NORTHWESTERN UNIVERSITY

The Pocket
A Theory of Beats as Domains

A DISSERTATON

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Music

By

Fred Hosken

EVANSTON, ILLINOIS

September 2021

Abstract

This dissertation proposes a theory that views beats as probabilistic domains that I term "pockets," taking a vernacular term commonly used by jazz, funk, and popular music performers to describe the state of being in a good groove and making it concrete through empirical methods. Pockets have three key properties: they are domains of time, whose membership is probabilistic, and the specific shape of the probabilistic distribution is associated the qualitative experience of a performance—its "feel." The theory of pockets that I advance can be utilized to provide new perspectives on rhythmic structures in music, as well as new approaches to understanding patterned microtiming in performance, the phenomenon of musical "groove," and musical "feel" more generally. Pockets are a complement to Danielsen's theory of perceptual "beat bins" (2010a), theorizing extended beats *in the sound signal* of performances. By importing the term "pocket," the technicalities of the shaping of music time can be made meaningful by drawing on real-world qualitative descriptors of "feel"—a pocket may be "loose" or "tight," a performer can "lay back," be "on top of the beat," or "push."

I first survey, in Chapter 1, the literature, drawing together the established theories of beats in music and cognitive research and arguing that there is a need for a theory of beats that prioritizes the sound signal produced by performance events. Chapter 2 introduces the theory of pockets, justifies why the label "pocket" is so important (rather than simply "domain"), and establishes the existence of pockets with a music information retrieval analysis of three corpora of drum performances. Chapter 3 considers what these pockets mean for the listener experience, introducing descriptive terms that might be used to make sense of timing variations between performances. Chapter 4 presents an empirical study that explores whether listeners are able to perceive subtle timing differences. And finally, Chapter 5 considers how differently shaped pockets interact with form in popular music, analyzing Stevie Wonder's "Superstition" to argue that musicians manipulate the shape, size, and location of the pocket across a song to enhance

the experiential qualia of different sections.

Acknowledgements

This dissertation is the capstone to five years of immense growth as a result of the mentorship and companionship of a multitude of people who have inspired, challenged, and helped me to thrive. First and foremost is my advisor, Ric Ashley, whose steady hand guided this project with a light and trusting touch that enabled me to discover my own identity while navigating new terrain safe in the knowledge that I shan't wander too far astray, and whose broad *and* deep knowledge of so many fields motivated a constant "big picture" view. On top of this, I am grateful for his sincere care and concern for all of our wellbeing, captured in his reassuring mantra of "All shall be well, and all shall be well, and all manner of thing shall be well" (Julian of Norwich). I am also thankful for the wisdom and guidance of my two other committee members: Anne Danielsen and Danuta Mirka. Anne Danielsen's groundbreaking research, particularly the theory of "beat bins," was the genesis of much of this project and I am beyond thankful for the opportunity to have musically and intellectually inspiring discussions with her. The present project would not have been possible without her. I am also deeply appreciative of the guidance and expertise that Danuta Mirka has leant this dissertation. Her theoretical and terminological precision is unequalled and I have learned a great deal about situating my research in the broad music-theoretic landscape. Additionally, Mark Butler played a significant role in laying the foundations of this project for which I am immensely grateful.

I have been blessed at Northwestern with a dream cohort. The students in music theory and cognition, across all year groups, have provided the most warm, supportive, and fun group one could hope for. My own year group, Aubrey Leaman and Cella Westray, have been fantastic partners throughout our time at Northwestern. Those in the years above me (Anjni Amin, Sarah Gates, Stephen Hudson, Miriam Piilonen, Rosa Abrahams, Bruno Alcalde, and Olga Sánchez-Kisielewska) led the way for us all, and shared so much experience and advice that has lifted the whole community up higher. Stefan Greenfield-Casas, Lena Console, Morgan Patrick, Sara

Bowden, Emily Schwitzgebel, and Michael Slattery have similarly been incredible allies who have taught me a great deal while also bringing light into the world (essential during a global pandemic!). I must thank the community around office 4-151, particularly Adam White and Anjni, who made going into work a joy and who held me together through the rollercoaster of graduate school. Outside of the music school, I am forever indebted to Nikita Salovich, Weston Twardowski, and Apoorva Shivaram who not only provided methodological, statistical, and philosophical advice, they also filled my time in Chicago/Evanston with love. They are a true second family.

Finally, I am immeasurably thankful for my parents, brother, and sister (-in-law) who supported me as I crossed an ocean and pursued my goals. Although I am nearly 4,000 miles from home, they have made the distance feel so much less. I would not have my passion for music, my drive to learn and explore, and anything like enough bravery to throw myself into the unknown without the life that they have all provided me.

Contents

List of Figures

List of Tables

Introduction

Contemporary R&B and Soul singer José James's song "U R the 1" from the album *While You Were Sleeping* (2014) takes familiar tropes—backbeat drum groove, arpeggiated bass, chord pads—and stretches them to near breaking point: The chords lunge forward, surging in dynamic up from nothing while the bass and drums seem untethered to any sense of regular musical time. The music seems to be perpetually on the cusp of falling apart. And yet it doesn't fall apart. Within its woozy atmosphere, it is possible to grab hold of some sense of certainty, and it is possible—to some extent—to predict what may come next. Despite temporal regularity being obscured in a fog, one can entrain and move along with the music. These movements may well be softer and less sharply defined than those used when entraining to other music, perhaps nodding slowly up and down rather than an ictic finger snap, but they are still there.

This fairly extreme example by José James poses significant questions about how performers construct musical time, how listeners hear musical time, and how the shaping of musical time impacts our personal aesthetic and embodied responses to a performance. Time in this dissertation is not defined as chronometric, absolute, neutral time, but instead is a phenomenological understanding of time. Musical time is of course informed by time on a clock, and this dissertation presents numerous empirical measurements of time, though the ultimate argument is how these objective measurements intersect with the lived experience of these *musical* durations. In this way, I follow Kozak (2020) and his definitions of "enacted time" (pp. 11-12), framing musical time in this dissertation as subjective, embodied, interactive, dynamic, emergent, affective, culturally and situationally informed, and experiential.

Returning to "U R the 1," an analysis of the timing details presents some phenomena that call into question whether our current canonic theories of musical time can meaningfully communicate what is going on. Looking at Figure 1, which illustrates the onset timings for the drum part of the track, it is evident that the archetypal "boom tish" backbeat drum pattern has been stretched and compressed both between metric positions (that is, the duration in seconds between beat 1 and 2, 2 and 3, etc. varies) and within divisions of a beat. The elements of the standard drum groove are all there, but the intervals between each bass drum onset are neither consistent (with the first eighth note of a pair ranging from 51% up to 58% of a beat), nor do they divide the main beats of the bar in a way that may be labelled ("straight," "dotted," "triplet swing," "quintuplet/septuplet swing," etc.). This elastic fluidity works against strict categorization—what could a prototype of the bass drum rhythm be?—yet we are still able to recognize the familiar, generic backbeat drum pattern and, despite the inconsistencies and lack of clear divisions, infer a temporal reference structure that scaffolds the track and our physical engagement with it. The boundaries of the categories a listener might use to perceptually organize the performance are fuzzy and this fuzziness is essential to the experience of the track.

Figure 2 presents an overly simplistic transcription in staff notation of the opening of "U R the 1." Here, all nuances of the performed timing have been quantized away, snapping all events that push and pull at time to a strict metric grid where everything aligns, is evenly spaced, and adheres to binary divisions.[1] The music this grid-locked transcription appears to represent would

[1]Straight eighth notes were selected as the most appropriate way to notate this as, though it might be possible to

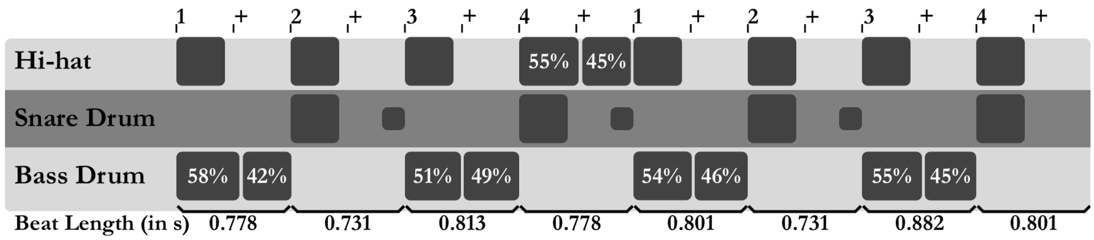


FIGURE 1: Graphic transcription and annotation of the timing details of the opening measures of the drums in José James's "U R the 1."

be flat, losing all that marks José James and his band's performance, and numbing any aesthetic or emotive responses. There is a parallel concern in the visual arts that grids present an abstraction from the "real." In a seminal essay, art historian Rosalind Krauss writes in response to Cubism and subsequent modernist art styles:

> [T]he grid states the autonomy of the realm of art. Flattened, geometricized, ordered, it is antinatural, antimimetic, antireal. It is what art looks like when it turns its back on nature. In the flatness that results from its coordinates, the grid is the means of crowding out the dimensions of the real and replacing them with the lateral spread of a single surface. In the overall regularity of its organization, it is the result not of imitation, but of aesthetic decree. Insofar as its order is that of pure relationship, the grid is a way of abrogating the claims of natural objects to have an order particular to themselves; the relationships in the aesthetic field are shown by the grid to be in a world apart and, with respect to natural objects, to be both prior and final. The grid declares the space of art to be at once autonomous and autotelic.
>
> (Krauss, 1979, pp. 50-2)

While Figure 2 is an intentional exaggeration that amplifies the antinatural, antimimetic, and antireal nature of symbolic transcription of rhythm, it illustrates the gap between what is happening in the performance—the surface-level nuances that bring life to a real-world, natural performance—and the more abstract, higher-level ways in which rhythmic phenomena tend to be understood in music-analytic terms.

use an elaborate set of tuplets—e.g., the first pair of bass drum onsets are close to a 3:2 ratio so could be approximated

Figure 2: Transcription of the opening measures of José James's "U R the 1."

Erykah Badu's "Fall In Love (Your Funeral)," from the album *New Amerykah Part Two (Return of the Ankh)* (2010), provides another opportunity to consider the subtleties of how performers and producers can impact the experience of a track through the way in which musical time is shaded. This track once again suggests limitations to current ways of understanding musical time. In "Fall In Love," Erykah Badu and programmer/producer Karriem Riggins take an existing sample—Eddie Kendrick's "Intimate Friends" (1977)—and add a synthesizer that chirps out eighth notes throughout. This additional layer, however, sounds out of kilter. It doesn't coincide with the rest of the instruments in the track apart from the hi-hat. In fact, upon closer inspection, these two layers are offset by 23% of a beat from the other instruments (Figure 3). Even with this sizable asynchrony, when listening to the track, the synthesizer chirps are somehow associated—at least to my enculturated ear—with the main beats of the bar. They are not perceived as exactly

with quintuplets, the third pair are roughly 11:9, which might be rounded to 4:3 and so notated using septuplets—this over-notation would give a false sense of fidelity to the original.

synchronous with the other events, but the difference is still enfolded within a broad sense of "beat." Why does this synth line not sound like it is in fact a distinct event that happens a sixteenth note after the main beats of the bar, functioning as a sort of displacement dissonance (Krebs, 1999)? The incessant asynchronous chirps transform Eddie Kendrick's original soul song into something new with a distinct experiential "flavor,"[2] yet they are not "wrong" or "out of time." "Fall In Love," therefore, suggests that there is a great elasticity in our abilities to perceive musical time in a holistic way, enfolding several potentially conflicting elements within a large category with generously and permeably defined inclusion criteria that results in a sense of musical time that has qualitative significance.[3]

The examples provided so far have drawn from "neo-soul" music, a genre that grew out of Detroit and was most dominant in the 1990s and early 2000s that is renowned for intentional, overt play with musical time. The figurehead of neo-soul production, J Dilla, was known for his anti-quantized style, and this can be heard to various degrees in numerous records produced by J Dilla and associates (see the recorded output of the music production collectives The Ummah and the Soulquarians). Expansions and contractions in the articulation of musical time that have significant ramifications for the qualitative experience of the track are by no means unique

---

[2]This "flavor," its qualitative and subjective perception, is even more noticeable when comparing Erykah Badu's usage of the Eddie Kendrick sample with Alicia Keys's use of the same sample in her song "Unbreakable" (2005). Alicia Keys does not have the off-kilter synth layer and so her track does not have the same ambience, feel, or character as Erykah Badu's.

[3]One song that presents a nice pairing with "Fall In Love" is Leela James's rendition of "I'd Rather Go Blind" (2012). This R&B cover of the Etta James's (no relation) 12/8 classic also has constant eighth-note chords, but, in this track's case, the eighths hold steady while the very prominent bass guitar arrives after the rest of the band and blurs the specific location of the onset further by sliding up to each pitch. The pairing of "Fall In Love" and "I'd Rather Go Blind" shows the instrumental and metric flexibility in creating these specific senses of time—a treble, harmonic instrument can act to transform the sense of time just as effectively as a bass instrument, plus tracks in both simple or compound time signatures can be imbued with time feels. Rounding out this group of tracks, alternative R&B group Moonchild's "The Truth" (2014) features regular eighth notes on a synth that fall behind the others, and the impact of this delay is compounded by a backbeat on the drums that *also* stretches the musical time. This track suggests that these asynchronies do not rely on having a mostly well-disciplined ensemble with just one or perhaps two layers artistically working against the metronome—somehow, even when all musical threads are being pulled in all directions, we are able to do far more than merely latch on to the performed musical time, we can also experience this shaping of musical time aesthetically and expressively.

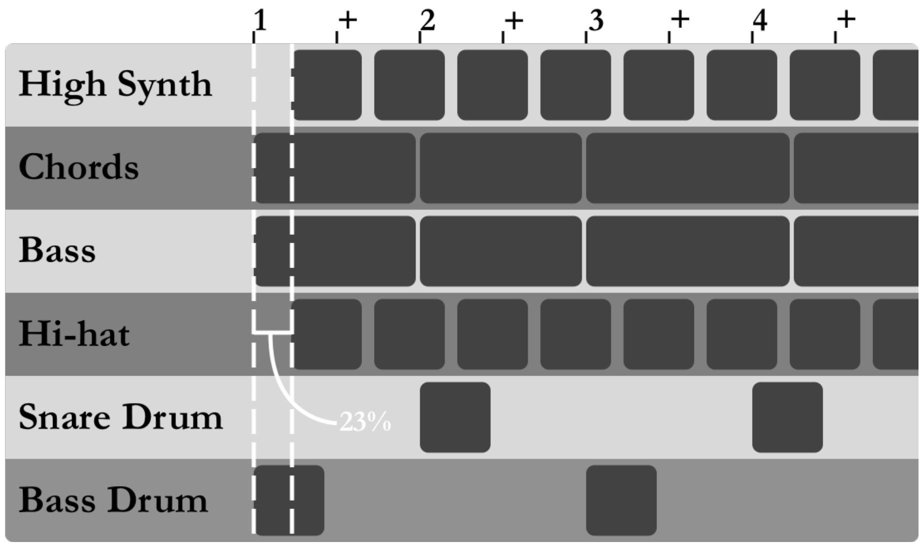


Figure 3: Graphic transcription and annotation of the timing details of a typical measure of Erykah Badu's "Fall In Love (Your Funeral)."

or exclusive to neo-soul and neo-soul inspired music,[4] however, and this dissertation is not constrained to any one musical genre.

Singer-songwriter Jesca Hoop, who combines a diverse and eclectic set of genres in her music (though is far from neo-soul), provides the final song that will be analyzed in this Introduction. Her song "Memories Are Now" (2017) takes minimal materials—electric bass playing eighths, metallic sizzles on the backbeats, and a sparse vocal line that has been overdubbed several times with heavy reverb—and creates a haunting atmosphere that is always in flux despite being constructed by repeating simple musical ideas. Just like in the José James example, there is a significant gap between what may be represented on a page (Figure 4—again, a highly simplified transcription) and the lived experience of the music. Jesca Hoop takes a simple and familiar way of delineating musical time—a bass line of repeated notes—and brings it close to, but somehow not beyond, the elastic limit.[5] Taking some measurements from the bass line in the first 12 bars of the song, the

[4]For just a few examples of neo-soul influences in other genres, see English pop rock band The 1975's inflections on "Sincerity is Scary" (2018), Australian jazz/funk group Hiatus Kaiyote's "Swamp Thing" (2015) or "Shaolin Monk Motherfunk" (from 4:10, 2015), American hip hop/R&B musician Anderson .Paak's "Heart Don't Stand a Chance" (2016), and just about any track within the lo-fi hip hop genre.

[5]"Elastic limit," here, is a metaphorical application of a fundamental concept from physics: You can stretch any

inter-beat interval (the time span between beat 1 and 2, 2 and 3, etc.) ranges from 0.673 seconds up to 0.790 seconds (a 17% increase), showing a very loose sense of tempo. The subdivisions of these quarter notes, as in "U R the 1," escape simple categorization, ranging from a 60:40 swing to a near equal division of the beat in bar 5 into bar 6. How are listeners—those enculturated to these ways of manipulating the flow of musical time and those less so—able to parse this temporal variability and how might we take aesthetic pleasure (and/or discomfort!) from an unstable musical surface?

FIGURE 4: Transcription of the opening measures of Jesca Hoop's "Memories Are Now."

This dissertation focuses mainly on "groove-based" music (i.e., predominantly American musical styles born out of the African diaspora that have developed since the 1970s that make use of short, repeated cells and which foreground rhythmic elements), though I strongly believe

material an amount and it will always return back to its original shape and size unless you stretch it too far—beyond the elastic limit—at which point it deforms and will not return. Jesca Hoop toys with regularity and normative expectations about musical time, stretching, compressing, and bending the temporal "Slinky" spring, but she does not break the Slinky by over-stretching it. Its form and function persist. For a similar description of limits to how far a performer may stretch time, see Danielsen's note, in the context of "beat bins" (explored and explained in great detail in Chapter 1 of this dissertation), that "even a big bin has a rim" (Danielsen, 2018, p. 187).

that many of the ideas explored throughout the chapters may substantively contribute to most musical styles from across the globe. Looking at some expressive timing research, they may, for instance, shed new light on the long history of analyses of between-hand asynchronies in piano performances. These were one of the earliest types of empirical performance analyses and involved investigating systematic asynchronies between pianists' hands (Vernon, 1936; see also C. E. Seashore, 1938, pp. 248–53). This discrepancy between hands and/or musical voices within a texture was (and, perhaps to a lesser extent nowadays, is) a notable and widespread performance technique utilized to affective means. One striking example is Alfred Cortot's performance style heard in recordings from the 1920s, such as his recording of Chopin's Prelude Op. 28, No. 4 in E minor (Cortot, 1926), where the right hand dances around the unbending, consistent left.[6] Repp has shown in analyses of modern performances of three stylistically rather different piano pieces (Schumann's "Träumerei," Debussy's "La fille aux cheveux de lin," and Chopin's Prelude in D-flat major) that there are pervasive between- and within-hand asynchronies where the lead/melody voice tends to lead lower-pitched notes that he argues are deliberate expressive strategies (Repp, 1996).[7] Similarly, in analyses of classical ensemble performances, meaningful and functional differences in beat locations have been found, such as Rasch's observation that lead players in classical ensembles (e.g. a trio) often lead by 30-50 ms (Rasch, 1979). And, lastly, in jazz performances (both solo drummers and ensembles), alterations to the articulation of musical time are prevalent and meaningful, such as the observation that jazz ensembles are described as "tight" and "interlocked" because of, not in spite of, their drummer's hi-hat onsets coming 2-26 ms ahead of the pulse of the music (Hofmann et al., 2017). This same analysis found that certain beats of the bar (the backbeats) have a higher degree of temporal precision (i.e. less variance between players' onsets) than beats 1 and 3. Butterfield (2006) has also described how different degrees

[6] This is but one instance of *rubato* performance in the earlier, original sense (R. Hudson, 1994). For a vivid collection of quotes by Chopin's peers, students, and critics about his approach to performance and rubato see Eigeldinger (1986, 49ff.) and Chapter 7 of R. Hudson (1994).

[7] Also see 20-50 ms "melody leads" in Palmer, 1996, replicated in Goebl, 2001, and extended in a large corpus of Chopin (Goebl et al., 2010) that also highlighted "bass anticipations".

of variance in beat location, particularly an anticipatory cymbal, engender different affective experiences, specifically enhancing the sense of pointing forward. Each of these examples from the field of expressive timing research have been explored from the mindset of finding *differences between* events rather than viewing them, as I argue for, as part of a larger single event that has dimensions to it—defining a particular moment as an origin and measuring events' temporal distance from this point rather than drawing a field around all the events of interest and analyzing the topography of this field.

The analyses and perceptual experiments presented in this dissertation elucidate a theory of *beats as domains* whose shape, size, and location can have significant and meaningful implications for the experiences of performers and listeners. In arguing for this perceptual and performed understanding of musical beats, I problematize the concept of *beats as instantaneous time points* that underlies many contemporary ways of understanding rhythm and meter in music. All of the musical examples that have been provided so far explore various scenarios in which the beats that the performances are founded upon conflict with the view that beats are instantaneous time points. Whether it is José James's drummer, Erykah Badu's synth part, Jesca Hoop's bass, or, indeed, Cortot's right hand, each example involves beat locations that are smeared, that are fuzzy, and that have indeterminate locations but can still function in ways that make musical sense and afford entrainment. As such, they challenge the idea that perceived beats are abstract instantaneous time points on a grid. The goal of this dissertation's argumentation for beats as domains is to strengthen our understanding of rhythm and meter as manifested in real, lived performances by building a theory of beats that occupies a position closer to the surface of the music and embraces the phenomenological experiences afforded by the surface. This lower-level specificity will bring to life the established higher-level cognitive and music theoretical ways of conceiving of beats that make valid use of the generalized view of beats as instantaneous points on grids. As Ito writes: "the solution [to shortcomings in theories of meter] is not to discard the theory but rather to augment it, understanding it as basic cognitive scaffolding around which more particular practices are anchored" (Ito, 2021, p. 110). My "augmentation" is to dwell nearer to the surface of the music at the place where the abstract musical phenomena that we label as beats are initiated by and for performers and listeners. The work I present looks at how a sense of musical time that has expressive implications can be created by variable beat placement over time (cumulatively creating a domain or "pocket") and/or by having multiple different layers within the texture suggest different beat locations at any one metric position (local pockets).

Chapter 1 surveys the literature, drawing together the established theories of beats in music and cognitive research and arguing that there is a need for a theory of beats that prioritizes the sound signal produced by performance events. Chapter 2 introduces the theory of pockets, justifies why the label "pocket" is so important (rather than simply "domain" or another term), and establishes the existence of pockets with a music information retrieval analysis of three corpora of drum performances. Chapter 3 considers what these pockets *mean* for the listener experience, introducing descriptive terms that might be used to make sense of timing variations between performances. Chapter 4 presents an empirical study that ascertains whether listeners are able to perceive subtle timing differences, an essential element in constructing a perceptually relevant theory of musical beats. And finally Chapter 5 considers how differently shaped pockets interact with form in popular music, analyzing Stevie Wonder's "Superstition" to argue that musicians manipulate the shape, size, and location of the pocket across a song to enhance the experiential qualia of different sections, for example adding to the intensity of a chorus section by markedly transforming the shape of the pocket.

CHAPTER 1

Beats in Theory & Cognition: Sounded and Perceived

"Beats" play many roles and have numerous definitions in music. There is the beat of a conductor's baton (physical motion as beats), the beat of a metronome (isochronous demarcations of time intervals), and the internal beat of a performer or listener (a perceived, unsounded reference structure sometimes likened to a heartbeat). There are patterns of sound called drum beats (standardized, familiar patterns such as the standard rock drum beat), and there is the creation of sound using a beater on a drum. The focus in this dissertation is on beats as organizing, structuring temporal entities that are sounded and/or perceived. I focus on the nature of beats, following scholars who see beats as having temporal extension rather than being instantaneous. Such temporal extension is the result of the variability in performers' actions, where even musical events notated to occur at the same time differ in their absolute time-locations of production. My solution is to deal with the nature and content of these beats as probabilistic domains rather than discrete points. These domains and their contents serve as inputs to perceptual and cognitive processes, from which listeners construct their understandings of rhythm, meter, and patterning in musical styles and in specific musical contexts. The perceptual and cognitive processes involved are categorical in nature, which will be explained as having between-category elements (such as those which can be represented in standard music notation) and within-category elements (encompassing such phenomena as "feel" and expressive timing). In this way, I extend and complement existing theories of meter and rhythm. By surveying canonical theories of beats from the modern era, I emphasize how one's choice of viewpoint on beats influences one's understanding of musical phenomena. I will also highlight a need in music theory for a perspective on beats that speaks to *performed* music, focusing on the sound signal, and which complements existing and prominent theories that tend to approach musical structure and the musical experience with the aid of notation and symbolic representations of music. In addition to a review of several established and more contemporary/alternative music-theoretic treatments of beats, I explore how categorical perception and empirical research into performers' and listeners' entrainment behaviors provide evidence for a more dynamic conception of sounding musical beats. In this way, I set the stage for presenting my theory of beats as domains in Chapter 2.

Beats as Instantaneous Time Points

The most influential lens through which contemporary theorists understand rhythm and meter is Lerdahl and Jackendoff's *Generative Theory of Tonal Music* (1983, abbreviated here to "*GTTM*"). This impressive theory, about the "musical intuitions of a listener who is experienced in a musical idiom" (p. 1), methodically models the rules and preferences of a musical grammar underlying such intuitions. Lerdahl and Jackendoff's focus is on "only the final state of [the listener's] understanding" since, they argue, "it would be fruitless to theorize about mental processing before understanding the organization to which the processing leads" (p. 4). This final-state stance informs their approach to the rhythmic structure of music and, as such, their explication of a listener's rhythmic intuition begins with discriminating between two types of perceptual segmentation: "grouping structure" (hierarchically organized "chunks" of musical time) followed by "metrical structure" ("the regular, hierarchical pattern of beats to which the listener relates musical events," p. 17).

Regarding beats, they write:

> It must be emphasized from the outset that beats, as such, do not have duration. Players respond to a hypothetically infinitesimal point in the conductor's beat; a metronome gives clicks, not sustained sounds. Beats are idealizations, utilized by the performer and inferred by the listener from the musical signal. To use a spatial analogy: beats correspond to geometric points rather than to the lines drawn between them. But, of course, beats occur in time; therefore an interval of time—a duration—takes place between successive beats. For such intervals we use the term time-span. In the spatial analogy, time-spans correspond to the spaces between geometric points. Time-spans have duration, then, and beats do not.[1]
>
> (Lerdahl & Jackendoff, 1983, p. 18)

Lerdahl and Jackendoff thus present an abstract theory of beats that defines them as idealizations and hypothetically infinitesimal points, drawing parallels to geometric points that have no size (i.e. no width, no length, and no depth) and are indivisible—beats as instantaneous time points. Indeed, they graphically represent beats as geometric dots (Figure 1.1); this way of conceiving of beats affords their theory clear and absolute locations of the beat. This certainty allows for the creation of their Metrical Well-Formedness and Preference Rules that specify the relation of beats to attack points.

Lerdahl and Jackendoff's metrical theory talks about beats at different hierarchical levels (see Figure 1.1) with one of these levels having primary significance—the level of the *tactus*. The tactus—"the level at which the conductor waves his [*sic*] baton, the listener taps his foot, and the

[1] They then add in an endnote: "Imbrie (1973, p. 53) and, in effect, Komar (1971, p. 52) precede us in observing that beats are durationless points in time" (Lerdahl & Jackendoff, 1983, p. 334). To elaborate on this, Imbrie insists on the importance of beats as instants or points in time to avoid semantic confusion and Komar defines beats as the initial time-point of a time-span.

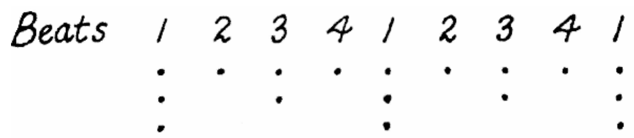


Figure 1.1: Lerdahl & Jackendoff's instantaneous time points (1983, p. 19).

dancer completes a shift in weight" (Lerdahl & Jackendoff, 1983, p. 21)—is intrinsically related to physical motion (see the Latin meaning of tactus as "having been touched"—perfect passive participle of *tangere*, to touch). In Figure 1.1, this level is most likely the top row of dots that indicates points on every beat of the bar. Depending on the musical context, there may well be faster, sub-tactus time points, too. The use of the term "tactus" links back to historical ideas of tactus as physical motion (for example the first type of tactus in DeFord's survey of Renaissance music theories [2015, pp. 51–2] or the Aristotelian embodied concept of *motus* [Grant, 2014, p. 27]). Even though the motions that Lerdahl and Jackendoff reference are continuous and have notable duration, they avoid any risk of confusion by referring only to the instantaneous points, by specifying "taps" and, in the case of more continuous movement like dance, the moment of completion of a movement rather than the movement itself.

Developments and Applications of Instantaneous Time Points

Lerdahl and Jackendoff are certainly not the first "modern" theorists to conceive of the beat as instantaneous time points—see, for a precursor, another important twentieth-century theory of rhythm and meter from Cooper and Meyer (1960)[2]—but the theory set out in *GTTM* is one of the most commonly used ways of conceiving of the beat in music in contemporary music analyses. *GTTM*'s isochronous points/dots may be seen through overt citation in analyses and also more tacit replication of dot-based theories of musical time—see, for example, Florestan and

[2] For Cooper and Meyer, the rhythmic structure of music is organized architectonically and there are three basic modes of temporal organization: pulse, meter, and rhythm. They define the lowest architectonic level—the pulse—as "one of a series of regularly recurring, precisely equivalent stimuli. Like the *ticks of a metronome or a watch*, pulses mark off equal units in the temporal continuum" (1960, p. 3, emphasis added).

Eusebius's exploration of metric dissonance using coffee beans (i.e., dots) to represent pulses in Krebs (1999, pp. 31–44). Beats as abstract time points are clear and provide stable foundations upon which metrical theories may be formed.

