniele, 2003; Daniele & Patel, 2004; VanHandel & Song, 2009; Jekiel, 2015). Temperley & Temperley (2011) go into more detail by comparing the prevalence of a certain rhythmic type and vowel length in several languages, finding that the prevalence of the rhythmic pattern "Scotch snap" (short accented note followed by a longer one) in Scottish and English songs but not in German or Italian songs is potentially related to the fact that British English song lyrics have many more very short stressed syllables compared to German or Italian song lyrics. The study of musical and linguistic structure gets even more complicated when one considers regional variation of one country's or culture's languages and musics. Gilbers et al. (2020) show that regional variation in African American English prosody and rap flows make patterns in similar ways, suggesting a connection between rhythm and melody in language and music. Furthermore, studies on both English dialects and folk music (McGowan & Levitt, 2011) and Slovenian dialects and folk music (Rebernik & Gilbers, 2017) have shown that undeniable regional differences exist in the speech and folk music across the countries. For example, in Slovenia, the speech and folk music of border regions show unique characteristics (Rebernik & Gilbers, 2017).

These comparative approaches, no matter how informative, fail to consider the breadth of languages and musics, also seen in the fact that they mostly focus on a single level (e.g. rhythm) and on a limited set of cultures (predominantly Western). They also face the fact that despite intriguing similarities between language and music, such as metrical structures and mechanisms processing pitch, there are also important differences. For example, music does not possess a lexicon with a conceptual system that gives rise to compositional meaning as language does. With a musical instrument you cannot communicate a sentence such as "let us convince you this is a very interesting research approach". Therefore, musical syntactic structures cannot be perfectly aligned with linguistic ones in a one-to-one manner. Hierarchical structures in language have referential,

[3] nPVI or the "normalized pairwise variability index" is a measure of the "degree of contrast between successive durations in an utterance" (Patel, 2010). Due to its nature, it can be used either for measuring durations in speech utterances or measuring durations in musical segments.

propositional meaning, whereas the relationship between essential and ornamental elements in music defines tension-relaxation patterns that encode affect. When studying music and language from a structural perspective, we must start seeing the forest as opposed to just individual trees. In this case, we must shift our efforts towards creating a comprehensive framework that could explain the structure of both linguistic and musical sounds.

Indeed, Asano & Boeckx (2015) suggest that a fruitful comparison of music and language needs to incorporate action-related components such as goal of action, action planning, motor control and sensory-motor integration. Language has a conceptual goal, i.e. organizing thought, and a pragmatic goal, i.e. communication. Music, on the other hand, has an affective-gestural goal, i.e. inducing emotion, and a socio-intentional goal, i.e. performing an enjoyable activity in a group. What they have in common is that they are temporally structured sequences. Hierarchical structures are considered as linking action and syntax. In this view, music and language share a planning component as an interface, adapting stored representations to achieve various domain-specific goals. Music and language use the same computation for these hierarchical structures, defining head and dependent elements in different domains. Accordingly, differences between language and music can be explained in terms of different goals reflected in the hierarchical plans. For example, in Western tonal music, the cadence can be seen as a kind of structural goal in the dynamics of tension and relaxation. However, the intended affect depends on the conventionally acquired knowledge of the musical idiom of the listener. We argue that this makes a constraint-based approach particularly suitable to study it.

The aim of this chapter is to discuss structure in terms of well-formedness conditions or constraints that identify essential and ornamental elements in the processing of music and language. We lean on and combine two constraint-based approaches: *Optimality Theory* (Prince & Smolensky, 1993), which has predominantly been used to explain linguistic structure, and the *Generative Theory of Tonal Music* (Lerdahl & Jackendoff, 1983), which was created to explain (Western tonal) music. We wish to quantify linguistic and musical characteristics in terms of the cognitive strategies that help people structure these two phenomena. While this chapter presents the early stages of our approach, mostly explaining it in terms of constraints that can be used for classifying inventories, harmony, rhythm and melody/intonation, the ultimate goal of the present research is to generate an account of the way language and music are structured across cultures and the degree to which differences in musical styles and languages can be explained by differently ordering general well-formedness conditions on structure depending on the culture. In other words, culture-specific differences are reflected in the relative salience of the well-formedness conditions for each culture.

In the next section, we briefly introduce Optimality Theory and the Generative Theory of Tonal Music. We follow by discussing the typology of universal well-formedness conditions. Subsequently, we discuss constraints on chord complexity and rhythm by presenting Optimality Theory-inspired tableaus. Finally, we conclude by considering the implications of our approach and the problems of describing two phenomena that are so similar yet differ to such a great extent across cultures.

# 2 Similarities between language and music

## 2.1 Constraint-based frameworks

Optimality Theory (OT) aims to explain structure in language. OT is an output-oriented theory of language and grammar that became a popular trend in linguistics after its introduction by phonologist Alan Prince and cognitive scientist Paul Smolensky (Prince & Smolensky, 1993). In OT a grammar consists of a set of well-formedness constraints on possible outputs, i.e. realizations of phonological forms. These constraints apply simultaneously to representations of structures, and they are soft, which means violable.

OT introduces several types of constraints. First, so-called "markedness" constraints ensure simple structures, as exemplified in, for example, a constraint on clusters of consonants within a syllable. Second, markedness constraints interact with so-called "correspondence" constraints, which establish relations between underlying, i.e. mentally stored (input) forms and the actually realized (output) forms. This is in line with Boersma (1998), who quotes Passy (1891), asserting that speakers will try to get their message across as *quickly* and *clearly* as possible. In a functionally oriented OT account of morpho-phonological processes, therefore, markedness constraints, which ensure articulatory easiness (*quickly*) for the speaker, are potentially in conflict with correspondence constraints, which ensure diversity of forms and meaning (*clearly*), which makes communication easier for the listener. Finally, OT also contains so-called "alignment" constraints, which function as domain boundary markers. They require that the edges of different domains, for example of morphological units and phonological units, coincide.

In OT, different constraints may lay down opposite requirements on the preferred structure. If so, conflicts are solved by assuming differences in weight between the different constraints. An optimal output may violate a certain constraint as long as this violation leads to the satisfaction of a more important constraint.

This can be likened to traffic rules: the constraint to wait for a red traffic light has more weight than having precedence on the main road, although both constraints are defined strictly. Eventually, the whole set of hierarchically ranked constraints determines the optimal realization of phonological forms.

Possible variations attested in the world's languages can be accounted for with reference to the different hierarchical ranking of the universal set of these constraints. In other words, not all universal constraints are equally important in each language. Indeed, individual languages rank the universal constraints in such a way that higher ranked constraints have total dominance over lower ranked constraints. By analysing the results arising from ranking the universal constraints in all possible dominance hierarchies, one can predict and explain which surface patterns are possible in natural languages (Gilbers & de Hoop, 1998).[4]

OT owes the idea of ranking soft constraints to the Generative Theory of Tonal Music (GTTM), introduced in 1983 by musicologist Fred Lerdahl and linguist Ray Jackendoff, who sought to explain structure in (Western) music. Lerdahl and Jackendoff describe how a listener constructs connections between different parts of a musical piece. In their music theory, the musical stream of sounds is hierarchically divided into domains. Each domain (e.g. a verse) contains some smaller domains (e.g. a phrase), which in turn contain smaller domains (e.g. a motif). In each domain, head and dependent parts are defined by the application of preference rules, comparable to constraints in OT. As in OT, the preference rules are not strict claims on the interpretation of a musical piece. It is possible for a head constituent to violate a certain preference rule as long as this violation leads to the satisfaction of a more important preference rule. By imposing this hierarchical structure on the entire piece and by distinguishing between important and ornamental parts by means of applying preference rules that are ranked for importance, the listener is able to understand the piece of music (Gilbers & Schreuder, 2000; 2002; Schreuder, 2006).

Gilbers & Schreuder (2002) mention that within existing theories of music and language structure, there is only one mentioned ranking of preference rules for music (in GTTM), whereas there are several rankings for language, as the ranking of universal constraints (which in themselves are unranked) has to be established for every individual language (in OT). Although Lerdahl & Jackendoff (1983) only offer one ranking, namely for tonal Western music, one can imagine that the dominance hierarchy of preference rules is different for, for example, Eastern music. The ultimate question in the current research approach is whether

[4] The appendix shows a summary table of all OT-constraints used in this chapter.

there is a relation between the relative importance of similar well-formedness conditions in the language and music of the same culture. For example, in the assignment of stress in most stress-timed languages, such as Dutch, syllable weight plays an important role, just like harmonic consonance (see description of the latter in section 2.2) in the culture's music.[5] In a tonal language, such as Sino-Tibetan Hmong, on the other hand, syllable weight is less important. The syllables in this language are less complex than in e.g. Dutch. Diversity in linguistic meaning is established by means of tonal differences in Hmong. Similarly, in the culture's music, prolongation of the melody line is more important than its harmonic consonance.

