Received: 24 June 2022 | Revised: 8 August 2022 | Accepted: 18 August 2022

DOI: 10.1111/lnc3.12472

REVIEW ARTICLE

WILEY

Metre, grouping, and event hierarchies in music: A tutorial for linguists

Jonah Katz

West Virginia University, Morgantown, West Virginia, USA

Correspondence
Jonah Katz, West Virginia University, Morgantown, WV 26506-6201, USA.
Email: katzlinguist@gmail.com

Abstract

This paper reviews some basic elements of musical structure, drawing from work in traditional and cognitive musicology, ethnomusicology, psychology, and generative textsetting. Music features two different hierarchical representational components that can both be visualised in grid notation: *metrical structure* and *event hierarchies* (also referred to as Time-Span Reduction). Metrical structure is an abstract pattern of stronger and weaker points in time that form a temporal 'scaffold' against which auditory events occur, but is partly independent from those occurring events. Event hierarchies encode the constituency (referred to as *grouping*) and prominence of actually-occurring musical events. Basic principles of these components are illustrated with examples from Western children's, folk, and popular song. The need for both metrical hierarchy and event hierarchy is illustrated using mismatches between metrical and rhythmic structure. The formal, conceptual, and empirical features of musical metre are quite different from linguistic stress and prosody, despite the frequent analogies drawn between them. Event hierarchies, on the other hand, are shown to resemble linguistic prosodic structure with regard to all of these features.

1 | INTRODUCTION

This paper offers an overview of basic musical structure for those familiar with linguistic theory. The structure and cognition of music has historically been of interest to linguists because it plausibly teaches us something about the nature of mental computation more generally. Music and language are both complex systems of implicit knowledge that involve conveying hierarchical structure in a linear way in some perceptual dimension (auditory or visual) over time. The kinds of representations, and relationships between those representations, necessary to describe the two domains have a lot in common, though they also differ in many ways (Katz, 2022; Lerdahl & Jackendoff, 1983; Liberman, 1975; Patel, 2008; Steedman, 1984). Just as in linguistics, however, accurately describing music as a cognitive system requires technical concepts and vocabulary that go beyond our everyday experiences and explicit knowledge. This article is an attempt to familiarise readers with some basic technical aspects of musical structure, a task that is made somewhat easier if the reader already understands linguistic notions of constituency, prominence, and hierarchy.

I assume no formal background in music, and examples are notated in a simplified way that doesn't require familiarity with Western musical notation. All examples are transposed to C major. This notation is somewhat impoverished, lacking information about the direction of pitch motion in melodies, but it won't matter much for the examples discussed here. I illustrate relevant principles and structures using the simplest and most widely-known musical pieces I could think of. Notably, I've avoided using complex examples from scholarly, notated Western Tonal (henceforth *classical*) Music. All of the examples here are pieces that readers in North America and many other parts of the world will likely recognise: children's songs, folk songs, and music from popular genres such as rock and pop. Some of these have roots in notated scores by professional composers, but are typically learnt and transmitted aurally in modern cultures, without many of their original notated details. Performances and written representations of all the pieces mentioned in the paper can be easily found with an Internet search, often in multiple versions. I strongly recommend that readers listen to some version of the examples while reading the paper: understanding the descriptions of music given here will be much easier with auditory input. I use the terms ‘song’ and ‘piece’ interchangeably here, to refer to whatever a listener or performer knows when they know a freestanding unit of music. What exactly this knowledge consists of is the main subject of the paper.

A terminological note: I discuss at length an aspect of music referred to as ‘metrical structure’. Linguists also use the term ‘metrical’ to refer to prosodic prominence and perhaps other aspects of prosody. As I will illustrate below, the metrical structure of music is very much *unlike* bracketed grids from language. And this means that the convention of referring to both as ‘metrical’ can be extremely confusing when we compare music and language. In this paper, I use the term ‘metrical’ to refer only to musical metre, and not to aspects of linguistic prosody or stress. I will claim in Section 4 that language has no counterpart to musical metre.

2 | MUSICAL METRE

Most pieces in most genres of music display *metrical structure*, an abstract repeating sequence of alternating strong and weak *beats* associated with particular moments in time.[1] While this has been most thoroughly explored in Western genres, it is also a basic property of music from Indonesia (Hughes, 1988), the Balkan countries (Hannon & Trehub, 2005; Kaliakatsos-Papakostas et al., 2014), Turkey (Mungan et al., 2017), China (Nan et al., 2009; Stock, 1993), Japan (Pasciak, 2017), Central Australia (Turpin, 2007, 2017), and diverse African cultures including Venda (Blacking, 1970), Ewe (Ekweume, 1975), Berber (Dell & Elmedlaoui, 2002), Dogon (McPherson & Ryan, 2018), and Hausa (Hayes & Schuh, 2019). While the *existence* of musical metre is widespread in human cultures, the precise nature and variety of repeating patterns differ to some extent between cultures. The detailed examples in this paper are from my areas of musical expertise, which are broadly North American and Western European.

Alternating strong and weak beats in musical metre are sometimes depicted in a grid visualisation, as in (1) (Lerdahl & Jackendoff, 1983; Liberman, 1975). The 'X' marks in the grid represent points in time. This grid depicts the most common Western metre, which is binary (referred to as *duple*) at all levels: every other beat at a given level projects up to the next, indicating its relative metrical strength. In most Western music, and all of the examples in this paper, the temporal interpretation of the grid approaches rigid *isochrony*: if the separation between the first X and the second one at a particular level is 147 ms, then the third X will also be roughly 147 ms after the second one. In some genres (notably classical music), isochrony can be relaxed or abandoned for sections of a piece. This is rare in the genres discussed here. In other rare cases, the *number* of low-level beats between higher-level beats can vary within a metrical pattern, for instance mixing duple and triple patterns at the same level. This is referred to as *odd metre*, and is possible but rare in pop, rock, and jazz music; it is famously common in Balkan music (Hannon & Trehub, 2005). Western musicians normally conceive of metrical patterns as 'beginning' on the strongest beat at some level and 'ending' just before the next strong beat at that level, a unit referred to as a *measure* or *bar*. That said, metrical patterns cycle continuously with no obvious beginning or ending, and songs need not begin on the strongest beat nor end on the 'final' beat of a cycle.

(1) Grid representation of duple meter

X								X							
X				X				X				X			
X		X		X		X		X		X		X		X	
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

This example depicts two repetitions of a metrical pattern with four levels of prominence, and each repetition thus contains 8 beats at the lowest level. Note that in principle, there is no upper limit to the number of levels that can be associated with a musical object: if we add an X mark above the first column, we will have one repetition of a duple metre with five levels of prominence; if we take away the top level of X marks, we have four repetitions of a duple metre with three levels of prominence. The number of levels of prominence that an individual may infer from a specific piece of music will in practice be limited, and so we usually notate only as many levels as we need to analyse whatever phenomenon we are examining.