One final development of beats as instantaneous points is the more recent and less well-known idea of "temporal atoms," abbreviated to "tatums." This term was coined in a masters thesis by Bilmes as "the lowest level of the metric musical hierarchy," a high-frequency subdivision of the tactus, that is used to help analyze and model expressive timing computationally (1993, p. 21). This term was taken up in a later, influential thesis by Iyer as "the smallest operative musical subdivision of the tactus" that is cognitively meaningful and used to describe expressive deviations at the sub-tactus level that do not perturb the tactus (1998, p. 17, see also p. 131ff.). The notion of tatums was also taken up by Jehan who defines the term in more detail:

> The tatum…can be defined as the lowest regular pulse train that a listener intuitively infers from the timing of perceived musical events: a time quantum. It is roughly equivalent to the time division that most highly coincides with note onsets: an equilibrium between 1) how well a regular grid explains the onsets, and 2) how well the onsets explain the grid.
>
> (Jehan, 2007, §3.4.3)

Jehan and others (e.g., Gouyon et al., 2002; Seppanen, 2001) use the idea of the time quantum/tatum in beat tracking and onset detection algorithms to define the lowest segmentation level or "tick" that is useful for their analyses. This concept of the tatum is more of a curiosity than a significant new addition to the theories of the beat and can probably be aligned with Lerdahl and Jackendoff's lowest level of instantaneous time points (the topmost line in Figure 1.1), but it is still valuable to observe that similar concepts have been arrived at and fruitfully utilized by theorists working with non-Western Art Music (the tatum is, after all, named after jazz pianist Art Tatum) and in the realm of signal processing.

The Musical Object of Instantaneous Time Points

As can clearly be seen across the rhythm and meter literature, theories of beats as instantaneous time points are well argued and widely utilized by theorists and analysts seeking to understand the metric construction of music. However, there is a fundamental ontological question that needs to be asked of this theory: What is the musical object being described? An alternative viewpoint is presented by Hasty who, in surveying existing theories of rhythm and meter and pushing back against the "rigid determinism" of homogenous meter (Hasty, 1997, p. 5), argues that notation plays a significant role in how the beat is conceptualized: "All our systematic theories of meter draw upon a conceptual framework grounded in the technology of metric notation" (p. 6). At the metric level, Lerdahl and Jackendoff are clearly cognitive and experiential, seeing meter as arising from the felt strength of a beat (1983, p. 19), which is phenomenal and not purely notational (as described in the Metrical Preference Rules). But at the foundational level, at the level of the beats that scaffold the perceived meter, the ontology of the beats as Lerdahl and Jackendoff frame them appears to be closely intertwined with the written score being analyzed. Staff notation represents musical time through the spatial arrangement of dots on a page and they write "each row of dots below the music symbolizes a level of metrical structure" (p. 68). And again, when they emphasize how they are not making claims of universality with their principles of grouping structure: Lerdahl and Jackendoff refer to a few examples of musics that may challenge their rules including "North Indian raga, and much contemporary music (*regardless of whether the notation is 'spatial' or conventional*)" (p. 18, emphasis added). Left unanswered is the question of how to apply *GTTM*'s theories to music (of any genre) that uses no notation and/or only exists in performance (see, for example, jazz pianist and scholar Vijay Iyer's observations about the "nontranslatability" of parts of these types of theory when seeking to understand African-American musics Iyer, 2002, p. 388).

Alternative Views of the Beat

Beats as Waves & Physical Motions

Theories that are not based upon isochronous instantaneous beats are attractive to some music theorists, either supplementing or acting in place of such a view of beats. For example, Cohn notes how "in some styles the pulses of meter are noticeably plastic" and so, "[i]n those situations, we adopt a sort of double consciousness: we recognize that they are both equal and not equal" (Cohn, 2021, p. 15). Additionally, Hasty observes that perfectly isochronous beats do not correspond to the real, lived world:

> The periodic events we encounter in the world are produced primarily (but not exclusively) by organisms—other organisms and our own. And such periodicities are often not very precise. Very precise periodicity in our world of "middle-sized" durations is encountered primarily in the workings of machines. But we have not evolved to respond to machines. We have evolved to respond to, among other things, creatures that we must capture and creatures that we must evade. Since our locomotion and the locomotion of many other creatures involve various periodicities, much of the information we need for our interactions with the environment comes from aural, visual, and kinesthetic perceptions of more or less equal durations. "More or less" is an important qualification.
>
> (Hasty, 1997, p. 94)

Thus, both Cohn and Hasty advocate for flexibility within theories of rhythm and meter, and ground this flexibility within the fuzziness of the real world. The following section will explore a range of alternative theories of the beat that strive to incorporate some of this flexibility into their views to foreground other facets of musical beats. These theories are less final-state than that described by Lerdahl and Jackendoff in *GTTM*, drawing on ideas from psychology to consider the experience of beats.

Attentional Behavior

In London's theory of meter and rhythm from *Hearing in Time* (2012), listeners attune or entrain to periodicities between musical events such that their attention is guided toward the most salient temporal locations for future events. His theory of meter is based on temporal expectations and is dynamic and cognitively informed, specifically drawing on Jones and colleagues' Dynamic Attending Theory (e.g., Jones, 1981, 1986, 2019; Jones & Boltz, 1989). Different hierarchical periods of attention combine to create and reinforce our sense of meter. In these ways, London sees meter as a mode of attending (2012, p. 4)—something that Gjerdingen had also previously observed as an early utilizer of Jones's theories (Gjerdingen, 1989).

London aims to show the beat as a dynamically attended *category*, an idea I strongly agree with and develop below; however, his representations—points at the peaks of the attentional waves—end up being idealized, fixed abstractions (Figure 1.2). Therefore, despite the cognitive underpinnings drawn from DAT that incorporate the spread and inconsistency of real-world music into a time-continuous, resonant view of musical time (drawing on, e.g., Large & Jones, 1999), London's theory can still be read as working with instantaneous beats (Figure 1.3a), even when allowing for not-perfectly-isochronous music (Figure 1.3b). He acknowledges this, writing:

> I have emphasized that meter is a fluid attentional process—a form of behavior rather than a musical or mental object—but these representations do seem to be akin to traditional metric analyses in that they are static abstractions from real-time processes. This is a fair enough assessment, for any printed two-dimensional picture of musical meter will have to be a kind of snapshot of a dynamic process. I would emphasize, however, that what is important here is not the structure per se, but the set of relationships (and the constraints under which the relationships may occur) that these metric diagrams aim to capture.
>
> (London, 2012, p. 96)

London's theory of the beat has two facets that are most pertinent to the present discussion: first, that peaks of attentional energy have temporal spread to them, and secondly, that there is a

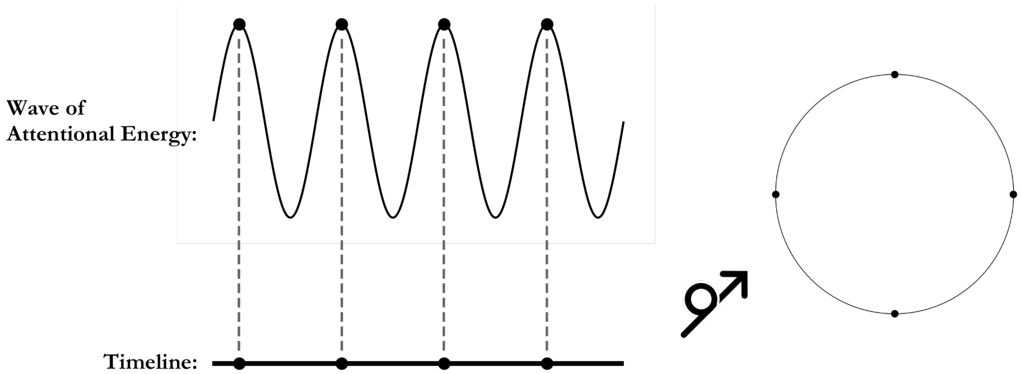
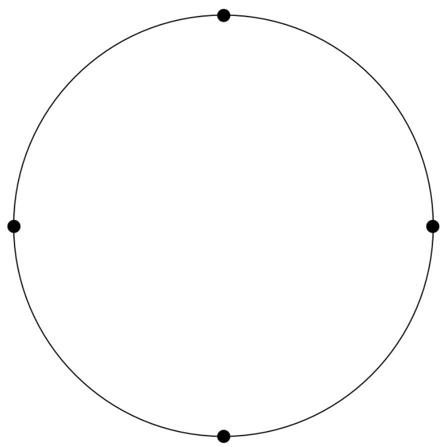


Figure 1.2: An illustration of how Dynamic Attending Theory's attentional waves (which have temporal spread) are transformed into London's "beat cycles" (e.g., p. 85) by picking the peaks. The reason this particular instance is a 4-cycle relies on the presence (not illustrated) of other attentional waves with longer periods—at least one with a period four times as long and most likely an additional wave with a period twice as long as that illustrated (see Yeston, 1974, p. 65ff. on meter arising from the interaction of multiple periodic strata).

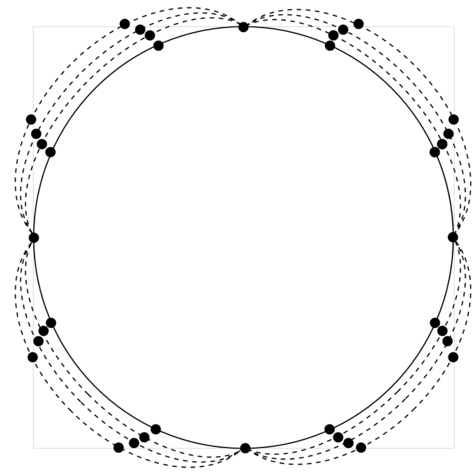
degree of elasticity to the cycles. Regarding the former (and the above discussion of waves being represented by dots), London writes:

> Although attentional peaks have some temporal spread, I will refer to them as time points in the sense that they serve to mark determinate temporal locations. Although the locations for attentional peaks are built into the structure of the cycle, they are subject to continuous timing modification and adjustment, so that the overall arrangement of the cycle remains stable.
>
> (London, 2012, p. 83)

Secondly, London explains that "Each component cycle has some degree of temporal elasticity both because the attentional peaks have some amount of temporal spread and because dynamic attending involves adaptive error correction (Large & Jones, 1999; Repp, 2005)" (London, 2012, p. 88). The system within which meter arises is *dynamic*—entrainment and attention, which involve anticipating the temporal location of the next event, are not entirely predetermined and set in stone. London talks of a listener's "evolving sense of beat" (p. 20) that results from tuning the periodicity of expectation-affording oscillators with the external musical rhythm. This is an ongoing process that adapts continuously to the new musical information as it is experienced.



(A) London's beat cycle.



(B) London's beat cycle with "petals" representing a set of possible expressively individuated timings for the meter in a jazz swing context (after p. 185).

FIGURE 1.3: London's beat cycles including expressive timings of a tempo-metrical type.

Despite the fact that beats are graphically represented as static abstractions, London's theory is a highly valuable development in theories of the beat and meter, presenting a theoretical platform that allows for non-isochrony and, most importantly here, cognitively based ideas of the beat as attentional peaks that have some elasticity/spread. To foreshadow my explication in Chapter 2 of the concept of pockets, if I were to suggest an evolution of London's graphic representation that emphasizes the dynamic sense of beat *with temporal spread* I might offer something like Figure 1.4.

Focal Impulses

Another alternative theory of beats is Ito's theory of felt beats, or "focal impulses," that explores the interrelations of the expressive shaping of sound, meter, and physical coordination in performance (Ito, 2004, 2012, 2021). This theory is a response (and complement) to traditional metrical theories that "have often appealed to the listener's experience, [but] have tended to deal with music

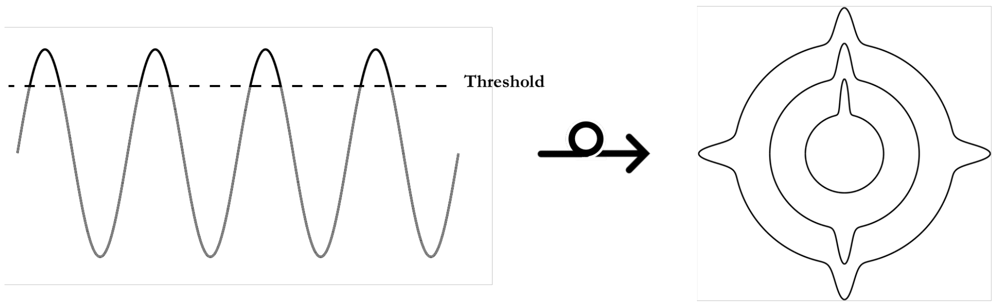


Figure 1.4: A possible alternative way of illustrating London's theory (right), showing beat locations not as abstract dots, but instead representing a 4-cycle as supra-threshold portions of nested waves of attentional energy (4-, 2-, and 1-cycles sketched). Note that the frequency and width of each cycle, as well as the alignment of all of the waves could, theoretically, be more flexible than depicted here.

in relatively abstract form, focusing on relationships among notes on the page more than on actual sound" (Ito, 2004, p. 4). The central idea of Ito's theory is that, based on the dynamic systems approach to motor control, motor behavior exploits stable resonance properties of musculoskeletal and neuronal systems. Motions with larger periods (involving larger portions of the body, e.g. spine) tend to align with metric periodicities in the music leading to a feeling of music being, for example, in two or in four. Complex motions are sequenced by combining together already existing and easily accessed physical resonances in a manner somewhat similar to a physical Fourier synthesis. In this way, hierarchically organized motions can combine to create expressive, felt periodic motions (that can shift in and out of phase with the beat levels they track).

The theory of focal impulses is illustrated by Ito with an analogy from halfpipe skating:

> The focal impulses are like the pushes off from the surface, as they set basic motional parameters for the aerial phase, especially momentum and angular momentum. In addition to determining flight time, these parameters determine which maneuvers can be done during the flight and which cannot. There remains much active motion to be made during the flight, but the push off from the surface establishes a crucial context of motion and imposes constraints on the motion to follow.
>
> (Ito, 2012, p. 482; also see Ito, 2021, p. 63ff.)

Here, impulses are moments of energy (initiating bursts) where the performer/skater launches themselves off and then, for the subsequent time span, they may be making all sorts of micropulsations or micromovements, but it is the impulse that initiates the sequence and controls much of how the motion will follow—they predetermine the "set of global motion constraints" (Ito, 2004, p. 33). The theory of focal impulses is more informative about the experience of meter than of beats per se, but I draw it into this survey as it presents an embodied perspective of the beat, paralleling the physical real motion of performers and listeners. Further, these movements are not framed as sharply defined (unlike Lerdahl and Jackendoff's view of dancer movements) and there are qualitative associations to do with the felt experience of these temporally organizing focal impulses. For example, Ito describes the feeling of anticipations (a note expected to arrive on a metrically strong beat that arrives early on a weaker beat) by outlining a physical exercise that involves pushing forward against resistance, then suddenly releasing the resistance so the push surges forward (Ito, 2004, pp. 155–6).

Performer Perspectives

Performers can offer a fresh perspective on the experience of the beat. Berliner, in interviews with jazz musicians about the compatibility of playing with specific other musicians, finds concern amongst musicians about "the failure to accommodate another's individual predilection for playing on *different parts of the beat*. 'When the bass player or the drummer is *right in the middle of a beat* and the other is not, there's going to be a little tug, and you're going to feel it,' [pianist] Tommy Flanagan insists" (Berliner, 1994, 396, emphasis added). He argues, using another interview, that this can threaten the cohesiveness of an ensemble:

> "There are times when I am playing with a drummer who wants to play more *on top of the beat* than I do," [bassist] Calvin Hill says. "I feel like he's rushing, so my reaction is to hold back. Since I like to play *on top of the beat* myself, if someone is playing even more on top of the beat, it usually means the tempo is going to pick

up, so I have to step back and hold the beat down."

(Berliner, 1994, p. 396, emphasis added)

Comments like these expand on Cohn's observations (above) about the "plastic" nature of beats in some styles. These world-class musicians evidently experience beats not as instantaneous time points (although their language indicates players' use of some kind of referential structure), but as a kind of collectively created category or domain within which individual musicians may construct their personal feel. This is not only relevant to jazz or even to live ensemble music as similar ideas about being "on top of" or "behind" beats may be seen in electronic musician Michael Stewart's "feel spectrum" (see Figure 1.5 from Stewart, 1987; featured in Prögler, 1995, p. 24).

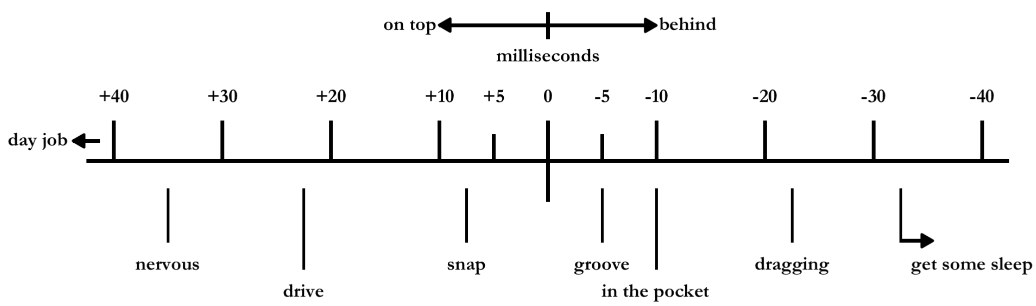


Figure 1.5: Reproduction of Stewart's "feel spectrum" (from Stewart, 1987, p. 64). The spectrum represents an attempt at putting a number of milliseconds ahead or behind the beat to subjective terms like "snap." Stewart personally adjusted a synth bass playing against a fixed drum machine set at 130 BPM until the desired "feel" was obtained (in his opinion).

Beat Bins

This performer perspective, as well as some of the foregoing theorists surveyed (particularly London), suggests some kind of categorical perception of beats—perceiving a continuously variable phenomenon as having distinct divisions (i.e., sound events located in time having membership of the category "the beat"). In the field of groove research, Abel has referred to culturally determined perceptual "capture zones" around beats that enable slightly early or late notes to be perceived as part of the beat (Abel, 2015, p. 34). This idea is more clearly expounded by Danielsen with her concept of "beat bins," in which the beat has "shape" and "where each beat in the basic pulse has been extended into a beat bin, which encompasses all of the different locations of the pulse at the micro level" (Danielsen, 2018, p. 185; see prior discussion in Danielsen, 2010a, pp. 29–33; Danielsen, Haugen, et al., 2015, pp. 139–40). The concept of beat bins arises from Danielsen's analyses of R&B artist D'Angelo's music, in which different instrumental layers (for instance drums vs. guitar) have slightly conflicting beat locations that, when explained through the lens of Dynamic Attending Theory, result in a larger, categorical sense of beat. The multiple temporal locations of the onsets are wrapped within a categorical bin that function as "a beat." Danielsen (2010a) illustrates three models of beat perception that explain why a more accommodating theory of beats is valuable: the first, the "metronome model" (pp. 21-26) "presupposes that there is one dominant or correct placement of the internal beat and, moreover, that the beats should be equally spaced" (p. 25; cf. Lerdahl and Jackendoff). Here, if there are any events in the performance that challenge the ideal metronome, for example the two competing beat locations in D'Angelo's "Left and Right" (2000), a listener would need to select one as the "correct" beat location (perhaps via something like *GTTM*'s preference rules, and most probably with some difficulty and cognitive effort). Danielsen's second option (pp. 26-29), an application of Honing's (2001) "local time-shift model," involves a listener's attention shifting dynamically between instrumental layers to define a sense of pulse that adjusts in the moment to new information about beat locations. In this model, pulse is an emergent property of the music rather than an unvarying metronome. Lastly, Danielsen arrives at the "beat bin model" (pp. 29-32) that suggests that performers and listeners have a (context dependent) degree of "rhythmic tolerance" in their *categorical* perception of beats. Beats are extended, so there is a period of time within which events fall into the category "part of the beat" (p. 30):

> With the beat-bin model, pulse is no longer a series of points in time, because each beat is thought to have both a shape and a duration. According to this view, differing rhythmic events may be regarded as located within the same beat, in turn contributing to the duration and shape of the beat's virtual counterpart, the beat category. This means that it is less the temporal relations between pulsations than the shape of the beats at a categorical level that is the dynamic feature of the groove.
>
> (Danielsen, 2010a, p. 33)

Using this, Danielsen explains the change in a listener's experience of D'Angelo's "seasick time-feel," from unsettled by the asynchronous events to experiencing them as a "distinctively organic, swaying musical whole" Danielsen (2010a, p. 21) as a change from a small beat bin to a wide one that can "absorb" competing events.

Danielsen's beat bins form part of what she describes as "virtual reference structures," which are defined as "schemes used by the performer/listener in their respective music-related acts" that, importantly, are *non-sounding* (Danielsen, 2010b, p. 4). The actual sounding events and the non-sounding virtual reference structure continuously inform each other—a relationship that this dissertation develops by focusing on the specific details of the actual sounding events. Danielsen frames the link and interaction between the musically sounded events and this virtual reference structure as core to musical rhythm. According to this perspective, we experience rhythm through how sound and an unsounding, but still present in the music, structure interact. This relationship is explored by Danielsen through a Bakhtinian lens, understanding the musical utterances as "gestures" that sound against virtual "figures," which are more abstract and schematic, and which provide the syntactical or grammatical scaffolding that affords understanding of the gesture (Danielsen, 2006, pp. 46–50).[3] Danielsen transfers Bakhtin's critique from linguistics to music, paralleling his concern that many linguists look at texts in the abstract, forgetting the actuality of the communication, the fact that sentences are spoken/"uttered" by people and made meaningful

[3] In Bakhtinian, linguistic terms, gestures are "utterances" while figures are "sentences." Danielsen highlights this difference by referencing multiple other, similar ideas from semiotics and linguistics, for example Saussure's "*lange*" and "*parole*" and Hjelmslev's "schema" and "usage."

in their expression. This is not an argument (by Danielsen or Bakhtin) for abandoning the structures and only caring about the utterances—quite the opposite, as it advocates for understanding the two as mutually dependent.

The theory of beat bins has evolved and clarified through further publications by Danielsen to be framed in more specifically metrical terms,[4] and is presented as a tool that can be used to describe the "perceived temporal 'width' or extent of a beat [category]" (Danielsen, 2018, p. 183). Building on earlier ideas about figure/gesture and the deeply enmeshed relationship between unsounding structures and the actuality of sounded rhythm, as well as drawing even more on the *dynamic* elements of dynamic attending theory, beat bin metric theory explains how metric expectations are emergent and changeable properties of musical performances, rather than the fixed result of top-down, external structures.

The preceding survey has considered music-theoretic ways of conceiving of the beat from a range of perspectives. Given the numerous ways that musical beats have been addressed, it is important to be precise and deliberate about the way in which we choose to conceive of the beat as this defines the frame of reference against which analyses and interpretations are made. For instance, the literature on groove is pervaded with analyses that are based upon systematic micro-deviations from a temporal grid composed of beats conceived of as idealized, abstract points. Framing beats in this way in this context *is* meaningful as it relates to how much modern-day music is synthesized in Digital Audio Workstations that default to quantized, 16-step grids that may then, intentionally, be deviated from; however, this certainly does not apply to all musical styles, may not relate to how live performers and listeners engage with the beat, and may not relate to how the sounds are heard (see the research into the "perceptual center/P-center" of sounds, e.g., Bechtold and Senn, 2018; Danielsen et al., 2019; London et al., 2019; Villing, 2010; Wright, 2008). It is vital,

[4]Meter here is defined as a "mode of experience" (after London, 2012) rather than meter as an analytic tool, and being a virtual scheme that corresponds to the relatively regularly recurring pulsations at different frequencies (tempi) in the listener (Danielsen, 2018, p. 180).

therefore, that an analyst or theorist be explicit about how the beat is being conceived when making arguments, for example in an analysis of groove, about "participatory discrepancies" or microtemporal deviations. In fact, Bengtsson contextualizes his idea of "systematic variations of durations"—a significant concept in the expressive timing literature—by stating: "In fact, we should avoid calling it 'deviations' when dealing with rhythm without stating clearly that we just mean deviations *from a mechanical norm* that we use as a sort of temporal ruler. We have no other ruler, mainly because we know far too little about such micro-structures" (Bengtsson, 1987, p. 78, emphasis in original). Before defining the theory of pockets in the next chapter, the cognition of beats will be considered with specific regard to categorical perception, as this supplies much of the justification for *how* sounding pockets may function musically and perceptually for performers and listeners, and demonstrating empirically the temporal flexibility of real -world situations.

Categorical Beat Perception

Categorical perception has a multitude of overlapping definitions, but here I take it as the tendency to perceive the environment in terms of the categories we have formed, with phenomena being classified into groups with more or less sharp boundaries (notably observed in the discrimination of speech sounds, see Liberman et al., 1957). Categorical knowledge is used to abstract away from perceptual differences between phenomena/items that belong to the same class, creating strong in-group identity while highlighting differences between classes (Harnad, 1987). Categorical perception has been identified in various modalities; a commonly mentioned example is how a rainbow is perceived as being comprised of bands of red, orange, yellow, . . . , etc. when it is actually a continuous spectrum (e.g., Beale & Keil, 1995; Collins & Olson, 2014; Livingston et al., 1998). In the famous Liberman et al. paper, /ba/ and /pa/ phonemes were heard by participants as having clear boundaries despite lying along a continuous spectrum of voice-onset times (the time between air release and vocal-cord vibration).[5]

Categorical perception is particularly pertinent to music as numerous dimensions of sound vary continuously (e.g., frequency and amplitude), but are divided into discrete classes perceptually and/or mechanically (as in the keys of a piano keyboard or the 16 buttons/grid positions on a drum machine or sequencer). Musical rhythm and meter are prime instances of this as they involve dividing continuous time into discrete, often nameable, classes. This has already been explored briefly through the discussion of Danielsen's "beat bins," but the following survey of core empirical and theoretical works will demonstrate how listeners and performers categorize time in music and, importantly for the theory of the pocket that will be presented in the next chapter, how these categories allow a range of distinct musical events to be drawn together into one entity.

First, in tasks where participants are asked to reproduce musical rhythms by tapping, they do not exactly replicate the rhythms; instead they compress and stretch duration intervals—dynamic processes that Fraisse, a significant figure in time and rhythm research, labels "*assimilation*" and "*distinction*" respectively (1947; see also a similar study reported in Fraisse, 1956, Chapter 4, pp. 47-59)—to produce whole-number ratios between durations (most favoring 1:1 and 2:1 ratios) and create stable rhythmic structures. Although there are an infinite number of possible divisions of a span of time, Fraisse's tapping participants tended to assimilate different patterns into the same category (1:1) or emphasize differences to place in a clearly distinguished category (2:1), thereby affording simple structures (*assimilation*) without the possibility of confusion (*distinction*).[6] This aligns with a key feature of categorical perception: that discrimination is better between groups than within. As Fraisse summarizes in an overview of his experimental findings:

> If two durations belong to the same category, there is a tendency to equalize these

[5] The phoneme /ba/ tends to have a faster onset time (below 30 ms) and /pa/ has a slower onset (above 30 ms) (Wood, 1976).

[6] *"Ils [les processus d'assimilation et de distinction] tendent à créer des structures simples (assimilation) et sans possibilité de confusion (distinction)"* (Fraisse, 1947, pp. 194–5).

> durations. We prefer to say that there is assimilation since this equalization is not absolute. Among durations of differing categories, there is a sharp distinction. Assimilation and distinction bring us back to the classical perceptual laws which correspond to a principle of economy in perceptual organization (Fraisse, 1947).
>
> (Fraisse, 1982, p. 167)

Also, in the original 1947 article that Fraisse cites in his English-language chapter from 1982, he points to the Gestalt law of proximity, in which elements that are close together tend to be perceived as a unified group, and writes that "retention becomes easy if there is assimilation of the different elements to each other and if one only has to retain the common character" (Fraisse, 1947, p. 190; cf. Lerdahl and Jackendoff's [1983] implementation of Gestalt principles in their theory, e.g., using the law of proximity in the Grouping Preference Rule 2).[7] He also shows how other Gestalt principles of resemblance (showing there is a shared similarity in the underlying structure of two objects) and contrast (exaggerating differences) facilitate the *assimilation* and *distinction* of rhythms (pp. 190-1). In short, Fraisse suggests that the imperfect tapping behavior of participants and preference for a couple of simple ratios supports the idea of categorical rhythmic perception, and that the use of categories affords a cognitive efficiency (Fraisse, 1982, p. 167).

This persistence of a category even when what is expressed or interpreted by the listener does not exactly match the sound (e.g., assimilating unequal durations into a 1:1 ratio in replication tasks) and then a sudden transition to another category—a hallmark of categorical perception—is also seen in Clarke's identification and discrimination tasks where he systematically varied the inter-onset intervals in metric contexts (Clarke, 1987a, pp. 23–6). For example, Clarke's results, reproduced here in Figure 1.6, take the characteristic S-shaped curve of categorical perception. Clarke observes that the precipitous drop from 80% identification (Type 5, 560:400 ms or 1.4:1) to only 50% identification for the next stimuli set (Type 6, 540:420 ms or 1.28:1) suggests the location

[7] *"[L]a rétention devient facile s'il y a assimilation des différents éléments les uns aux autres et s'il s'agit uniquement d'en retenir le caractère commun."*

of the categorical boundary with rhythms above this being perceived with the 1:1 category.[8] Schulze (1989) responded to Clarke's paper with more fine-grained stimuli at multiple tempi (however, recruiting only two participants to Clarke's already small 10). He at once concurred with Clarke, finding a nonmonotonic discrimination function for one of the conditions, but also found that his highly trained participants could discriminate more precisely the different rhythms along the continuum. Clarke reads Schulze's results as proposing "that categorical perception is a function of perceptual learning: if sufficient training is provided, perceivers may learn to identify and discriminate between rhythmic categories which without training might have been part of a single more undifferentiated category" (Clarke, 2000). In this way, categories are not absolute and pre-existing, but arise through experience and learning (see also the discussion of Polak, Jacoby, and collaborators' work, below).

In an approach different from tapping and discrimination tasks, Desain and Honing asked highly trained musicians at conservatories in the Netherlands and Japan to transcribe various rhythms into musical notation (Desain & Honing, 2003). The semi-open nature of the responses collected meant that category boundaries cannot be observed in the same way as in Clarke (1987a, see Figure 1.6 above). Instead, Desain and Honing used "time clumping maps" to show the "coagulation" of rhythmic categories around simple integer ratios (Figure 1.7). They suggest that the areas contained in grey borders represent rhythmic categories perceived and used by the participants to understand the heard rhythms. In the same paper, Desain and Honing also conducted a second, near-identical experiment, but this time with metric priming incorporated into the stimuli. The metric contexts changed the shapes of the time clumps, demonstrating that rhythmic categorization depends on a pre-established cognitive framework structuring a listener's perception of time. They also argue that the changed shape of the categories suggests that meter has an effect on the expected expressive timing, as certain deviations from the metric time structure become more or less appropriate.

[8] Analysis is limited only to observations derived from the visualizations. No further statistical tests are reported.