## 2.2 General well-formedness conditions

How can we use the two approaches introduced above in order to find musico-linguistic similarities? Both GTTM and OT are underlined by a simple fact: listeners construct connections between the sounds they perceive. Mostly subconsciously, the listener is capable of recognizing the construction of a piece of music by considering some notes/chords as more prominent than others. If listeners cannot recognize what is essential and what is ornamental, they will "lose contact" with the piece, and it will become a meaningless sequence of unrelated sounds to them. Similarly, with language, if listeners, for example those learning a second language, cannot recognize how the stream of sounds is structured, they will have problems with comprehension. Well-formedness conditions, defined as, respectively, preference rules or constraints, identify prominent elements in music and language, e.g. in terms of "prominence of strong elements" or "domain marking". These conditions in turn can help us explain the structure of both language and music.

According to Ball (2010), the brain is a pattern-seeking organ; it looks for patterns in sound to make sense of what we hear.[6] Consider two examples of well-formedness conditions in music and language. First, one that refers to the differences in weight between syllables in language and between chords in music.

[5] Harmonic consonant intervals in music are characterized by ratios of frequencies of lower integers: 2:1 (octave), 3:2 (fifth), 4:3 (fourth). The lower these integers are, the less tension there is in the music. Musical preference rules are based on harmonic stability, following Lerdahl & Jackendoff, 1983.

[6] Some modernist composers, such as Schönberg, intentionally undermine this cognitive aid for making music easier to understand, which makes it harder for the brain to find structure.

In language, syllables may differ in weight, which may be an important cue in order to find out which syllable is the stressed one, i.e. the prominent one, in a word. Prominent parts may be characterized acoustically by a higher pitch, a longer duration and/or more intensity, which makes perception of the structure of the message and thereby communication easier for the listener.

For example, in Hindi, the strongest, i.e. most complex, syllable available in the word is stressed (Hayes, 1995). In the word *kidhar*, the closed syllable *dhar* is heavy and the open syllable *ki* is light. Accordingly, *dhar* will be stressed. Likewise, in the word *reezgarii*, the syllable *reez* (tense vowel and closed) is super heavy, *ga* (lax vowel and open) is light and *rii* (long vowel and open) is heavy, hence *reez* will be stressed.

Similar to the way the smallest linguistic building blocks, phonemes, can be combined into a next higher domain, syllables, the smallest musical building blocks, tones, can be combined with each other into a next higher domain, chords. Similar to syllables in language, chords in music can also be ranked according to differences in weight. In GTTM, the preference rule specifies that the head of a domain is the chord (or the note) which is relatively harmonically consonant. This preference rule is connected to a hierarchy of chords based on harmonic stability. A triad tonic-tierce-fifth (c-e-g) is more stable than a diminished chord C0 (c-$e_b$-$g_b$). The latter chord is to be used ornamentally as a transition from one prominent chord to another. The preference rule indicates that a chord C is preferred to C0 as the head of a domain. Just compare the notes c and g in a tonic-fifth combination of a triad tonic-(tierce)-fifth to the combination of c and g flat in a diminished chord C0 (c-($e_b$)-$g_b$). The upper picture in Figure 1 (below) shows that two cycles of the sound wave c have the same duration as three cycles of the sound wave g (ratio 3:2). At the point indicated by the arrow the pattern repeats itself. These waves harmonize in such a way that the combination is easy to process for us. On the other hand, the lower picture in Figure 1 (below) shows that the waves of c and $g_b$ (ratio 64:45) do not easily coincide in a periodic pattern: tension remains and needs to be solved for listeners in a combination of waves that harmonize better.

Therefore, a diminished chord is perceived as a chord building up tension, as a transition chord. This is not a matter of taste: c and g are strongly related. The fact that c and g harmonize better than c and $g_b$ follows from physical principles. The combination of frequencies as in Figure 1 (upper) will be preferred to the combination of frequencies as in Figure 1 (lower). The universal cognitive strategy in structuring language and music we identify is prominence of strong elements. Humans focus on strong elements in order to detect structure in a sequence of different sound events.

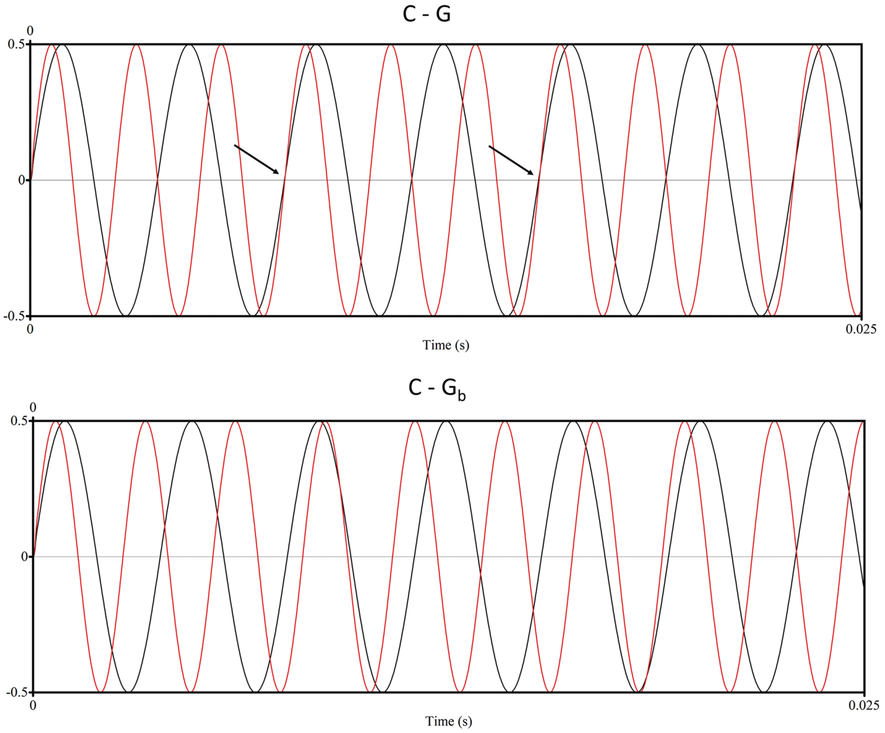

**Figure 1:** Sound waves of two note combinations.

As a second example of well-formedness conditions, boundary marking effects can be observed in both language and music. Prosodic cues such as intonation contours, pitch accents, stress shifts, pauses and rhythm patterns help listeners to detect the structure, to understand where a domain begins and ends. As we have seen in the previous section, OT makes use of alignment constraints to account for this effect of boundary marking. In music, not all chords are suitable for domain boundary marking. Just as there is a hierarchy in strength of chords, not all chord combinations are equal; a logical sequence of chords is predictable. Usually, the optimal chord is the final chord, a chord which generally is built on the tonic, preceded by a dominant chord. In the key of C, the dominant chord is G. This chord is suitable for a cadence: G7 creates a kind of tension in music that has to be solved by a subsequent tonic chord (see Figure 2). The cadence chord often concludes a phrase or section.

This preference of a harmonically more consonant chord in the chord sequence marking the boundary of a domain is defined as a GTTM-preference rule which chooses the chord which emphasizes the end of a group as a cadence as the

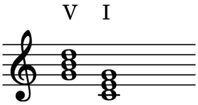

**Figure 2:** C major perfect authentic cadence (dominant to tonic).

head of the domain. We hypothesize domain marking to be an important universal cognitive strategy in structuring both language and music. This is where our search for universal cognitive well-formedness conditions begins.

## 3 Typology of universal well-formedness conditions

The typology of universal well-formedness conditions in language is based on the phonological parameters in Dryer and Haspelmath (2013). They include parameters such as “vowel and consonant inventory” (small, average, large), “syllable structure” (simple, moderately complex, complex), “stress system” (unbounded, fixed stress, weight-sensitive, weight-insensitive) “location stress” (left-edge, right-edge), “rhythm type” (trochaic, iambic), “tone system” (no tone, simple or complex tone system). These linguistic parameters, in turn, can be compared to music characteristics such as melodic shape, interval range, rhythm, and more.

The typology of musical universals is based on Alan Lomax’s ambitious Cantometrics project (Brown & Jordania, 2011). It includes parameters on pitch (e.g. discrete pitches vs. portamentos, musical scales, intervals), rhythm (e.g. predominance of isometric rhythms, metre types such as duple or triple, use of few durational values), melodic structure and texture (organization into phrases, melodic archetypes: descending, ascending, undulating contours, small or large intervals, harmonizing), form (beginning, middle, end, internal repetition), vocal style (predominance of syllabic singing, use of embellishment: melisma, vibrato, glides), expressive devices (tempo, amplitude, mode/emotion association).[7] Although the

[7] Musical universals can be split into five different categories: tautological universals that are true for every culture’s music (i.e. music as a system with its physical and sensory properties); conserved universals that are governed by biology and are true for every musical utterance (e.g. every utterance uses discrete pitches); predominant patterns that are true for every musical system but can have outliers in individual utterances (e.g. every scale has seven or fewer pitches per octave); common patterns that appear in many cultures but not all (e.g. singing is syllabic); and, finally, range universals that refer to the diversity in different categories (e.g. a culture with a free rhythm vs. a metric one, monophony or polyphony) (Brown & Jordania, 2011).