Other common Western metrical structures are like (1), but with one or more levels, generally lower in the grid, displaying ternary rather than binary subdivisions. Two examples are shown in (2). On the left, the lowest level is ternary; on the right, the second lowest. It is also possible for both of the two lowest levels to be ternary.

(2) Mixed duple/triple metrical structures

X											
X						X					
X			X			X			X		
X	X	X	X	X	X	X	X	X	X	X	X

X											
X						X					
X		X		X		X		X		X	
X	X	X	X	X	X	X	X	X	X	X	X

The basic form of musical metrical structure, then, is a repeating pattern of points in time that are associated with higher and lower levels of strength. Of course, in order for a listener to infer such a representation, or for a performer to implement it, particular events must *occur* in particular metrical positions. For instance, if I hit a stick on a table at isochronous intervals, with loudness proportional to the metrical prominence in one of the examples from (1–2), listeners are likely to infer the corresponding metrical representation. That said, the series of stick sounds is not itself a metrical structure: it is a series of musical events that implicates a particular metrical structure. For occurring musical events, I use the term *rhythmic*, and distinguish it from *metrical*. This is because music generally does *not* consist of isochronous, repeating sequences of events isomorphic to a metrical structure. Instead, musical event onsets are placed in some metrical positions but not others, listeners use the occurring rhythms to infer a metrical structure, and the correspondence can sometimes be ambiguous or can fail (Povel & Essens, 1985). I refer to a metrical position associated with the onset of an event as being *filled* by that event, and a metrical position not associated with an event onset as *unfilled*, even if it is contained in some occurring event.

Example (3) illustrates simple association of a rhythmic phrase with the metrical structure in (1): this represents the first part of the children's song ‘Frère Jacques/Brother John’. In place of music notation, I use the English letter names of pitches. In this example and some of the following ones, I include the conventional lyrics for songs to help readers orient themselves. There are clear and obvious correspondences between linguistic structures in the lyrics and musical rhythm, which I briefly discuss in Sections 3 and 4. But all of the aspects of rhythmic organisation discussed here hold for non-vocal music as well, and pertain to music-internal features and properties. I would invite the reader to ignore the linguistic structure of lyrics as much as possible in the first two sections of the paper.

(3) Rhythm: events aligned with metrical structure

X								X							
X				X				X				X			
X		X		X		X		X		X		X		X	
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
C	D	E	C	C	D	E	C	E	F	G		E	F	G	
are	you	slee	ping	are	you	slee	ping	bro	ther	John		bro	ther	John	
fre	re	Ja	cques	fre	re	Ja	cques	dor	mez	vous		dor	mez	vous	

The most basic principles of rhythm-metre correspondence are fairly straightforward. Palmer and Krumhansl (1990) show that in classical music, stronger metrical positions are more likely to be filled by an event than weaker positions are, and that Western listeners judge rhythmic sequences as being better when stronger metrical positions are filled (see also Jongsma et al., 2004). Povel and Essens (1985) demonstrate that rhythms are easier to learn and imitate when more strong positions are filled, and when the strongest positions are filled by longer events. Dawe et al. (1995) find that acoustic salience also contributes to the probability of perceiving an event onset as filling a strong metrical position, including the long duration just mentioned, as well as the presence of more simultaneous notes (referred to as a *chord*) or a change in *harmony* (roughly the tonal properties of a chord).

None of these tendencies are hard rules: they are frequently violated, even in rhythmically simple music. In the example in (3), the two unfilled metrical positions are both weak, which accords with the preferences summarised above. The two long notes in the example, however, are at an intermediate level of prominence. This does not interfere with inferring a metrical structure. Another factor relevant to (3) is *tonal stability*: there is a strong tendency in most Western genres for more tonally stable pitches or pitch combinations to occupy stronger metrical positions. The theory of tonal stability is complex (e.g. Krumhansl, 1990; Lerdahl, 2001), but the basic idea is that tonal music is perceived within one or more tonal *spaces*, and certain pitch classes within a space are more central and stable than others. The example above is in the tonal space of C major, with implied tonic harmony, which means that C, E, and G will be the most stable tones in this context. It's easy to see that every prominent beat in (3) is associated with one of these tones, and all less-stable tones occupy weak beats. In the Dawe et al. (1995) study mentioned above, listeners were more likely to infer a strong metrical position where the chord behind a melody changed, but the effect was considerably weaker when the novel chord was highly unstable in the tonal space established by the stimuli.

To understand why most pieces of music are metrically unambiguous despite sometimes violating rhythmic alignment preferences, consider what would happen if we inferred one of the triple metres

(4) A bad metrical interpretation of ‘Frère Jacques’

X												X			
X						X						X			
X			X			X			X			X			X
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
C	D	E	C	C	D	E	C	E	F	G		E	F	G	
are	you	slee	ping	are	you	slee	ping	bro	ther	John		bro	ther	John	
fre	re	Ja	cques	fre	re	Ja	cques	dor	mez	vous		dor	mez	vous	

from (2) and tried to align ‘Frère Jacques’ to it. This is shown in (4).

One obvious issue here is that lexical and phrasal stress in the lyrics badly mismatch metrical prominence. We focus here, though, on several other factors that would make this a bad metrical analysis of the song *even if it had no words*. In the structure in (4), the two longest notes are both on weak beats. One of the unoccupied metrical positions (the final one) is stronger than an adjacent filled position. Musicians use the term *syncopation* for situations where a strong beat is unfilled while an adjacent weaker one is filled; syncopation is systematic and pervasive in many genres (Tan et al., 2019; Temperley, 1999; Waller, 2016), but fairly unusual in children's songs. In (4), one of the intermediate-strength beats is filled by a less tonally stable pitch (the 10th note, F) while adjacent weaker beats are filled by more stable events (E and G, which are in the tonic chord); mismatch between tonal stability and metrical strength is referred to as *appoggiatura*. Perhaps most strikingly, though, the repeated *cells* or *motifs* in the song, C-D-E-C and E-F-G, have a different periodicity (4 beats) than the triple metre and are therefore associated with different metrical positions each time they occur. This type of phase mismatch, referred to as *hemiola*, is rare in most genres and almost unattested in children's songs. For all of these reasons, ‘Frère Jacques’ is extremely unlikely to be associated with the metrical structure in (4).[2]

The phenomena just introduced, syncopation, appoggiatura, and hemiola, have profound implications for modelling musical metre and rhythm. While they are vanishingly unlikely to be associated with ‘Frère Jacques’, there are many Western songs and genres where they unambiguously occur. Example (5) shows all three phenomena occurring in ‘Turkey in the Straw’, a traditional American folk song. In (5a), local repetitions of melodic thirds (E-G) occur in three-beat intervals against a duple metre. The metrical prominence of the two notes is reversed on each repetition (this rhythm continues into the following measures, not shown here). The relatively strong 3rd and 11th metrical positions (shaded in grey) are unfilled, while both adjacent weaker beats are filled. In (5b), the harmony shifts from the tonic I to the dominant V chord on the shaded beat (harmony is shown below the example). This position is filled by a relatively unstable melody note E, not contained in the local harmony, while the following more stable D notes, which are contained in the dominant chord, occupy weaker metrical positions.