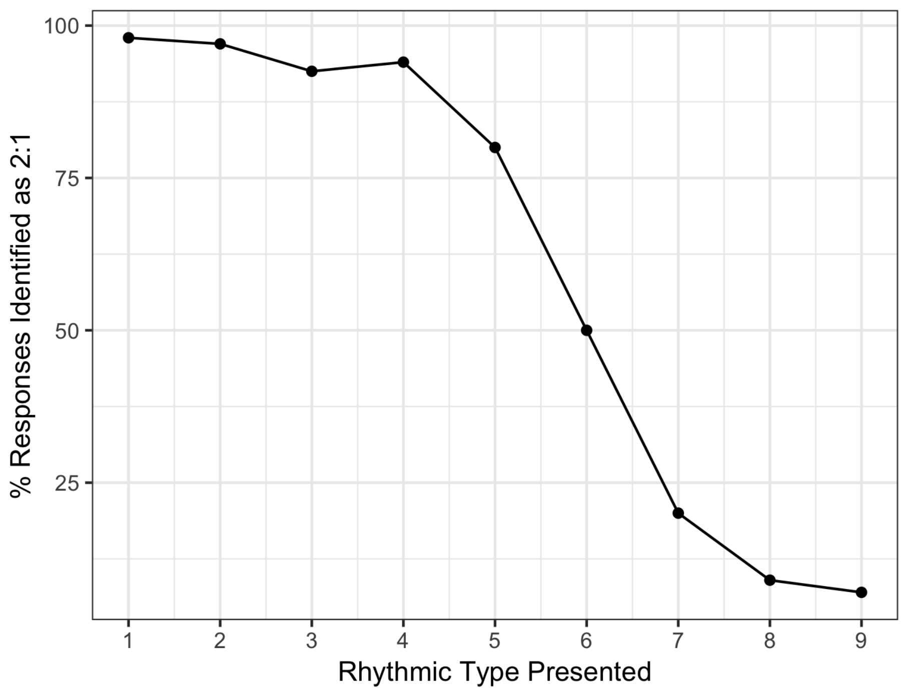


FIGURE 1.6: Reproduction of identification results from Clarke's Figure 2 (1987a, p. 25). "Rhythmic Type Presented," on the *x* axis, represents the durations of the nine different rhythmic stimuli Clarke presented that ranged from a perfect 2:1 ratio ("Type 1," 640:320 ms durations) to a perfect 1:1 ("Type 9," 480:480 ms) with methodical shortening of the first duration by 20 ms and lengthening of the second duration by 20 ms each step.

Researchers at the Max-Planck-Institut für empirische Ästhetik have been leading a project that investigates the perception and performance of musical rhythm in different cultures. Polak, Jacoby, and colleagues from other institutions have developed our understanding of rhythm prototypes and categories through conducting studies with participants from various countries, including Mali, Bulgaria, Bolivia, Germany, and Uruguay. They challenge the implicit (and sometimes explicit e.g., Drake and Bertrand, 2001, pp. 24–5), suggestion that 1:1 and 2:1 proportions are universal rhythmic prototypes, instead arguing that "humans' learning of rhythmic categories is a dynamic and ongoing process" (Polak et al., 2018, p. 1), and that someone's music-cultural

[9] For interactive time clumping maps and audio examples of stimuli, see the "Music, Mind, Machine" website: http://www.nici.ru.nl/mmm/. Also see the "rhythm space"/"chronotopological map"/"rhythm simplex" in Jacoby and McDermott (2017) where participant tapping moved iteratively towards integer ratios in a similar way to that seen in Desain and Honing (2003).

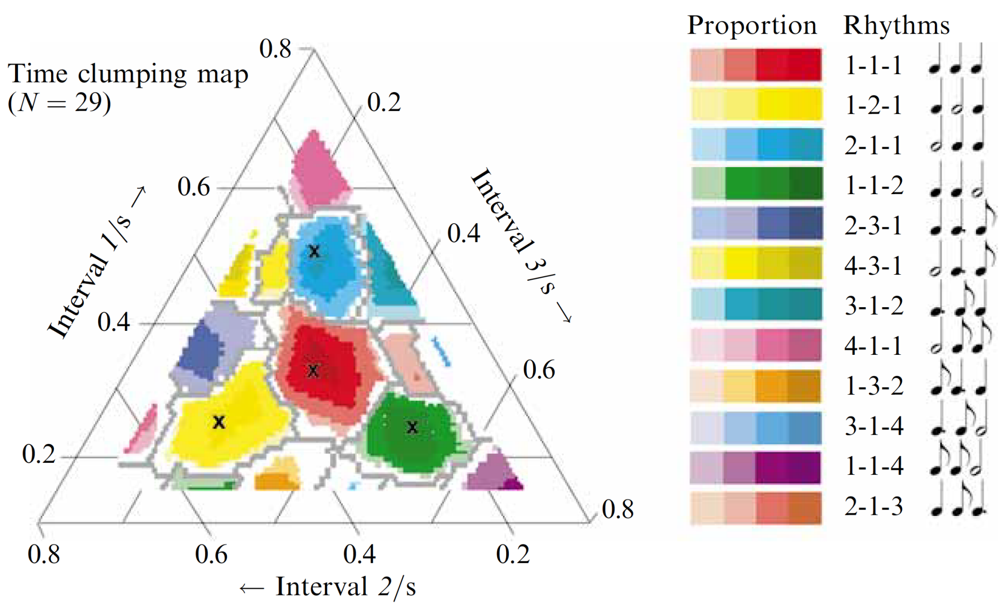


FIGURE 1.7: "Time Clumping Map" from Desain and Honing (2003, p. 357) showing participants' transformation of continuous time intervals (physical time) into rhythmic categories (perceived time). Colors show which rhythm corresponds to which area and darker shades of colors show higher proportions of responses. Grey lines show category boundaries with black **X** marks to represent the centroid (which differs from perfect, mechanical integer ratios).[9]

environment—either in the short term, as in Schulze (1989), where participants were trained over the course of the study, or in the long term, over a lifetime's experience—can specify links between music performance patterns and perceptual prototypes. This research group's output suggests that categorical rhythm perception depends on culture as "musicians whose cultural background did not include [a] prototype would not have such a corresponding cognitive prototype at their disposal" and so could not synchronize their tapping behavior (p. 13; see also Jacoby and McDermott, 2017 and Witek et al., 2020).

The above-cited studies, with their evidence from tapping tasks and discrimination judgements as well as notational tasks and the self-reported instincts of performers, appear to support categorical perception of rhythms, which maps from a temporal continuum onto learned structures that simplify and make more efficient the task of perceiving a rhythm. Clarke argues that perception *always* has a categorical component and a non-categorical component (Clarke, 2000).

It is worth contemplating the non-categorical elements he mentions. In the aforementioned 1987 study, Clarke questions the fate of the "remainder": the perceptual information that is discarded in the process of categorizing something as, for example, "a 1:1 rhythm" or, in his pitch-based examples, "a perfect fifth" (p. 22). He writes that the deviations from a perfect categorical fit are in fact not discarded, but are instead assessed as expressive information. This leads him to conclude that "the separation of temporal information into structural and expressive dimensions raises the possibility that rhythm perception can be regarded as picking up the best fit for an array of perceptual information in a multi-dimensional space" (Clarke, 1987a, p. 31). The two dimensions—structural and expressive—consist respectively of simple, whole-number duration ratios of categorical perception that relate to the "invariant" of meter (p. 33) and then the expressive "remainders" that allow for the realities of human performance. In fact, Desain and Honing build on Clarke's work to claim that "expressive timing is only perceivable *because* there is categorisation, the categorisation functioning as a reference relative to which timing deviations are perceived. In this view, both types of information are available to the listener, with categorisation determining the expressive timing perceived" (Desain & Honing, 2003, p. 343, emphasis in original).

This section has mostly discussed the categorical perception of *rhythms* as opposed to *beats*. This rhythm-focused research provides vital support for categorical ways of conceiving of beats. Research on the identification, discrimination, and reproduction of musical rhythm informs us that our perception of musical time is not always exactly the same as the time represented in the sound signal—we make use of learned categories to simplify the complex, continuous temporal domain and, by doing so, we are able to generate our internal reference structures that guide, for example, dance and entrainment more generally. With specific regard to beats in a metric context, Kvifte has argued (with anecdotal support from performers) that complex, additive meters such as 7/8 (2+2+3) can be perceived as three, non-isochronous beats rather than seven subdivisions (Kvifte, 2007, p. 70; also see S. S. Hudson, 2019, pp. 81-8 on "The Status of 3+3+2 Rhythms as

Meter" and S. S. Hudson, 2022 for a similar discussion in the context of metal music). In order to experience the meter in this way, the unequal durations must fall into a perceptual category of "beat." Further, imperfect subdivisions of beats (for example in jazz swing or Scandinavian *springar* dance music) suggest that listeners do not perceive rhythms against a framework of fast, isochronous units (Kvifte, 2007, p. 70ff.) but instead perceive categories that are subject to expressive (or random) variation (p. 74). As a result of perceiving beats categorically, we are able to make practical use of "untidy" (p. 72) divisions of time. This can be seen in genres with non-isochronous meters, not just additive meters, being associated with dance (see, for example, waltzes), which requires reliable expectations to be generatable for the coordination of movement with music.

Conclusion

This survey has shown that the beat may be thought of as a means of organizing time and/or as a sounding, perceptual experience (these views are not mutually exclusive). How we conceive of the beat depends on what role it performs and what use we are making of it. Boone, writing in the context of mensural time, observes:

> Of course, one's intuitive response to the idea of pulse framework, as to pulse itself, depends in no small measure on one's musical orientation. For ears steeped in common practice and related musical traditions, the rhythmic intuition of pulse and pulse framework tends to be tightly bound to those of ictus and meter. This orientation has posed thorny problems for the interpretation of rhythmic systems in other cultures, such as those of traditional sub-Saharan African music; the intuition of meter by American and European listeners in such music seems often to be problematic, if not simply wrong, from the standpoint of native musicians, even where pulses, pulse frameworks, and rhythmic hierarchies would seem clearly to be present.
>
> (Boone, 2000, p. 5)

The way we choose to conceive of the beat has significant implications for which features of the music are foregrounded and which are relegated. There is no single "best" way of defining musical beats—each theoretical position affords a frame through which to understand a facet of the complex phenomenon. However, in a given context, there may be more or less appropriate frames. As Boone highlights above, the tools developed for understanding common practice music sometimes cannot translate effectively to other musical cultures, styles, situations, and experiences. Echoing Ito's critique of Lerdahl and Jackendoff's theory of rhythm and meter, and broadening it to encompass the field of rhythm research more generally, "the solution is not to discard the theory but rather to augment it" (Ito, 2021, p. 110). The next chapter offers just such an augmentation, presenting a theory of musical beats that views them as sounding, dynamic, experiential, participatory, and interactive musical elements.

CHAPTER 2

The Theory of Pockets

In this dissertation, I advance a theory of beats as domains of time that I describe as "pockets," taking a vernacular term commonly used by jazz, funk, and popular music performers to describe the state of being in a good groove. The theory of pockets that I propose can be utilized to provide new perspectives on rhythmic and metric structures in music, as well as new approaches to understanding patterned microtiming in performance, the phenomenon of musical "groove," and musical "feel" more generally. Viewing beats as domains acknowledges the importance of microtemporal performance details while reframing them as stable reference structures. By importing the term "pocket" from performers in groove genres, the technicalities of the shaping of music time can be made meaningful by drawing on real-world qualitative descriptors of musical "feel"—a pocket may be "loose" or "tight," and a performer can "lay back," be "on top of the beat," or "push." In drawing this emic language into systematic analyses of musical performances, I present a theoretical perspective that provides avenues into understanding some musical phenomena that are loosely described in metaphorical language by performers.

According to the theory of pockets that I define in this chapter, beats are (short) spans of time—not instantaneous time points located on an isochronous grid—during which an onset may be categorized by performers and/or listeners as having the function of "beat X" (i.e., an onset, or set of onsets in the same domain of time, can function as "beat 1," "beat 2," etc.). These domains are not simply featureless windows of time, but instead have probabilistic distributions associated with them such that there are temporal locations in these spans of time at which an event is more or less likely to be categorized as that beat. Furthermore, as I will demonstrate here with an analysis of three corpora of drum performances (and in greater detail in Chapter 3), these probability distributions have shapes, which may not be perfectly symmetrical or centered on a metronomic time point, and these shapes develop through time as more information about how onsets are performed is learned. I associate the shape of the pocket with the qualitative experience of a performance, arguing that individual performer styles as well as genre conventions may be captured by this shape.

Etymology of "The Pocket"

This dissertation presents a theory of "pockets," rather than "domains," "spans," "periods," or other similar terms. The word "pocket" has significant meaning to a wide variety of musicians in a number of musical styles and means so much more than simply a span of time. As such, it is informative to understand how musicians have used the term and also to consider *which* musicians were using the term at what points in history. Drawing on this vernacular, slang language embeds musicians' lived experiences into this dissertation's systematic and computational methods, humanizing the empirical findings and rendering them musically meaningful. Even when language evolves with time, and even if the slang term falls out of common usage, there is still much to be gained in drawing on the concepts derived from tthe on-the-ground experience of people. One famous example of this is how the word/concept of "signifyin' " (Gates, 1988) is not necessarily a common topic in Black communities in the present day, yet there is still immense value and utility in drawing on this vernacular idea. Here I do not present a complete archival investigation of the etymology of the word "pocket" or the phrase "in the pocket" as used by performing musicians ("completeness" is impossible when investigating cultural phenomena),[1] but rather present various etymological threads that paint a helpful picture of why musicians might have settled on the word "pocket" and why they might have found it useful for capturing some of the qualities of performances that they value.

First, there are a number of mid-century jazz songs that feature the word "pocket" somewhere in the title. In these cases, this may be a very literal usage of the word. For example, "Corner Pocket" by Freddie Green (1955, made famous by Count Basie on the album *April in Paris*, 1955) refers to the pocket of a billiard table. To read metaphorically into the billiard pocket and to lay the first of the etymological threads, musicians may associate the pocket of a billiard table with satisfying accuracy, success, and also, perhaps, the pleasant sound the balls make as they go in. "The pocket" may also refer to pockets in clothing and the act of getting paid, as in "Money in the Pocket" by Cannonball Adderley (1966), Freddie Hubbard's album *Put It In the Pocket* (1975), and (most likely) Stanley Turrentine's album *In the Pocket* (1975).[2] This financial meaning of "pocket" could feed into an aspect of the musicians' usage of the word: if something is "in the pocket" it is guaranteed and secure. These musicians were very concerned with getting something "in the pocket" for their labors, and so the word was likely heard fairly often around the rehearsal

[1]Keir Keightley's thorough and insightful case study of the history of the term "Tin Pan Alley" serves as a reminder of how complex origin stories behind vernacular turns of phrase may be, as well as how these kinds of terms and products of culture rarely ever have a singular origin story, rather they often manifest a coming together of numerous threads that are co-occurring (Keightley, 2012).

[2]Outside of jazz, the exact same album title, with the connotation of getting paid (track 3 is "Money Machine"), is used by folk/soft rock artist James Taylor for his seventh studio album just one year later in 1976. Likewise, James Brown sings "In the pocket, sock it in the pocket" on the track "I Refuse To Lose" (1976) about not letting financial hard times get him down. The use of"pocket" in referring to paying someone has long been established with, for example, a letter to the editor of *The Observer* in London in 1792 quoting the Queen as saying "let us...put a few guineas in the pocket of the artist" (Hopner, 1792).

studio or in the green room before a performance.

One interesting usage of the phrase "in the pocket" comes from a 1967 review in the *Washington Post* of jazz trumpeter Charlie Shavers's performance at the Blues Alley in D.C. In this review, the critic, William Rice, draws on a football quarterback metaphor to chide the band leader for walking about the auditorium while taking a solo, instead advising him to "stay in the pocket provided by the musicians on stage" (Rice, 1967). This is not a rhythmic or metric use of the word "pocket," however, it does link with the notion of a musician playing in a defined and expected region.[3] It also shows that "pockets" are being used metaphorically in other cultural realms, realms the musicians undoubtedly would have some familiarity with due to the importance of football in American culture.

The earliest use of the word "pocket" to describe something to do with how the musicians are playing time that I can find in jazz's flagship magazine, *DownBeat*, is from 1978[4] where jazz pianist and organist Bobby Lyle describes his experiences playing with funk pioneer Sly Stone: "He gave everybody a really simple pocket to play into; when these pockets were all synchronized and syncopated, the music became compelling" (Underwood, 1978, p. 40). This use of the word "pocket" to describe something rhythmic, something to do with the alignment and interaction of the musicians' performances, and the positive aesthetic outcome of these has no accompanying definition or unpacking of the term in the interview, perhaps suggesting that *DownBeat* readers, by 1978, should be *au fait* with the term.[5]

[3] A separate, similar sports usage comes from baseball where the "pocket" refers to the concave area of the glove in which the ball is caught.

[4] *DownBeat* magazine's archive is only partially digitized and searchable so this is not guaranteed to be the earliest mention; however, from searching the available issues from earlier years, back to Volume 24 in 1957, the only other uses of "pocket" are very literal: for example in discussion of performances at New York City's "Pocket Theatre," slipping some money into a doorman's pocket, reviewing Freddie Hubbard's 1975 album *Put It In the Pocket*, or describing a performance on a pocket trumpet.

[5] At this point in time, the Commodores released their ninth studio album *In the Pocket* (1981), which I would argue *does* evoke this performance-related meaning of the word "pocket," unlike the Turrentine and Taylor albums from just five years earlier. None of the track titles or lyric themes stress money, rather the album is just a final celebration of the original Commodores line up (Lionel Richie left the group soon after the recording date) and their famed sophisticated funk sound.

In the trade magazine *Modern Drummer*, there is no mention of "pockets" until 1981, and that is only in one interview with the drummer of FAME Studios' house band (often called the Muscle Shoals Rhythm Section), famed for their "Southern soul" sound. In this interview, Roger Hawkins says "It's very important to know the lyrics so you can phrase them properly also. If a singer has a certain style of phrasing, I want to get into that style so I can compliment it and put the drum right in the pocket, so to speak" (Flans, 1981, p. 58). Here, Hawkins is describing the importance of aligning oneself with the phrasing of a singer and thinking about how to perform in a style that matches a particular artist. Again, there is no further elaboration on the word—just casual, conversational usage. This may herald the entry of the term into the drum magazine lexicon as more and more articles in ensuing issues—though by no means consistently in every following issue—involve at least one mention of "pocket" (see, for example, the next appearances: Nixon, 1981 and Flans, 1982; present day issues of the magazine are awash with the term and there was even a multi-issue instructional column titled "In The Pocket": Adamo, 2013a, 2013b, 2013c, 2013d, 2013e, 2013f. The latest issue (#300, 2021) of another drum magazine, *Rhythm*, features the word "pocket" 51 times over just 100 pages). The musicians who use the term in the earliest instances found in *Modern Drummer* magazine all come from the worlds of soul, funk, and R&B.

From searching in newspaper archives, particularly review sections, I cannot find any uses of the word "pocket" that relate to how the musicians are playing until 1984, when Neil Young describes his shift from playing in the rock genre into a more country style for the *Boston Globe*. He describes his excitement at playing with this ensemble of old friends, saying "Everybody feels like we just fell in the pocket. We're in the right place at the right time" (Sullivan, 1984, A11). The term is still rarely used from this point in mainstream news sources, suggesting that it had not entered the vocabulary of broad-public print publishing and took a while longer to be adopted. The next two instances that can be found through searching ProQuest's historical newspaper database are notable for describing white soul singer Lisa Stansfield's performances (Willman, 1990, NB: six years after the preceding usage of "pocket") and a way of praising the style of rock band The Black Crowes's second album (Graff, 1992). Though still very rare in mass media print, by the 90s, the concept of the "pocket" was now seemingly something that could be applied to a variety of musical genres.

One final thread that can add context and situate the term "pocket" culturally is the use of the term to describe the music heard in Black churches. Writing about his own experiences growing up in Chicago in the 1980s, Guthrie Ramsey describes how the Liberty Temple Full Gospel Church's Sanctified Band "prided itself on professionalism, slickness, and most important, being 'in the pocket' " (Ramsey, 2003, p. 13). The directionality of the movement of the phrase "in the pocket" is hard to disentangle—whether Black musicians and audiences from funk and associated genres (Ramsey notes how the Sanctified Band sounded like Chaka Khan or the Gap Band) imported this lingo into the church or whether church bands may have been a point of origin—but I draw this into the current etymological exploration to underline and make explicit how the term was closely associated with Black musicians in metropolitan America.

Overall, it is hard to trace the exact lineage of the term "pocket" as it grew out of of performing musicians' parlance; this vernacular language takes a while to seep into print, our only source of documentary evidence.[6] The challenge of locating a point or time of origin, even generally, is especially difficult as magazines, and particularly newspaper reviewers, have a more elitist (and perhaps more white?) tenor and so do not adopt the language of the musicians immediately. Additionally, as with many turns of phrase, it is likely that there is no single point of origin, rather musicians' use of "pocket" may have emerged in multiple places and times, drawing on multiple literal meanings of pockets described earlier, and gradually coalesced into a musical term. The early uses in *DownBeat* and *Modern Drummer* both suggest that these magazines'

[6]While interviews with still-living musicians of this era would undoubtedly be fascinating, their vocabulary would have evolved with the passing years.

enthusiast readers, by 1978 and 1981, respectively, will already have encountered the term and be fully comfortable with it. As such, I would conjecture that Black musicians from the worlds of funk, soul, and R&B, and a short while later in jazz, were widely using the concept of the "pocket" by the early or mid-1970s. From here, the concept spread across popular music styles. The word may have entered the musicians' vocabulary on the bandstand from its positive association with security (from the financial meaning of money in the pocket—if you've got something in your pocket, you're doing well) and its use to describe a location within an expected region (imported from the quarterback's position among his teammates). Once established in the vernacular of performing musicians, the "pocket" and playing "in the pocket" was used to describe a positive, interpersonal, aesthetically desirable quality of music with particular emphasis on how time is articulated.

Three Key Properties of Pockets

From surveying some of the usages of the word "pocket," it is evident that pockets are a concept that is highly meaningful to an "in group" of performers, journalists, and enthusiast listeners. The following sections take this jargon and explore three facets of this complex concept, focusing on the performance and experience of the temporal dimension.[7] I define three key properties of pockets: they are *domains* of time, that have *probabilistic* profiles such that events within the domain are more or less likely to be described as/take on the function of "the beat," and that these domains of time have *qualitative associations* that are expressed through a language of musical "feel."

[7] "Pockets" are also used for other purposes by musicians, for example describing a drummer as a "pocket player" might suggest that they play simple drum patterns and play them well (like drummer Nate Smith, featured in the Loop Loft corpus that is explored here, below, and in the next chapter), rather than having a preference and notoriety for playing more complex patterns (like Jeff Porcaro or David Garibaldi, featured in the Lucerne corpus).

Domains

The first key element to the theory of pockets is viewing beats as domains of time, not instantaneous time points. The idea of beats as domains arises from occupying a different theoretical position to that of point-based theories, one that is much closer to the real, in-the-moment sound signal. When we zoom in on performances in the real, lived world, we recognize that they are inherently noisy and imprecise in many regards, including in the temporal realm. This imprecision is present when performances are by humans on physical instruments and even can be found in performances created by programing computers. This can be seen on a number of scales: first we can consider the asynchronies that happen when multiple people try to coordinate their sound actions; anyone who has heard an audience clap along with a performance will have experienced the cloud of claps that do not perfectly align. The imperfect alignment of group actions shown by the claps is not simply an artifact of the time it takes sound to propagate across a large concert theater or the unskilled/unpracticed actions of an audience, it can also be seen when an individual, expert musician is performing, as when one hand on the piano tends to play just slightly ahead of the other (Goebl, 2001; Goebl et al., 2010; Palmer, 1996; Vernon, 1936). On an even smaller timescale, that of the individual sound or note, musical events are complex and do not have a perfectly clear moment of starting and stopping. As Danielsen et al. (2019), Lartillot et al. (2021), London et al. (2019), and Nymoen et al. (2017) have shown, the P-center ("perceptual center," the moment at which the sound is perceived to occur) of a sound can vary widely through time depending on the specific qualities of the sound, for example, how rapid the attack phase of the sound is. Viewing beats as domains embraces the fuzziness of the real world instead of minimizing any timing differences.

The concept of beats as domains is built on viewing beats categorically. The pocket that can be labelled linguistically as, and assigned the musical function of, "beat 1" is a category within which many sound events may be contained, all of which can be associated with the identity "beat

1.” A categorical perspective draws together like events (within-category compression), bringing together the cloud of sound events associated with the identity “beat 1” under one identity. This drawing together under one categorical umbrella, however, does not erase the differences between sound events. A penguin and an eagle are both recognizable under the high-level category “bird,” but we are fully aware of the differences between the two animals. Even looking at just one family of eagles, no two bald eagles are the same, with differences in size, coloration, and behavior. Likewise, though the label and function “beat 1” may encompass a number of sound events, the members of this categorical pocket are still distinct. This is illustrated in Figure 2.1, which sketches a scenario where four individual instruments sound onsets. None of these four onsets perfectly align in time, but all can take on the identity “beat 1” because they are folded into that categorical pocket. Despite all sharing the identity “beat 1,” they are four distinct sounds that we can differentiate. Instrument 3’s second onset, however, falls outside of the domain and so does not get included within the pocket of “beat 1.”

FIGURE 2.1: A sketch that illustrates the *domain* quality of pockets.

Probabilistic

The pocket is a continuous category and, as such, elements may be more or less likely to be a member of the category encompassing one beat. There are degrees of membership and typicality in the functional category. Returning to the penguin and the eagle: both are in the "bird" category, but they have different degrees of membership in the "bird" category. An eagle is a more typical member than a penguin (for most people), but some people may actually have a dove, or a pigeon, or some other bird as an even more prototypical member of the category than eagle—for some person, the dove might be "more bird-y" than an eagle.

Accordingly, it is necessary to define two more important pieces of the theory of pockets: first, that membership of the beat category is graded—sound events may be more or less likely to function as part of "the beat" depending on where they lie in time and what syntactical function they express—and, second, that the boundaries of the pocket category are highly variable. It is not possible to prescribe strictly, for example, that a pocket may only last 5%/10%/25% of the inter-beat interval. The limits would be dependent on a number of factors such as context, style, genre, and instrument, and, for the sake of analysis, often a line must be intentionally drawn in the sand to say that, in this particular musical context, having done some preliminary analysis and reflection about the musical function of the individual onsets, the boundaries of the pocket will be X% of the inter-beat interval.[8]

In Figure 2.1, Instrument 3 has a second onset that fails to meet some criteria for inclusion within the beat category; here, likely, it is too far away in time. To explain this, Figure 2.2 provides an updated illustration of how pockets are both domains *and* have probabilistic properties such that an element is more or less likely to be contained within the beat category. The x axis continues to represent time, different colors are different instruments, and position on the y axis represents the probability that an onset is captured within the beat category. Thus, given this curve as a hypothetical probability distribution, the higher on the curve an event is, the more likely it is to be included int the pocket. The four events that are all contained in the pocket in Figure 2.1 all have high probabilities of being included in the pocket in this updated Figure 2.2 (represented by

[8] This parallels a quality of perceptual beat bins. Danielsen writes that "there must still be a limit: even a big bin has a rim" (Danielsen, 2018, p. 187). Just as the scale of "rim" of the beat bin is highly context- and perceiver-dependent, the "seams" of the pocket are also context- and performer-dependent.

the tall dashed grey lines that meet the probability curve). The fifth onset, from Instrument 3, which fell outside of the pocket in Figure 2.1's initial sketch, only has a very low probability (a short grey dashed line) of being captured within the pocket. Because of the graded membership in the categorical pocket, there is a chance that someone may choose to define that fifth onset as being part of the pocket, as the boundaries of a pocket are not strictly prescribed (perhaps doing so in a way that is informed by Lerdahl and Jackendoff's Grouping Preference Rule 2: Proximity [1983, pp. 45 & 345]); however, the probability of this occurring is very low.

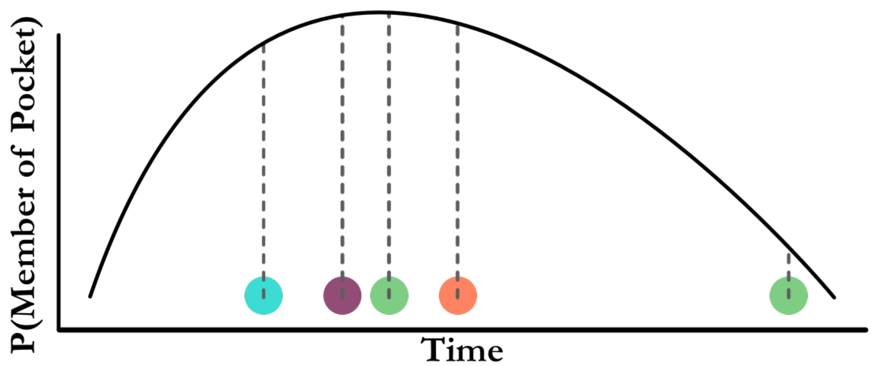


FIGURE 2.2: A sketch that illustrates both the domain and the probabilistic properties of pockets.

Qualitative Associations: "Feel"

The final element that the theory of pockets adds to our understanding of musical beats is that the shape, size, and location of the pocket can be associated with qualitative descriptors. In the etymological exploration of the word "pocket" as used by musicians, it was made clear that pockets are not simply some technical descriptor, but rather that they have musically meaningful significance. There are aesthetic qualities to pockets. A term that is intimately connected to pockets is "feel." Feel is explored in Chapter 3 as part of an analysis of individual drummers' performance styles, but briefly, for now: feel provides metaphorical language drawn from performing musicians that captures the experience of certain ways that musicians shape time. A performer might play with a "loose" feel, or a "pushed" feel, for instance. Stewart (1987) presents a whole range of feel descriptors in his "feel spectrum" (see Figure 1.5 in Chapter 1 for a reproduction of this spectrum). It is important to distinguish that this is not the same as "half-time feel" or "double-time feel," a different meaning of the word that has been used by performers and researchers with regard to a multitude of genres to describe a sensation of tempo change (half or double speed) without actually changing tempo (see, for example, its usage in bebop [Deveaux, 1997], contemporary pop music [de Clercq, 2016; Osborn, 2010, pp. 123–4], punk [Pearson, 2019], and metal [Garza, 2021]).

I argue that the aesthetic experience of "feel" arises at the meeting point of pockets and what Danielsen describes as "beat bins" (2010a, 2010b, 2015, 2018, see Chapter 1 for an overview). That is to say that feel is the unique perceptual experience—for performers and for listeners—that is afforded by the intersection of the domain of time that is articulated in the sound signal (the pocket) and the auditor's internal and unsounding "virtual reference structures" (Danielsen, 2006, 2010b). The perceptual dimensions of the beat bins are constantly informed by the pockets that are presented in the musical sound so the intersection is not a static, one-way relationship. Rather, the potential for meaningful "feel" is dynamic. A beat bin, in some ways, describes a span of time during which a listener's attention is elevated (drawing on Dynamic Attending Theory) and, as such, there is some expectation related to this period of time. The classic theory described in Meyer (1956) proposes that emotional and aesthetic responses to music derive from the thwarting or fulfillment of expectations. Feel is an aesthetic response that is tied to the shape of the pocket produced by sounding musical events and how this pocket fulfills or violates the expectation that the listener's internal beat bin describes.