Cantometrics project has been criticised heavily with respect to, e.g., song sample, classification scheme and statistical analysis (see Savage, 2018, for a critical review), we agree with Savage that Lomax's map and song coding provides a useful starting point, keeping in mind that no method of cross-cultural comparison will be completely perfect.

Below we show exemplary analyses of how language and music characteristics can be described in a constraint-based OT manner. We can compare intonation patterns in a language, with long or short (descending or undulating) melodies in the culture's music, for example. Using similar constraints/well-formedness conditions makes comparison between music and language easier. The first example concerns variation in tone and segment inventories between cultures. Cultures may differ in the number of categories they display in the language as phonemes or in their music as note differences. In our approach comparing bigger units is done by using similar well-formedness conditions for language and music. It enables us to compare large or small inventories of phonemes in a language with the number of tone categories in the culture's music (3.1). The second example concerns cross-cultural variation in the way tones and segments can be combined to bigger units, such as, chords and syllables, respectively (3.2). The third example concerns the prosodic difference between descending and undulating melodies in music and in intonation patterns in language (3.3) and the fourth example concerns rhythmic variation in language and music (3.4).

## 3.1 Segment inventories in language and music

> *Music and language are 'particulate' sound systems, in which a set of discrete elements of little inherent meaning (such as tones or phonemes) are combined to form structures with a great diversity of meanings.* (Patel, 2010)

In this section we will introduce similar conditions for the description of language and music in order to make comparison of the two disciplines more fruitful. In language, the biggest contrast in 'particles', speech sounds, is between those produced with "mouth closed", e.g. plosives like /p/, versus those produced with "mouth open", e.g. vowels like /a/. No language lacks the contrast of voiceless plosives and vowels (see WALS, Dryer & Haspelmath, 2013; Jakobson, 1972). While this contrast is necessary, it is not sufficient. Segment inventories of languages need to be more complex in order to be an adequate vehicle of communication.[8]

[8] There are indeed various ways to couple enough differences in meaning to sound events. Some languages exhibit complexity in the structure of syllables (e.g. English), whereas others

This complexity can be achieved in different ways, e.g. through recourse to classes of segments between the extreme plosives and vowels in the segment inventory. Dryer & Haspelmath (2013) describe segment inventories in languages ranging from small to large. For example, the number of steps in sonority between voiceless plosives and full vowels varies in the world's languages.[9] Indeed, not every language exhibits a meaningful contrast between the so-called liquid speech sounds /l/ and /r/. In Japanese, [l] and [r] are variants of the same segment category, whereas they are contrastive in Indo-European languages. We can observe that Japanese monolinguals have difficulty perceiving and producing a distinction between those two sounds. For them, 'lake' and 'rake' may seem identical. Put differently, while the contrast between plosives and vowels is universal, the meaningful distinction between different categories of speech sounds is, broadly speaking, language-specific.

Table 1 depicts the acoustic differences of liquids /l,r/ and glides /j,w/ schematically. These differences can be related to their relative second and third formant locus frequencies. Ainsworth & Paliwal (1984) found that in a perceptual-identification experiment sounds having a mid F2 locus frequency were classified as /r/ if they had a relatively low F3 locus frequency and as /l/ if they had a relatively high F3 locus frequency. The sounds were identified as /w/ if they had a low F2 locus frequency and as /j/ if they had a high F2 locus frequency.

In our constraint-based framework, "Parse as category (PARSECAT)" is a correspondence constraint that classifies acoustically available features into a category and is in conflict with markedness constraints such as MaxContrast, which establishes dispersion, and "No category (*CAT)". Since children can learn to perceive any category, they start with constraints against acquired categories (Boersma, 1998).[10] The number of perceptual dimensions increases with the number of categories. With respect to /l/ and /r/, the *CAT constraints are ranked gradually for

keep syllables simple and exhibit complexity in e.g. the tone system (e.g. Mandarin Chinese) or in morphological operations such as reduplication (e.g. Hawaiian).

[9] Sonority is a challenged concept because there are languages that exhibit counter-examples. Nevertheless, satisfaction of sonority slopes in syllable structures is attested in most languages. This is where the merits of a constraint-based approach are evident. The OT-constraints are soft, which means violable. Satisfying the constraints describes unmarked structures, violating the constraints results in marked structures and counter-examples.

[10] However, children are quick to lose this flexibility, as perceptual categories for a particular native language are formed before one year of age and it might be that universal perception of speech sound occurs even before birth (see review article by Chládková & Paillereau, 2020).

**Table 1:** Typical set of responses obtained from listening to glide/liquid-vowel synthetic stimuli (adapted from Ainsworth & Paliwal, 1984).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3160 Hz | w | w | w | l | l | l | l | j | j | j |
| ↑ | w | w | w | l | l | l | l | j | j | j |
| F3 locus freq. | w | w | w | r | r | r | r | j | j | j |
| ↓ | w | w | w | r | r | r | r | j | j | j |
| 1540 Hz | w | w | w | r | r | r | r | j | j | j |
| | 760 | Hz | | ← | F2 | locus | freq. | → | 2380 | Hz |

the locus frequency of the third formant given the value of F2 as shown in Table 1: *CAT (F3 1500 Hz – 1700 Hz) >> . . . >> *CAT (F3 1500 Hz – 2200 Hz) >> . . . >> *CAT (F3 1500 Hz – 3200 Hz). If PARSECAT is dominated by *CAT (F3 1500 Hz – 1700 Hz) and *CAT (F3 1500 Hz – 2200 Hz), [l] and [r] will be allophones as in Japanese. If PARSECAT intervenes between *CAT (F3 1500 Hz – 2200 Hz) and *CAT (F3 1500 Hz – 3200 Hz), /l/ and /r/ are contrastive in the language system as in English. In other words, the position of PARSECAT is determined by the number of categories, i.e. the phonemes that the language displays in this frequency range.

Table 2 shows an OT-table of /r/-categorization. Assume the input sound, the perceived sound segment, has all the acoustic characteristics liquids and glides share and a third formant of 2000Hz, shown in the highest-leftmost cell. The constraints are depicted horizontally in dominating order from left to right and the candidate outputs are depicted vertically. "*" indicates violation of a constraint and "*!" means the violation is fatal, i.e. there is a better candidate given the ranking of constraints.

**Table 2:** /r/ as phoneme, a contrastive category (English system).

| **F3 2000 Hz** | ***CAT (F3 1500 Hz – 1700 Hz)** | **PARSECAT** | ***CAT (F3 1500 Hz – 2200 Hz)** | ***CAT (F3 1500 Hz – 3200 Hz)** |
|---|---|---|---|---|
| /r/1 (F3 1500–1700Hz) and /r/2 (F3 1700–2200Hz) | *! | | | |
| ☞/r/ (phoneme) (F3 1500–2200Hz) | | | * | |
| [r] (allophone) (F3 1500–3200Hz) | | *! | | |

Table 3: [r] as allophone (in the same category with [l]) (Japanese system).

| F3 2000 Hz | *CAT (F3 1500 Hz – 1700 Hz) | *CAT (F3 1500 Hz – 2200 Hz) | PARSECAT | *CAT (F3 1500 Hz – 3200 Hz) |
|---|---|---|---|---|
| /r/1 and /r/2 | *! | | | |
| /r/ (phoneme) | | *! | | |
| ☞[r] (allophone) | | | * | |

In Table 2 the dominating constraint *CAT (F3 1500 Hz – 1700 Hz) is satisfied by both /r/ as a phoneme, which has a range of approximately 1500–2200 Hz, and [r] as an allophone, because it falls within the range of 1500–3200 Hz. The first candidate violates the dominating constraint since the input F3 = 2000 Hz falls outside the range of 1500–1700 Hz. The candidate [r] as an allophone violates PARSECAT because this candidate is not categorized and although the candidate violates *CAT (F3 1500–2200 Hz), it is the optimal candidate given the constraint ranking of English. Reranking the constraints, as in Table 3, shows the dominance hierarchy of a different language system, as in Japanese. Another reranking could establish a system with two r-like sounds as separate categories. The constraints are universal, but the ranking is culture-specific. The number of categories is of course limited by undominated physical constraints that indicate so-called "just noticeable differences" humans can perceive.[11]

Similar to phonemes as building blocks in language, the building blocks in music are notes and tones. A musical universal is the octave, a doubling of frequencies between tones (Brown & Jordania, 2011). Comparable to the way in which languages divide the scale between plosives and full vowels into categories of segments differently, the way in which the octave is divided into different steps is also culture-specific. In Western music the octave is divided into 12 equal-sized pitch intervals, whereas e.g. Javanese Gamelan music divides the octave into 7 pitch intervals (Perlman & Krumhansl, 1996; Patel, 2010). Just as it was difficult for a Japanese native speaker to discriminate between /l/ and /r/ in a Western language, leading to difficulties with semantic processing, it may be difficult for someone who is only familiar with Western music to process a Javanese song, for example. Previous research has shown that adults detect mistuned tones for familiar (major, unequal-step) scales more easily than for unfamiliar

[11] For more elaborated functionally-oriented OT-accounts of segment inventories, see Flemming, 1995 and Boersma, 1998.

scales (e.g., Lynch et al., 1990; Trehub et al., 1999). Furthermore, aesthetic *preferences* in music perception are not purely biologically conditioned: McDermott et al. (2016), for example, reported that consonance and dissonance were perceived equally aesthetically pleasing by members of a native Amazonian society compared to individuals living in the city who found consonance more pleasing (even though both groups could discriminate between the sounds themselves).