(5) Syncopation, hemiola, and appoggiatura in ‘Turkey in the straw’

(a) Beginning of second section: syncopation, hemiola

X								X							
X				X				X				X			
X		X		X		X		X		X		X		X	
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
(E	G)		(E	G)		(G)		(E	G)		(E	G)			

(b) Excerpt from first section (half-cadence): appoggiatura

		X								X					
		X				X				X				X	
X		X		X		X		X		X		X		X	
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
C	D	E		E		E	D	C	D	E		D		D	
I (C)										*V (G)*					

Each of these phenomena is somewhat unusual in folk and children's melodies; they tend to be more common in pop, rock, hip-hop, and classical music. It's hard to think of popular songs from the last 50 years that *don't* have some form of syncopation or appoggiatura (Temperley, 1999); picking three examples from songs that my children happen to have been listening to today, for instance, the chorus sections of Michael Jackson's ‘Don't Stop 'Til you Get Enough’, Weezer's ‘Say it Ain't So’, and Carly Rae Jespen's ‘Call me Maybe’ all feature prominent syncopation and appoggiatura. Local hemiola, where an out-of-phase rhythmic cell repeats once or twice (as in 5a), is reasonably common. The iconic horn line from Sam & Dave's ‘Soul Man’, for instance, consists of four notes spanning six low-level beats against a duple metre, repeated with a slight rhythmic change. Extended hemiola, where out-of-phase rhythms are repeated many times, is less common, perhaps because it can eventually overwhelm the inferred metrical structure and disorient the listener. Famous examples include the melody from Irving Berlin's ‘Puttin' on the Ritz’, the bass/guitar/horn interlude from Kool & the Gang's ‘Jungle Boogie’, and the final verses of the guitar solo from the Eagles' ‘Hotel California’.

Even in genres that include these complex rhythmic elements, they are always a departure from a more basic norm. Example (3) and the metrical alignment preferences discussed above illustrate a default for rhythms: strong musical events (whether through occurrence, acoustic salience, or tonal stability) are associated with strong metrical positions. Syncopation and appoggiatura, then, are cases where metrical prominence dissociates from the prominence of actual occurring musical events (sometimes referred to as *accent* or *phenomenal accent*), and as such, they show that the prominence of musical events must be represented independently from the prominence of metrical positions. Hemiola is a case where some constituent in a musical piece dissociates from periodicities in the metrical structure. As such, it shows that rhythmic constituents exist and are partly independent from metrical structure. These complex phenomena are particularly important in the context of music-language comparison, because they have no linguistic counterparts and constitute an important source of evidence about which types of linguistic and musical representations bear similarities to one another.

The nature of metrical 'prominence' in music is rather different from linguistic analogues such as stress and accent. One of the core properties of musical metre is its abstract nature: metrical positions are points in time that need not be associated with any particular kind of event, or with any event at all. 'Strength' and 'prominence', however, are terms that we normally associate with auditory events and particular acoustic properties, including in linguistics. So what is metrical strength *for* if it isn't invariantly marked by particular forms of acoustic salience? One common answer is that metrical strength guides the allocation of implicit *attention* in music perception.

The metrical grids in the preceding section are perfectly regular, repeating over and over again, and in most cases are isochronous. The time between 'X' marks is equal at each level (with limited exceptions in odd metre), these temporal intervals are longer by factors of 2 or 3 at higher levels, and as such it is possible to think about metrical grids as 'hierarchies of intersecting periodicities', in Liberman and Prince's (1977) famous formulation. This dovetails nicely with psychological accounts of auditory attending as a dynamic process in time (Jones, 1976; Jones & Boltz, 1989; Large & Kolen, 1994; Large & Palmer, 2002). While details differ depending on the implementation, the basic idea behind these theories is that attention can be modelled as a quantity of 'energy' that oscillates over time periods, and multiple levels of periodicity can correspond to multiple levels of attentional oscillators. Metrically prominent positions are then points in time when maxima coincide at several layers of hierarchically nested oscillators, 'adding up' to increase the attentional energy at and around that point in time.

This Dynamic Attending approach to metre has generated a wealth of studies demonstrating that auditory perception tracks or *entrains to* the details of metrical structure. Using loudness and duration discrimination tasks, for instance, Repp (2010) shows that auditory sensitivity is greater on the strong beat of a triple metre than on other beats. Listeners are better at categorising pitch differences (Jones et al., 2002) and faster at detecting auditory 'oddball' events (Penel & Jones, 2005) on beats than they are in positions not corresponding to a beat. Katz et al. (2015) show consistent within-listener effects of higher versus lower metrical strength across different metrical structures on an oddball detection task, but the *direction* of the effect differs between listeners; Penel and Jones (2005) report similar results for beat versus non-beat positions. Bolger et al. (2013, 2014) find faster oddball detection and categorisation in stronger metrical positions than weaker ones, but with the caveat that positions immediately preceding a beat seem to inherit the strength of that beat. While the details of which tasks are enhanced for which listeners in which precise temporal positions are somewhat complicated, the overall literature clearly establishes that a detailed theory of metrical structure is crucial for any attempt to model dynamic attending.

3 | RHYTHM, CONSTITUENCY, AND ACCENT

Metrical structure describes the prominence and possibly the grouping of abstract beats. The theory of music also requires a description of the prominence and grouping of actually-occurring musical events, an *event hierarchy*. If all music were like the first phrase of 'Frère Jacques', this would be trivial: occurring events are grouped according to which metrical measure they occur in, and they bear the prominence associated with their metrical position. This is shown in (6), which copies the metrical structure for each occurring event in (3) and adds brackets corresponding to individual events and metrical groupings. Prominence marks here have been changed from 'X' to pitch-class names; as noted earlier, tonal properties matter for both structural importance and metrical setting.