Types of Pocket

Before providing evidence for viewing beats as pockets with a music information retrieval study on a corpus of drum performances, there are two types of pockets that need to be differentiated between: local pockets and cumulative pockets.

Local Pockets

Local pockets are pockets associated with a single metric position: for example, "the pocket associated with beat 2 of bar 6" or "the 'and' of 4 in bar 1's pocket." For non-metrical or un-metered music, a local pocket would refer to a single musically meaningful slice of time—not an instant, but a brief domain that has some musical significance. There may be any number of onset events that are associated with this moment in time, from no sounding events whatsoever to some maximum dictated by performance forces. For example, the first beat of the first full bar of Stevie Wonder's "Superstition" (analyzed in detail in Chapter 5) has a bass drum onset and a hi-hat onset. These two sound events make up the local pocket associated with metric position 1 of "Superstition." Figures 2.1 & 2.2 likewise sketch local pockets as they describe a single moment in time.

Cumulative Pockets

Cumulative pockets are the amalgamation of multiple local pockets. Which local pockets are brought together to define the cumulative pocket depends on what identity and function the cumulative pocket takes. There is a cumulative pocket that is the amalgam of every single pocket in an entire track. There is also the keyboard player's cumulative pocket for the entire track, or their cumulative pocket for just the verses. Or there is the rhythm section's backbeats, and many other possible cumulative pockets including a general pocket for a musician, that is, across time, over many performances, a sense of an individual musician's personal pocket. This personal cumulative pocket is what is described earlier when pianist Bobby Lyle describes how Sly Stone "gave everybody a really simple pocket to play into" (Underwood, 1978, p. 40)—Sly's pocket, which persists across performances, was renowned and a notable facet of his performance style for his band members. I propose that, if any slice of time is musically meaningful and of interest, it is possible to define a cumulative pocket that informs us of how its musical events articulate the domains of time that are the musical beat. Cumulative pockets become more and more clearly defined as a musical performance plays with each new local pocket, providing more and more information about the dimensions of the overall cumulative pocket. Most of the analyses in the chapters to follow consider cumulative pockets, though numerous local pockets are also evaluated in Chapter 5.

Evidence for Pockets: Drum Groove Corpora

To make the theory of pockets concrete and to begin to explore the real-world, sounding existence of them, I present a music information retrieval analysis of corpora of drum performances.[9] I use descriptive graphs that show the density of events to describe how these drum onsets are distributed through time. The spread and location in time of these distributions are not simply the result of noise in the human motor system or from any performance errors; rather they are evidence of consistent, conscious or subconscious musical behaviors by performers.

I focus on drum performances for a few reasons: drummers are often thought of as the "timekeepers" (Sheila E., 2020) or the "beating heart" (C. Schwartz, 1985, pp. 11–2) of a band, of being responsible for defining a temporal structure through playing a standard "boom tish" pattern against which other musical layers may be woven, and also because drummers are actually

[9] This set of performance data is an alternative version of one published in *Empirical Musicology Review* (Hosken et al., in press). The *EMR* data report and accompanying data repository (https://doi.org/10.17605/OSF.IO/3SEJT) are processed using a regression model to approximate metronomic beat locations. Here, the data sets are utilized in their raw, unprocessed format. For far more detail about the performances and the nature of the data, see the *EMR* report.

articulating numerous layers themselves, so it is interesting to investigate the role of different instruments and patterns that are all under one person's control. Drummers also are specialists in the temporal domain, not really having a strong melodic or harmonic side to their playing (although drum parts most definitely also make musical use of pitch, timbre, and sustain). And lastly, the envelope of the sound of a drum onset has a well-defined attack, which affords consistent measurement of onset times by music information retrieval techniques and, as will be discussed, closely relates to the perceptual onset time of the musical sound. As such, nuances in the shape of a pocket can be attributed to the musician's placement of the sound, not merely an artifact of the specific instrument's rise time or because of some inconsistency in the measurement method.

The Corpora

This analysis draws on three corpora of drum performances: one novel corpus of drum grooves collected by myself (The Loop Loft) as well as a repackaging of two existing corpora (Lucerne Groove Research Library and Google Magenta's Groove MIDI Dataset). Timing data about the three corpora have been obtained using different methods and each corpus's drummers performed under different conditions.

1. **The Loop Loft** is a commercial vendor that provides short recordings of musical performances ("loops") for DJs and producers to use in their creative work.[10] The company invites world-class performers into the studio to record short drum patterns while listening to a click track. Audio files for each microphone placed on each instrument within the drum kit are available allowing for clear identification of which drum was struck at what time. All audio is provided dry (no EQ, compression, reverb, or other processing). Information about microphone types or proximity of microphones to the drum heads is unavailable, so there may be very slight variances in the responsiveness of the microphone to sound and in

[10] https://www.thelooploft.com/

the time taken for sound to travel from the drum head to the microphone (though sound travels 1.13 feet/34.3 cm per millisecond and differences in microphone placement would be a matter of inches). The audio of the click tracks is not available, so the exact timing of click track events cannot be guaranteed; however, seeing as the intended use of the loops is to drag and drop them into Digital Audio Workstation workflows without any need to edit the timing, it seems safe to assume that no significant timing manipulations (e.g., adding time to or cutting time off the start of the track) are included that may corrupt the ensuing analyses. Each track title contains tempo information in BPM. Here, 1,467 tracks performed by four top-tier session musicians were purchased and were analyzed using the MIRtoolbox in MATLAB (Lartillot et al., 2008a, 2008b). Onsets below a threshold of 10% the maximum amplitude were discarded to remove bleed from other drums.[11] This provides a precise and replicable means of measuring timing, which would vary somewhat with parameter settings or with using different tools.[12]

[11]Onset times were recorded and filtered using the MIRtoolbox function: `mirevents('FILENAME', 'Threshold', 0.1);`.

[12]In the time since the Loop Loft corpus of onset times was originally created, a significant update to the MIRtoolbox has been developed and beta-released: the MiningSuite (Lartillot, 2019). Very generally, with default parameters, the onset detection function of the MIRtoolbox, `mirevents`, uses the envelope and not the waveform of the sound signal, which can lead to some incorrect values (Nymoen et al., 2017). The improved attack function in the MiningSuite uses the waveform for peak detection. To see whether there are any significant differences between the two measurement techniques, a subset of the Loop Loft corpus—Nate Smith's "Big Bass" package, comprising 551 hi-hat, snare drum, and bass drum onsets—was reanalyzed with both the MIRtoolbox and the MiningSuite (`a = aud.events('FILENAME', 'Detect', 'Attack'); b = sig.peaks(a, 'Threshold', 0.25); get(b, 'PeakPrecisePos');`). Pairwise t-tests show that there *are* significant difference between the two methods' onset times for each instrument (see below); however, the effect size is negligible for each comparison and the differences are small in absolute terms. Therefore, while the MiningSuite is a better-designed onset detection tool and would increase confidence in the findings related to the Loop Loft corpus, the mirtoolbox function is perfectly adequate for the present analysis and, since analyses in this and the following chapter are almost entirely within-instrument, the differences in precision by instrument are not a great concern.

- Hi-hats: $t(229) = -10.211, p < .0001, d = 0.00117$. Mean difference 5.045 ms (i.e., the onset found by the mirtoolbox is, on average, 5.045 ms after the MiningSuite).
- Snare drums: $t(169) = -66.445, p < .0001, d = 0.000670$. Mean difference 2.887 ms.
- Bass drums: $t(150) = -11.498, p < .0001, d = 0.00182$. Mean difference 8.169 ms.

In addition to computational methods, manual inspection of a number of waveforms in Logic Pro X confirmed that the `mirevents` function was not returning spurious results.

2. **The Lucerne Groove Research Library** is a corpus of 251 drum grooves drawn from commercial recordings played by 50 highly acclaimed drummers in the fields of pop, rock, funk, soul, disco, R&B, and heavy metal.[13] Two professional musicians transcribed the drum patterns by ear and manually identified each drum onset using spectrograms and oscillograms in LARA software.[14] Onset measurement is estimated to be accurate to ±3 ms for most of the music excerpts and, even in the most problematic cases, the timing measurement error is expected to rarely exceed ±10 ms (see Senn et al., 2018 for a full description of the method). Drum patterns are provided in MIDI- and MP3-format. Since these drum performances are part of full-band recordings (i.e. not just the drums in isolation) drawn from 1956 to 2014, it is not knowable whether a click track was used in the performance, nor the precise location in time of a click track if one was used. Despite this, it is still possible to analyze the temporal position of each drum onset relative to the others.

3. **Google Magenta's Groove MIDI Dataset** is a corpus of 503 drum patterns performed on a Roland TD-11 electronic drum kit by five professional drummers and four amateur players (Google employees).[15] Drummers, who are anonymized in the data set and referred to only by ID number, played on the MIDI drum kit to a click track. The TD-11 has a temporal resolution of 480 MIDI ticks per quarter note, so the lowest resolution (for a performance recorded at 50 BPM) is 2.5 ms and the mean resolution of all performances is 1.17 ms. The drummers performed drum patterns and solos for as long as they desired. This corpus was initially created as training data for a machine learning project into expressive drum performances (Gillick et al., 2019). The audio of the click tracks is not available, so the exact timing of click track events is unknown, though, given the MIDI data available, it is presumed that the files begin at time point 0 (i.e., the provided MIDI data aligns

[13] https://www.grooveresearch.ch/
[14] https://www.hslu.ch/lara, version 2.6.3
[15] https://magenta.tensorflow.org/datasets/groove

perfectly with the unavailable click track). The titles of the individual recordings provide information about the tempo of each track.

Data Filtering

The onsets of all corpora were filtered according to the exclusion criteria summarized in Figure 2.3. The differences in how the filtering process affected different corpora may be attributed to the unique nature of the performances captured by each set. For example, the Lucerne corpus is comprised entirely of the "core grooves" of famous songs in typical rock/pop styles and so no complete tracks were excluded from the set. The Loop Loft corpus, however, features numerous tracks explicitly labelled "Fill" and numerous tracks in Latin American styles, so a number of complete tracks were excluded. The Magenta performances are in a range of styles (hence the drop in complete tracks), often feature extended drum solos during the course of a track (so several bars are filtered out), and the grooves sometimes develop into patterns that are based around the tom-toms (these bars are filtered out).

Overview of Performance Data

In-depth analyses of one corpus of performance data—the Loop Loft—will be presented in Chapter 3, but a general overview of all of the performance data is presented here to illustrate some of the fundamental properties of pockets that are explicated above. First, I concentrate on *cumulative* pockets, aggregating onsets across multiple metric positions and across multiple performances. It is still important, however, to explore what a *local* pocket looks like in real performance. Homing in on a randomly selected example from one of the corpora: in Nate Smith's "Chorus 1" performances contained in the "Pure Pop" package from the Loop Loft, he plays a bass drum at 1.101 seconds and a hi-hat at 1.102 seconds. These two events are associated, functionally, with metric position 3 of the first bar (i.e., the third beat of the first bar) and so the

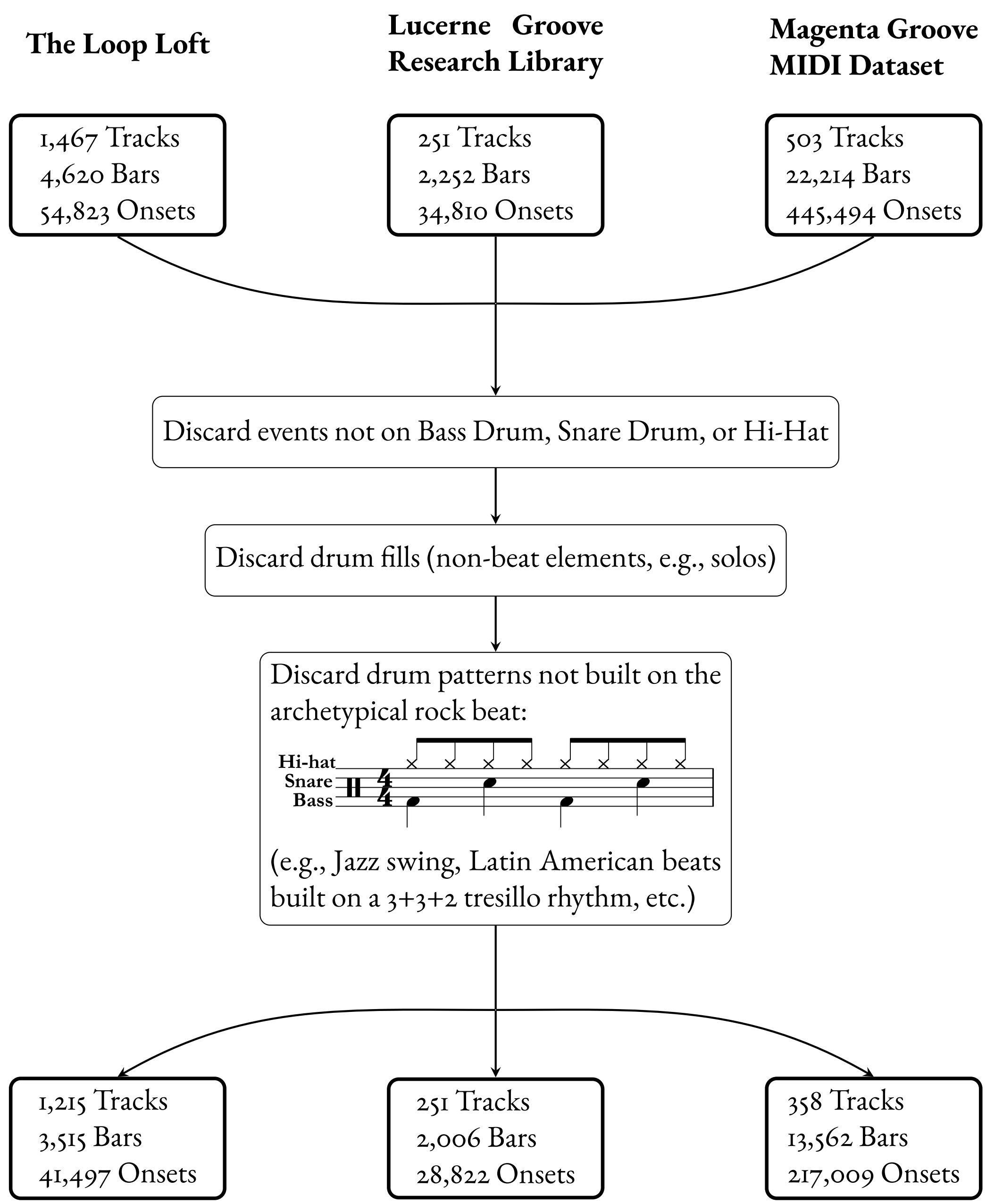


FIGURE 2.3: Schematic representation of the data filtering process.

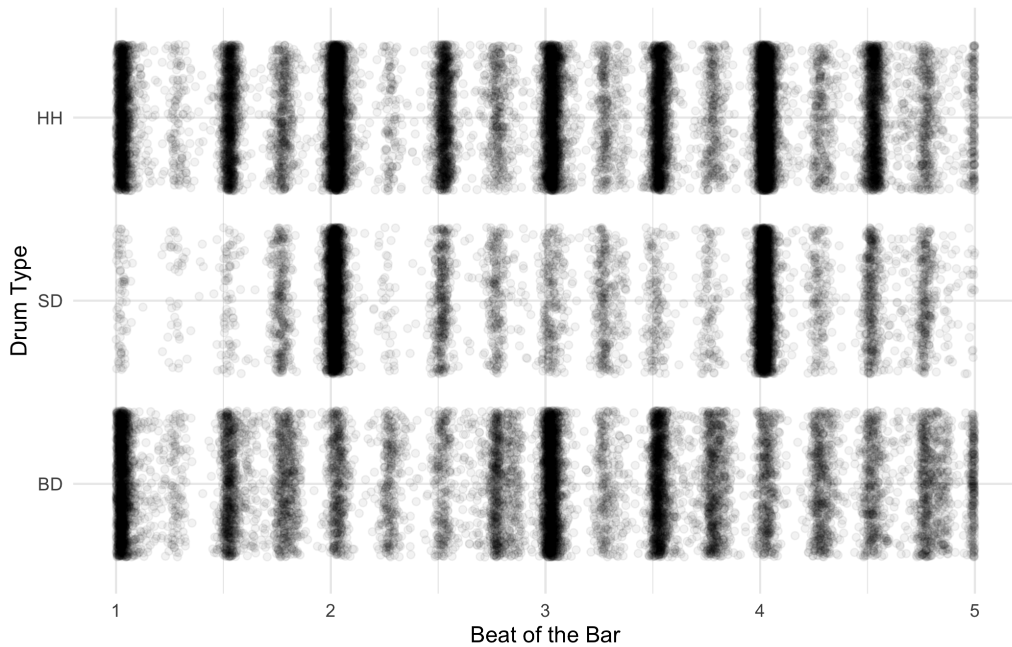
associated local pocket is a domain that encompasses these two onsets. We do not have any further information, so cannot say that the local pocket starts at 1.101 and finishes at 1.102 seconds, just that the domain of time that constitutes the local pocket includes these time codes. A visualization of this would look like Figure 2.1, though the left and right bounds of the pocket are undefined and therefore the pocket's shape and temporal limits are unknown.[16]

Looking more globally at the collected performance data, Figure 2.4 presents three plots with all of the onsets for each corpus. Each small, faint black dot on the strip plot represents an onset. The vertical height within each strip is functionally meaningless, being simply a way to spread the data out visually and allow events that would lie on top of one another to be more easily seen. Even so, darker areas in the plots are the result of numerous faint dots lying on top of one another. As such, darkness may be used as an indication of how many events are occurring in that region of time. Using this interpretation, we can observe that the typical backbeat drum pattern is clearly visible in each corpus's strip plot, with a large number of bass drums occurring around beats 1 and 3, a large number of snare drums around beats 2 and 4, and eighth-note hi-hats throughout the bar. Most importantly, we can observe that there is *width* to these important locations in the bar. Onsets occur *in the region of* metric positions, but are not absolutely perfectly in line with them or with one another.[17]

The various strip plots presented in Figure 2.4 provide a general impression that all of these performers (62 unique drummers across the three corpora) sound out the core drum patterns in such a way that there is *width* to the beat. It can be seen that the beats as performed in the three corpora are not intersubjectively agreed upon points in time, but instead are small domains of time. However, these strip plots, containing every single onset, contain too many details of what is occurring at these key metric positions to read and interpret clearly. As such, I frequently visualize pockets in this dissertation using "density plots," which are constructed by collecting the time code for every individual drum onset, measuring how densely packed events are at each moment

[16] A number of events that can be interpreted through the lens of local pockets are visualized in Chapter 5. This describes a full band situation—the opening measures of Stevie Wonder's "Superstition"—and so there are more musical events to help refine the dimensions of the local pocket.

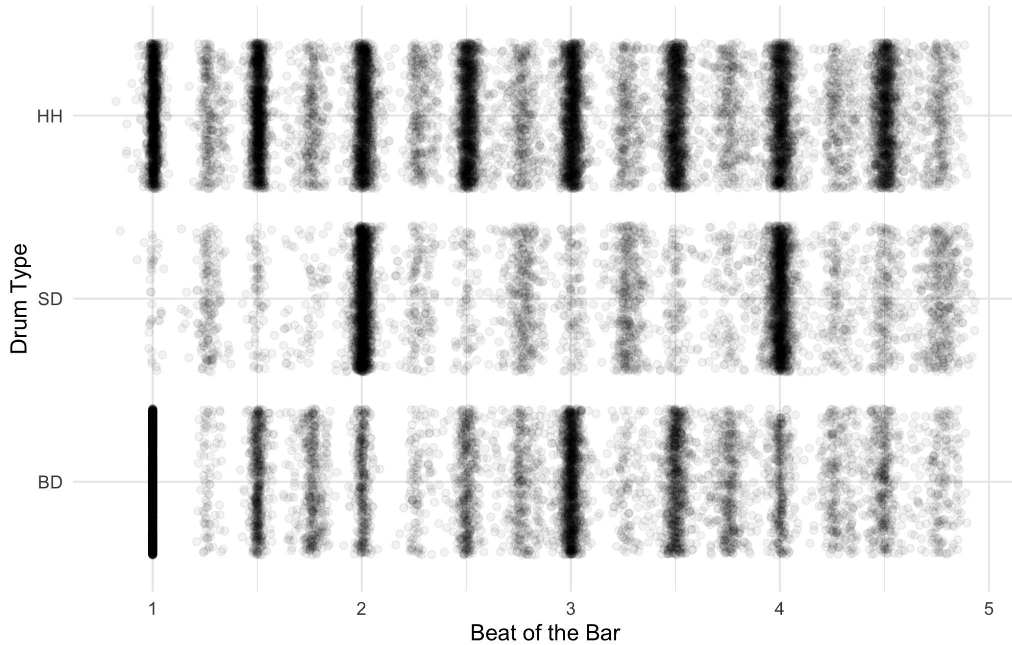
[17] Lucerne's bass drums on beat 1 are an exception to this (see the bottom row of Figure 2.4.2). This black line suggests perfect alignment with metric position 1, most likely due to how the data were collected and represented in the Lucerne set (there is no click track information available for this entire corpus so the researchers appear to have made a decision to define the first bass drum onset as the start of the track, time code 0:00. This is not discussed in the methodology described in Senn et al., 2018).



(2.4.1) Loop Loft Corpus.

in time, and then plotting a smooth curve that represents this (see illustration in Figure 2.5). If the drummer plays a lot of events around beat 3, for instance, the graph will show this as a higher density and so a higher curve. An added benefit of visualizing the drummers' performances in this way is that the area under the curve is always 1, and so the different number of onsets available for each drummer and each drum does not matter as the data are scaled. This facilitates easy visual comparison between drummers and drums (a property that is utilized in the analyses presented in Chapter 3). It is important to remember, however, that density plots are merely a way of representing the sound and providing a visualization of the concepts expressed by the theory of pockets. They are representations—one of many possible ways of conceptualizing sound events.

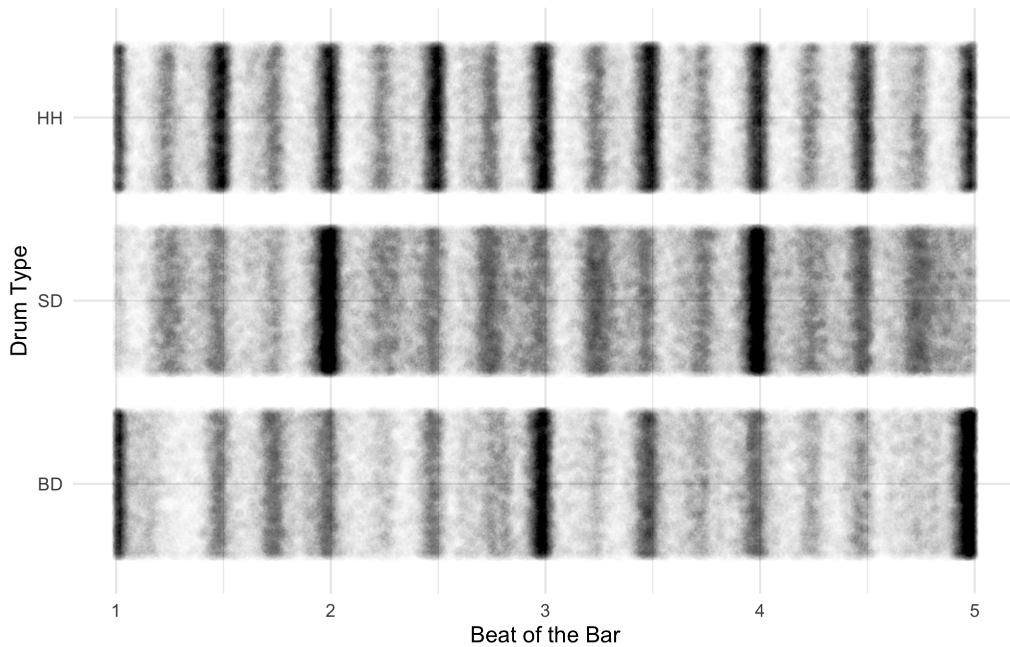
An example will show how a density plot allows us to zoom in on the details of how the beat is performed and so reveal the construction of the pocket. Figure 2.6 focuses on the density of snare drum onsets in the Loop Loft corpus that occur around metric position 2. This summary of 2,772 onsets gives a well-defined image of the details of the pocket that is performed by these



(2.4.2) Lucerne Corpus.

drummers. Starting with a general look at this figure, we can see that the snare drum events are not randomly dispersed throughout time (this would result in a plot that looks more rectangular than triangular); instead, there is a clear picture of attack's organization. The drummers are articulating beat 2 of the bar with their snare drum, though there is spread to where they actually sound this event: snare drum 2 occurs within a domain of time. If all attacks were perfectly synchronized with a metronome, we would just see a vertical line at position 2. The width to the distribution is also not a side effect of, for example, a sixteenth note being played just before the main onset on beat 2, as may be the case in a shuffle drum groove. Rather, this density plot only looks at events ±0.1 beats. A sixteenth-note anticipation would be in the vicinity of 0.25 beats away from the metronomic position and thus is excluded from consideration here.

Looking for the moment at a few details of the shape of this pocket (full analyses of the shape of numerous pockets are presented in Chapter 3), the peak of the distribution—the most dense area where the largest number of snare drum onsets occur—is located slightly after what I infer



(2.4.3) Magenta Corpus. (There are over 200,000 onsets represented here, hence the hazy appearance)

Figure 2.4: Strip plots showing all onsets in each of the corpora. HH = Hi-hat, SD = Snare drum, BD = Bass drum.

as the metronomic beat 2, suggesting that a large number of snare drums are not locked to the metronome, but instead sound just after it. While the snare drum beat 2 pocket has some width to it, it is notable how steeply it rises and falls. The domain that is articulated as beat 2 is less than 0.1 beats long. The swift increase and decrease in the density of events associated with beat 2 aligns with a key feature of categorical cognition—that there is strong within group identity and between group distinctiveness (Clarke, 2000, §3). Another detail that is visible in Figure 2.6 is that the pocket looks relatively symmetrical, though there is a "kink" in the slope up to the peak (around 1.975) that is not found on the mirroring side. This "kink" is the result of a number of onsets occurring around this position, suggesting that a portion of the onsets actually precede the click track's beat 2. This is not simply the result of occasional inaccuracies by the performers; even if we allow for some errors by the top-tier professional musicians included in the Loop

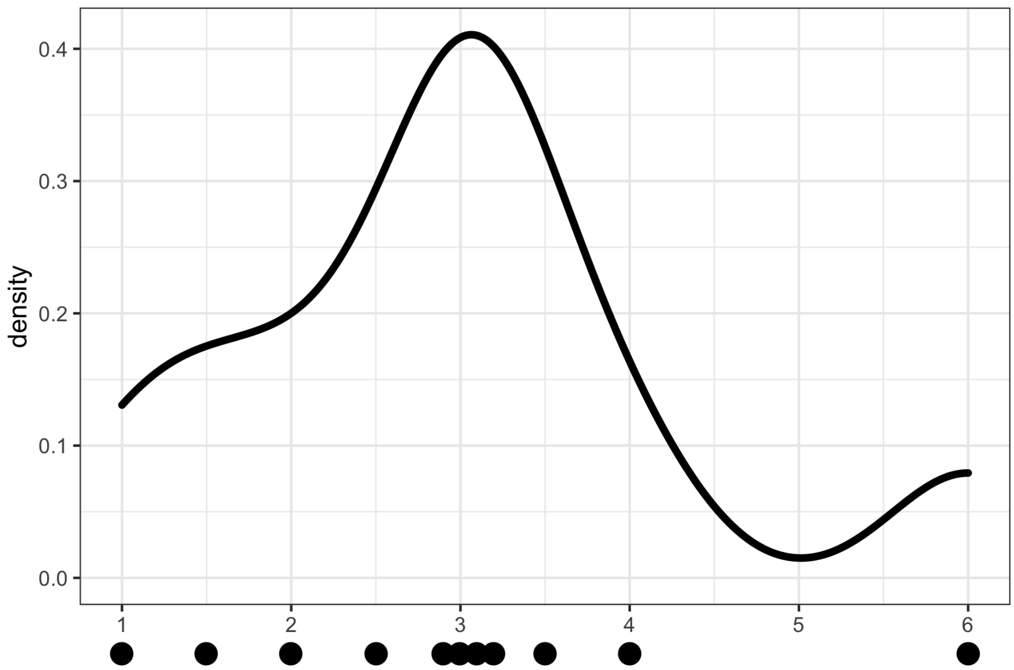


Figure 2.5: An illustration of how density plots are constructed. Beneath the *x* axis, the black circles each represent an event (here, it would be the onset of a drum). An event was recorded at time 1, 1.5, 2, etc. Where there are more circles—for example the more dense collection of events around 3 on the *x* axis—the density plot is taller. Where the density of events is more sparse—no events between 4 and 6—the curve is lower.

Loft corpus, these would nearly vanish when looking across 2,772 events. The domain of time, although it has width to it, has a clear identity and function as "beat 2." If we constructed a similar density plot for any other key metric location of the bar, for any instrument in the corpus, for any drummer, from any of the three corpora featured here, a pocket with somewhat similar shape would be found (see examples in Figure 2.7). However, depending on the performer, tempo, genre, intention of the performer, instrument(s) being evaluated, and a host of other factors, the density plot would of course vary in a number of dimensions. If the performance of a Linn LM-1 Drum Computer (a highly coveted drum machine from the 1980s) was analyzed, the width of the domain would be far smaller because of the accuracy of the computer behind the performance, though the pocket found would not be a singularity; technology, especially older technology, does exhibit variations in timing depending on temperature, electrical supply, and condition of the instrument. More modern digital instruments also often intentionally include "jitter" that varies timing ever so slightly.