We assume the octave to be present in all musical cultures. In the key of A, for example, intervals with a ratio 1:2 (440Hz:880Hz) can be observed. MAX-CONTRAST is a constraint that evaluates the harmonic value of intervals with a 1:2 ratio as optimal. The second-best interval has a ratio 2:3 adding E (660Hz; the fifth in the key of A) to the possible intervals. Similar to the linguistic constraints, *CAT constraints are ranked gradually for the fundamental frequency, F0: *CAT (F0 440–441 Hz) >> . . . >> *CAT (F0 440Hz – 660 Hz) *CAT (F0 440–880 Hz). If a correspondence constraint PARSECAT intervenes between *CAT (F0 440Hz – 660 Hz) and *CAT (F0 440–880 Hz), only octaves and intervals of fifths exist in the music system.[12] The position of PARSECAT in the gradually ranked, acoustically defined *CAT constraint determines the number of categories in the segment inventory, similar to linguistic categories as depicted in Tables 2 and 3.

Once the number of segments in the octave is defined this way, different musical scales can be described by different positions of PARSECAT within a gradually ranked series of *CAT constraints for semitone steps. For example, given a 12 steps division of the octave, (1a) shows the chromatic scale (see notation in Figure 3), with PARSECAT dominating all *CAT constraints. (1b) shows the constraint ranking for a pentatonic scale (see notation in Figure 4) with steps of two semitones within the octave.

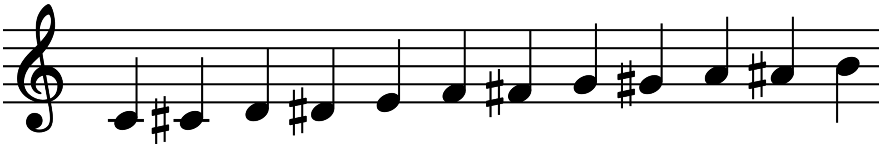

**Figure 3:** Ascending chromatic scale, starting on C.

[12] As in language systems, the number of categories in music is of course also limited by undominated physical constraints that indicate the “just noticeable differences” humans can perceive in order to identify frequency differences as belonging to different categories.

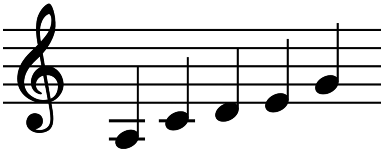

**Figure 4:** Minor pentatonic scale, starting on A.

The constraint ranking in (1) is comparable to the constraint ranking in the top row of Tables 2 and 3. The different positions of PARSECAT in (1a) and (1b) can be compared to the different positions of PARSECAT in Tables 2 and 3. The different positions describe different inventories of phonemes in language and different tone steps within an octave in music.

(1) a. Chromatic scale in OT:
PARSECAT >> *CAT 1 semitone >> * CAT 2 semitones, etc.
b. Pentatonic scale in OT (simplified):
*CAT 1 semitone >> PARSECAT >> *CAT 2 semitones, etc.

## 3.2 Syllables and chords

Linguistic segments, phonemes, are combined in units called syllables and musical particles, notes, can be combined in units called chords. In (2) some exemplary OT-markedness constraints for syllable structure are shown (Prince & Smolensky, 1993; Archangeli, 1997).

(2) Markedness constraints on syllable structure
ONSET: syllables begin with a consonant
*CODA: syllables end with a vowel
*COMPLEX: syllables have at most one consonant at an edge
PEAK: syllables have one vowel as nucleus

These markedness constraints ensure simple structures. If all constraints are satisfied, the result will be a syllable that consists of a consonant followed by a vowel (CV), the optimal syllable that is attested in all languages. These markedness constraints interact with correspondence constraints that warrant diversity in structure and thereby in meaning. Different rankings of the same set of constraints describe the possible variations attested in the world’s languages. For example, Hawaiian does not allow consonant clusters or codas, as exemplified in words such as *kanaka* “man” and *wahine* “woman”. In other words, *COMPLEX and *CODA are high-ranked constraints in that language. In English, *COMPLEX and *CODA are low-ranked as exemplified in words such as *sprint* with an initial complex CCC-cluster and a CC-coda. The differences between the systems can be seen in loan words from English in Hawaiian, such as *weleweka* “velvet” which is adapted to the Hawaiian system satisfying *COMPLEX and *CODA by inserting vowels.

Archangeli (1997) nicely shows that languages solve conflicts between these constraints after morphological operations differently. In Yawelmani, syllables cannot be more complex than CVC (consonant-vowel-consonant). Therefore, the morphologically complex form *logwen* (*logw* + *en*) “will pulverize” is not problematic: *log.wen* (with the dot indicating the syllable boundary), but *logw* + *hin* “pulverized” is problematic. In Yawelmani the attested word is *lo.giw.hin*, indicating that the correspondence constraint DEP-IO, “no insertion”, is violated in order to satisfy the dominant markedness constraint *COMPLEX, “no clusters of consonants within a syllable”. In Spanish, a similar conflict is solved differently. In Spanish *absorber* is unproblematic since it can be syllabified as *ab.sor.ber*. However, suffixation with *to* instead of *er* leads to a violation of the dominant syllable constraint *COMPLEX within a syllable: neither **ab.sorb.to* nor *ab.sor.bto* satisfies *COMPLEX. Spanish solves the conflict by means of violation of the correspondence constraint MAX-IO, “no deletion”. The attested form is *ab.sor.to* in which the root ends with the single consonant /r/. In English, the combination *limp* + *ness* results in *limp.ness*, indicating that the markedness constraint *COMPLEX, which is so dominant in Yawelmani and Spanish, is violable and thus less important in the ranking of universal constraints for this language. In other words, the universal constraints can be ranked differently in different languages constituting variability between language systems.

Complexity in musical chords can be described in a similar way. The notes of a musical system can be combined horizontally, as in a melody, or vertically as in chords or harmony. Just like linguistic differences in syllable structure, musical cultures may differ in the complexity of chords and chord sequences that is used in the music. The notes in triads harmonize better than in a $7^{th}$ chord, which is more harmonic than e.g. sus4 or diminished chords. This can be described as a parse note combination correspondence constraint, which interacts with an intervening set of ordered *Combi markedness constraints on parsing note combinations. The ordering of these constraints on note combinations is based on (Krumhansl, 1979), who describes relatedness between tones as obtained from a perception task using multidimensional scaling. The closer the perceived relatedness, the closer the tones are in the graph in Figure 5. The possible combinations of tones are also dependent of the tone inventory, of course. In other words, the inventory constraints mentioned in section 3.1 dominate the constraints introduced in this section.

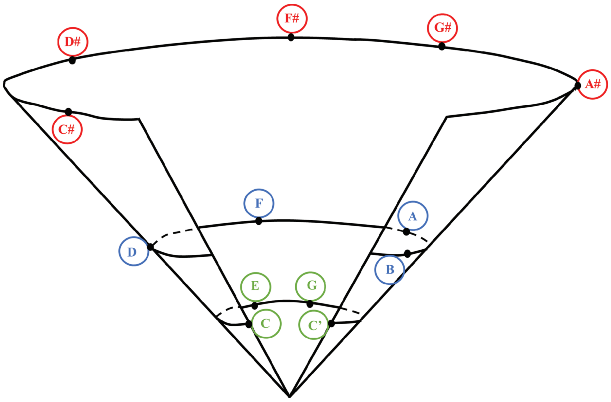

**Figure 5:** Tone distances in the key of C (adapted from Krumhansl, 1979).

The markedness constraint *Combi in (3) is defined as restrictions on combinations of frequencies in an octave of 12 equal-sized pitch intervals. *Combi is a gradually violable constraint in a similar way sonority constraints in language are. Therefore, *Combi is split up into a set of restrictions on combinations of fundamental frequency ratios. For example, *Combi 2:1 means no combination of notes that form an octave are allowed, e.g. 440 Hz ($A_4$) plus 220 Hz ($A_3$). The order of ratios in (3) is strict. *Combi 16:15 (diminished 2nd) is always higher ranked than *Combi 3:2 (perfect 5th), just as /r/ is always less sonorous than /a/ in language. The order in (3) reflects the observation by Pythagoras: ratios of lower simple numbers are more consonant than those that are higher.[13] The chords in question are depicted in Figure 6.