(6) Musical phrasing projected from metrical structure

(C							)	(E					)
(C			)	(C			)	(E		)	(E		)
(C	)	(E	)	(C	)	(E	)	(E	)	(G)	(E	)	(G)
(C)	(D)	(E)	(C)	(C)	(D)	(E)	(C)	(E)	(F)	(G)	(E)	(F)	(G)
C	D	E	C	C	D	E	C	E	F	G	E	F	G
[are	you	slee	ping]	[are	you	slee	ping]	[bro	ther	John]	[bro	ther	John]
[fre	re	Ja	cques]	[fre	re	Ja	cques]	[dor	mez	vous]	[dor	mez	vous]

This structure is a pretty good description of 'Frère Jacques'. The repeating melodic motifs are characterised as constituents. The two elongated notes are characterised as constituent-final. More tonally-stable notes are generally assigned higher prominence than less tonally-stable ones. The higher-level musical constituents picked out by this analysis are precisely the ones to which linguistic constituents are aligned in both the French and English lyrics. It is a shame, then (at least for the analyst), that this is not how musical rhythm, constituency, or prominence actually work. As syncopation, appoggiatura, and hemiola show, musical constituents frequently mismatch 'metrical constituents', and the structural prominence of musical events can mismatch their relative metrical prominence.

There are many children's, folk, or traditional songs that pattern like 'Frère Jacques' in starting musical phrases at the strongest metrical positions, for instance 'Row your Boat' and 'Twinkle Twinkle Little Star'. There are songs that have a short *pickup* or *anacrusis*, an initial event in some constituent that occurs before the first beat of a putative metrical unit, for instance 'My Bonnie Lies over the Ocean' or 'The Farmer in the Dell'. And there are songs whose misalignments with putative strong-initial metrical units goes beyond a short pickup note, such as 'Miss Mary Mack', 'The Chicken Dance', and 'Michael Row the Boat Ashore'. All of these configurations are also attested in non-vocal music. Two examples of non-alignment are shown below in (7). A single putative metrical unit is shaded in grey, and likely rhythmic constituents are shown as brackets in the bottom line. For the moment, we assume that linguistic constituency and rhyme are unambiguous cues to constituent endings; there are purely musical properties, described in the next section, that would converge on the same constituents in these specific examples. The observation is simply that metrical and rhythmic units mismatch in various ways.

(7) Misalignment in simple songs

(a) 'My Bonnie Lies over the Ocean'

	X												X									
	X						X						X						X			
	X			X			X			X			X			X			X			X
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
G	E	D	C	D	C	A	G	E				G	E	D	C	C	B	C	D			

[my bon nie lies o ver the o cean] [my bon nie lies o ver the sea]

(b) 'Miss Mary Mack'

			X								X				
			X				X				X				X
	X		X		X		X		X		X		X		X
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
G	A	B	C		C		C	G	A	B	C		C		C

[miss ma ry mack mack mack] [all dressed in black black black]

The segmentation of occurring musical events into constituents like those in (6) is referred to as *grouping* (Cambouropoulos, 2001; Lerdahl & Jackendoff, 1983; Narmour, 1990; see Katz, 2022 for review). The term is also used to describe the segregation of simultaneously sounded musical sequences into component 'voices' or 'streams' (Huron, 1991), which I do not address in this paper. Musicologists sometimes use the term *phrasing* for this type of segmentation, and sometimes that term is limited to relatively high-level constituents. In children's songs, linguistic constituency is a clear cue to high-level musical segmentation. But music without words is also marked by broad regularities in the relationship between surface acoustics and perceived constituents (Deliège, 1987; Deutsch, 1980; Dowling, 1973; Peretz, 1989), some of which extend across cultures (Mungan et al., 2017; Nan et al., 2009; Popescu et al., 2021).

There is broad agreement that some properties of musical grouping relate to general Gestalt principles of perception, in particular *proximity* and *similarity* (Wertheimer, 1938). There are also culture-specific conventions that can affect grouping in various ways, but I begin here with more basic principles. Grouping principles are probabilistic and violable, and may be weighted differently in different songs, cultures, auditory dimensions, and individual listeners. Nonetheless, evidence for their existence is quite robust within and across human cultures (Katz, 2022). The proximity principle for audition entails that successive events that are closer in time are more likely to be associated with coherent units than events that are more distant in time. In music, it essentially says that a silence or long note is likely to mark the end of a constituent or *group*. The similarity principle holds that for any acoustic/auditory dimension *d*, successive events that are more similar in *d* are more likely to be associated with coherent units than events that are less similar in *d*. Pitch, *timbre* (spectral properties pertaining to the identity of voices and instruments, among other things), and loudness are examples of auditory dimensions that play a role in similarity-based grouping.

As an illustration, consider the beginning of the standard melody for the Jarabe Tapatío, referred to in the US as the 'Mexican Hat Dance', in (8). This version of a traditional dance melody from Jalisco was composed and notated by a professional musician in the 19th century, but most North American children learn it aurally through a variety of performances that don't preserve the details of the original notated version.

(8) 'Jarabe Tapatío/Mexican Hat Dance'

		X												X								
		X						X						X						X		
		X			X			X			X			X			X			X		
X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
G	F#	G	E	D#	E	C	B	C	G			E	F	G	A	B	C	D	E	F	D	

In terms of grouping cues, the most obvious one is the silence or sustain between the 'G' halfway through the excerpt and the 'E' that follows it. For pitch, we find that the first half consists of triplets of notes that are relatively close in pitch internally, but not as close to the preceding or following triplets. In the second half, we find a durationally identical phrase but without the pitch 'leaps'. If you search for an actual performance of this melody, you will find more often than not that the second half is also divided acoustically into groups of three notes, with the third note performed shorter and a bit louder than the first two. This is optionally indicated in Western music notation with an arc referred to as a *slur*, and many written representations of this tune feature slurring around these three-note sequences. Because these groups of three are often present in performances and scores, and other logically possible groupings are not, I will assume that they are part of the representation of the song. So the low level grouping is (G F# G) (E D# E) (C B C) (G) (E F G) (A B C) (D E F) (D).

The minimal claim of grouping theory with regard to this melody is that, when presented with a typical performance, listeners will segment it into the groups indicated above. This, in and of itself, is not that interesting, although of course other segmentations are logically possible and it is worth noting that they are less likely than this one. The real power of grouping theory comes from the fact that the constituents we single out through the application of low-level psychoacoustic Gestalt principles are claimed to be the *same* constituents that are relevant to other components of musical cognition such as combinatorics, harmonic interpretation, representation in memory, or setting a linguistic text. Part of this argument for 'Jarabe Tapatío' is relatively simple and requires little musical insight; a second part involves tonal analysis. I attempt to outline both parts of the argument below, with the simple part first.

If we were going to build a theory of how songs like 'Jarabe Tapatío' are composed, or how a listener assigns structure to them, one of the basic insights we would need to incorporate is that various bits of music that are *equivalent* to one another on certain dimensions tend to *repeat*. The Gestalt principles singled out a lot of three-note constituents in the analysis above. It turns out that these three-note *motifs* or *cells* have identical internal pitch relationships within each half of the excerpt, and are normally phrased together in the manner represented by written slurring. They also have identical durations. The pitch contours, but not the durations or slurring, contrast between the two halves of the excerpt. That halfway point is marked by the most obvious Gestalt disruption in the excerpt, the long pause between the first and second halves. So the basic repeating units that one might use to describe what 'Jarabe de Tapatío' consists of are precisely the units that are separated from one another by Gestalt principles. Another important aspect of such 'building blocks' is their tonal and harmonic content.