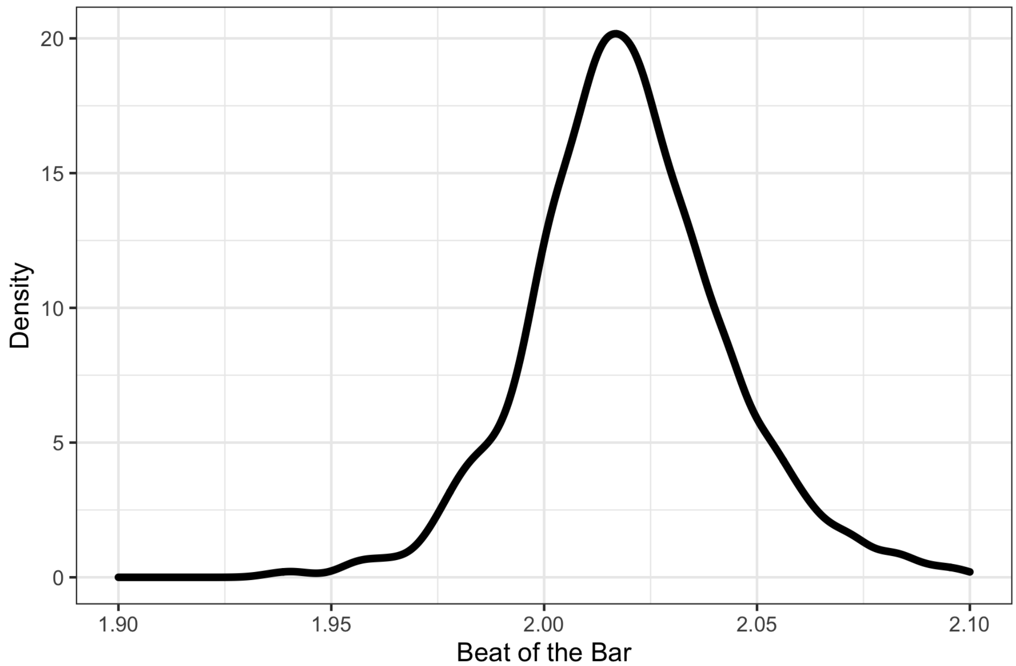
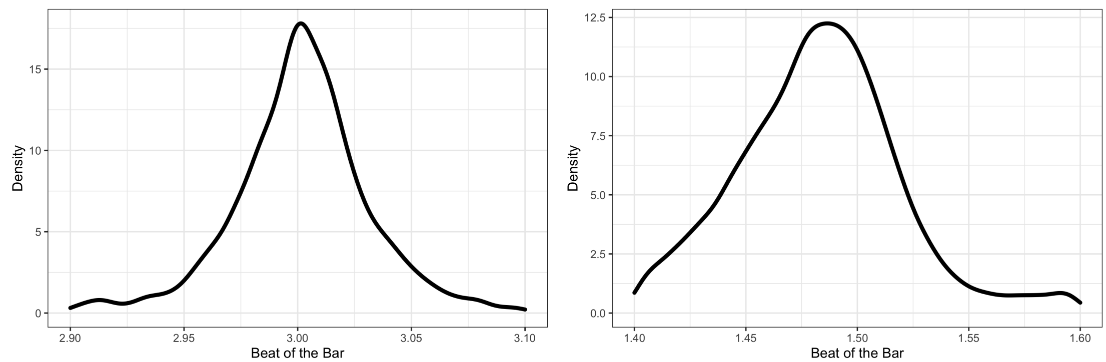


FIGURE 2.6: Density plot of Snare Drum 2 in the Loop Loft Corpus. For this illustration, all onsets between metric positions 1.9 and 2.1 (as defined by the click track) are included.

Summary

This chapter has introduced the theory of pockets—a way of viewing beats as domains by focusing on sounding performances. The choice to call this a theory of "pockets" emphasizes how the ideas presented here draw on the experience of performers, performers who already have a sophisticated vocabulary for describing the nuances of sounded musical time. I have described three key properties of pockets in this chapter: that pockets are domains of time, that these domains are categorical and membership in the category is graded (i.e., there is a probabilistic element to the pocket such that some events are more or less likely to be contained within the



(2.7.1) Lucerne Bass Drum Beat 3.

(2.7.2) Magenta Hi-Hat Beat 1.5.

FIGURE 2.7: Example density plots from the Lucerne and Magenta Corpora.

beat domain), and that the dimensions of the domains—their size, shape, and location—can be associated with elements of the qualitative experience of music (described as the “feel” of the performance). Each of these properties is investigated in far greater detail in the following chapters where I present analyses of a variety of musical performances, on top of the overview of three corpora of drum performances presented in this chapter.

The theory of pockets is important as it offers a view of beats that hews closer to the acoustic signal than many canonic theories of rhythm and meter. This proximity to the sounding musical experience compliments and augments existing theories by presenting a perceptually driven theory of musical time that embraces the variability of human activity. By zooming in on the variability and fuzziness of performed musical time and placing the subtle nuances of sounded music into the spotlight, fresh perspectives on much-theorized musical phenomena may be presented. Providing a foretaste of the analyses and examples to come in the following chapters, the theory of beats as domains offers a means by which the contradictory status of (temporal) theories of groove may be resolved: by reframing attitudes towards participatory discrepancies, syncopation, and swing away from deviations from an absolute, metronomic temporal location, and instead viewing the performance of musical time as probabilistic and a phenomenon that develops with experience through time. I propose that in this way the idiosyncrasies and variability in the groove experience may be better explained. By conceiving of beats as domains, as argued here, nuances in performed time (e.g., participatory discrepancies and expressive timing) are reframed away from "deviations from the exact" (H. G. Seashore, 1936, p. 155), and instead positively appreciated for the subtle, but important, ways they shape the sounding musical experience.

CHAPTER 3

Individual Performer's Pockets & Feels

> Time is one of the most defining elements of a player's style. Your own unique sense of time is as capable of setting you apart as your tone, vocabulary, gear, or speed…Pocket is really about style.
>
> DeCaro, 2017, p. 5

> Feel…relates specifically to the one-of-a-kind stamp and personality of a musician's groove or pocket.
>
> Drummer Van Romaine, interviewed in Sturgis, 2021, p. 94

Performers, fans, and journalists in the popular music and jazz scenes have a well-established language that points to the musical significance of differently shaped beat pockets. They compliment a bassist for "laying back in the pocket" or note that a hallmark of a player's style is their ability to "play on the edge" of a pocket (for interviews with several musicians who use this language, see Haid, 2007a, 2007b). Implicitly, this vernacular language recognizes that the way that a performer constructs the shape of their pocket—the probabilistic domain of time that is a musical beat—impacts the qualitative "feel" of a performance. Music producer Michael Stewart, back in 1987, implying this language is long-established, suggested a way of formalizing these musical intuitions, connecting qualitative descriptors to millisecond deviations from a metronome to define a "feel spectrum" (replicated in Chapter 1, Figure 1.5), and in so doing implied that this language was already established. Stewart's model of musical feel looks far more like the traditional, instantaneous model of beats I have critiqued in earlier chapters with its quantified deviations from an exact moment and is limited in its generalizability, yet there is something appealing about how Stewart has made his subjective experience somewhat concrete. This chapter uses the pocket as a vehicle for discussing connections between performance nuances and the subjective, qualitative experience of individual performers.

Certain musicians are famous for, and indeed sought out for, their signature pockets. For instance, Bernard "Pretty" Purdie, self-described as the most recorded session drummer in the world ("Bernard "Pretty" Purdie Official Website", 2021), has a pocket that one journalist believes adds "force, nuance, and forward locomotion," that inspires dancing, and has "define[d] what the great American R&B-rooted groove feels and sounds like" (Potter, 2012, p. 56). More generally, as seen in this chapter's epigraphs, active musicians Dru DeCaro (guitarist and producer) and Mike Sturgis (drummer) both point out that a performer's sense of time is a key, defining feature of their style and is capable of setting one performer apart from another, a sentiment concisely mirrored by Prince's drummer Michael Bland: "your feel is like your face" (bdrum50, 2008, 2:37). This is a central motivation for developing and utilizing the theory of pockets: What is it about performer X that distinguishes them from performer Y? What causes someone to hire or want to play with X rather than Y when, speaking very simply, a four-on-the-floor drum beat is a four-on-the-floor drum beat, a *son clave* is a *son clave*, a dembow riddim is a dembow riddim, and so on? The concept of "feel" is ever present for drummers whose role is often to play repeated standard patterns, though it is also highly relevant to any performer in styles that involve playing repeated riffs or ostinati.[1] In all of these cases, it is not just *what* is performed, but *how* it is performed. The theory of pockets that I present provides an analytical lens through which performer's personal time feels can be understood.

This chapter uses computational methods to analyze drummers' pockets and empirically describe individual performance styles and time feels. Broadly, I ask: How does one particular drummer articulate musical time and how does this differentiate them from another drummer? I investigate this using the Loop Loft corpus described in the previous chapter, looking at all drum events for each of the four drummers (Matt Chamberlain, Omar Hakim, Nate Smith, and Joey Waronker). This will then be examined further to consider how the pocket's shape is affected by the way each drummer utilizes the different instruments of the drum kit: Specifically, can a drummer's feel be codified through a preference for, e.g., locating the snare drums in the middle of the pocket while bass drum onsets are laying back? Digging one step further into the analysis, I will also consider how pockets are arranged throughout the bar by locating *intra*-pocket details within *inter*-pocket structures. Moving outside of individual beats and looking at how pockets interact with meter develops the ability to describe a performer's unique feel more specifically and to engage with a number of questions such as:

- How does a particular drummer's One (beat 1 as an essential point of stability that initiates the groove, defined by James Brown— Danielsen, 2006; R. J. Smith, 2012) differ from the rest of the bar?

- What role does the shape of backbeats (beats 2 and 4) play in defining a feel or individual's

[1]Consider, for example, the detail of the constant sixteenth-note strumming that is the "fingerprint" of "bona-fide virtuoso rhythm guitarist" Cory Wong (Gardner, 2019). Or the "thumping foundation" and "signature sound" (Grimes, 2011) that Johnny Cash's bassist Marshall Grant provided when performing bass lines that were often constrained to root–fifth alternation. These signature sounds certainly include facets of the artists' performances beyond their rhythmic playing (Wong's sound is also a result of setting his pickup selector switch on his Fender Stratocaster to fourth position, which takes the sound signal from the middle and neck pickups), but each of these musicians build their feel foundations in the temporal domain.

> style (see Biamonte, 2014, §6.1 vs. Butterfield, 2006)?

The purpose of these analyses is to produce a thorough and empirical methodology that enables the colloquial qualitative descriptors of performance style to be made quantitatively concrete in the timing domain.

Background and Motivation

Music theorists have well-developed languages to describe the "primary" parameters of music: melody, rhythm, and harmony (Meyer, 1989, p. 14). That these theoretical languages are well-developed also tends to privilege their use. This is the central tenet of Charles Keil's opposition to Leonard Meyer's *Emotion and Meaning in Music* (1956), reflected in his article's title "Motion and Feeling through Music" (1966). Keil argues that Meyer's privileging of primary parameters and his binding of form with expression falters when considering musical traditions beyond Western Art Music. He writes: "When, however, this equation [of form and expression] and the corresponding evaluative criteria are applied to non-Western styles or to certain Western compositions *in performance*, we often find that something is missing" (Keil, 1966, p. 338, emphasis in original). Keil is echoed by several more recent theorists and musicians who seek answers to questions about their experiences of music beyond the Western Art Music canon. Zbikowski writes that we need to break out of "limiting ourselves to conventional music-theoretical constructs (and their representations)" (2004, p. 273) in order to understand groove-based musics properly. Zbikowski's solution is to try to understand musical feel from a cultural anthropology perspective. He explores the interpersonal, dynamic network of bodily knowledge that provides people with a conceptual model, which allows them to understand, without necessarily being able to articulate, such a complex phenomenon as groove or feel. Iyer has a critique similar to Keil's and Zbikowski's, writing that:

> linguistics-derived musical grammars do not apply well to the vast majority of other

> genres of music. This nontranslatability is quite glaring in the cases of African-American forms such as jazz, rumba, funk, and hip-hop. In these cases, certain salient musical features, notably the concept of groove, seem to have no analogue in rational language.
>
> Iyer, 2002, p. 388

He argues that there are "disparities [in] the status of the body and physical movement in the act of making music" (Iyer, 2002, p. 388) in different musical cultures and so these should be reflected in the tools used to understand the organization, production, and cognition of music.

Analyzing music through alternative lenses values facets of the musical event that are neglected or underserved by "traditional," linguistics-derived analytic tools. As such, and sticking with Meyer's terminology, we should turn our attention to the "secondary," processual parameters of music. These are parameters that "cannot be segmented into perceptually proportional relationships" and "tend to be described in terms of amounts rather than in terms of classlike relationships" (Meyer, 1989, pp. 14–15). Very generally, these are features that cannot be written down on a score precisely, such as louder/quieter, faster/slower, thicker/thinner. Prioritizing that which cannot be notated on a staff allows, for example, the 10-minute long vamp of a two-bar idea that makes up The J.B.'s "Doing It to Death" (1973) to be understood not for its paucity of primary parameters, but for the enticing, bodily, *subjective* detail of its performance nuances (Hosken, 2020).

Kvifte (2004) offers an important nuancing of this supposed divide between the syntactical and processual elements of the musical experience. He argues that "processual descriptions must be understood in relation to syntax" (p. 54). This is informed by his earlier work (e.g., Kvifte, 1989, p. 110ff.) on categorical perception where he conceives of both "analog" and "digital" components to perception that are roughly equivalent to Meyer's syntax and process (respectively): A listener may comprehend that they are hearing a D (analog/syntax) while simultaneously being aware that it is sharp (digital/processual).[2] The theory of pockets presented here offers a way to decenter traditional music analysis by emphasizing performances and providing a methodology that facilitates insight and appreciation of the fine details of a performance, placing the spotlight firmly on the processual details in ways that make meaningful the analog/syntactical structures they are located within.

There is a long-established community of music theorists, music cognition researchers, and musicologists who do take performed music—live or on record—as their principal focus.[3] Dating back to Binet and Courtier (1895), properly commencing with Carl Seashore (1938), and growing significantly since the 1980s, secondary, continuous parameters of music such as pitch, dynamics, timbre, and particularly time have been systematically explored using various technological means. Such research on what may also be described as "processual" features includes work led by Ingmar Bengtsson and Alf Gabrielsson on "SYVARD" at the rhythm research project in Uppsala.[4] This theory hypothesizes that the live performance of musical rhythm is characterized by various consistent, systematic variations that deviate from the mechanical "norm" (key citations include Bengtsson & Gabrielsson, 1974, 1977, 1980; Bengtsson et al., 1978; Bengtsson et al., 1969; Gabrielsson, 1982; Gabrielsson et al., 1983). The Uppsala research group focused principally on Swedish folk songs and waltz performances, though they also claim that the stylistic rhythmic "dialects" (special ways of nuancing the timings of patterns of sound) "have been shown to exist in all cases investigated to date" (Bengtsson et al., 1978, p. 15). Outside of SYVARD, others have analyzed the fine details of performance timing such as the specific metric patterning in Norwegian folk dance styles (*gangar*, Blom and Kvifte, 1986 and *springar*, Groven, 1971; Haugen,

[2] See also Clarke (1987b) on the importance of the "remainder" in categorizing musical time.

[3] Cook (2012, §1-2) notes how scholars have taken different approaches to the study of musical performance as a result of nations' differing institutional priorities. Of particular note is the impact, in the early era of performance analysis, of the absence of a music-theoretical mindset in the UK. Meanwhile, researchers in the US built on established approaches to scores from the well-entrenched music theory tradition (see, for example Berry, 1989). This helps understand the history behind the critiques of approaches to the analysis of music by Keil, Zbikowski, and Iyer (described above).

[4] Systematic VARiation as regards tone Duration.

2014; Johansson, 2010; Kvifte, 2004), the inter-onset-intervals of jazz swing (Butterfield, 2010, 2011; Honing & Bas de Haas, 2008; Prögler, 1995; Waadeland, 2001), the (provocative) theory that there are composer-specific patterns of expressive microstructure (Clynes, 1983, 1986; Repp, 1989, 1990a), and a large body of work on expressive timing in Western Art Music, particularly piano performances (Gabrielsson, 1999; Repp, 1989, 1990b; Rink, 2002; Shaffer, 1981).

While these performance studies researchers have much to offer and are drawn into the music-theoretic fold on occasion, they do seem to have to continually make the case for their relevance in calls to action (Cook, 2013; Cook & Pettengill, 2013; Leech-Wilkinson, 2012), suggesting that their demonstrations of the rich value of analyzing performances are not being recognized and drawn into music-theoretic discussions. We also find that, most typically, the spotlight of performance analysis is on Western classical performance traditions and the findings are often interpreted (overtly or tacitly) through the lens of deviations from or variations of a discrete, syntactical "norm" (see, for example, the considerable number of papers on performances of just one piece—Chopin's Prelude Op. 28 No. 4 in E minor: Clarke, 1995; Rink, 2001; Senn et al., 2009; Sloboda & Lehmann, 2001; Thompson & Luck, 2012). This reveals an epistemology in which there is a fixed musical object, which is denoted symbolically in notation by a musical score. From this score, a performer seeks to realize the composer's intentions (criticized by Cook as "from page to stage" [2012, §2]). However, even within the confines of Western Art Music, analyses of historic recordings foreground how, while the musical notation may be the same, performance practice changes over time. There is considerable scope for expressive freedom even when there is evidence that some syntactic features of a score do constrain the expressive patterns observed in individual performances (Repp, 1992, 2000). For instance, analyses have shown how the processual means by which a harmonic crux or pitch-height zenith are expressed can and do change over time (Leech-Wilkinson, 2009, Chapter 4, especially §16–19). Likewise, Sapp (2006) has shown how performances of Chopin Mazurkas changed across the 20th century with expressive performance decisions typically operating at the moment-level in the early part of the century and at the phrase-level in the latter part. And Leech-Wilkinson (2006, 2007, 2009) has illustrated how performers' use of portamento and vibrato changes across the 20th century as listener apprehension and understanding of these techniques transform with cultural shifts.

This chapter focuses on, and provides methodologies to analyze, the processual timing details of musical styles that are rarely realizations of a written text and whose syntactical elements are defined and prioritized differently to Western Art Musics. Recorded sound does affix a performance into a kind of musical object whose ontology has been debated (S. Davies, 2001; Doll, 2018; Gracyk, 1996; Kane, 2018), but the focus here is on aural—not written—culture. As such, and to echo Keil, Zbikowski, and Iyer, we need to place renewed emphasis on "secondary" parameters of music in order to engage meaningfully with music of the African diaspora (and most other cultures outside the Eurocentric Western art music tradition).

Groove research is a field that has done a significant amount of work to recenter attention on performance, on process, and on diverse listener experiences by dwelling on *how* the music is performed and, vitally, how it may be perceived. Books including Butler (2006), Danielsen (2006), and Ohriner (2019), as well as the collection *Musical Rhythm in the Age of Digital Reproduction* (Danielsen, 2010c), have shown how artists in genres like electronic dance music, funk, and hip hop can make relatively scant, repetitive materials come alive in ways that engender pleasurable, corporeal engagement.[5] For example, when Butler transcribes EDM tracks (2006, Appendix C), the entire contents of a 6+ minute track can be represented in staff notation in just a page and a half; however, Butler explains how the groove is unlocked in the listening experience as a result of how a listener engages with the fine details of the musical layers, with these layers' interactions, and also how the performed music interacts with listener expectations. He explores the multiplicitous musical experiences crafted from minimal musical materials, concluding that

[5]This pleasurable urge to move in response to music is now a widely accepted, canonic definition of groove (Janata et al., 2012; Madison, 2001, 2006; Senn et al., 2020). While important, it is important to remember that there are numerous dimensions to the groove experience (Hosken, 2020; Pfleiderer, 2010).

repetition and processive development "invite[s] listeners to seek out diverse ways of hearing and to experiment with these interpretations as a piece is going on" (Butler, 2006, p. 256).

In a manner that prefigures much of this chapter's analyses, Anne Danielsen (2010a) has analyzed the beat locations afforded by the instrumental layers in D'Angelo's "Left and Right" (2000) to show how the band created the "seasick time-feel" that is one of the central stylistic features of the whole album.[6] Danielsen's analysis shows how the guitar and percussion parts at the start of the track clearly define for the listener a stable sense of where the beat is. When the bass guitar and bass drum enter, however, they define a beat location that is significantly earlier than that which the guitar and percussion had defined. Rather than being perceived as a mistake or frustrating the listener who seeks resolution of this difference, Danielsen argues that these potentially competing definitions of beat locations eventually get absorbed within "extended beats" and "merge into a distinctively organic, swaying musical whole" (Danielsen, 2010a, p. 21).[7] Danielsen's analysis shows how D'Angelo's "seasick time-feel" that listeners recognize and enjoy is created through events that, despite being associated with the "same" syntactical positions in the listener's internal metric reference structure (beat 1, 2, etc.), are performed asynchronously and define conflicting locations for the metric anchors of the bar. In this way, Danielsen has laid the groundwork for the present chapter by illustrating that it is possible to meaningfully link the microtiming details of a performance, particularly with regard to how beats are sounded, with one's expressive qualitative *experience* of a piece of music.

[6]This was an intentional artistic choice for D'Angelo who sought a "loose, way back in the pocket feel," also described as a "rubber band feeling" (King, 2013). D'Angelo was inspired by producer J Dilla, renowned for how he "loosened his beats from their rhythmic bedrock; they were not rigid, but gambled forward with a woozy lilt" (D. Schwartz, 2019)—see, for example Dilla's "E=MC2" from *The Shining* (2006), his work on The Pharcyde's "Runnin' " (1995), and on albums like Slum Village's *Fantastic, Vol. 2* (2000) and Erykah Badu's *Mama's Gun* (2000). To achieve this feel on his *Voodoo* album, D'Angelo told the musicians how he wanted them to articulate time, for example instructing drummer Questlove (Ahmir Khalib Thompson) that "I need you to keep the pocket but don't drag behind me, but play a little crooked" (King, 2013).

[7]This is part of her concept of "beat bins" that the present theory of pockets extends (see Chapter 1).

Analysis

Câmara et al. (2020a, 2020b), Danielsen, Waadeland, et al. (2015), Kilchenmann and Senn (2011), and Sioros et al. (2019) have demonstrated that musicians who are fluent in popular music and jazz performance styles can operationalize different feel terms, exercising deliberate control over their performances in a consistent way to suggest to a listener or band member a particular feel. These studies provide in-depth analyses of drum kit, electric guitar, and electric bass performances in three conditions: laid back, on the beat, and pushed. They show how performers nuance the timing, timbre, and loudness of onsets depending on the intended performance style. The methodology of instructing a performer to play laid back/on the beat/pushed parallels that used in expressive performance studies that direct performers to play deadpan or mechanical/normal/exaggerated (for example, Bengtsson, 1987; Bengtsson and Gabrielsson, 1983; Palmer, 1989; C. E. Seashore, 1938, 244ff. and also in audio-visual paradigms Davidson, 1993, 1994, 1995, 2007; Vuoskoski et al., 2014). These studies have all shown how musicians can intentionally and artistically deviate from a literal interpretation of a written text.

However, this methodology, whether applied to expressive piano performance or to drum/guitar grooves, requires a performer to play in an affected manner. The analyses presented here investigate whether microtiming correlates of musical feels can be found in "natural" performances, with a specific focus on whether *individual* performers' pockets can be discerned. That is, when looking at how individual performers arrange their onsets to create their personal sense of time, do these pockets meaningfully differ in location, scale, and/or shape? The following analysis asks, statistically, whether we may make a musical argument that, for example, drummer X is more "laid back" than drummer Y or that Y has a "looser" pocket than X.

Materials and Method

I will focus on the Loop Loft corpus for the present analysis, though my intention is to demonstrate a replicable methodology for understanding and differentiating between unique performer's pockets. To recap briefly, the Loop Loft corpus features numerous performances by four world-class session drummers: Matt Chamberlain, Omar Hakim, Nate Smith, and Joey Waronker. Each drummer went into a recording studio and played drum grooves to a click track that was set at whatever tempo they chose. Each drum in the kit had at least one microphone directed at it, so onset timing information can be extracted using Music Information Retrieval techniques (MIRtoolbox, Lartillot et al., 2008a).[8] The Loop Loft data set provides the highest quality data of the three corpora described in the previous chapter and in Hosken et al. (in press), the other two sets being the Lucerne Groove Research Library and Google Magenta's Groove MIDI Dataset. This is because the Loop Loft recordings were created in pristine studio environments that allow for precise onset timing data to be collected (the Lucerne corpus, as a result of being derived from full-band recordings, loses some precision here) and the drummers involved are all world-class professionals (the Magenta set comprises five professionals and four amateurs, and data is anonymized so it is impossible to discern professional from amateur data).

To reiterate and extend a point made when introducing the Loop Loft corpus in the previous chapter, the exact location of the external reference (the click track) that the drummers performed in relation to is not known. In order to make empirical measurements of performance timing, a reference point needs to be decided upon from which relative measurements may be made. In the case of the Loop Loft corpus, the beginning of the audio file is defined as 0.00—each file begins with less than one millisecond of silence, with the sound then beginning at an amplitude

[8]Onset times for each instrument were recorded and filtered using the MIRtoolbox function: `a=mirevents('FILENAME', 'Threshold', 0.1);`. In the previous chapter, this measurement method was compared with the newer, waveform-based function in the MiningSuite (Lartillot, 2019). Significant differences were found between the two measurement methods, though the effect size of this was near-zero and the absolute size of the difference between the two measurement methods was small.

zero-crossing, presumably the first in the recording—and all measurements are made relative to the beginning of the track. Locations of each of the click track's beat locations are then calculated by using information about the performance BPM contained in each track's metadata. To give an example: if a performance in the corpus is at 120 BPM, the location of the click track's beat 1 is defined as 0.00 and the location of the second click track event (beat 2) would be at 0.5 seconds (500 ms). I believe that the Loop Loft data is provided in a way that 0.00 in the audio file *is* aligned with the click track time 0.00 because the intended usage of the recordings is for people to be able to add, for example, Omar Hakim to their own creative work without having to do any further processing.

It is important to control for tempo in the following analyses as tempo is known to have a significant impact on performance timing details (see Friberg and Sundström, 2002; Honing and Bas de Haas, 2008; Moelants, 2011; Repp, 1995; Repp et al., 2002). As such, I split the corpus into three bands:

- Slow (< 80 BPM)
- Medium (80–112 BPM)
- Fast (> 112 BPM)[9]

There is a wide range of tempi in the Loop Loft performances (see Figure 3.1) and the Medium tempo band captures the majority of the performances, which suggests that the tempo bands are well-suited to the data. Note that Waronker has no performances in the "Slow" category and Smith has only one set of performances (described in the corpus as one "Package"), recorded at 79 BPM. Since this set of Smith performances is titled "Halftime," it may be argued that the BPM is

[9] After Danielsen, Waadeland, et al., 2015, which asked drummers to perform at three tempi 64, 96, and 148 BPM corresponding to slow, medium, and fast descriptors. Since the drummers involved in this study set the click track at whatever tempo they chose, there is a wide variety of tempi. To create bands based on Danielsen et al.'s definitions, the halfway point between the medium and fast tempo was calculated (112 BPM) and defined as the boundary between medium and fast. Similarly, the lower bound of the medium category (80 BPM) is the halfway point between Danielsen et al.'s medium and slow definitions.

158 and so, to avoid confusion and minimize imbalance in the statistical tests, this set is dropped from all analyses.

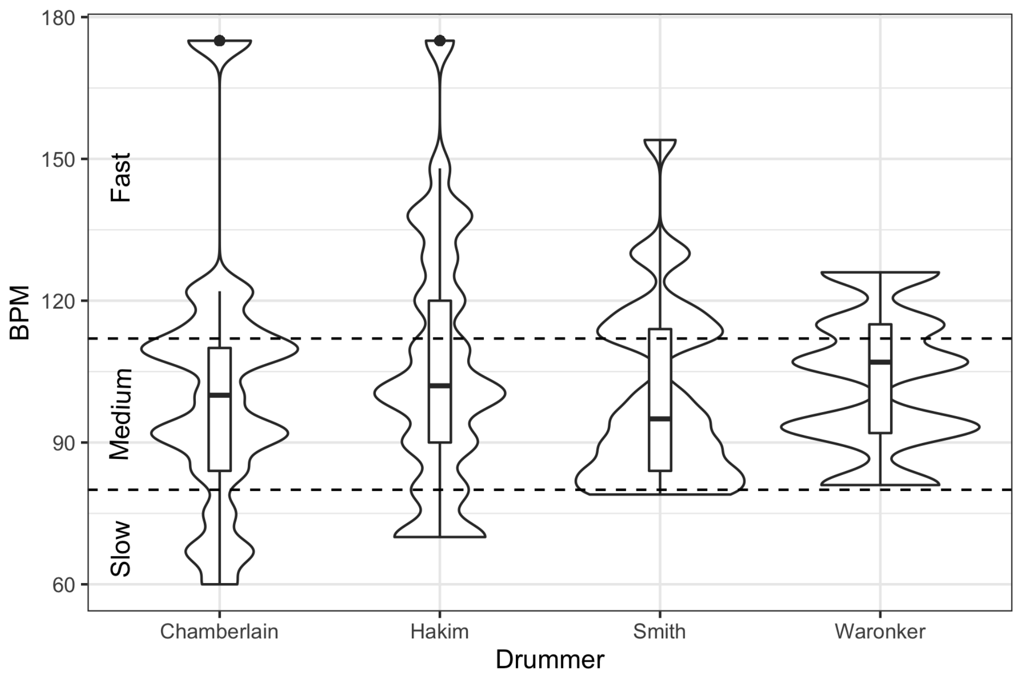


FIGURE 3.1: A violin and box plot showing the range in tempi for performances contained in the Loop Loft corpus. The box plots provide summary information about the median and interquartile ranges of the data, while the violin plot shows the full distribution of the data (wider = more data). Dashed horizontal lines show the splits at 80 and 112 BPM.

To investigate the potential uniqueness and significance of performer pockets, I will analyze how each of the drummers locates their bass drum and snare drum onsets in time. The various patterns that these drummers play all, arguably, have the archetypal "boom tish" drum pattern in the background (see Figure 3.2 and also the dark black bands showing concentrations of onsets in the previous chapter's strip plots, Figure 2.4) and so the following analyses investigate the drum onsets associated with key locations: bass drum beat 1, snare drum beat 2, bass drum beat 3, and snare drum beat 4. Syntactically, these performances are all very similar (particularly at the schematic/background level), but the processual details of each performance, as we shall see, are what distinguish these performers.