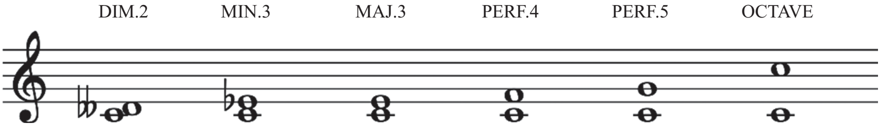

**Figure 6:** Chords mentioned in example (3), in C major.

[13] There is no general consensus on the distinction between consonance and dissonance in the history of music. Unisons, octaves, perfect fifths and perfect fourths are often regarded as perfect consonances; major and minor thirds as examples of imperfect consonances and diminished seconds and diminished fifths as examples of dissonance, but this categorisation varies in time. The order in (3) depicts a gradual change from dissonance to consonance, similar to the gradual change in sonority between segments in language.

(3) Combinations of notes
*Combi 16:15 (dim. 2nd) >> . . . >> *Combi 6:5 (min. 3rd) >> *Combi 5:4 (maj. 3rd) >> *Combi 4:3 (perfect 4th) >> *Comb 3:2 (perfect 5th) >> *Combi 2:1 (octave)

The position of the correspondence constraint PARSE COMBI MAX(imally) determines the complexity of harmonic structures that appear in the music. For example, if PARSE COMBI MAX intervenes between *Combi 2:1, e.g. C8 – C, and *Combi 3:2, e.g. G – C, only octaves are allowed in the music culture, characterized in Brown and Jordania (2011) as a lack of polyphony. The position of PARSE COMBI MAX is culture-specific and determines whether or not certain note combinations are used. The higher the ranking of an intervening PARSE COMBI MAX in the sequence in (3), the more complex harmony can be observed in the music culture. Low-ranked PARSE COMBI MAX describes a music culture of simple harmonic structures which might of course be accompanied by more complex melodies or rhythms, resembling a language with simple syllable structure which might be characterized by a more complex tone system.

In Table 4, PARSE COMBI MAX is ranked between *Combi 6:5 and *Combi 5:4, which means that the culture only allows for monophony, octave combinations, power chords, e.g. C-G combinations, perfect fourths and major triads. The output candidate monophony shows the most violations of PARSE COMBI MAX, given the possible output candidates presented here, whereas a diminished 2nd shows the least violations, because monophony rules out all possible combinations and a diminished 2nd, being the least harmonic combination with the tonic within an octave, allows for all note combinations in Table 4.

On the other hand, Table 5 depicts a more complex system in which PARSE COMBI MAX is promoted one step, which reflects a system with minor and major chords, but e.g. no diminished second. Notice that this account implies that a culture that allows for minor and major chords also allows power chords, perfect fourths and octave combinations, similar to language: if a language system allows for consonant clusters in syllables, it also allows for lesser complex CV-combinations.

The relation between chords as in chord progressions can be described in a similar way. The tonic I (e.g. C) is the most central chord, followed by the dominant V (G) and the subdominant IV (F). Krumhansl et al. (1982) investigated perceived relatedness between chords, as shown in Figure 7. Indeed, chords V and IV are most closely related to chord I, the tonic.

**Table 4:** Chord complexity (simplified) (tableau of a simple harmonic system).

| **Input:**<br>**All possible fundamental frequency combinations** | ***Combi 16:15** | **. . .** | ***Combi 6:5** | **PARSE COMBI MAX** | ***Combi 5:4** | ***Combi 4:3** | ***Combi 3:2** | ***Combi 2:1** |
|---|---|---|---|---|---|---|---|---|
| Monophony | | | | ***!*** | | | | |
| ≤Octave | | | | ***!** | | | | * |
| ≤Power Chord | | | | ***!* | | | * | * |
| ≤Perfect Fourth | | | | ***! | | * | * | * |
| ☞≤Triad/Major 3rd | | | | ** | * | * | * | * |
| ≤Minor 3rd | | | *! | * | * | * | * | * |
| . . . | | | | | | | | |
| ≤Diminished 2nd | *! | | * | | * | * | * | * |

**Table 5:** Chord complexity (simplified) (tableau of a slightly more complex harmonic system).

| **Input:**<br>**All possible fundamental frequency combinations** | ***Combi 16:15** | **. . .** | **PARSE COMBI MAX** | ***Combi 6:5** | ***Combi 5:4** | ***Combi 4:3** | ***Combi 3:2** | ***Combi 2:1** |
|---|---|---|---|---|---|---|---|---|
| Monophony | | | **!**** | | | | | |
| ≤Octave | | | **!*** | | | | | * |
| ≤Power Chord | | | **!** | | | | * | * |
| ≤Perfect Fourth | | | **! | | | * | * | * |
| ≤Triad/Major 3rd | | | **! | | * | * | * | * |
| ☞≤Minor 3rd | | | * | * | * | * | * | * |
| . . . | | | | | | | | |
| ≤Diminished 2nd | *! | | | * | * | * | * | * |

If a music culture only makes use of the three most related chords I, V and IV (tonic, dominant, subdominant), it can be described as in (4), in which the high ranked restrictions on combinations define the less related chord combinations and the low-ranked restrictions on combinations the closest relatedness between chords.

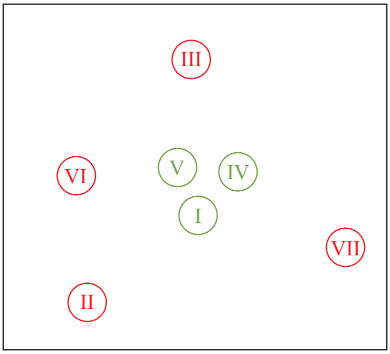

**Figure 7:** Psychological relatedness of different chords (adapted from Krumhansl et al., 1982).

(4) Combinations of chords (simplified). Ranking for a culture in which the music isn't more complex than tonic dominant subdominant harmonically, as in most blues and country music
*Combi I-II >> *Combi<I-III >> PARSE COMBI >> *Combi<I-IV >> *Combi< I-V >> *Combi

Again, low-ranked PARSE COMBI describes a music culture of simple chord progressions which might be accompanied by more complex melodies or rhythms. Our research approach is interested in these kinds of correlations. The more markedness (*COMBI) constraints are dominated by the correspondence PARSE COMBI constraint, the more complex chord progressions can be found in the music of a certain culture, just like the more linguistic markedness constraints, such as *COMPLEX, are dominated by linguistic correspondence constraints, such as MAX-IO, the more complex syllables can be attested in the language of the culture. Therefore, this constraint-based approach makes comparison of the music and language within a certain culture or between cultures possible in a straight-forward way.

## 3.3 Intonation/melody

In section 2, we mentioned "prominence of strong elements" and "domain marking" as important well-formedness conditions in structuring music and language. The latter becomes especially prominent in the constraint-based account of intonation and melody patterns. Prosodic phrasing is described in (Selkirk, 1995;

2000; Truckenbrodt, 1999; Eisenberg, 1991; Antilla & Bodomo, 1996; Lacy, 2002; Zhang, 2004; Ghamdi, 2006; and Gussenhoven, 2004), among others. In these approaches, edges of prosodic domains need to be aligned with morpho-syntactic constituents, new information needs to be associated with a prosodic constituent, and the length of the prosodic domains is variable. All these characteristics of prosodic constituents can be captured in OT by the interaction of prosodic structure constraints.

For our comparison of intonation patterns in language and melody in music, alignment constraints are in conflict with a so-called WRAP-constraint. Alignment constraints function as phrase boundary markers, they require that the edge of a phrase coincides with a tone. For example, satisfying ALIGN-L (High tone) and ALIGN-R (Low tone) constitutes a descending pattern for each intonation phrase (IP). WRAP requires that a minor phrase, e.g. a preposition phrase is contained in a single major phrase, e.g. an intonation phrase (IP), as illustrated in (5) below.

In (5), two possible intonation patterns of the sentence *Deep Purple stole the melody from Bombay's calling* are depicted. We recorded, respectively, a descending and an undulating intonation pattern of the same sentence. If WRAP dominates ALIGN, all phonological phrases, e.g. the verbal phrase (VP), the noun phrase (NP) and the preposition phrase (PP) in (5a) are incorporated in one IP, creating a descending pattern. If, on the other hand, ALIGN dominates WRAP, all phrases constitute their own IP (5b), creating an undulating intonation pattern. The different, simplified intonation patterns (only H- and L-marks for tone) are also depicted visually in Figures 8 and 9.