Although we gain some insight into (8) through the analysis of grouping and motifs, it is impossible to understand how the song is structured without considering its orientation in tonal space. The entire excerpt up until the last two notes emphasises and elaborates the *tonic* chord in the tonal space of the piece, the most central and stable harmony. It does this by outlining each individual pitch-class contained in the tonic chord through a particular melodic device called a *neighbour-tone* motion: a structurally-important pitch-class contained in the tonic chord is sounded, a minimally lower 'neighbour' is sounded, and then the structurally important pitch repeats. This melodic device repeats three times at the beginning of the phrase, and the 10^{th} note is then a repetition of the first (*modulo* octave equivalence) with no neighbouring tone. The second half approaches the same tones of the tonic chord from below via step motion up until the crucial last two notes, which are not in the tonic chord but are in its *dominant* neighbour. At this point, the harmony underlying this excerpt changes to dominant rather than tonic (this could be inferred from listening to the melody alone, but in most performances it is also explicitly played by accompanying instruments).

Generative harmonic theories (Johnson-Laird, 1991; Katz, 2017; Rohrmeier, 2011 *et seq.*; Steedman, 1984 *et seq.*) make a strong case that these kinds of tonal structures empirically pick out the difference between occurring and non-occurring pieces of music in Western traditions including jazz, pop, and 'classical' music. Lerdahl and Jackendoff (1983) claim that this type of analysis is crucial to the interpretation of Western tonal music and that harmonic domains bear a law-governed relationship to grouping much like that illustrated here (but with some additional complexity). The entire initial phrase of 'Jarabe Tapatío', whose second half is not shown in (8), forms a type of harmonic and rhythmic structure referred to as an *antecedent-consequent period*. This structure has a name because it occurs over and over again throughout Western tonal art music, popular music, and traditional folk and children's song. The crucial tonal 'ingredients' used to build a period are cleanly separated by Gestalt principles in this and many other examples. So the interesting fact about Gestalt grouping is not so much that it exists (because we can group any auditory stimulus, musical or not), but that it tends to segment musical surfaces into exactly those units that are crucial to understanding the tonal structure of a piece.

While Western harmony is complex and atypical, the general relationship between rhythm and tonality is robust across human musical cultures. The tendency for passages that emphasise different locations or centres in tonal space to be separated by Gestalt grouping is attested in Japanese Edo-period music (Pasciak, 2017), Arab taqsim (Ayari & McAdams, 2003), Turkish makam (Mungan et al., 2017), and Javanese gamelan (Hughes, 1988), for instance. More generally, the most prominent musical events in metrical, harmonic, and psychoacoustic terms play an outsized role in people's ability to identify (Dalla Bella et al., 2003), perform (Dell & Halle, 2009; Large et al., 1995; Repp, 1992, 1998), and judge the similarity (Bigand, 1990; Dibben, 1994) of musical pieces. Listeners are much better at remembering musical sequences that have hierarchical elaboration structures of tonally-stable events, as in (8), than those that do not (Deutsch, 1980). We model this relative structural importance using *headedness* within musical groups, as in (6). One plausible event hierarchy (also called *Time-Span Reduction*) of the excerpt from 'Jarabe Tapatío' could be represented as in (9)[3]:

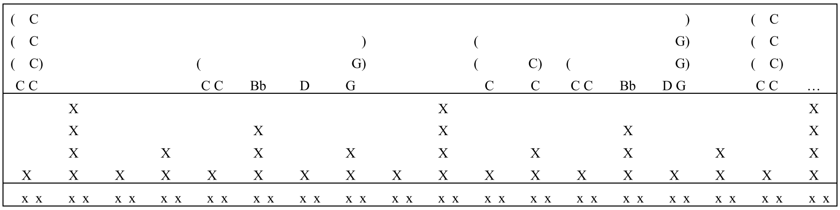
(9) Event (Time-Span) hierarchy of 'Jarabe Tapatío'

(								C											)
(								C	)	(								F	)
(		G			)	(		C	)	(		G			)	(		F	)
(		G)	(		E)	(		C	)	(		G)	(		C)	(		F	)
(G)	(F#)	(G)	(E)	(D#)	(E)	(C)	(B)	(C)	(G)	(E)	(F)	(G)	(A)	(B)	(C)	(D)	(E)	(F)	(D)

This structure captures several important aspects of the excerpt: repeated neighbour-note figures and slurred triplets are constituents; the higher levels of the piece outline a tonic chord then land on a dominant seventh; the largest division in the excerpt is between the two halves separated by a silence; the most structurally important pitch events are those that eventually will be aligned with prominent metrical positions; and the central tone in the entire excerpt is the tonic C, which will eventually be 'prolonged' by another return to tonic at the end of the phrase (not shown in 9).

Event hierarchies also allow us to describe *dissociations* between metrical prominence and event prominence, such as syncopation and appogiatura. Example (10) shows the melodic, metrical, and event structure of the chorus from Michael Jackson's 'Don't Stop 'til you Get Enough'. This melody repeats several times against alternating compound Bb/C and tonic C chords. We focus on the simpler tonic chord here. The most prominent filled metrical positions in the excerpt are occupied by a Bb that is not in the tonic chord, while larger groups end with a more stable G in a weaker metrical position (in some repetitions this pitch is a C instead, which is even more stable). Some of the notes in (10) (e.g. the second C) occur on beats at a lower metrical level, which is boxed off separately and printed in smaller type for reasons of space. The analysis in (10) is one feasible event hierarchy for this melody; other analyses are possible, especially with regard to lower levels in the hierarchy. In any viable analysis, however, the event prominence of the Bb will not match its metrical prominence. Beyond this, the strongest metrical positions in (10) are not filled by *any* event onset. These are arguments for distinguishing between event hierarchies and metrical hierarchy.