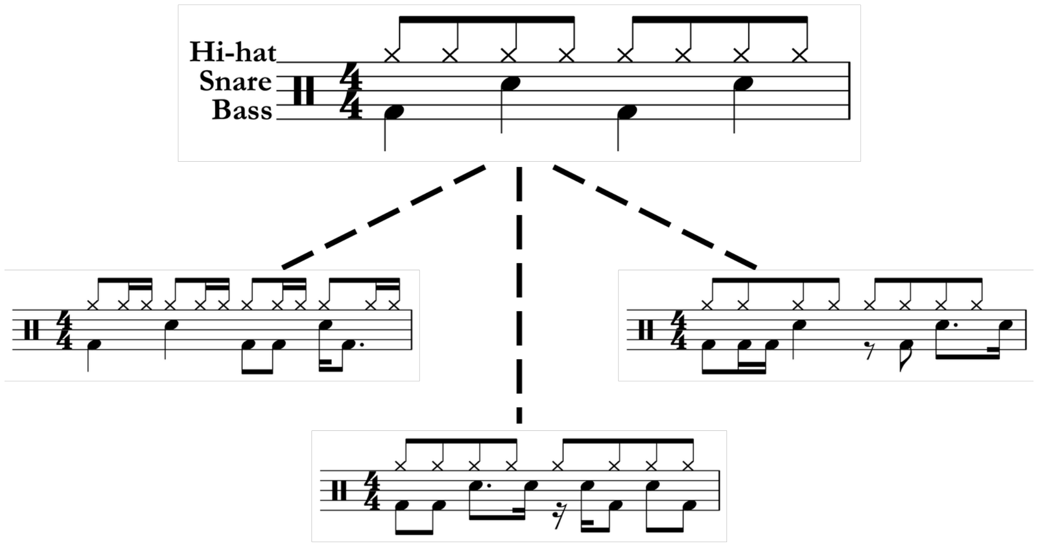


FIGURE 3.2: A schematic showing the archetypal Anglo-American popular music drum groove (top) and, connected via dashed lines, examples of drum patterns that derive from it.

In order to describe any differences between drummers' pockets, I will analyze the location, scale, and shape of the distributions of onsets through time. In my interpretation of pockets, location corresponds, musically, to how "laid back" or "pushed" the drummer is and can be described by one of the two averages presented later in the summary tables, either by the mean or median. Here, the degree of pushed/laid back describes a tendency to sound out the musical beat before or after the reference position (respectively). Scale is how "tight" or "loose" a drummer is and can be described by the width of the distribution, either the standard deviation or the median absolute difference.[10] This corresponds to the amount of variance in a pocket with a "tight" pocket sounding out the beat in consistent locations (with respect to the metronomically defined positions). The shape of a pocket is harder to describe musically and statistically. Statistically, this can be captured by the skewness and kurtosis. Simply, skewness describes the symmetry of the distribution while kurtosis describes how much data is contained in the tails of

[10] These values are remarkably similar in the ensuing analyses, so only standard deviations are reported.

the distribution. Musically, it could be argued that the shape of a drummer's pocket can describe another facet of how consistent the performer is, emphasizing how many outliers they have to their performances. If a performer has a lot of leftwards skew ("positive skew") to their pocket, they have an asymmetrical tendency to articulate the beat before the metronome, though, unlike a wholesale change in location ("pushed"), they do still play a number of events after. If the performer plays onsets only within a very tight band around the metronomic locations, but also with a notable number of outlier events outside this band, then they have "excess kurtosis" or a "leptokurtic" pocket that is rather more pointy with long tails on each side. There are no convenient vernacular feel descriptors for these properties (perhaps outliers may be recognized as a degree of "sloppiness"), but these may prove to be important dimensions in the construction and description of pockets.

These attributes of location, scale, and shape can be seen as changes in the density plots, as illustrated in Figures 3.3 and 3.4. When a drummer plays "laid back," the location of the pocket changes (Figure 3.4a): the whole plot is pushed backwards in time as the majority of their onsets now occur at a later time. Likewise, if a drummer plays "tighter," the scale of the pocket changes (Figure 3.4b): the majority of their onsets now occur within a narrower interval of time. If a drummer has a tendency to play before the metronomic location, but as a result of asymmetric distribution of their onsets (rather than an entire shift forward, as in the "pushed" plot), then the shape of the pocket is positively skewed (Figure 3.4c). Lastly, if a drummer has a strong tendency to play in one small region (therefore a high density of events in one area), but also has a number of events—potentially outliers—far outside this region, they would have a distribution with a lot of information contained in the far left and right tails of their pocket (a "leptokurtic" distribution—see Figure 3.4d).

All analyses use two-sample Kolmogorov–Smirnov tests, which quantify a distance between two empirical distributions, analyzing whether it is likely that they come from the same underlying

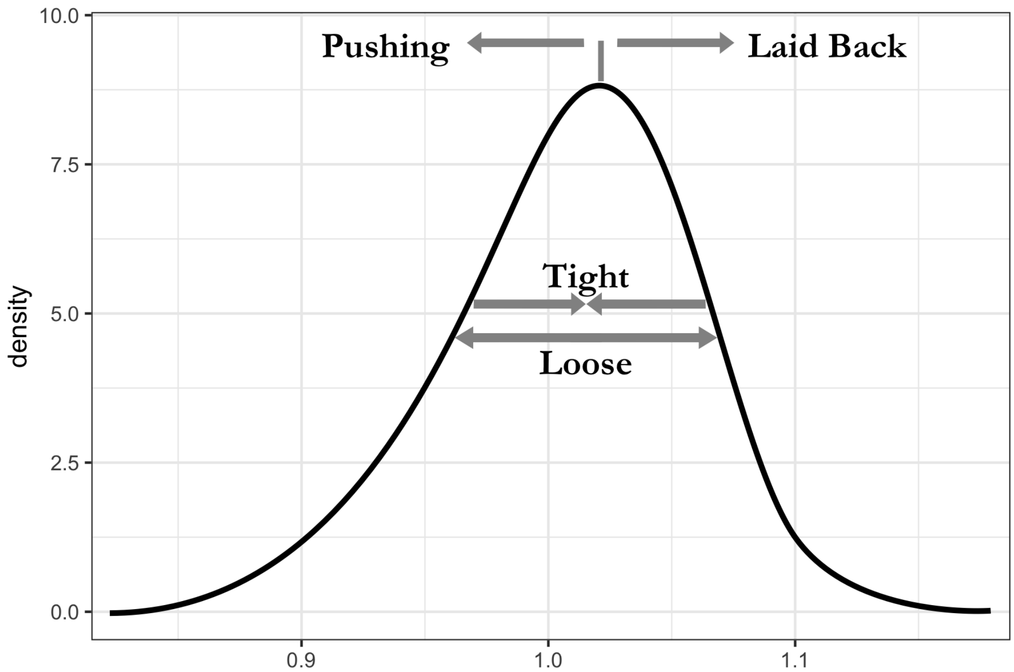
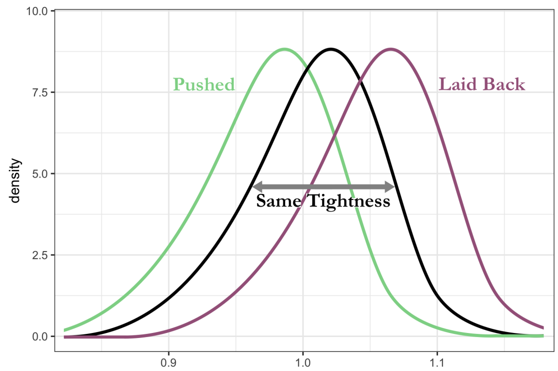
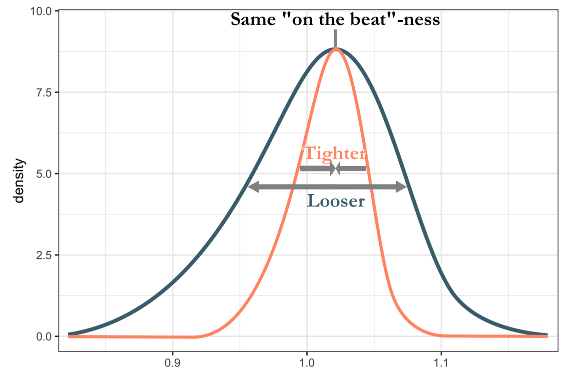


FIGURE 3.3: A hypothetical onset distribution annotated to show how feel descriptors can apply to statistical descriptions of a distribution's moments.

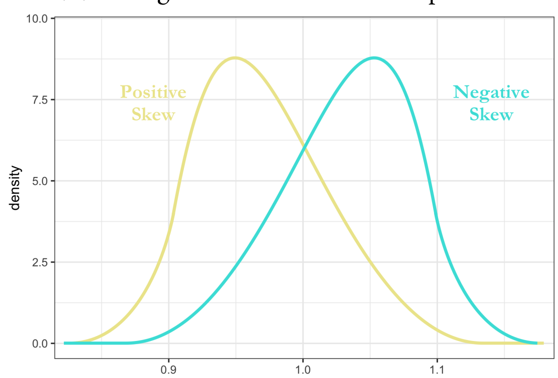
distribution. The Kolmogorov–Smirnov test is sensitive to differences in location, scale, and shape of the distribution functions of the two samples, which will prove important in evaluating the differences, if any, between each of the drummers in the Loop Loft corpus. It is a non-parametric test, so does not assume that the two distributions it compares come from Gaussian/normal distributions (as, for example, an ANOVA would). This is important as some of the drum onsets are distributed non-normally ($p < .05$ in the Shapiro–Wilk test of normality). Another assumption of ANOVA testing is also violated in the drum timing data, with the variances of the populations that the samples come from not equal being ($p < .05$ in Bartlett's test of homogeneity of variances), exacerbated by unequal group sizes in the data set. While some simulation studies have shown that false positive rates are not affected too much by violations of these assumptions (Glass et al., 1972; Harwell et al., 1992; Lix et al., 1996), the Kolmogorov–Smirnov tests are deliberately chosen here as the K–S test is a statistical tool specifically intended for presumption-



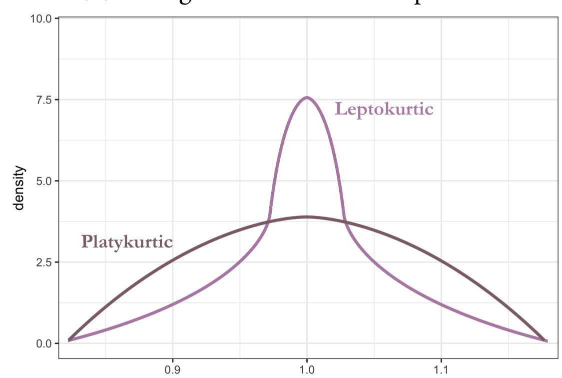
(A) Changes in the *location* of the pocket



(B) Changes in the *scale* of the pocket



(C) Changes in the *shape* (skew) of the pocket



(D) Changes in the *shape* (kurtosis) of the pocket

FIGURE 3.4: Examples of different hypothetical pocket shapes and qualitative, musical labels that can be applied to describe them.

free comparisons of distributions and is sensitive to several types of differences (as opposed to, for example, just being a comparison of means). For all Kolmogorov–Smirnov tests presented here, the significance level is set at $\alpha = 0.05$, i.e., the risk of incorrectly rejecting a true null hypothesis is 5%.

Comparison of Pocket Shapes: Slow Tempo Band

Looking first at the Slow data (Figure 3.5 and summarized in Table 3.1), which only compares Chamberlain and Hakim, the onset distributions are remarkably different: Chamberlain appears to play far tighter—with less spread—than Hakim. Note how the measure of the spread of data

(SD) is lower for Chamberlain than for Hakim in all instances (Table 3.1). The peaks of Hakim's distributions are also later than Chamberlain's, suggesting that he is a more "laid back" performer. In Table 3.1, this difference can be seen in both measures of the average (mean and median) for all drum onset locations. Statistically, Kolmogorov–Smirnov tests show that these two distributions are significantly different from one another for all four drum locations (Table 3.2). As such, we can argue that Chamberlain and Hakim, when playing at tempi below 80 BPM, have unique pockets.

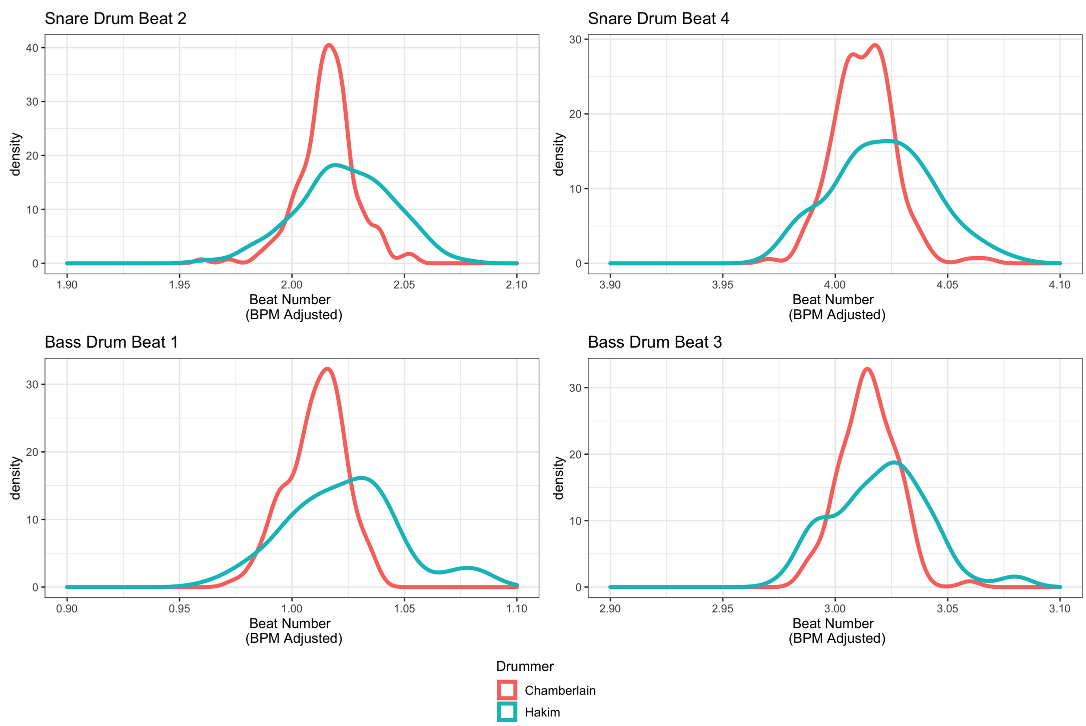


FIGURE 3.5: Density plots showing the distribution of onsets in Chamberlain and Hakim's slow tempo (< 80 BPM) performances. The area under each curve is exactly the same, so the differences in number of onsets performed by each drummer does not matter.

In order to apply qualitative descriptors to these differences with any certainty, however, we need to know whether the differences lie in the location, scale, or shape of each drummer's pocket. Currently, we are only able to say that there is some general distance between the distributions of the two drummers' onsets, but the quality of this difference—whether Hakim's snare

	Drummer	N	Mean	Median	SD	Skewness	Kurtosis	Normal p
SD2	Chamberlain	187	2.016	2.016	0.013	-0.424	5.418	<.0001
	Hakim	102	2.024	2.023	0.021	-0.207	2.888	0.887
SD4	Chamberlain	175	4.013	4.013	0.013	0.374	4.363	0.011
	Hakim	79	4.021	4.023	0.022	0.109	2.602	0.717
BD1	Chamberlain	111	1.011	1.012	0.013	-0.380	2.806	0.162
	Hakim	64	1.023	1.024	0.025	0.285	3.176	0.284
BD3	Chamberlain	107	3.015	3.014	0.012	0.213	3.774	0.278
	Hakim	72	3.020	3.022	0.021	0.319	3.192	0.150

TABLE 3.1: Descriptive statistics for the Slow tempo band. Normal p is the p value of a Shapiro–Wilk test of normality. The null hypothesis for this test is that the data are normally distributed. If the p-value is less than the chosen alpha (0.05), then the null hypothesis that the data are normally distributed is rejected. If the p-value is greater than 0.05, then the null hypothesis is not rejected. I.e., data are significantly non-normal if $p < .05$.

	D	p
BD1	0.392	<.0001
SD2	0.316	<.0001
BD3	0.261	.004
SD4	0.333	<.0001

TABLE 3.2: Summary table of Kolmogorov–Smirnov test statistics for Chamberlain and Hakim in the Slow tempo band.

drums and bass drums differ from Chamberlain's because of a difference in location, scale, or shape—is unknown. To gain more insight into which dimensions the performances differ along, Kolmogorov–Smirnov tests are performed once more, but this time on Z-scored data. Z-scoring is a way of standardizing data by taking each data point (x), subtracting the distribution's mean (μ), and dividing by the distribution's standard deviation (σ):

$$Z = \frac{x - \mu}{\sigma} \tag{3.1}$$

Here, each onset's time (in beats) is Z-scored by drum and by drummer, for example, all of Chamberlain's snare drum onsets that are associated with beat 2 are grouped together and standardized. Likewise for Hakim's snare drums on beat 2, and so on for all other applicable subgroups of timing data. Each of the distributions now, as a result of being Z-scored, all have a standard distribution of 1 and a mean of 0. Performing Kolmogorov–Smirnov tests on this Z-scored data will therefore only show a significant result if there is a difference in the *shape* of the distribution as the location (μ) and scale (σ) have been controlled for. As seen in Table 3.3, there are now no significant differences between the two drummers, therefore the two drummers differ in the location (mean/median) and scale (variance) of their pockets, but not the shape (skewness and kurtosis). Returning to the qualitative descriptors, these Z-scored tests confirm the descriptions from earlier—that Hakim is more "laid back" and Chamberlain is "tighter"—as the uniqueness of their onset distributions is not attributable to the consistency of their performances.

	D	***p***
BD1	0.049	.989
SD2	0.109	.416
BD3	0.072	.662
SD4	0.054	.902

Table 3.3: Summary table of Kolmogorov–Smirnov test statistics for Chamberlain and Hakim in the Slow tempo band after Z-scoring.

To conclude this first set of comparisons and gain a fuller picture of drummer styles, it is important to consider how the above-described pockets are arranged throughout the bar. As such, further Kolmogorov–Smirnoff pairwise comparisons will reveal whether, for example, the way Hakim places his bass drum attacks on downbeats differs from his bass drum attacks on beat 3 or whether Chamberlain lays back more with his snare drums on beat 2 than on beat 4. These comparisons are limited to bass drum 1–bass drum 3 and snare drum 2–snare drum 4 comparisons as the onset envelope for the two types of drums do differ and so the perceptual centers of these events likely are not the same, leading to fundamental differences in how the musicians perform them (Danielsen et al., 2019; Danielsen et al., in review; London et al., 2019). As reported in Table 3.4, only Chamberlain's snare drum performances have pockets in the first half of the bar that are significantly different to the pockets in the latter half, likely due to the bimodal nature of Chamberlain's snare drum beat 4 pocket (see the twin peaks in the Snare Drum Beat 4 panel of Figure 3.5). There is a remarkable consistency in the nature of each pocket in these comparisons, suggesting that, if a drummer is for example "laid back" (at a slow tempo), they are consistently laid back, not just locating their pocket later in time at one beat location.

	Comparison	***D***	***p***	
Chamberlain	BD1–BD3	0.133	.289	NS
	SD2–SD4	0.211	<.001	***
Hakim	BD1–BD3	0.127	.596	NS
	SD2–SD4	0.113	.564	NS

TABLE 3.4: Summary table of Kolmogorov–Smirnov test statistics for beat-pair comparisons of Chamberlain and Hakim's performances in the Slow tempo band. * $p < .05$, ** $p < .01$, *** $p < .001$, NS = Not Significant.

Comparison of Pocket Shapes: Medium Tempo Band

Turning to the Medium tempo performances (80–112 BPM), these analyses now compare all four drummers in the Loop Loft corpus. The Kolmogorov–Smirnov test can only compare a maximum of two sample distributions at a time, so, for the Medium and Fast bands of data, multiple pairwise comparisons are used. All comparisons will therefore be Bonferroni corrected to minimize the likelihood of incorrectly rejecting a true null hypothesis.

Looking at the onset distributions (Figure 3.6 and summarized in Table 3.5), there appear to be two groups of drummers: Smith and Waronker may perhaps be described as "tighter" since their distributions appear less spread out than those performed by Chamberlain and Hakim. Additionally, Hakim consistently has the widest distribution, suggesting a "looser" performance style. Lastly, Chamberlain appears to have a remarkably "laid back" bass drum on beat 3, which is particularly noteworthy when considering that his performance distributions for the other beats of the bar are not especially later than the others. This highlights the importance of looking at these onset distributions metrically as each beat of a standard 4/4 bar plays a different role.

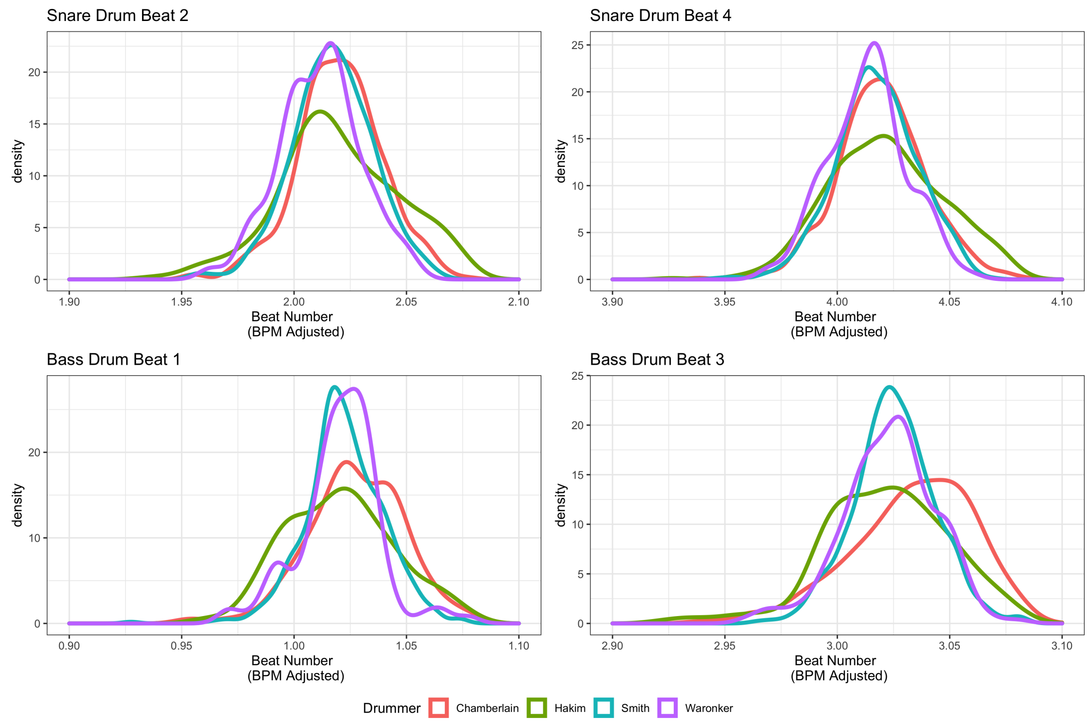


FIGURE 3.6: Density plots showing the distribution of onsets in Chamberlain, Hakim, Smith, and Waronker's medium tempo (80–112 BPM) performances.

To see if the differences between drummers' performances are statistically meaningful and ascertain whether these are unique distributions, Kolmogorov–Smirnov tests are conducted between all pairs of drummers for each metronomic beat (Table 3.6). The first observation to make is that there are some overlaps in the results—we do not have four distinct, unique onset distributions for each beat of the bar, which would indicate that each drummer has a unique pocket. For instance, while Chamberlain and Waronker are statistically different from one another for all beat locations, the properties of Waronker's bass drum pockets for beats 1 and 3 overlaps with both Hakim and Smith's. This is interesting because, looking across beat locations, Hakim

	Drummer	N	Mean	Median	SD	Skewness	Kurtosis	Normal p
SD2	Chamberlain	505	2.020	2.020	0.019	-0.154	3.484	0.054
	Hakim	345	2.020	2.017	0.027	-0.091	2.923	0.014
	Smith	613	2.017	2.017	0.018	-0.168	3.311	0.104
	Waronker	139	2.012	2.013	0.018	-0.082	2.943	0.662
SD4	Chamberlain	509	4.020	4.019	0.020	-0.44	3.686	0.061
	Hakim	397	4.022	4.021	0.026	-0.011	2.873	0.075
	Smith	768	4.017	4.017	0.018	-0.168	3.236	0.043
	Waronker	182	4.014	4.015	0.017	-0.004	2.828	0.667
BD1	Chamberlain	333	1.028	1.027	0.022	-0.333	3.640	0.019
	Hakim	296	1.021	1.019	0.024	0.168	2.632	0.143
	Smith	560	1.022	1.022	0.017	-0.279	4.715	<.0001
	Waronker	101	1.020	1.023	0.018	-0.098	4.399	0.003
BD3	Chamberlain	202	3.033	3.037	0.027	-0.653	3.259	<.0001
	Hakim	222	3.022	3.023	0.028	-0.432	3.720	0.003
	Smith	552	3.025	3.025	0.018	0.021	3.355	0.327
	Waronker	194	3.023	3.024	0.020	-0.330	3.408	0.136

Table 3.5: Descriptive statistics for the Medium tempo band.

and Smith have statistically different pockets. This might suggest that Waronker's bass drum pocket falls somewhere between the other two, a conjecture that makes sense when looking at the relative shapes and sizes of the pockets in Figure 3.6. Returning to some of the hypotheses about the pocket shapes mentioned earlier, the suggestion that Chamberlain has a more "laid back" bass drum on beat 3 is borne out as the distribution of onsets is significantly different from all other drummers. Hakim has a significantly different pocket for nine of the 12 comparisons, supporting to a degree the suggestion that he has a "looser" performance style when compared to the others.

As previously, we need to know whether the differences lie in the location, scale, or shape of each drummer's pocket and so Kolmogorov–Smirnov tests are performed again on Z-scored data for pairs of drummers that were previously found to be significant (i.e. dropping any pairs marked "NS" in Table 3.6). When the location and scale of the distributions are standardized by the Z-score, there are no significant differences between any pairs of drummers for any of

SD2	**Ha**	**Sm**	**Wa**
Ch	0.118 (.009) *	0.088 (.029) NS	0.227 (<.001) ***
Ha		0.144 (<.001) **	0.199 (<.001) **
Sm			0.163 (.005) *

SD4	**Ha**	**Sm**	**Wa**
Ch	0.107 (.012) NS	0.062 (.197) NS	0.179 (<.001) **
Ha		0.137 (<.001) ***	0.215 (<.001) ***
Sm			0.133 (.010) NS

BD1	**Ha**	**Sm**	**Wa**
Ch	0.194 (<.001) ***	0.168 (<.001) ***	0.275 (<.001) ***
Ha		0.166 (<.001) ***	0.160 (.041) NS
Sm			0.113 (.220) NS

BD3	**Ha**	**Sm**	**Wa**
Ch	0.229 (<.001) ***	0.257 (<.001) ***	0.272 (<.001) ***
Ha		0.173 (<.001) ***	0.135 (.046) NS
Sm			0.080 (.312) NS

TABLE 3.6: Comparison matrices of Kolmogorov–Smirnov test statistics for all pairs of drummers in the Medium tempo band. Ch = Chamberlain, Ha = Hakim, Sm = Smith, and Wa = Waronker. The topmost number in each cell reports the D statistic, beneath it in parentheses is the p value, and, finally, asterisks refer to significance level-equivalents *after* Bonferroni corrections (* $p < .05$, ** $p < .01$, *** $p < .001$, NS = Not Significant).

the beat locations (Table 3.7). This confirms that the differences noted earlier—that Hakim has the "loosest" distribution, that Chamberlain has a "laid back" bass drum on beat 3, and that, generally, Smith and Waronker are "tighter" than the other two drummers—are valid as there are no discernible differences between the drummers outside of observations to do with location and scale of their distributions.

	Drummers	***D***	***p***
SD2	Chamberlain–Hakim	0.053	.605
	Chamberlain–Waronker	0.053	.919
	Hakim–Smith	0.057	.470
	Hakim–Waronker	0.073	.668
	Smith–Waronker	0.051	.933
SD4	Chamberlain–Waronker	0.090	.800
	Hakim–Smith	0.057	.470
	Hakim–Waronker	0.073	.668
BD1	Chamberlain–Hakim	0.086	.888
	Chamberlain–Smith	0.113	.290
	Chamberlain–Waronker	0.230	.094
	Hakim–Smith	0.072	.956
BD3	Chamberlain–Hakim	0.078	.930
	Chamberlain–Smith	0.076	.797
	Chamberlain–Waronker	0.124	.542
	Hakim–Smith	0.089	.781

Table 3.7: Summary table of Kolmogorov–Smirnov test statistics for Z-scored data from all pairs of drummers that were previously found to be significantly different in the Medium tempo band.

Assessing the performances metrically, the only significant differences that exist within each drummer are between Chamberlain's two bass drum pockets and between Smith's bass drum pockets (see Table 3.8). The preceding analyses have already shown that Chamberlain's performance in general is distinctly laid back, and this metric analysis shows further that his bass drum beat 3 pocket stands out as especially late (a mean lateness of 0.033 beats as opposed to bass drum beat 1's already high lateness of 0.028 beats) compounded by having a statistically "moderate" negative skewness (skewness between -0.5 and -1), which means that there is a longer "left tail" to the distribution, i.e. that the distribution is asymmetric. Smith's bass drum pockets were previously found to be not so distinct, being statistically inseparable from Hakim and Waronker's performances, but this metric analysis does show some difference between how Smith nuances the bass-drum pocket around beat 1 and around beat 3. Looking back to Table 3.5 and Figure 3.6, the difference appears to lie in the kurtosis of Smith's beat 1 bass drums. Smith's beat 1 performances appear "pointier" than the other drummers', but also with longer tails either side, reflected in an excess kurtosis of 1.715 (excess kurtosis is the amount the kurtosis value exceeds 3, which suggests that there are more outliers than found in a normal distribution, which has a kurtosis of 3, and an excess, by definition, of 0). Interpreting this musically, we could say that Smith does have a tight pocket associated with beat 1, but that there are also a number of outliers, suggesting a degree of variability in how Smith performs beat 1.

	Comparison	***D***	***p***	
Chamberlain	BD1–BD3	.200	<.001	**
	SD2–SD4	0.044	.717	NS
Hakim	BD1–BD3	0.065	.651	NS
	SD2–SD4	0.070	.333	NS
Smith	BD1–BD3	0.097	.011	*
	SD2–SD4	0.029	.942	NS
Waronker	BD1–BD3	0.152	.092	NS
	SD2–SD4	0.094	.491	NS

Table 3.8: Summary table of Kolmogorov–Smirnov test statistics for beat-pair for all drummers' performances in the Medium tempo band.