(5) Different intonation patterns based on the position of an Alignment constraint
[[Deep Purple]$_{\text{NP}}$ [[stole]$_{\text{V, in focus}}$ [the melody]$_{\text{NP}}$ [from Bombay's calling]$_{\text{PP}}$]$_{\text{VP}}$]$_{\text{S}}$
a. WRAP Phrase >> ALIGN-R:
(Deep Purple stole the melody from Bombay's calling)$_{\text{IP}}$
b. ALIGN-R >> WRAP Phrase
(Deep Purple)$_{\text{IP}}$ (stole)$_{\text{IP}}$ (the melody)$_{\text{IP}}$ (from Bombay's calling)$_{\text{IP}}$

As in language, in music alignment constraints parse boundary tones that mark the beginning and end of musical phrases and as in language various melodic patterns can be obtained by ranking the association constraints: WRAP >> ALIGN describes descending melodies and ALIGN >> WRAP describes undulating melodies. Of course, there is much variation in the shape, range and length of melodies in each culture, but the Global Jukebox (Wood & Arèvalo, 2018) – an online repository of music from all over the world based on Alan Lomax's field recordings, collected for his Cantometrics project (Lomax, 1976) – describes the dominating

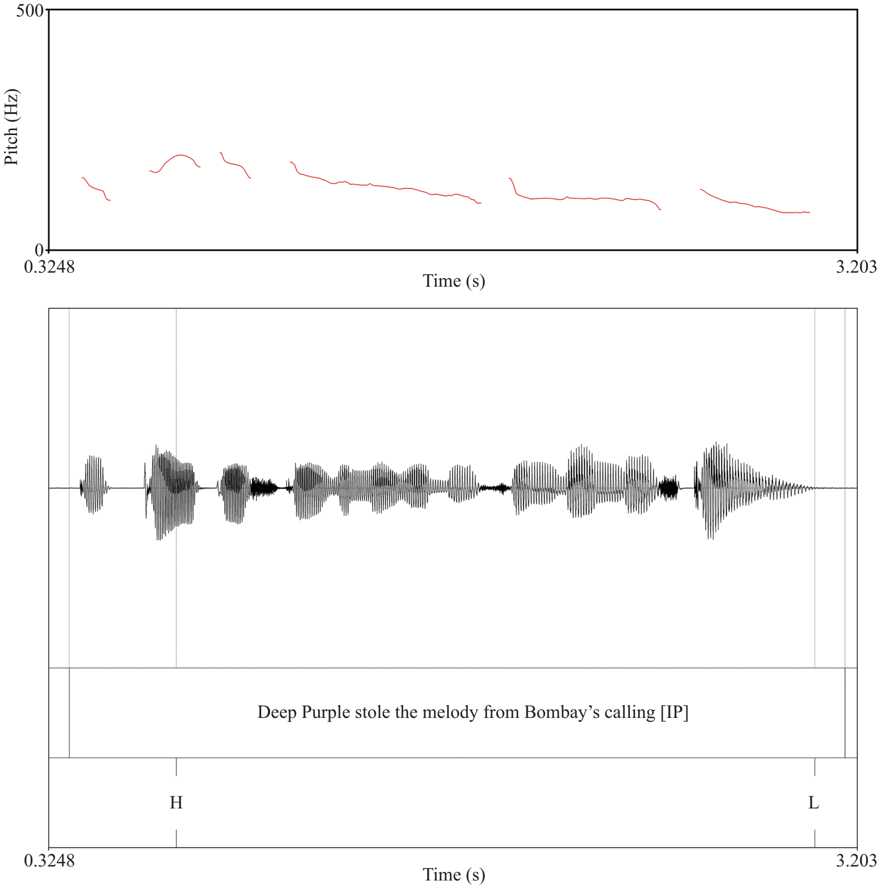

**Figure 8:** Descending intonation pattern (based on the ranking in 5a).

patterns in each culture as characteristic for its music, e.g. undulating short melodies are dominant in African Shilha music (cf. the ranking in 5b), whereas Southern American Kalina music is characterized by predominance of descending medium length melody lines (cf. the ranking in 5a).

## 3.4 Linguistic and musical rhythm

As a final example of our constraint-based approach, we consider variation in rhythm types. Over again, we introduce similar conditions for the description of language and music in order to make comparison of the two disciplines possible.

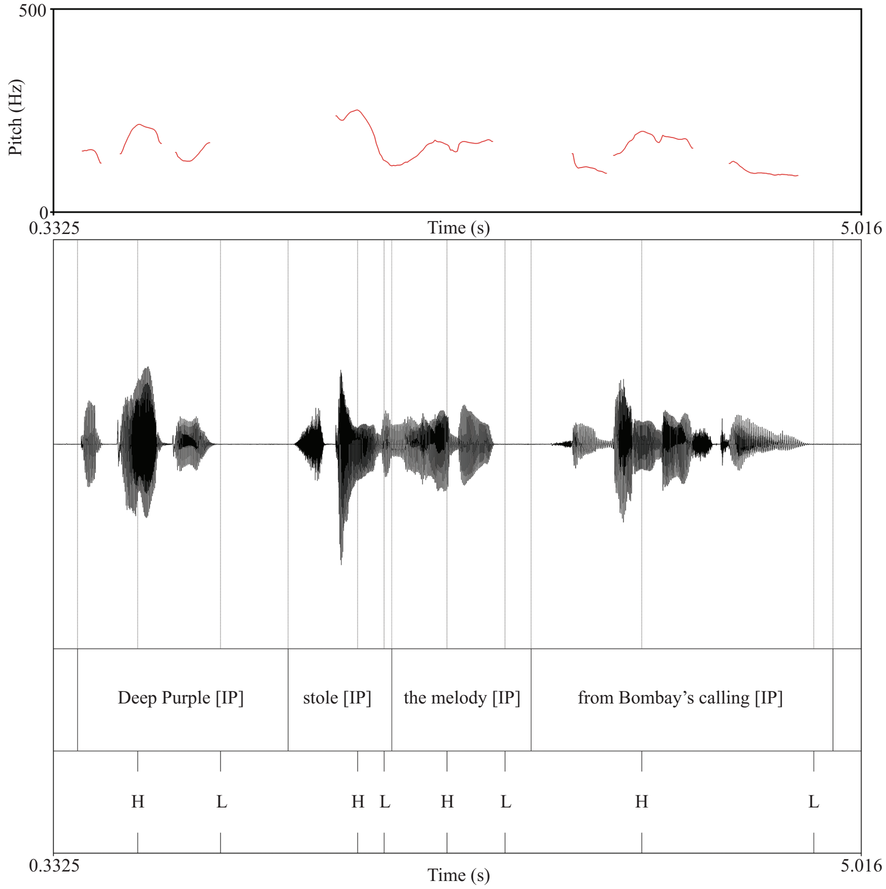

**Figure 9:** Undulating intonation pattern (based on the ranking in 5b).

This means that we define our music constraints similar to already established constraints in linguistics. In (6) some exemplary linguistic constraints on rhythm types are shown (Hayes, 1984; Prince & Smolensky, 1993; McCarthy & Prince, 1993; Nouveau, 1994; Gilbers & Jansen, 1996).

(6) Some constraints on linguistic rhythm structure

PARSE-σ: syllables are footed

ALIGN-L/ALIGN-R: The left/right edge of every foot is aligned with the left/right edge of a Prosodic Word (PrWd) (it forces all feet to be adjacent at an edge)

PEAK PROMINENCE: the heaviest syllable in a PrWd has main stress

FOOT BINARITY (FtBIN): a foot consists of two syllables
RHYTHM TYPE = TROCHAIC/IAMBIC: the first/last syllable in a foot is stressed

Again, different rankings of the universal constraints in (6) describe culture-specific variation in rhythmic structures. For example, if PARSE-σ dominates ALIGN the optimal pattern will be a bounded stress system as in stress-timed languages, such as English: (σ σ)(σ σ). If, on the other hand, ALIGN dominates PARSE-σ, unbounded systems are preferred as in syllable-timed languages such as French: (σ σ σ σ). If ALIGN-L dominates PARSE-σ the pattern of leftmost systems such as Khalkha Mongolian is optimal, if ALIGN-R is dominant, the rightmost system of e.g. Yawelmani is optimal (McCarthy & Prince, 1993; Prince & Smolensky, 1993). Furthermore, the interaction of ALIGN and PEAK PROMINENCE enables us to describe the difference between fixed stress systems as in Yawelmani (ALIGN-R dominates PEAK PROMINENCE) and weight-sensitive systems as in Hindi (PEAK PROMINENCE dominates ALIGN-R).

Possibly the most striking characteristic of rhythm is alternation, formulated in linguistics as the Obligatory Contour Principle (OCP; Leben, 1973), which prohibits two adjacent unstressed syllables (*LAPSE) or two adjacent stressed syllables (*CLASH) in (7). If these constraints are dominant, alternating rhythm patterns of strong (s) and weak (w) syllables occur: (s w)(s w)(s w). (Hayes, 1984) introduces eurithmicity rules for optimally alternating patterns: the disyllabic rule ensures (s w)-patterns and the quadrisyllabic rule (S w)(s w)-patterns in which S is the head of a larger domain. The former is captured in OT with the combination of FtBIN and RHYTHM TYPE = TROCHAIC, the latter is reformulated here as a constraint that requires alternation of strong and weak feet. ALIGN forces all feet to be adjacent at the left or right edge.

(7) OCP-markedness constraints on linguistic rhythm structure
ALTERNATE (OBLIGATORY CONTOUR PRINCIPLE) (OCP)
Examples: *LAPSE: two adjacent weak syllables are forbidden
*CLASH: two adjacent stressed syllables are forbidden
*FtFt: feet must not be adjacent
QUADRI-SYLLABIC constraint: alternate strong and weak feet: (S-w) (s-w)

These constraints enable us to describe rather complex systems as well. In a ternary rhythm system, as in Estonian, *FtFt dominates ALIGN-L and *LAPSE is ranked low in the hierarchy of constraints allowing for (s w) w (s w) w-patterns (Kager, 1994).