(10) Syncopation and appoggiatura: ‘Don’t Stop ‘til you Get Enough’



keep on with the force don’t stop don’t stop til you get e nough keep on…

The event hierarchies in (9–10) include information about the constituency, relative prominence, and tonal structure of musical elements. They do not include metrical structure, and the reason why gets at something important about the nature of pieces of music. As reviewed above, there is evidence that grouping and event prominence play a central role in encoding music in memory, generating performances, and assigning structure to a performance one hears. While most pieces are associated with clear conventions about which structurally-important events occur on which metrical beats, there is also a fair bit of flexibility in this regard for many pieces. Common-practice ‘classical’ music from the Baroque and Classical periods can be performed with un-notated flourishes and alterations. Folk, rock, and children's songs allow for most notes, especially structurally less prominent ones, to be performed in a variety of metrical configurations, omitted, or given new words with different rhythmic properties (Halle & Lerdahl, 1993; Hayes & Kaun, 1996; Kiparsky, 2006). Jazz performances freely change the rhythm, melody, harmony, and even metrical structure of existing song-forms (see any introductory text, e.g. Berliner, 1994, pp. 65–71). While writing this paper, I examined a wide variety of sheet music representing various pop, rock, and soul recordings. I found that syncopation and appoggiatura in such recordings is frequently ‘undone’ when people transcribe them. When performers make up a new version of an existing tune, set new words to an existing tune, compose a new piece with a standard song-form, or compose variations on a theme in the Western classical tradition, they are more likely to retain more structurally important events and less likely to change or omit them (e.g. Halle & Lerdahl, 1993; Katz, 2017; Large et al., 1995; Lerdahl & Jackendoff, 1983). This raises the question of what a piece of music actually *is*, in mental terms.

It is undoubtedly the case that listeners can represent fine details of a musical piece in long-term memory, down to the absolute pitch of a canonical performance (Levitin, 1994), fine-grained differences in whether a musician plays a bit ahead of or behind the metrical beat, or a scratch at a particular location in a record. The fact that we still have no trouble identifying and producing novel instances of a song, even with changes in melody, rhythm, harmony and virtually all other musical parameters, shows that such fine-grained details are not a *necessary* part of the representation of a piece of music. The implicit knowledge of a piece that people draw on when they perform, recognise, or change a piece of music is a hierarchy of more and less essential musical events, where lower levels in the hierarchy may be altered in a variety of ways. That is to say, what a person knows when they know a song is an event hierarchy for that song, along the lines of the ones explored above.

4 | SOME LESSONS FOR MUSIC AND LANGUAGE

From the linguist's perspective, one obvious point of interest here is that the event hierarchies in (9–10) look a lot like *prosodic structure* in language. Musical event hierarchies encode constituency, which is associated with Gestalt grouping principles on one hand and with tonal and motivic interpretation on the other. Prosodic structure encodes constituency, which is associated with Gestalt-relevant sound patterns (Jeon & Nolan, 2013; Katz & Fricke, 2018; Kentner & Féry, 2013) on the one hand and syntactic and information-structural interpretation on the other (e.g. Selkirk, 1984; Steedman, 2000). Musical event *prominence* relates to tonal interpretation and psychoacoustic salience. Phonological prominence above the word relates to information structure and psychoacoustic salience. Structures in both domains can be represented visually as bracketed grids.

The picture that emerges is that while phrasing and event hierarchies are broadly similar in music and language, musical metre has no clear linguistic analogue.[4] Musical metre consists of abstract positions in time that are usually isochronous, iterate identically ad infinitum, and need not be filled by any occurring event. None of these properties inhere to phonological prominence. The point is important because several prominent reviews of music and language have proposed a deep cognitive parallel between phonological stress and musical metre (e.g. Jackendoff, 2009; Jackendoff & Lerdahl, 2006; Temperley, 2022). This is generally based on the fact that both can be notated as grid structures and that there is a strong correlation between linguistic stress and musical metre when words are set to a piece of music. The inevitable conclusion given this premise is that while both systems are represented in a grid structure, the nature of the grids in the two domains is quite different.

My claim here is that the analogy between musical metre and linguistic stress is a category error: it is musical *event* hierarchies that fulfil the function in music that prosodic structure does in language. It is true that musical metrical structure can be represented as a grid, but event hierarchies must be represented separately from metre, and these hierarchies are similar to linguistic prosody in both their formal details and their substantive interpretation; musical metre is not. The correlation between linguistic stress and musical metre is an indirect consequence of the correlation between *event prominence* and musical metre, which is independently attested in non-vocal music (discussed in Section 3). Where metrical structure and event hierarchies dissociate, linguistic constituent edges are generally aligned with rhythmic constituent edges and not putative metrical ones (as in example 7 above), and linguistic stress or weight corresponds to event prominence rather than metrical prominence. Where linguistic texts *do* dissociate from rhythmic groups, they are not particularly likely to align with any putative metrical 'constituent'. These tendencies can be observed in complex textsetting in many Western genres (see examples in Dell, 2015; Duinker, 2021; Heetderks, 2022; Murphy, 2016; and Temperley, 1999), and they mirror the properties of rhythm-metre mismatch in non-vocal music.

Ontologically, musical metre and linguistic stress are very different kinds of systems. Musical metre repeats strictly and ad infinitum at multiple levels. Once a metre is established, it generally repeats independently of how musical events are phrased against it (though in rare cases metre can change internal to a piece). Stress systems, on the other hand, don't even exist in many languages, and when they do, they do not iterate endlessly. Even in languages with highly-regular alternating stress, the domains over which patterns do or do not repeat are determined by morpho-syntactic or prosodic constituency (see Baker, 2014 for an overview of Australian languages). One cannot compute stress in any language without knowing the constituent structure of that language. This informational dependency is similar to musical event prominence, which depends on information about musical grouping, but is different from musical metre.

One obvious question that arises at this point is *why* music has separate metrical and rhythmic components while language lacks the abstract 'timekeeper' component. We do not know the answer. It is easy to come up with possible explanations, however, perhaps so easy that we should be sceptical any of them are correct. One possibility is that music is more likely to be performed in tandem by groups of people: many cultures have some form of group chant, singing, or instrumental music, while 'normal' speech rarely involves multiple speakers saying the same thing at the same time. A rigidly isochronous temporal scaffold should help multiple agents coordinate their vocal or motor actions to be simultaneous in a way that is not necessary for speaking. That said, this is highly speculative and I know of no direct evidence bearing on the issue.

5 | CONCLUSION

Music includes a number of different representational elements that vary in details across musical cultures, but in constrained ways. This paper has introduced metrical structure, grouping, and event hierarchies (Time-Span Reduction). Metrical structure is most often rigidly iterative, approximately isochronous, and partially independent of occurring auditory events. Phrasing and event hierarchies represent the constituency and prominence of occurring musical events; they do not necessarily iterate and are not necessarily isochronous. In the default case, both the constituency and prominence of occurring musical events correlate with metrical properties, but there are also clear mismatches between the two structures in many genres. Musical metre has no clear linguistic analogue, and is notably different from linguistic stress in both conceptual and empirical terms. Musical grouping and event hierarchies, on the other hand, bear a strong formal and substantive resemblance to prosodic structure in language, and it is these structures to which linguistic texts are matched in vocal music, with the same consequences for metrical alignment observed in non-vocal musical genres.

ACKNOWLEDGEMENTS

Many of the ideas in this work initially emerged from an interdisciplinary seminar I ran at the Institut Jean-Nicod in Paris from 2010 to 2012. This work would not have been possible without the freedom afforded by my post-doctoral position there, which was made possible by Philippe Schlenker and the European Science Foundation. In addition to Philippe, my French colleagues Emmanuel Chemla, François Dell, Christophe Pallier, and Benjamin Spector greatly influenced my understanding of this material through detailed and critical discussion during that time and continuing afterwards. Merci à tous.