Comparison of Pocket Shapes: Fast Tempo Band

Lastly, all four drummers' performances at the Fast tempo (>112 BPM) are compared using the same analytical process as in the Medium tempo band. From inspecting the onset distributions

(Figure 3.7 and summarized in Table 3.9), the first element that calls for further investigation is the shape of Waronker's onset distribution for bass drums associated with beat 1. This distribution has one large, tight peak as well as several other smaller peaks. The shape of this distribution, and, to lesser degrees, the shape of Waronker's other pockets in the Fast category is symptomatic of the relatively few data points available (see the "N" column of Table 3.9) and is not necessarily indicative of something musically meaningful. That being the case, the peaks in Waronker's distribution are spaced out remarkably consistently, with clear peaks around 0.975, 1.00, 1.025, 1.05, and 1.075. This regularity and alignment with binary divisions of the metronomic grid, divisions that are spaced roughly 13 milliseconds or less apart (13 milliseconds corresponds to 0.025 beats at 112 BPM; faster tempos would reduce the time difference as measured by a clock) does warrant further consideration. Digging into Waronker's performance, the smaller peaks are each the result of just one, two, or three onsets that fall outside of the main peak at 1.025. While they do cluster near the clear grid divisions in the smoothed density plot of Figure 3.7, the reality is less precisely aligned. As such, we may conclude that, in fact, these multiple peaks are more the result of few data points than anything musically meaningful.[11]

Regarding the other drummers' performances, Hakim once again appears to have the most "loose" pocket, having the widest variance of all the drummers. Additionally, Chamberlain's onset average, whether mean or median, is often the latest, continuing the trend seen in the previous tempo bands for Chamberlain to be the most "laid back" of the drummers. The distribution of Smith's beat 3 bass drums are not necessarily later than the other drummers', but there is a "tightness" to them with less variance and so fewer onsets occur at timepoints outside of the moment around 3.025.

[11] The relatively few data points available for Waronker is important to bear in mind when performing the statistical comparisons as the critical value of the Kolmogorov–Smirnov D statistic—the value D must be greater than for us to reject the null hypothesis at significance level α—is intrinsically tied to the number of observations available. As the number of data points available increases, the value that D must exceed decreases. This is why, in Table 3.10, although the D statistic for the comparison of Hakim and Waronker's bass drum beat 1 is greater than the D statistic for the comparison of Hakim and Smith's bass drum beat 1, the Hakim–Smith comparison (combined $N = 417$) is

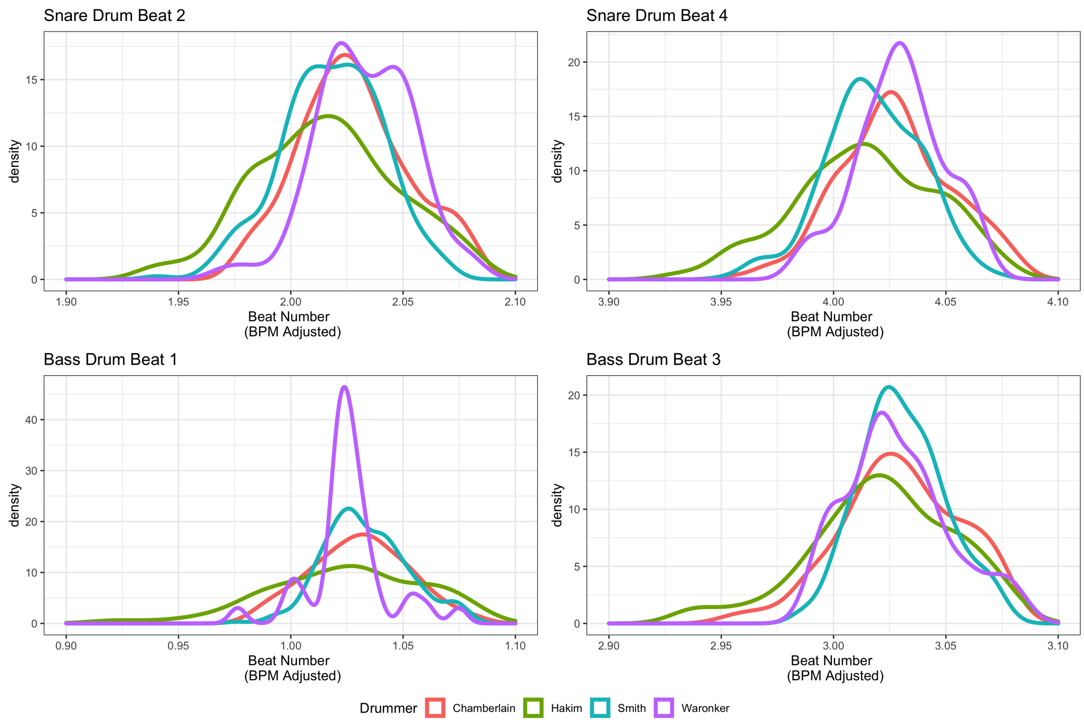


FIGURE 3.7: Density plots showing the distribution of onsets in Chamberlain, Hakim, Smith, and Waronker's fast tempo (>112 BPM) performances.

Looking at the results of pairwise Kolmogorov–Smirnov tests (Table 3.10), there are far fewer significant differences between drummers than in the other tempo bands. This may be attributable to the fact that, at higher tempi, the span of time associated with the beat is shorter and these drummers have a window of time that is only a few tens of milliseconds long within which they can nuance their performances. Previous studies into jazz drummers' swing have shown how their long–short swing pattern straightens out at higher tempi because of this exact reason (Friberg and Sundström, 2002; Honing and Bas de Haas, 2008; also see the disappearance of microtiming details at higher tempi in Danielsen, Waadeland, et al., 2015, though contrast with Haugen and Danielsen, 2020, which found an *exaggeration* in short/long duration differences at higher tempi in pandeiro drumming, perhaps due to the pattern of short/long durations' core structural role in the samba "groove template"). The differences between the Loop drummers that

statistically significant and the Hakim–Waronker comparison (combined N = 219) is not.

	Drummer	N	Mean	Median	SD	Skewness	Kurtosis	Normal p
SD2	Chamberlain	229	2.028	2.026	0.024	0.231	2.584	0.013
	Hakim	204	2.016	2.016	0.032	0.006	2.620	0.194
	Smith	247	2.018	2.019	0.022	-0.262	3.011	0.270
	Waronker	103	2.033	2.032	0.021	-0.217	3.233	0.367
SD4	Chamberlain	206	4.027	4.027	0.025	0.013	2.782	0.226
	Hakim	241	4.014	4.015	0.032	-0.207	2.590	0.046
	Smith	305	4.018	4.016	0.021	-0.169	2.997	0.332
	Waronker	72	4.030	4.029	0.019	-0.160	2.698	0.225
BD1	Chamberlain	152	1.031	1.031	0.022	0.026	2.566	0.645
	Hakim	183	1.024	1.026	0.034	-0.456	3.064	0.003
	Smith	234	1.033	1.030	0.019	0.252	3.009	0.027
	Waronker	36	1.026	1.025	0.017	0.186	5.033	0.006
BD3	Chamberlain	197	3.030	3.029	0.027	-0.249	2.774	0.041
	Hakim	216	3.022	3.024	0.032	-0.434	3.080	0.003
	Smith	221	3.031	3.029	0.019	0.215	2.701	0.217
	Waronker	60	3.029	3.023	0.023	0.499	2.670	0.045

Table 3.9: Descriptive statistics for the Fast tempo band.

are significant include Hakim's snare drum onset distributions, which differ from Chamberlain and Waronker, but not Smith. This is inverted when looking at bass drum onset distributions, where Hakim differs from only from Smith and *not* Chamberlain or Waronker. Qualitatively, this makes the earlier-hypothesized characterization of Hakim as "looser" than the other drummers problematic. It may be fair to say that his performance on snare drum *is* looser than Chamberlain and Waronker, but this is not consistent across all drums and distinct from all of the drummers. Similarly, the pairwise Kolmogorov–Smirnov tests suggest that it is not straight forward to describe Chamberlain as "laid back": while his performances do differ from Hakim and Smith's for both snare drum locations, the only bass drum difference is with Waronker on beat 1 (a potentially problematic set of onsets, as described above). The last difference hypothesized earlier is that Smith has a "tight" distribution of bass drum onsets associated with beat 3, but his performance only differs with any significance from Hakim. So, while he may well be playing in a manner that

SD2	Ha	Sm	Wa
Ch	0.217 (.009) ***	0.161 (.004) *	0.155 (.065) NS
Ha		0.141 (.024) NS	0.339 (<.001) ***
Sm			0.287 (<.001) ***

SD4	Ha	Sm	Wa
Ch	0.243 (<.001) ***	0.202 (<.001) ***	0.138 (.265) NS
Ha		0.150 (.005) NS	0.330 (<.001) ***
Sm			0.290 (<.001) ***

BD1	Ha	Sm	Wa
Ch	0.165 (.022) NS	0.115 (.178) NS	0.314 (.005) *
Ha		0.250 (<.001) ***	0.271 (.019) NS
Sm			0.303 (.006) *

BD3	Ha	Sm	Wa
Ch	0.141 (.033) NS	0.111 (.151) NS	0.106 (.685) NS
Ha		0.204 (<.001) **	0.178 (.103) NS
Sm			0.138 (.330) NS

Table 3.10: Comparison matrices of Kolmogorov–Smirnov test statistics for all pairs of drummers in the Fast tempo band.

is tight, it is only distinct from Hakim and not the others.

When the location and scale are standardized using Z-scores, there are no significant differences between any pairs of drummers (Table 3.11). In a similar way to the findings from the Z-scored Medium and Slow tempo bands, this lends credence to the qualitative descriptors suggested earlier—the differences between drummers' pockets are to do with location ("pushing" or "laying back") and scale ("tight" or "loose"), and not to do with the shape of the onset distributions.

Finally, comparing each drummer's bass drum beat 1 pocket with their bass drum beat 3 pocket and snare drum beat 2 with their snare drum beat 4, there are no significant differences whatsoever (Table 3.12). That is, pockets from the first half of the bar do not statistically differ from pockets

	Drummers	*D*	*p*
SD2	Chamberlain–Hakim	0.062	.806
	Chamberlain–Smith	0.069	.621
	Hakim–Waronker	0.080	.776
	Smith–Waronker	0.073	.839
SD4	Chamberlain–Hakim	0.054	.902
	Chamberlain–Smith	0.058	.805
	Hakim–Waronker	0.062	.983
	Smith–Waronker	0.073	.914
BD1	Chamberlain–Waronker	0.177	.284
	Hakim–Smith	0.105	.211
	Smith–Waronker	0.209	.130
BD3	Hakim–Smith	0.139	.325

Table 3.11: Summary table of Kolmogorov–Smirnov test statistics for Z-scored data from all pairs of drummers that were previously found to be significantly different in the Fast tempo band.

in the second half. This lack of difference within each drummer's performances, as well as the relatively fewer differences when comparing *between* drummers, builds more evidence towards a hypothesis for future study that microtiming performance style becomes more homogenous at higher tempi. This parallels the previously mentioned finding from the analysis of jazz swing that saw a "straightening out" of the long–short swing pattern at higher tempi. At higher tempi, there simply is less time available to personally stylize and performers are more limited—physically/biomechanically—in what they can do.

Additional Findings Across Performances

Several of the key results have already been discussed and contextualized in situ, but there are several other findings that are seen when taking a holistic view of the many individual analyses presented in the previous section. One of the most easily seen results is that, looking at the results with specific reference to the external 0.00 reference point as I have defined it (see discussion on page 73), the location of the pockets across all of the analyses tends to be later than the metronomic

	Comparison	*D*	*p*	
Chamberlain	BD1–BD3	0.109	.265	NS
	SD2–SD4	0.041	.994	NS
Hakim	BD1–BD3	0.084	.481	NS
	SD2–SD4	0.062	.789	NS
Smith	BD1–BD3	0.079	.478	NS
	SD2–SD4	0.048	.909	NS
Waronker	BD1–BD3	0.233	.150	NS
	SD2–SD4	0.147	.284	NS

Table 3.12: Summary table of Kolmogorov–Smirnov test statistics for beat-pair for all drummers' performances in the Fast tempo band.

reference point (mean and median values found in the summary tables are all greater than the beat location they refer to, which is also seen in the location of the peaks of the density plots). I approach this finding by noting that this runs contrary to the frequently observed finding that, when asked to tap in synchrony with a metronome, people tend to tap systematically *earlier* than the actual pulse ("negative mean asynchrony": Repp, 2005, a finding also found specifically with drummers by Fujii et al., 2011). The present study's finding might suggest that, when playing artistically and not simply performing a synchrony task, drummers have a tendency to play *later* rather than earlier. Repp (2005, p. 973) notes, in the survey of NMA literature, an important point regarding NMAs in musically trained participants' tapping behavior when contrasted with untrained: "The NMA is small or absent in musical contexts: Musically trained participants tend to show a smaller NMA than do untrained participants (Aschersleben, 2002) and, sometimes, none at all (see, e.g., Repp, 2004)." The lack of NMA here also aligns with the perspectives of expert musicians such as Iyer, who describes "on top" of the beat playing as "stiff" (Iyer, 2002, p. 406), and Stewart, whose feel spectrum locates "in the pocket" as being 10 ms behind the metronomic beat (1987, p. 64). There is a chance that the lack of NMA finding in the data may also be a result of how the mirevents function operationalizes the onset of a sound event, though the secondary analyses using the MiningSuite functions show that the differences in the measurements using the mirtoolbox, while statistically significantly later, are of negligible size when compared with the improved MiningSuite function.[12] The consistent delay in the location of the pocket also potentially complicates the findings of some groove studies that show listener preference for quantization/exactitude (M. Davies et al., 2013; Frühauf et al., 2013; and also Senn et al., 2016, which found similar groove ratings for quantized performances and performances with their natural microtiming profiles).

A separate finding involving this general delay in the pocket location is that further Kolmogorov–Smirnoff pairwise comparisons between downbeats and backbeats—bass drum beat 1 and snare drum beat 2—show that the nature of the pockets at these two significant metric pillars do vary significantly most of the time. Briefly, in the Slow band: Chamberlain's BD1 and SD2 differ significantly, but not Hakim's. In the Medium band: Chamberlain, Smith, and Waronker's downbeat and backbeat pockets differ, but not Hakim's. And in the Fast band: Hakim, Smith, and Waronker differ significantly, but not Chamberlain. This supports, to an extent, Iyer's emic assertion that snare drums are often played slightly later than bass drums, a small difference he actually labels as playing "in the pocket" (Iyer, 2002, p. 406). These differences may simply be an artifact of the bass drum and snare drum's different sound envelopes' impact on the detected onset location and it is also vital to remember that these two instruments are used for different purposes in the performance of a drum groove. That being said, this finding still builds support for challenging presumptions of isochrony in meter and reminds us of the importance of situating individual beat pockets within larger metric frameworks.

To dwell briefly on meter here, the start of this chapter questioned whether empirically

[12] It should be noted that the MIRtoolbox did, on average, locate the onset time *later* than the MiningSuite's improved method (a mean difference of 5.235 milliseconds for performances at 90 BPM, which is 0.00785 beats), though the effect size of these differences was near-zero. To contextualize this difference between the two measurement methods, the average distance the Loop Loft drummers played behind the external metronomic reference they were listening to (the degree to which the pocket was laid back) is three or more times this amount. As such, no matter which of the two measurement methods is used, the pocket would still be located after the click track.

understanding the interaction of pockets with metric structures would enable one to describe a performer's unique feel more specifically. This was raised with particular regard to performances of the One and backbeats, both integral elements of a drum groove whose subtle nuances have been associated with meaningful musical and experiential phenomena.

The One is a characteristic rhythmic gesture in funk (and R&B, soul, and related genres) that James Brown championed and his band, particularly bassist Bootsy Collins and drummer Jabo Starks, spread throughout the world of funk performance. It relates to the manner in which beat 1 is played, both syntactically and processually. Starks says: "dancers dance to beats that's right on top of everything. A dancer is always on top. So that's what I equated the 'on the one' to mean" (Y. Smith and Walsh, 1996, 13:20 ff.—full interview available in Deane, 1995a). His description and performance demonstration for the interviewer stresses that the performance should be "on top"—"right on time" as opposed to pushed or laid back—and there should be a strong accent on beat 1 (which he contrasts with the backbeat accent in the blues). A demonstration by Collins shows that the downbeat should be clearly, simply, and heavily articulated, freeing up the rest of the bar for syncopation, and, in Collins's case, it should land ever so slightly early/pushed (Lent, 1983). Building on Starks and Collins's descriptions, Danielsen theorizes and explores the One further in a chapter titled "The Downbeat in Anticipation" (Danielsen, 2006, Chapter 5), explaining that the fundamental temporal quality of the One is that "the One should be played on top" (p. 73), "on top" here referring to a slight anticipation that elides the upbeat with the succeeding downbeat. Danielsen's ensuing analyses demonstrate events that are also *syntactically* on top, highlighting syncopations that are so *late* that they combine with the next beat, extending the strong beats (see, for example, the bass line to James Brown's "Sex Machine" (1970), transcribed and annotated in Danielsen, 2006, p. 81, Figure 10).

All of the analyses presented in this survey of drum performances show that beat 1 (as performed on the bass drum) tends to come *after* the metronomic beat location, not on top (whether that is “right on time” or before—as can be seen, musician’s usage of these metaphorical descriptors can lead to ambiguity of meaning). This could be a result of these patterns being derived from a certain style-constrained model (the basic rock groove), or might be linked to the previously discussed concern as to where the original click track is located and whether the start of the recordings as provided by the Loop Loft is exactly in line with the click track used in the recording studio. What can be said without ambiguity is that very few of the bass drum beat 1 distributions actually differ from the bass drum beat 3 performances suggesting that the One, in this corpus, does not have a unique temporal profile. The set of performances included in the Loop Loft corpus represents a variety of musical styles, not just funk, though Danielsen points out how it became “a distinguished, ‘funky’ trait of many styles other than funk” (Danielsen, 2006, p. 139) and I believe the concept can still be meaningfully applied to the styles contained in the present corpus. This quantitative finding does not threaten the existence of the concept of the One: Danielsen clearly highlights the difference between metronomic time and performed time, so the downbeats here, while not chronometrically early, may still be *phrased* early, whether through dynamics, timbre, interaction with other instrumental layers, or the rhythmic design of the pattern (particularly if there is a syncopation that is just before the window of time being analyzed here). The drum performances analyzed from the Loop corpus are also in isolation. A bassist could, for example, play against the drummer’s beat 1 pocket, anticipating the downbeat/the One, stretching the ensemble pocket earlier.

Lastly, the near-total absence of significant differences for beat-pair comparisons (between each drummer’s BD1 and BD3, SD2 and SD4) presents an interesting situation where the first and second halves of each 4/4 bar of a drum groove are not notably different. As such, the data suggest that you could hypothetically swap the two halves to no noticeable change—a swap that would likely violate the standard metrical hierarchy of beat 1 being strongest, then 3, then 2 and 4 (see, for example, Figure 2.7a of Lerdahl & Jackendoff, 1983, p. 19) as well as diminishing the significance of the One for these patterns.[13] In fact, there is a very niche sub-genre/musical meme called "Beat Edits" where people take famous tracks and swap, for example, beats 2 and 4 by slicing apart every bar of the track. The result makes no lyrical or melodic sense, but the rhythm section's parts often end up working remarkably well.[14]

Conclusion

This chapter has presented a methodology that can be applied to recorded music to understand how performers create a personal "feel" through how they cumulatively locate their onsets in time. The methodology involves collecting timing data for onsets associated with particular, meaningful musical locations (here, metric landmarks), aggregating them across multiple bars and multiple performances, and then analyzing the distribution of events around this landmark. By analyzing the aggregate, we can understand the probabilistic tendencies of the individual performers—where they are likely to locate their onsets in time—and thereby define the location, scale, and shape of their pockets. Additionally, it is possible to statistically compare two performers or two unique locations by the same performer to see whether the pockets associated with these moments differ and, if they differ, in what ways do they differ (whether in location, scale, or shape).

I have tied qualitative descriptors to these quantitative analysis of performances, showing how quantifiable differences in the location, scale, and shape of pockets can be made musically meaningful through precisely applying performers' colloquial language of "tight" or "loose," "pushed" and "laid back." To reiterate what was illustrated in Figures 3.3 and 3.4: Changes in where the peak/center/average of a pocket (these are not one and the same) relate to how pushed

[13] This is not a concern for Danielsen as she argues that downbeats in anticipation can actually apply throughout the bar (see, for example, the section on the "Way to Treat Strong Beats," Danielsen, 2006, p. 79ff.).

[14] Probably the most popular example of this involves swapping beat 2 and beat 4 of every bar of Daft Punk's "Get Lucky": https://www.youtube.com/watch?v=dSvvlu5zTDQ (Emond, 2018).

or laid back a performer is; changes in how spread out/how wide/the variance of the distribution can describe how tight or loose a performer is; and any differences remaining after controlling for location and scale are to do with the asymmetry (skewness) and consistency (kurtosis) of a performance.

The theoretical discussion that pervades this chapter works to integrate primary and secondary, syntactical and processual, linguistic and pre-symbolic features of music into something that adds meaningful insight into understanding what musicians are doing when they create a performance. While bearing in mind the former member of each of these pairs, I recenter attention onto the latter member to prioritize the elements of music that are often underserved by "traditional" music-analytic tools and to provide a way to engage with facets of music that emic performers in the popular music and jazz scenes value and intentionally work to create. In doing so, we understand a musical style on its own terms.

A note of caution: the timing nuances in the performances being analyzed are measured in milliseconds. For example, in the Medium tempo band, Chamberlain's mean location for his bass drum beat 3 is 3.030 while Hakim's is 3.022. At 96 BPM (halfway between the band's bounds of 80 and 112 BPM), this difference of 0.011 beats would be a difference of just 6.875 milliseconds. While there may be statistically significant differences that supported the hypothesis that Chamberlain's bass drum beat 3 pocket is "laid back," 6.875 ms is potentially below the threshold for distinguishing between two events in musical contexts. The following chapter presents experimental work to explore this exact concern and provides perceptual context for the analytic claims, illustrating the complex picture when considering the perception of real musical performances, but overall validating much of the analytic work of this chapter. The differences described in this chapter, which are legitimate differences in how the individual performers shape time, can be accepted, though we must always include in microtiming studies the caveat that some of the differences may potentially fall beneath the threshold at which these differences are perceived for some listeners.

Finally, the analyses presented here have been focused on a corpus of drumming in Anglo-American popular music styles (that are heavily influenced by musical cultures of the African diaspora), but the methodology described here is not limited to these genres or instruments. It would be just as meaningful and insightful to apply these methods in considering how individual beat locations are performed and stylized in numerous other musical settings. Of course, in each setting, it is important to listen to the language emic performers use to describe their playing in order to find appropriate and meaningful descriptors. It may, for example, be interesting to apply this methodology to understand the unequal subdivisions of beats in various genres of Brazilian samba, particularly with regard to the anticipated third and fourth subdivisions (when each beat is divided in four). It is well established that this non-isochronous phenomenon occurs (Gerischer, 2006; Gouyon, 2007; Haugen & Danielsen, 2020; Haugen & Godøy, 2014; Lindsay & Nordquist, 2007; Naveda et al., 2011; Naveda & Leman, 2009), but is there meaningful variance in how different performers construct these divisions? Far less well understood is the Gnawa music from sub-Saharan West Africa (principally Morocco) that also features asymmetric subdivisions of beats, performed on qraqab/krakeb (large metallic castanets), in two fundamental rhythmic patterns that are meant to imitate trotting and galloping horses (Sum, 2012, 138ff.). A culturally sensitive application of this chapter's methodology would enable one to see whether there are unique timing profiles—on top of syntactical differences—that are more reflective of trotting than galloping style performing. Both of these examples relate to the timing profiles of percussion performances as drums are often the key articulators of musical time, though there is nothing to limit the methodology outlined in this chapter from being applied in a variety of other musical contexts.

CHAPTER 4

The Perception and Discrimination of Different Musical "Feels"

In order to talk meaningfully about the qualitative experience of musical time that is encapsulated by the term "feel" and its relation to the nuancing of musical events at the millisecond level, we need to understand whether listeners have the ability to perceive these subtle performance details. It is straightforward to make measurements of a performance and make claims about how "tight" or "loose" a performer is (how small or large the variance to their microtiming is), or comment on how "pushed" or "laid back" they are (how far ahead of or behind the metronomic point of reference their average onsets are), but these analytic findings are only meaningful if they relate to actual perceivable phenomena. This need is far from new—decades ago, Pressing's microstructural analysis of a keyboard improvisation referred to "the issue of the meaningfulness of resolution" (Pressing, 1987, p. 153)—yet there has been little music-specific research, even in the years since, that provides analysts with concrete guidance as to what is perceivable with complex musical signals. Pressing, in fact, concludes that "a straightforward answer is not easy" (p. 153). In the previous chapter, the analysis of pocket locations, scales, and shapes, and the description of their potential experiential qualia using the language of "feel" was predicated on the idea that small differences in how the individual drummers cumulatively create their beat pockets are musically meaningful and not simply a coincidence or consequence of some kind of biomechanical or technological idiosyncrasies in the performances.

To begin to explore these issues, this chapter investigates experimentally whether listeners are able to discern different musical feels using an ABX paradigm (described below) with synthesized recordings of drum performances that have carefully defined timing profiles. This methodology enables us to understand whether listeners can accurately recognize timing profiles without relying on their ability to communicate their awareness linguistically (as expert performers, producers, or listeners may be able to). By validating the perceptibility of microtemporal phenomena in the context of modern drum performance styles—and not just in lab settings with non-musical streams of tones and clicks—analyses of feel can be made with confidence that they reflect perceptual experiences.

Performers and creators are seemingly able to operate at the "micro-" level when creating musical feels. Previous research shows that performers can operationalize feel descriptors like "laid back" and "pushed," nuancing the loudness, timbre, and timing of their performances in systematic ways (Câmara et al., 2020a, 2020b; Danielsen, Waadeland, et al., 2015; Kilchenmann & Senn, 2011; Sioros et al., 2019). Additionally, music producers are known to make intentional use of delay and pre-delay/anticipation for aesthetic means (see articles in trade magazines, e.g., Anon., 2016; Hobbs, 2020), so much so that some Digital Audio Workstations and third-party plug-ins even have "feel" settings built in. For example, Pro Tools' Beat Detective uses "Feel Injector Templates"[1] that modify the timing of quantized MIDI; Reason has a "ReGroove Mixer" whose "Slide" dial adjusts the (pre-)delay of all onsets; Soundtoy's "EchoBoy" plug-in has a "Feel" dial that ranges from "Rushin" to "Draggin"; and AIR's "Strike 2" plug-in has a "Feel" dial that

[1]Now called "DNA Groove Templates" by the original plug-in maker Numerical Sound.

ranges from "Ahead" to "Fat." Evidently, microtemporal modifications can be utilized in the act of creating a specific performance with the intention of communicating a specific feel. However, creators and performers who operate at the microtemporal level are highly specialized listeners, highly experienced at focusing on subtle musical details, and so they may be utilizing listening strategies different from those of typical listeners. They may also benefit from kinesthetic and/or visual information about timing. For example, a drummer may feel and/or see that their left hand hit is slightly ahead of their right, which can enhance their discrimination of the sounds through multisensory integration (Spence et al., 2003; Zampini et al., 2003, though see also Zampini et al., 2005). What remains to be understood, and what needs to be understood in order to truly make analytic arguments about musical "feel," is whether "normal," non-specialist listeners can perceive these differences and comprehend the aesthetic consequences.

Matthew Butterfield (2010) explores the perception of "participatory discrepancies" ("semi-conscious or unconscious slightly out of syncnesses," Keil, 1987), investigating whether listeners can detect onset asynchronies between bass and drums in a jazz setting. While a key part of the expressive qualities of swing is the swing ratio/Beat Upbeat Ratio (see, for example, Benadon, 2006; Collier & Collier, 1996; Friberg & Sundström, 2002; Reinholdsson, 1987), Butterfield also stresses the importance of asynchronous timing events between bassists and drummers, which, in his argument, is a negotiation of the beat that creates an expressive tension. To explore this, Butterfield conducted a perception study in which a synthesized walking bass line was heard against a standard jazz swing drum pattern. The two instruments' performances were each synthesized perfectly in time with the metronome; however, the alignment of the two parts was systematically varied across seven conditions such that the bass part was consistently ahead of the drums by 10, 20, or 30 ms, the two parts were perfectly in synchrony, or the drums led the bass by 10, 20, or 30 ms. The participants (students in a music fundamentals class and students in an interdisciplinary course on time and rhythm) were asked to identify which instrument was ahead.

Butterfield found participants performed barely better than chance at this task, concluding that "ordinary listeners, irrespective of music training or stylistic preference, are unable to discern a discrepancy of 30 ms or less between bass and drums" (p. 165).[2]

Butterfield's article, and others related to it (e.g., Chor & Ashley, 2006; Matsushita & Nomura, 2016), raises the issue of how discriminating listeners may be about different pockets and feels. These evaluations of the perceptibility of asynchronies would appear to be highly problematic for the analyses in the preceding chapter, as the differences between drummer pockets are far below this 30 ms threshold and therefore, according to Butterfield, are not readily available to perception and so potentially of no musical meaning. However, as important as Butterfield's study is to understanding the role of participatory discrepancies in music, its methodology requires participants to perform a fairly sophisticated musical task: accurately recognize precisely what microtemporal phenomenon they are hearing. This does not reflect typical, everyday listening strategies and would be especially challenging for participants who are less experienced at perceptually separating the layers of a musical performance, a skill required to hear the bass part separate from the drum part and thereby make the judgement of which is ahead of the other. Participants may well have been able to recognize that there was *some kind* of asynchrony, but accurately naming it proved too challenging. As such, the question of whether microtemporal musical phenomena are perceptible still remains unanswered.

Aside from Butterfield's music perception study (and related research), there is a separate, but in ways related, body of research that helps to inform us as to whether listeners are able to perceive, judge, or make use of millisecond differences in the location of sounds. This is the literature on "Just Noticeable Differences" (JND)—the amount something must be changed in order for a difference to be (statistically significantly) perceptible to a perceiver. Although the

[2] This finding that participants *cannot* recognize and articulate what type of profile they heard is also seen in Ganis et al. (2021), which asked participants in a "danceability" study to assess whether the heard drum performances were laid back, pushed, or on beat (using stimuli from Câmara et al., 2020b). Both musicians and non-musicians could only correctly label the timing profile that they heard at chance levels.

present study is not asking about JNDs, this well-established field of psychophysical testing gives a wide range of results for how fine-grained listeners' perceptual acuity is, variously finding that listeners have a temporal resolution of just over 1 ms (Fujii & Schlaug, 2013), up to 10 ms (Friberg & Sundberg, 1993; Madison, 2004), and even down to 10 μs (Brughera et al., 2013). The majority of these types of study look at whether a sequence of clicks is perfectly isochronous or whether one of the events occurs slightly earlier or later, though other methodologies include detecting whether two sounds are perfectly synchronized, judging the temporal order of two simple tones, and detecting small gaps in a continuous sound.