Musical preferences on rhythmic structure can be formulated as OT-like well-formedness conditions as well, including alignment constraints, OCP-constraints. In order to make comparison with language easier, we introduce the musical domain "unit" as a combination of two beats, analogous to two syllables united into a linguistic foot.

(8) Some similar constraints on musical rhythm structure
PARSE-BEAT: every beat is united
UNIT BINARITY (UnitBIN): a unit consists of two beats
RHYTHM TYPE = TROCHAIC: the first beat in a unit is stressed
PARSE-UNIT (PARSE-UNIT): every input unit is parsed equally strong in a phrase
ALIGN-L: The left edge of every unit is aligned with the left edge of a phrase
ALTERNATE (OBLIGATORY CONTOUR PRINCIPLE) (OCP)
Examples: *LAPSE: two adjacent weak beats are forbidden
*CLASH: two adjacent stressed beats are forbidden
*UnitUnit: units must not be adjacent
QUADRI-BEAT: alternate strong and weak units: (S-w) (s-w)

The constraints in (8) enable us to describe musical metre. Just like FtBIN in language, UnitBIN is an undominated constraint with respect to the description of the most common metre patterns, such as 3/4 and 4/4. Just as in the ternary rhythm patterns in Estonian, musical odd metres can be described by ranking *UnitUnit above ALIGN-L. Dominant ALIGN-L, on the other hand, is satisfied in even metres. The correspondence constraint PARSE-UNIT demands all feet to be equally strong, whereas the markedness constraint QUADRI-BEAT requires alternation between strong and weak feet. If PARSE-UNIT dominates QUADRI-beat, 2/4 will be optimal. If QUADRI-beat dominates PARSE-UNIT, 4/4 will be optimal. Some examples of isometric, regular rhythm types are given in (9). The ternary pattern in (9) consist of binary feet followed by a stray beat: (s w) w (s w) w, etc. With respect to language prosody, Selkirk (1984) describes these combinations as superfeet.

(9) Typology of metre based on re-ranking constraints
3/4: *UnitUnit >> Align-L >> PARSE-UNIT >> QUADRI-beat
2/4: Align-L >> * UnitUnit >> PARSE-UNIT >> QUADRI-beat
4/4: Align-L >> * UnitUnit >> QUADRI-beat >> PARSE-UNIT

If the constraints in (9) are low-ranked, rhythmic stability may be overruled by e.g. the influence of dominant melodic or harmonic conditions. If, for example, in a certain music culture parsing long notes of unequal length in a melody is more important than the rhythm constraints, the rhythm might become irregular or free.

Keep in mind that we might have to add a caveat to the description of rhythm in this section. Despite the attempt to speak of universal language and music characteristics, the focus is Western-centric, based on distances between beats. In other music cultures, e.g. in Indian Raga music, the idea of what rhythm is might be different. With "Western ears" we might fail to perceive a rhythm in the sequence of tones as a repetitive pattern of beats, but with "Eastern ears" the differences in note density in the melody structure may possibly be perceived as the rhythm of the song. For example, a sequence of predominant quarter notes in the melody followed by predominant sixteenth notes followed by predominant quarter notes again can be felt as a rhythmic flow without isochronic accented events.

In this section, we provided examples of universal cognitive strategies formulated as well-formedness conditions that can be ranked for importance culture-specifically. Identification of these conditions enables us to describe the language and music of different cultures and their mutual relation. We now turn to broader implications of the constraints and approaches we discussed in section 3.

## 4 Conclusion and perspective

The research approach presented in this chapter concerns the cognitive mechanisms underlying the human capacities to learn and structure language and music (universal well-formedness conditions) and the possible variations in languages and music cultures (the ranking for importance of these conditions). The first aim of the framework we suggest is to find out what kind of musical features can be matched with linguistic ones. The well-formedness conditions identified in this chapter will make it possible to relate typologically different languages to their matching musical cultures. This part of the research is aided in its aim by two freely available databases. For language, Dryer & Haspelmath (2013) present the World Atlas of Language Structures, including linguistic analyses of the sound systems of more than two thousand languages and language varieties, belonging to different language families. For music, ethnomusicologist Alan Lomax's field recordings, collected for his Cantometrics project (discussed above), served as the basis for the Global Jukebox (Wood & Arèvalo, 2018), a survey of the world’s music styles including structural analyses. Both databases are meant to show similarities and differences between cultures with respect to, on the one hand, the inventories of distinctive sound systems in both music and language, and, on the other hand, temporal ordering in both disciplines for harmony, melody and rhythm, as discussed in the previous section.

In our approach, similar well-formedness conditions for music and language enable us to answer questions such as: are the rhythm patterns in the music of a certain culture related to the rhythm type of its language? Does the complexity in harmonic patterns in music relate to the complexity in syllable structure in language? And what are the relevant correlates? Indeed, the second aim of our approach is the culture-specific ranking of the universal conditions on rhythm, melody, harmony and delivery. Lerdahl & Jackendoff (1983) show that in Western tonal music harmonically optimal conditions dominate metrical ones, i.e., harmony is more important than rhythm. Our first analysis of the data gives birth to the hypothesis that rhythmic conditions are more important in African music and language, whereas in Asian music and language melodic conditions predominate. For example, a preliminary analysis of Vietnamese music in South-East Asia in the Global Jukebox reveals that it is predominantly characterized as monophonic, the overall rhythm of the music is one-beat rhythm (meaning all notes are of the same length) or free rhythm, the melodic shape is arched or undulating, the phrases are long and the intervals between tones are narrow. The tempo of the music is quite slow or very slow and the singing is characterized by much melisma, maximal glissando, and extreme embellishment in general. Interestingly, the languages of the related cultures exhibit a complex tone system (Dryer & Haspelmath, 2013). Hausa music in West Africa, on the other hand, is predominantly characterized as polyphonic and very rhythmic. A common melodic shape is descending and very short phrases predominate. The intervals are wide or very wide and the tempo of the music is fast or very fast. The singing shows little or no embellishment nor glissando. We can see, then, that African music shows quite the opposite characteristics to Asian music. Similarly, if we compare African languages to Asian languages, African languages in general have a simple tone system, and Asian languages have a complex tone system. We intend to study patterns of this kind in the language and music families within the context of one general constraint-based framework for analysing linguistic and musical structure.

Liberman (1975) already assumed every form of temporally ordered behaviour to be structured the same way. The findings from our investigations will reveal 1) to what extent language and music are structurally similar, and 2) whether the world’s musical typologies and the corresponding regions’ linguistic typologies are related to each other on a structural level. If it is found that the rankings of universal well-formedness conditions in both the language and the music of a particular culture are related, this research will provide fundamental insight into the cognitive mechanisms that underlie the way people learn and structure language and music (i.e. universal well-formedness conditions) and provide basic insight into the possible variation in language and music of different cultures (i.e. the ranking of these conditions).

## Appendix: OT-Constraints used in this chapter

**A. Markedness constraints** (all these constraints ensure simple forms)
*COMPLEX: consonant clusters in syllables are forbidden
*Coda: a coda segment in a syllable is forbidden
ONS: every syllable begins with a consonant
Hnuc: the most sonorant segment is chosen as nucleus of a syllable
Hons/Hmar: the least sonorant segment is chosen syllable-initially
MaxContrast: establishes dispersion in a segment inventory
*CAT (no category)/*Combi, etc: prohibits new categories in the inventory, combinations of chords or notes in a chord, etc. that make inventories, output structures or combinations of events more complex
WRAP (Phrase) combines different categories into one. For example, it requires that a minor phrase, e.g. a preposition phrase is contained in a single major phrase, e.g. an intonation phrase (IP)
PEAK PROMINENCE: the heaviest syllable in a PrWd has main stress
FOOT BINARITY (FtBIN): a foot consists of two syllables
UNIT BINARITY (UnitBIN): a unit consists of two beats (musical equivalent of FtBIN)
ALTERNATE (OBLIGATORY CONTOUR PRINCIPLE) (OCP), examples:
RHYTHM TYPE = TROCHAIC/IAMBIC: the first/last syllable in a foot is stressed
RHYTHM TYPE = TROCHAIC: the first beat in a unit is stressed (musical equivalent)
*LAPSE: two adjacent weak syllables are forbidden
*LAPSE: two adjacent weak beats are forbidden (musical equivalent)
*CLASH: two adjacent stressed syllables are forbidden
*CLASH: two adjacent stressed beats are forbidden (musical equivalent)
*FtFt: feet must not be adjacent
*UnitUnit: units must not be adjacent (musical equivalent of *FtFt)
QUADRI-SYLLABIC constraint: alternate strong and weak feet: (S-w) (s-w)
QUADRI-BEAT: alternate strong and weak units: (S-w) (s-w) (musical equivalent)

**B. Correspondence/Faithfulness constraints** (ensure diversity by linking segments on different strings, e.g. a segment in the underlying form and the realized form)

MAX-IO: every (underlying) input segment/feature has a correspondent in the output (no deletion) (cf. PARSE in earlier OT-literature: prohibits underparsing)

DEP-IO: every (realized) output segment/feature has a correspondent in the input (no insertion/epenthesis) (cf. FILL in earlier OT-literature: prohibits overparsing)

PARSECAT (Parse as category): classifies acoustically available features into a phoneme category

PARSE COMBI MAX(imally): determines the complexity of harmonic structures that appear in the music

PARSE-σ: all syllables are footed in the prosodic structure

PARSE-BEAT: every beat is united (musical equivalent of PARSE- σ)

PARSE-UNIT: every input unit is parsed equally strong in a phrase

**C. Alignment constraints** (function as domain boundary markers, they require e.g. that the edge of a phrase coincides with a high or low tone).