ORCID

Jonah Katz https://orcid.org/0000-0002-1057-7052

ENDNOTES

[1] Some theorists use the term 'beat' to single out a single prominent level of metrical structure, mainly for a convenient way of referring to that level. I instead refer to *any* metrical position at any level as a beat.

[2] The terms *syncopation*, *appoggiatura*, and *hemiola* don't have a single rigid or precise definition; some musicians or musicologists may use them to mean something more general or specific than the definitions here. In particular, 'syncopation' is sometimes defined in terms of 'beats' and 'off-beats' a distinction that I've avoided in this paper for reasons of both comprehension and empirical inadequacy. 'Appoggiatura' is bound up in classical music with the notation system and specific labels for metrically strong non-chord tones; my definition is more general, but captures the usual meaning of the word reasonably well. 'Hemiola' is used in classical music mainly for one specific case of the kind of phase mismatch described here, so it may be helpful for knowledgeable readers to think of my definition as 'generalised hemiola'; my impression is that this generalised usage is common in jazz.

[3] The term *Time-Span Reduction* comes from Lerdahl and Jackendoff's (1983) theory, and the term *reduction* more generally is a central part of traditional Schenkerian music theory. The idea is that the analyst recursively *reduces* out surface events that are less structurally important and in this way arrives at the structural core of a piece.

[4] This statement is true of *almost* all linguistic theories, but Hyde (2016) proposes a 'hidden' metrical structure for linguistic stress based on that of music. The idea, while intriguing, hasn't been the subject of much empirical investigation as far as I am aware.

REFERENCES

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (Taqsim). *Music Perception*, *21*(2), 159–216. https://doi.org/10.1525/mp.2003.21.2.159

Baker, B. (2014). Word structure in Australian languages. In H. Koch & R. Nordlinger (Eds.), *The languages and linguistics of Australia: A comprehensive guide* (pp. 139–213). De Gruyter.

Berliner, P. F. (1994). *Thinking in jazz*. University of Chicago Press.

Bigand, E. (1990). Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music*, *18*(1), 45–59. https://doi.org/10.1177/0305735690181004

Blacking, J. (1970). Tonal organization in the music of two Venda initiation schools. *Ethnomusicology*, *14*(1), 1–56. https://doi.org/10.2307/850292

Bolger, D., Coull, J., & Schön, D. (2014). Metrical rhythm implicitly orients attention in time as indexed by improved target detection and left inferior parietal activation. *Journal of Cognitive Neuroscience*, *26*(3), 593–605. https://doi.org/10.1162/jocn_a_00511

Bolger, D., Trost, W., & Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychologica*, *142*(2), 238–244. https://doi.org/10.1016/j.actpsy.2012.11.012

Cambouropoulos, E. (2001). The local boundary detection model and its application in the study of expressive timing. In *Proceedings of the international computer music conference* (pp. 17–22). ICMA.

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, *65*(7), 1019–1028. https://doi.org/10.3758/bf03194831

Dawe, L., Platt, J., & Racine, R. (1995). Rhythm perception and differences in accent weights for musicians and nonmusicians. *Perception & Psychophysics*, *57*(6), 905–914. https://doi.org/10.3758/bf03206805

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception*, *4*, 325–360. https://doi.org/10.2307/40285378

Dell, F. (2015). Text-to-tune alignment and lineation in traditional French songs. In T. Proto, P. Canettieri, & G. Valenti (Eds.), *Text and tune* (pp. 183–234). Peter Lang.

Dell, F., & Elmedlaoui, M. (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Kluwer.

Dell, F., & Halle, J. (2009). Comparing musical textsetting in French and in English songs. In J.-L. Aruoi & A. Arleo (Eds.), *Towards a typology of poetic forms* (pp. 63–78). John Benjamins.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*(5), 381–389. https://doi.org/10.3758/bf03204881

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Perception*, *12*, 1–25. https://doi.org/10.2307/40285753

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, *14*(1), 37–40. https://doi.org/10.3758/bf03198614

Duinker, B. (2021). Segmentation, phrasing, and meter in hip-hop music. *Music Theory Spectrum*, *43*(2), 221–245. https://doi.org/10.1093/mts/mtab003

Ekwueme, L. E. N. (1975). Structural levels of rhythm and form in African music: With particular reference to the West Coast. *African Music*, *5*(4), 27–35. https://doi.org/10.21504/amj.v5i4.1615

Halle, J., & Lerdahl, F. (1993). A generative textsetting model. *Current Musicology*, *55*, 3–23.

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, *16*(1), 48–55. https://doi.org/10.1111/j.0956-7976.2005.00779.x

Hayes, B., & Kaun, A. (1996). The role of phonological phrasing in sung and chanted verse. *The Linguistic Review*, *13*(3–4), 243–303. https://doi.org/10.1515/tlir.1996.13.3-4.243

Hayes, B., & Schuh, R. G. (2019). Metrical structure and sung rhythm of the Hausa rajaz. *Language*, *95*(2), 253–299. https://doi.org/10.1353/lan.2019.0043

Heetderks, D. (2022). Norms of textual scansion and rhyme in Beatles AABA forms. *Music Theory Spectrum*, *44*(1), 41–62. https://doi.org/10.1093/mts/mtab018

Hughes, D. W. (1988). Deep structure and surface structure in Javanese music: A grammar of Genghing Lampah. *Ethnomusicology*, *32*(1), 23–74. https://doi.org/10.2307/852225

Huron, D. (1991). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception*, *9*(2), 135–154. https://doi.org/10.2307/40285526

Hyde, B. (2016). The role of phenomenal accent. In J. Heinz, R. Goedemans, & H. van der Hulst (Eds.), *Dimensions of phonological stress* (pp. 49–78). Cambridge University Press.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception*, *26*(3), 195–204. https://doi.org/10.1525/mp.2009.26.3.195

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*(1), 33–72. https://doi.org/10.1016/j.cognition.2005.11.005

Jeon, H.-S., & Nolan, F. (2013). The role of pitch and timing cues in the perception of phrasal grouping in Seoul Korean. *Journal of the Acoustical Society of America*, *133*(5), 3039–3049. https://doi.org/10.1121/1.4798663

Johnson-Laird, P. (1991). Jazz improvisation: A theory at the computational level. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 291–326). Academic Press.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355. https://doi.org/10.1037/0033-295x.83.5.323

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491. https://doi.org/10.1037/0033-295x.96.3.459

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*(4), 313–319. https://doi.org/10.1111/1467-9280.00458

Jongsma, M. L. A., Desain, P., & Honing, H. (2004). Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. *Biological Psychology*, *66*(2), 129–152. https://doi.org/10.1016/j.biopsycho.2003.10.002

Kaliakatsos-Papakostas, M., Katsiavalos, A., Tsougras, C., & Cambouropoulos, E. (2014). Harmony in the polyphonic songs of Epirus: Representation, statistical analysis and generation. In A. Holzapfel (Ed.), *Proceedings of the 4th international workshop on folk music analysis* (pp. 21–28). Bogazici University.