ALIGN-L (High tone): aligns the left boundary of the domain with a H-tone

ALIGN-R (Low tone): aligns the right boundary of the domain with a L-tone

ALIGN-L/ALIGN-R: the left/right edge of every foot is aligned with the left/right edge of a Prosodic Word (PrWd) (it forces all feet to be adjacent at an edge of the directly higher prosodic domain)

ALIGN-L: The left edge of every unit is aligned with the left edge of a phrase (musical equivalent)

# References

Ainsworth, William A. & Kuldip K. Paliwal. 1984. Correlation between the production and perception of the English glides /w, r, l, j/. *Journal of Phonetics* 12(3). 237–243.

Antilla, Arto & Adams Bodomo. 1996. Stress and tone in Dagaare. Ms. Stanford University and Norwegian University of Science and Technology.

Archangeli, Diana. 1997. *Optimality Theory: An introduction to linguistics*. Oxford: Blackwell.

Asano, Rie & Cedric Boeckx. 2015. Syntax in language and music: What is the right level of comparison? *Frontiers in Psychology* 6. 942.

Ball, Philip. 2010. *The music instinct: How music works and why we can't do without It*. Oxford: Oxford University Press.

Boersma, Paul. 1998. Functional phonology: Formalizing the interaction between articulatory and perceptual drives. [Doctoral Dissertation]: University of Amsterdam.

Brown, Steven & Joseph Jordania. 2011. Universals in the world's musics. *Psychology of Music* 41(2). 229–248.

Chládková, Katerina & Nikola Paillereau. 2020. The what and when of universal perception: A review of early speech sound acquisition. *Language Learning: A Journal of Research in Language Studies* 70(4). 1136–1182.

Daniele, Joseph R. & Aniruddh D. Patel. The interplay of linguistic and historical influences on musical rhythm in different cultures. Proceedings of the 8th International Conference on Music Perception and Cognition; 2004; Sydney, Australia. Causal Productions. p 759–762.

Dryer, Matthew S. & Martin Haspelmath. 2013. The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Eisenberg, Peter. 1991. Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. *Zeitschrift für Sprachwissenschaft* 10(1). 37–64.

Fenk-Oczlon, Gertraud & August Fenk. 2009. Some parallels between language and music from a cognitive and evolutionary perspective. *Musicae Scientiae* 13(2). 201–226.

Flemming, Edward S. 1995. Auditory representations in phonology [Doctoral Dissertation]: UCLA.

Ghamdi, Ahmed Al. 2006. A preliminary analysis of the intonation of Riyadh Saudi Arabic. Rutgers Optimality Archive. [#812 – 0306].

Gilbers, Dicky & Helen de Hoop. 1998. Conflicting constraints: An introduction to Optimality Theory. *Lingua* 104(1–2). 1–12.

Gilbers, Dicky & Wouter Jansen. 1996. Klemtoon en ritme in Optimality Theory, deel 1: Hoofd-, neven-, samenstellings- en woordgroepsklemtoon in het Nederlands. *Tabu* 26. 53–101.

Gilbers, Dicky & Maartje Schreuder. 2000. Taal en muziek in Optimaliteitstheorie. *Tabu* 1–2. 1–27.

Gilbers, Dicky & Maartje Schreuder. 2002. Language and music in Optimality Theory. Rutgers Optimality Archive [#571–0103].

Gilbers, Steven, Nienke Hoeksema, Kees de Bot & Wander Lowie. 2020. Regional variation in West and East Coast African-American English prosody and rap flows. *Language and Speech* 63(4). 713–745.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15(1). 33–74.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago, IL: The University of Chicago Press.

Jackendoff, Ray. 2009. Parallels and nonparallels between language and music. *Music Perception* 26(3). 195–204.

Jakobson, Roman. 1972. Why 'mama' and 'papa'? In: Bertil Malmberg, editor. *Readings in modern linguistics*, 313–320. The Hague: De Gruyter.

Jekiel, Mateusz. 2015. Comparing rhythm in speech and music: The case of English and Polish. *Yearbook of the Poznan Linguistic Meeting* 1. 55–71.

Kager, René. 1994. Ternary rhythm in alignment theory. Utrecht University.

Krumhansl, Carol L. 1979. The psychological representation of musical pitch in a tonal context. *Cognitive Psychology* 11(3). 346–374.

Krumhansl, Carol L., Jamshed J. Bharucha & Edward J. Kessler. 1982. Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance* 8(1). 24–36.

Lacy, Paul de. 2002. The interaction of tone and stress in optimality theory. *Phonology* 19(1). 1–32.

Leben, William R. 1973. Suprasegmental phonology [Doctoral Dissertation]. Cambridge, MA.

Lerdahl, Fred & Ray Jackendoff. 1983. *A generative theory of tonal music*. Cambridge, MA: The MIT Press.

Liberman, Mark. 1975. *The intonational system of English*. New York: Garland Publishing Inc.

Lomax, Alan. 1976. Cantometrics: An approach to the anthropology of music [Doctoral Dissertation]. Berkeley: University of California Extension Media Center.

Lynch, Michael P., Rebecca E. Eilers, D. Kimbrough Oller & Richard C. Urbano. 1990. Innateness, experience, and music perception. *Psychological Science* 1(4). 272–276.

McCarthy, John J. & Alan S. Prince. 1993. *Prosodic Morphology I: Constraint interaction and satisfaction*. New Brunswick, NW: Rutgers University Center for Cognitive Science.

McDermott, Josh H., Alan F. Schultz, Eduardo A. Undurraga & Ricardo A. Godoy. 2016. Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535. 547–550.

McGowan, Rebecca W. & Andrea Levitt. 2011. A comparison of rhythm in English dialects and music. *Music Perception* 28(3). 307–314.

Nouveau, Dominique. 1994. Language acquisition, metrical theory, and optimality: A study of Dutch word stress [Doctoral Dissertation]: Rijksuniversiteit Utrecht.

Passy, Paul. 1891. *Étude sur les changements phonétiques et leurs caractères généraux*. Paris: Librairie Firmin – Didot.

Patel, Aniruddh D. 2010. *Music, language, and the brain*. Oxford: Oxford University Press.

Patel, Aniruddh D. & Joseph R. Daniele. 2003. An empirical comparison of rhythm in language and music. *Cognition* 87. B35–B45.

Perlman, Marc & Carol L. Krumhansl. 1996. An experimental study of internal interval standards in Javanese and Western musicians. *Music Perception* 14. 95–116.

Prince, Alan S. & Paul Smolensky. 1993. *Optimality Theory. Constraint interaction in Generative Grammar*. Malden, MA: Blackwell.

Rebernik, Teja & Dicky Gilbers. 2017. Morebitni medsebojni vpliv prozodije slovenskega govora in slovenske ljudske pesmi. *Slavistična Revija* 65(4). 577–952.

Savage, Patrick E. 2018. Alan Lomax's cantometrics project: A comprehensive review. *Music & Science* 1. 1–19.

Schreuder, Maartje. 2006. Prosodic processes in speech and music [Doctoral Dissertation]: University of Groningen.

Selkirk, Elisabeth O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.

Selkirk, Elisabeth O. 1995. Sentence prosody: Intonation, stress, and phrasing. In: John A. Goldsmith, editor. *The handbook of phonological theory*, 550–569. Cambridge, MA: Blackwell.

Selkirk, Elisabeth O. 2000. The interaction of constraints on prosodic phrasing. In: Merle Horne, editor. *Prosody: Theory and experiment*, 231–261. Dordrecht: Kluwer Academic Publishers.

Temperley, Nicholas & David Temperley. 2011. Music-language correlations and the "scotch snap". *Music Perception* 29(1). 51–63.

Trehub, Sandra R., E. Glenn Schellenberg & Stuart B. Kamenetzky. 1999. Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance* 25(4). 965–975.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2). 219–255.

VanHandel, Leigh & Tian Song. Influence of linguistic rhythm on individual compositional style in 19th century French and German art song. Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music; 2009; Jyväskylä, Finland. p 553–557.

Wood, Anna L. & M. Jorge Arèvalo. 2018. The global jukebox.

Zhang, Jie. 2004. *The role of contrast-specific and language-specific phonetics in contour tone distribution*. In: Bruce Hayes, Robert Kirchner, Donca Steriade, editors. *Phonetically-based phonology*, 157–190. Cambridge: Cambridge University Press.