Katz, J. (2017). Harmonic syntax of the 12-bar blues: A corpus study. *Music Perception*, *35*(2), 165–192. https://doi.org/10.1525/mp.2017.35.2.165

Katz, J. (2022). Musical grouping as prosodic implementation. *Linguistics and Philosophy*. https://ling.auf.net/lingbuzz/003938

Katz, J., Chemla, E., & Pallier, C. (2015). An attentional effect of musical metrical structure. *PLoS One*, *10*(11), e0140895. https://doi.org/10.1371/journal.pone.0140895

Katz, J., & Fricke, M. (2018). Auditory disruption improves word segmentation: A functional basis for lenition phenomena. *Glossa*, *3*(1), 38. https://doi.org/10.5334/gjgl.443

Kentner, G., & Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review*, *30*(2), 277–311. https://doi.org/10.1515/tlr-2013-0009

Kiparsky, P. (2006). A modular metric for folk verse. In B. Elan Dresher & N. Friedberg (Eds.), *Formal approaches to poetry: Recent developments in metrics*. Mouton de Gruyter.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.

Large, E., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*, 1–37. https://doi.org/10.1207/s15516709cog2601_1

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*(2–3), 177–208. https://doi.org/10.1080/09540099408915723

Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced memory representations for music. *Cognitive Science*, *19*(1), 53–96. https://doi.org/10.1207/s15516709cog1901_2

Lerdahl, F. (2001). *Tonal pitch space*. Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. https://doi.org/10.3758/bf03206733

Liberman, M. (1975). *The intonational system of English*. PhD dissertation. MIT.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.

McPherson, L., & Ryan, K. M. (2018). Tone-tune association in Tomma So (Dogon) folk songs. *Language*, *94*(1), 119–156. https://doi.org/10.1353/lan.2018.0003

Mungan, E., Yazici, Z. F., & Kaya, M. (2017). Perceiving boundaries in unfamiliar Turkish Makam music: Evidence for Gestalt universals? *Music Perception*, *34*(3), 267–290. https://doi.org/10.1525/mp.2017.34.3.267

Murphy, N. E. (2016). *"The times they are a-changin'": Flexible meter and text expression in 1960s and 70s singer-songwriter music*. PhD dissertation. University of British Columbia.

Nan, Y., Knösche, T. R., & Friederici, A. D. (2009). Non-musicians' perception of phrase boundaries in music: A cross-cultural ERP study. *Biological Psychology*, *82*(1), 70–81. https://doi.org/10.1016/j.biopsycho.2009.06.002

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realisation model*. University of Chicago Press.

Palmer, C., & Krumhansl, C. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728–741. https://doi.org/10.1037/0096-1523.16.4.728

Pasciak, K. J. (2017). *A transformational approach to Japanese traditional music of the Edo period*. PhD Dissertation. UMass Amherst.

Patel, A. (2008). *Music, language, and the brain*. Oxford University Press.

Penel, A., & Jones, M. R. (2005). Speeded detection of a tone embedded in a quasi-isochronous sequence: Effects of a task-irrelevant temporal irregularity. *Music Perception*, *22*(3), 371–388. https://doi.org/10.1525/mp.2005.22.3.371

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology*, *24*(1–5), 157–178. https://doi.org/10.1080/00207594.1989.10600040

Popescu, T., Widdess, R., & Rohrmeier, M. (2021). Western listeners detect boundary hierarchy in Indian music: A segmentation study. *Scientific Reports*, *11*(1), 3112. https://doi.org/10.1038/s41598-021-82629-y

Povel, D., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440. https://doi.org/10.2307/40285311

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's 'Träumerei'. *Journal of the Acoustical Society of America*, *92*(5), 2546–2568. https://doi.org/10.1121/1.404425

Repp, B. H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, *104*(2), 1085–1100. https://doi.org/10.1121/1.423325

Repp, B. H. (2010). Do metrical accents create illusory phenomenal accents? *Attention, Perception, & Psychophysics*, *72*(5), 1390–1403. https://doi.org/10.3758/app.72.5.1390

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*(1), 35–53. https://doi.org/10.1080/17459737.2011.573676

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press.

Steedman, M. (1984). A generative grammar for jazz chord sequences. *Music Perception*, *2*(1), 52–77. https://doi.org/10.2307/40285282

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*(4), 649–689. https://doi.org/10.1162/002438900554505

Stock, J. (1993). The application of Schenkerian analysis to ethnomusicology: Problems and possibilities. *Music Analysis*, *12*(2), 215–240. https://doi.org/10.2307/854273

Tan, I., Lustig, E., & Temperley, D. (2019). Anticipatory syncopation in rock: A corpus study. *Music Perception*, *36*(4), 353–370. https://doi.org/10.1525/mp.2019.36.4.353

Temperley, D. (1999). Syncopation in rock: A perceptual perspective. *Popular Music*, *18*(1), 19–40. https://doi.org/10.1017/s0261143000008710

Temperley, D. (2022). Music and language. *Annual Review of Linguistics*, *8*(1), 153–170. https://doi.org/10.1146/annurev-linguistics-031220-121126

Turpin, M. (2007). The poetics of Central Australian song. *Australian Aboriginal Studies*, *2*(2), 100–115.

Turpin, M. (2017). Parallelism in Arandic song-poetry. *Oral Tradition*, *31*(2), 535–560. https://doi.org/10.1353/ort.2017.0020

Waller, A. (2016). *Rhythm and flow in hip-hop music: A corpus study*. PhD dissertation. University of Rochester.

Wertheimer, M. (1938). Laws of organization in perceptual forms [English translation of 1923 essay]. In W. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). Routledge.

AUTHOR BIOGRAPHY

Jonah Katz received his PhD from the Massachusetts Institute of Technology in 2010, and is currently an Associate Professor of Linguistics at West Virginia University. His linguistic research explores the relationship between acoustic phonetics, speech perception, and cross-linguistic phonological tendencies. He completed Bachelor's degrees in linguistics and music (jazz piano performance) at the University of Massachusetts, and has developed a secondary research focus on the structure and cognition of music and its relationship to language. He has published research on musical metre, grouping, harmony, and the relationships amongst them.

How to cite this article: Katz, J. (2022). Metre, grouping, and event hierarchies in music: A tutorial for linguists. *Language & Linguistics Compass*, e12472. https://doi.org/10.1111/lnc3.12472