The wide range in JND threshold values may be attributable, at least in part, to the different methodologies utilized. This is clearly seen in the results of the various subtests of the Harvard Beat Assessment Test (Fujii & Schlaug, 2013): The Beat Interval Test (detecting tempo change in an "isochronous" sequence of woodblock tones) found a mean JND of 1.83 ms (SD = 1.48) while the Beat Finding Interval Test (detecting tempo change in a repeated simple rhythmic pattern on a woodblock) found a mean JND of 1.18 ms (SD = 0.95). What is more, these JND threshold values change when participants tap the beat/rhythm themselves, instead of simply reporting that they perceived a tempo increase or decrease. When tapping, the BIT JND mean is 0.48 ms (SD = 0.25) and the BFIT mean is 0.76 ms (SD = 0.57). It is evident, therefore, that the temporal resolution of aural perception is highly context and task dependent. As such, if we are to talk meaningfully about feel and its relation to microtemporal musical events, we cannot rely on JND research that uses pure tones/woodblocks/white-noise clicks; instead, we need music-specific—and perhaps even genre- and/or instrument-specific—research. This present study focuses on the complex, multi-instrument texture of drum kit performances to provide drum-specific measurements of what kinds of drummer "feels" are perceivable by American listeners. The methodology presented here may easily be replicated with other musical instruments to establish microtemporal perceptual baselines for other musical contexts.

Experiment Overview & Summary of Findings

Here, I present a series of experiments that investigate, in an exploratory manner, whether listeners can hear timing differences that are of the magnitude found in the preceding computational analyses of performances. This set of experiments progresses from a general evaluation of whether any overall differences are perceivable, through the more "obvious" difference of early vs. late, to the much more subtle question of whether listeners can aggregate numerous onsets over time and use this aggregate to gauge the consistency of a performer's timing (whether they have minimal variance—tight—or more variance—loose—in how they locate their onsets in time). The detailed findings of each of the three exploratory studies, as well as a full description of the methodology and analytic process, can be found below, but I first provide a general précis of findings here.

Experiment 1: Summary

1A—General Test of Discriminability of Performance Profiles

Can listeners detect differences between drum performances that have different timing profiles?

Null hypothesis: listeners' performance in a discrimination task is no different from chance—they cannot detect differences between drum performances that have different timing profiles.

Experiment 1 found that participants, who were presented with all possible pairings of pushed tight, pushed loose, laid back tight, and laid back loose performance timing profiles, could, for at least some participants and in some combinations, successfully discriminate between drum performances with different timing profiles at above chance rate (as measured by both an exact binomial test and signal detection theory—explained below).

1B—Effect of Tempo

Are listeners able to detect differences in timing profiles for performances both at slow (80 BPM) and moderate (110 BPM) tempi? (These tempi are chosen to align with the most salient data in the previous chapter's analysis of Loop Loft data)

Null hypothesis: listeners' performance is no different from chance at 80 BPM *and/or* 110 BPM.

There was a significant difference in the ability of participants to discern different timing profiles at different tempi with participants performing better (i.e., more success at discriminating between the two profiles) at 110 BPM than at 80 BPM.

Experiment 2: Summary

2A—Pushed vs. Laid Back Timing Profiles

Can listeners detect differences between drum performances that have tendencies to locate onsets *before* and *after* the metronomic beat (i.e., between *pushed* and *laid back* timing profiles; see previous chapter, Figure 3.4 for a full illustration)?

Null hypothesis: listeners' performance is no different from chance—they cannot detect differences between drum performances that have *pushed* and *laid back* timing profiles.

Individual participant performance varied widely in this set of performance profile contrasts and participants generally exhibited large confidence intervals and high false alarm rates. While participants generally are able to hear differences above chance performance between drum performances that have pushed and laid back timing profiles, they find this task rather challenging and success at the task is highly idiosyncratic.

2B—Effect of Tempo

Are listeners able to detect differences in pushed vs. laid back timing profiles for performances both at slow (80 BPM) and moderate (110 BPM) tempi?

Null hypothesis: listeners' performance is no different from chance at 80 BPM *and/or* 110 BPM.

There was no significant difference in the ability of participants to discriminate between pushed and laid back timing profiles heard at 80 BPM and at 110 BPM.

Experiment 3: Summary

3A—Tight vs. Loose Timing Profiles

Can listeners detect differences between drum performances that have *low* and *high* amounts of variance (i.e., between *tight* and *loose* timing profiles)?

Null hypothesis: listeners' performance is no different from chance—they cannot detect differences between drum performances that have *tight* and *loose* timing profiles.

Experiment 3's investigation of whether participants can discriminate between tight and loose performance profiles find results that are even more mixed than the previous two studies. While the success rate is above chance when the participants are taken as a whole, when this is investigated more thoroughly per participant and with more sophisticated analytic tools, participant performance becomes no different from chance. Additionally, because of wide variability in participant performance and poor performance overall, a lot of participant data is unanalyzable. As such, this study does *not* support the hypothesis that listeners are able to detect differences between drum performances that have tight and loose timing profiles.

3B—Effect of Tempo

Are listeners able to detect differences in tight vs. loose timing profiles for performances both at slow (80 BPM) and moderate (110 BPM) tempi?

Null hypothesis: listeners' performance is no different from chance at 80 BPM *and/or* 110 BPM.

As in Experiment 2, there was no significant difference in participant performance at the task between performances heard at 80 BPM and 110 BPM.

Additional Findings

Participants in all three exploratory studies also answered a number of questions about their musical training and self-reported listening habits. While this survey is nowhere near as thorough as, for example, the Gold-MSI (Müllensiefen et al., 2014), the goal was to get an impression of whether the ability to successfully discriminate between drum performance timing profiles is a specialist skill associated with musical training or close involvement with music as a listener, or whether it is a more general ability. In each of the three studies, there was no correlation between success at the discrimination task and musical training, musical activity (i.e., whether participants perform music), or number of hours of listening per week.

* * *

Experiment 1: Comparisons of all Performance Profiles

Method

Participants

36 participants (18 female), aged between 24 and 69 years ($M = 41.6$, $SD = 13.8$), took part in the experiment. Participants were recruited through Amazon MTurk in early 2021. MTurk is

a useful tool for research because it enables the researcher to obtain more diverse samples than traditional "WEIRD" samples of college students (Buhrmester et al., 2011; Follmer et al., 2017; Paolacci & Chandler, 2014). Additionally, MTurk workers provide high-quality, reliable data (Bartneck et al., 2015; Buhrmester et al., 2011; Holden et al., 2013; Rouse, 2015) and are somewhat representative of the general population across various psychological dimensions (McCredie & Morey, 2019). There are concerns about automated "bots" providing junk data that are not without merit; however, this risk can be attenuated with careful validity checks and screening (Chmielewski & Kucker, 2020). To mitigate against bots, fake accounts, and multiple submissions, participants in the current studies had to have an approval rate at, or above, 98%, had to have completed more than 500 tasks, and had to be accessing the study from an IP address in the United States of America. Additionally, the questionnaire software (Qualtrics) monitored who was accessing the study using cookies and prevented anyone from taking the study multiple times (to obtain multiple payments). Participants were compensated for their time and incentivized to pay attention by offering a 25% bonus payment if they passed all attention checks. All of the experimental procedures received the relevant approval from the Northwestern Institutional Review Board.

Procedure

This study makes use of an ABX/"match to sample"[3] paradigm in which participants listen to three stimuli each trial: two reference stimuli **A** and **B**, followed by a third stimulus **X** that shares the properties of, but is not identical to, either **A** or **B**. Participants then make a choice as to whether **X** sounds more like **A** or **B**. Over the course of all trials, participants are given stimulus sets where **X** pairs with **A** and where **X** pairs with **B** to allow full analysis of their ability to perceive differences. This paradigm is ideal for testing naïve participants as they do not need

[3]There is a slight difference between ABX and match to sample in that match to sample is an XAB task. The experiment presented here is an ABX with match to sample mentioned only to provide an alternative label that may facilitate comprehension of the goal of the task.

to have any knowledge about the physical dimensions along which the stimuli vary, nor do they need to be able to verbally articulate what the differences/similarities are (cf. Butterfield, 2010).

The study was presented online using the Qualtrics survey platform and delivered using Amazon MTurk. Participants first provided informed consent, successfully completed a CAPTCHA (to distinguish human from machine input and thwart spam/bots), and reported demographic information. Participants' musical training and listening habits were surveyed using a brief questionnaire that asked about listening quantity, listening style (active/passive), listening choices (genres), and musical training amount (years of training and years since training) and type (formal training/self-taught/none). To assess self-reported hearing ability, participants answered the following screening question: "Which sentence best describes your hearing status (while not using hearing aids)?" Response choices were: (1) Don't feel difficulty at all, (2) A little bit of difficulty, (3) Very difficult, or (4) Can't hear at all. Self-reported hearing difficulty was indicated when participants selected (2), (3), or (4), and only participants who reported no difficulty at all (1) were included in the analysis (Choi et al., 2019). Before the true trials, participants were presented with a demonstration trial that walked through how to successfully complete the task to ensure task comprehension. Lastly, attention and comprehension checks were interspersed with the experiment trials (see below for full details of data screening).

The core of the experiment involved 24 trials that were split into two blocks with an optional break in the middle. In each trial participants were presented with three independent media players labelled "Drummer 1," "Drummer 2," and "Drummer 3" that they could play in any order and as many times as they wished (Figure 4.1). Participants could not advance to the next trial until a duration equivalent to playing each track once had elapsed; this requirement was intended to encourage participants to listen to each track and not simply click through the study. Participants responded to the question: "Who does Drummer 3 sound most like?" by clicking a radio button for "Drummer 1" or "Drummer 2." For a full overview of the experiment design, see Figure 4.2.

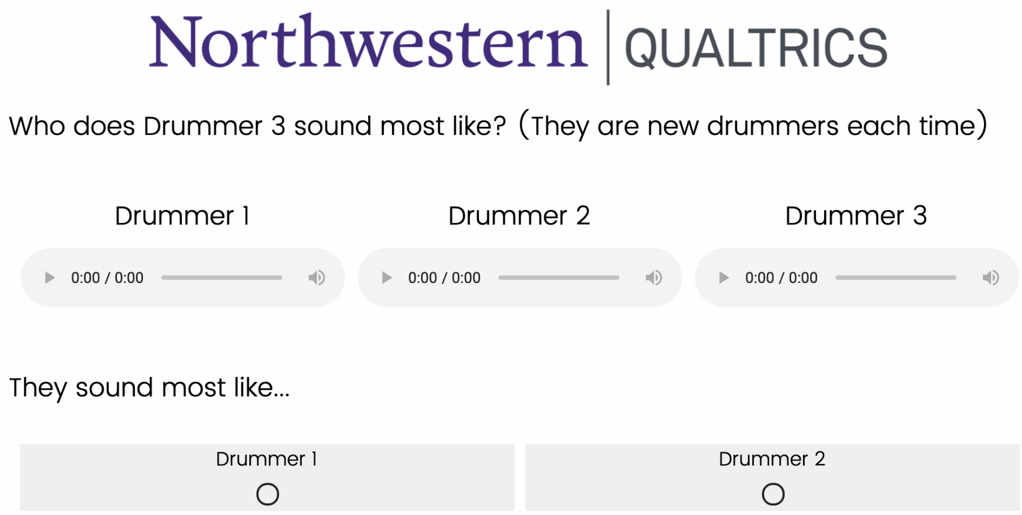


FIGURE 4.1: A screenshot of a standard trial in the experiment.

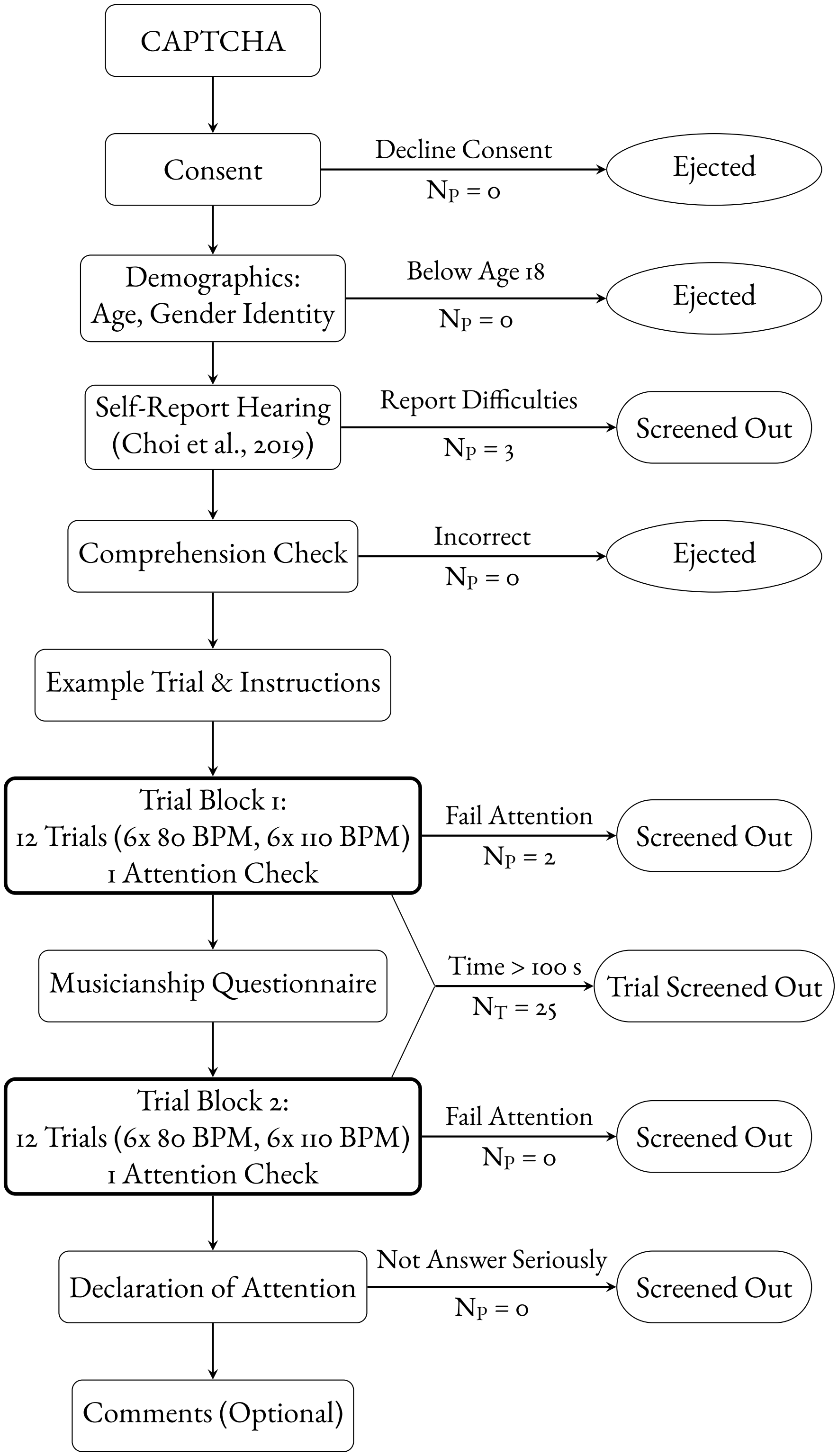


FIGURE 4.2: Flowchart of the experiment including data screening processes. N_P refers to the number of participants screened out or ejected in Experiment 1, while N_T refers to the number of trials.

Data Screening

The quality of the data provided by MTurk samples is known to be quite high and has been found to approximately equal that obtained from traditional college student samples (Buhrmester et al., 2011; Goodman et al., 2013; Paolacci & Chandler, 2014). However, it is especially important to have stringent data screening criteria when conducting an unmonitored, online study. Several methods were used to screen out any participants who may be bots (i.e., a computer program rather than a real human) as well as those who are simply proceeding through the study as quickly as possible without actually participating. These methods ranged from a CAPTCHA ("Completely Automated Public Turing test to tell Computers and Humans Apart") at the start of the study (whose function is to block automated bots), to attention check trials, to a simple question at the end: "Did you take the study and your answers seriously?" (Aust et al., 2013). Attention checks involved playing participants words instead of musical performances: For example, an attention check trial's **A** stimulus would be a recording of someone saying the word "cat," **B** would be a recording of the word "dog," and **X** would be "cat" again. To pass the attention check and not have all data screened out of the study, the participant would have to select the "Drummer 1" button as **X** matches the **A** stimulus. A comprehension check was also utilized to ensure participants had a grasp of the instructions and involved asking participants to select the word that describes something that you use to cook from a list of unrelated items, e.g. "rock," "shoe," "guitar," and the only relevant word, "stove." Participants who successfully completed all attention checks were given a 25% financial bonus as an incentive to make a concerted effort in the study. In order to minimize the role of any hearing difficulties on a participant's ability to perform the experimental task, all participants who self-reported any hearing challenges were removed from the final data. Lastly, any trial that lasted longer than 100 seconds was screened out from the analysis as it is highly likely that participants were distracted on this trial.[4] See Figure 4.2

[4] 100 seconds was arrived at as the cut off by inspecting the distribution of trial times.

for precise details of the outcomes of the data screening process in Experiment 1.

Stimuli

72 standard "boom tish" drum patterns (bass drums on beats 1 and 3, snare drums on 2 and 4, and eighth note hi-hats throughout) eight bars in length with unique timing profiles were synthesized for use in the study. This set of stimuli included 36 performances each at 80 BPM and 110 BPM. In each trial all stimuli were at the same tempo. The timing profiles for the drum patterns were generated by drawing random onset times for each drum strike from normal distributions that were precisely defined to have specific characteristics (see Table 4.1 and Figure 4.3).[5] These onset times were then converted into MIDI note events, which were subsequently rendered into audio by creating a software instrument in Logic Pro X that was loaded with samples of Matt Chamberlain's "Vintage Gretsch" drum kit from the Loop Loft. There were five different profiles that could possibly be assigned to each of the drums:

- Metronomic
- Laid Back (i.e., mean onset location is after the metronomic beat) and Loose (i.e. wide variance)
- Laid Back and Tight (i.e., small variance)
- Pushed (i.e., mean onset location before the metronomic beat) and Loose

[5]For the Python source code used to generate the profiles and create the MIDI files used in the study, see the OSF deposit: https://doi.org/10.17605/OSF.IO/C6QXN. Broadly, this code defines a set of 1000-point normal curves with the properties defined in Table 4.1. It then pulls random timing data from the assigned profile, with replacement, for each drum at each specified metric position for the full eight bars (every onset is uniquely pulled from the distributions). Finally, having assigned a time code (measured in beats) to each note event, a MIDI file is then created. This drum machine is necessarily simplistic—sampling randomly from the distribution—to ensure control over the performance parameters. More sophisticated, and perhaps more "real," drum performances could be synthesized by picking timings from the distributions according to priors, for example incorporating Markovian or Bayesian logic that selects the next bass drum's onset time depending on the time code of the previous one, though the rules that would have to be defined for these enhanced drum machines could result in the addition of other confounding variables to the stimuli generated.

- Pushed and Tight

For experimental control and to limit the number of trials required, hi-hats were defined as metronomic throughout, while snare drums and bass drums both took on the same one of these profiles (e.g., metronomic hi-hats with snare and bass drums that are *both* "laid back and loose"). Building on the findings of the previous study (Chapter 3), the location and size (mean and variance) were set to be within the range of real life performances.

Profile Name	**Mean**			**SD**		
	Beats	**80 BPM**	**110 BPM**	**Beats**	**80 BPM**	**110 BPM**
Metronomic	0	0	0	0	0	0
Laid Back Loose	0.03	22.5	16.36	0.04	30	21.82
Laid Back Tight	0.03	22.5	16.36	0.02	15	10.91
Pushed Loose	-0.03	-22.5	-16.36	0.04	30	21.82
Pushed Tight	-0.03	-22.5	-16.36	0.02	15	10.91

TABLE 4.1: Summary of the parameters used to construct the normal distributions for each timing profile. Parameters are presented in terms of beats (e.g., 0.03 beats before/after the metronome) and also in milliseconds at 80 and 110 BPM (e.g., 22.5 ms after the metronome). Each distribution would be adjusted for each specific metric position, for example an event assigned to beat 2 would have a mean of 2 + the designated profile's mean (so the distribution for an event at metric position 2 in a laid back performance would be centered about 2.03).

As an example of how a one-bar virtual drummer's performance is created:

- Hi-hats are set to be perfectly *metronomic*, so for metric positions 1, 1.5, 2, ... 4, and 4.5 it simply samples those exact same values from the "metronomic" timing profile.

- Snare drums are to be *laid back and loose*, so for metric positions 2 and 4 it randomly pulls numbers from the appropriate distribution and returns: 2.02 and 4.04.

- Bass drums, due to a deliberate constraint of the study design, share the same *laid back and loose* profile as the snare drums, so for metric positions 1 and 3 it randomly pulls numbers from the appropriate distribution and returns: 0.99 and 3.03.

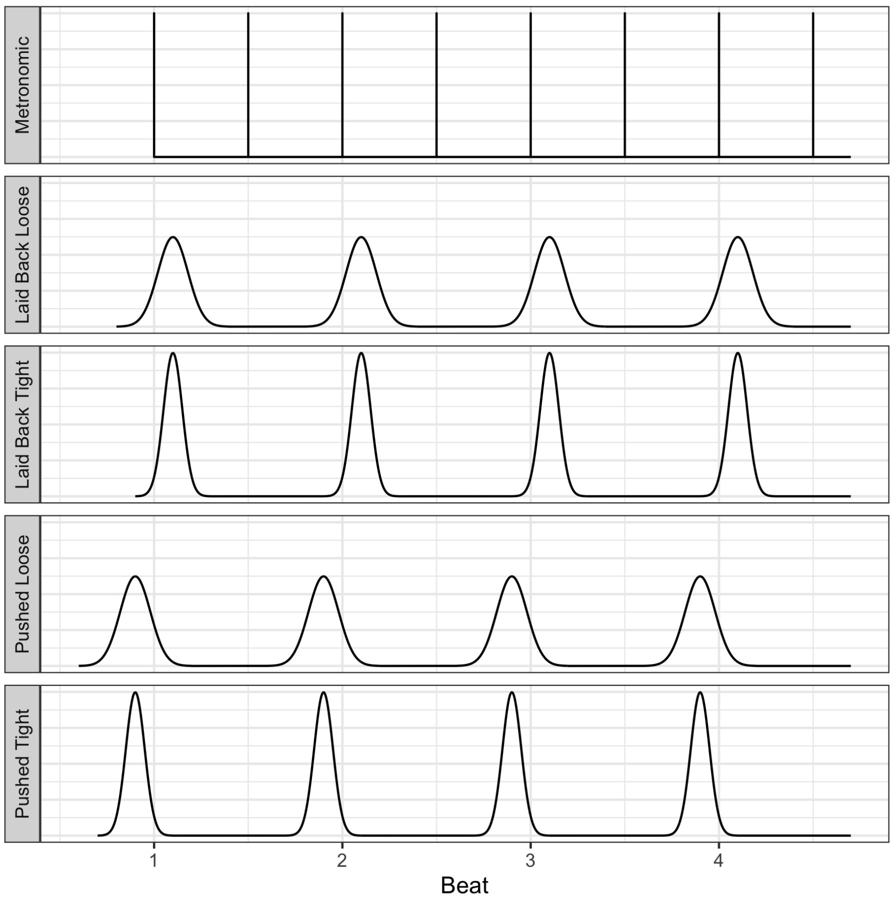


FIGURE 4.3: An illustration of the five timing profiles used in the present study. NB: The unique characteristics of each profile are exaggerated in this visualization to make the differences more readily perceived.

This generation process would result in the drum pattern illustrated in Figure 4.4.[6]

The 24 trials in Experiment 1 were constructed such that participants heard all possible pairings of the four timing profiles that could be assigned to the snare drum and bass drum, and that participants heard two instances of each pairing so that **X**, at different times in the experiment, could be both **A** and **B** (see Table 4.2). As will be explained below, this duplication is essential to facilitate the use of signal detection theory in the analysis. The ordering of stimuli was randomized.

[6]NB: The performances in the stimuli heard by participants involved eight unique bars, not the same one bar repeated eight times. All selections from the distributions are random with replacement.

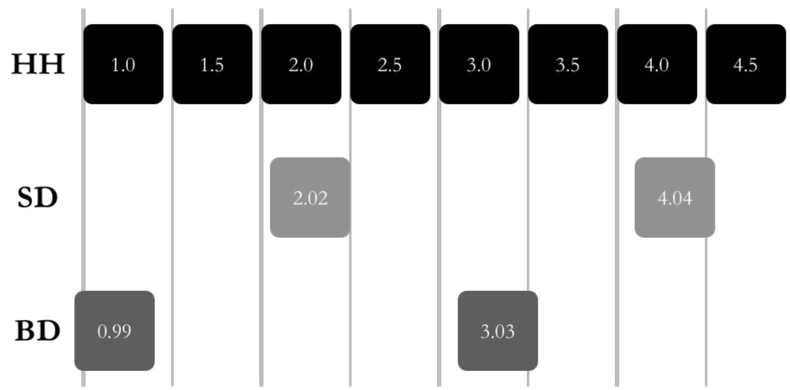


Figure 4.4: A graphic realization of the example virtual drum performance. Hi-hats (HH) are metronomic, Snare Drums (SD) and Bass Drums (BD) are laid back and loose. Onset times (in beats) are annotated on each note event.

Trial	**A**	**B**	**X**
1	80 BPM: Laid Back Loose	80 BPM: Laid Back Tight	80 BPM: Laid Back Loose
2	80 BPM: Laid Back Loose	80 BPM: Laid Back Tight	80 BPM: Laid Back Tight
3	80 BPM: Laid Back Loose	80 BPM: Pushed Tight	80 BPM: Laid Back Loose
4	80 BPM: Laid Back Loose	80 BPM: Pushed Tight	80 BPM: Pushed Tight
⋮	⋮	⋮	⋮
23	110 BPM: Pushed Loose	110 BPM: Laid Back Tight	110 BPM: Pushed Loose
24	110 BPM: Pushed Loose	110 BPM: Laid Back Tight	110 BPM: Laid Back Tight

Table 4.2: Table illustrating the experiment trial design. Braces show pairs where **A** and **B** stay the same while **X** changes. Ordering is randomized in the experiment.

Analysis

Statistical Analysis Method

Analysis of participant responses through the lens of Signal Detection Theory allows one to separate *sensitivity* (i.e., the ability to discriminate between **A** and **B**) from *bias* (e.g., a tendency, due to the experimental setup or a participant's answering strategy, to prefer selecting **A**). This more detailed approach does not simply consider the binary "Correct/Incorrect" scoring of each trial, but instead probes deeper, analyzing the two independent psychological processes that determine response proportions in detection or discrimination tasks: bottom-up sensitivity and top-down bias. The theory was first outlined in Green and Swets (1966), was updated and expanded in Macmillan and Creelman (2004), and more recently was reiterated and tailored for the specific requirements of auditory ABX paradigms by Boley and Lester (2009).

The core statistic in detection theory is d', which is a standardized estimation of the distance between distributions representing the two stimuli. Figure 4.5 illustrates what d' measures and Figure 4.6 clarifies the meaning of d' values with a second case where the two types of stimuli being compared are far more easily distinguished between and so the value of d' is much larger.[7] A helpful feature of d' is that, because it is calculated in standardized units, if one participant or condition's d' score is twice that of another, the sensitivity is exactly double, too. d' is calculated using the "hit rate" and the "false alarm rate," which is, respectively, the proportion of each participant's answers that are correct (hear **A**, select **A**) and that are false alarms (hear **B**, select **A**—see Table 4.3 for the full classification of participant responses).

The dashed vertical line in Figures 4.5, 4.6, and 4.7 represents the "response criterion," a measure of the willingness of a respondent to make a decision in an ambiguous situation. In

[7] ABX paradigms are actually slightly more complex than the case illustrated in these figures as they have *four* distributions in a 3D space representing each possible ordering and outcome of stimuli S_1 and S_2 ($<S_1S_2S_1>$, $<S_1S_2S_2>$, $<S_2S_1S_2>$, $<S_2S_1S_1>$). See Macmillan and Creelman (2004, p. 231) and Hautus and Meng (2002, p. 91) for more. That said, the fundamental concept is exactly the same and more easily grasped with these simplified, 2D illustrations.

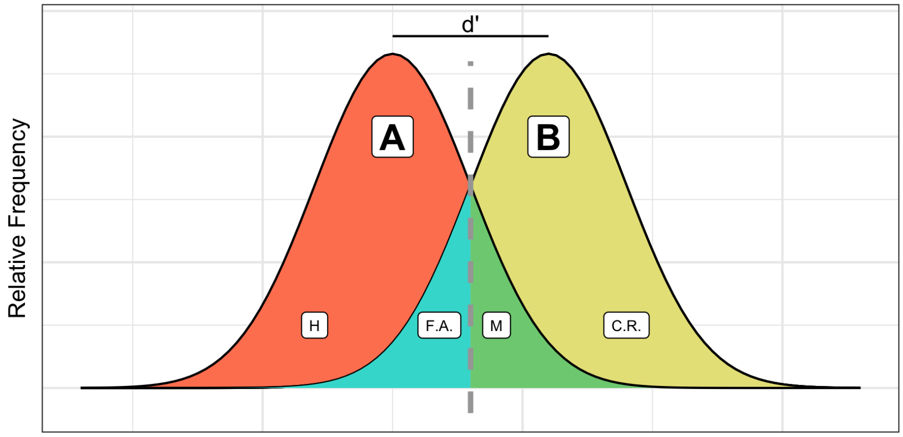


FIGURE 4.5: An illustration of what d' is a measure of: the distance between (standardized) distributions representing each possible choice: **A** and **B**. This figure is from the perspective of **X** = **A** so "H" stands for "Hit," "F.A." for "False Alarm," "M" for "Miss," and "C.R." for "Correct Rejection." (From the perspective of **X** = **B**, the value of d' would be identical, but the 'Hit," "Miss," etc. labels would be mirrored about the middle.) The dashed vertical line is a representation of the "decision criterion" and its location here, perfectly centered between the two distributions, shows an unbiased, neural observer.

	Choose A	**Choose B**
Hear A	Hit	Miss
Hear B	False Alarm	Correct Rejection

TABLE 4.3: Matrix showing definitions of the key terms used in signal detection theory. These are from the perspective of detecting **A**, so hearing **B** but selecting **A** is a false alarm, though the statistical results would be exactly the same if doing it from the perspective of **B**.

Figures 4.5 and 4.6, the hypothetical participant is neutral—equally likely to select **A** as they are to select **B**. If, however, they have a tendency to prefer selecting **A**, then this line—the response criterion, otherwise described as bias—would shift right (as in Figure 4.7). The two distributions would remain as they are and the participant's sensitivity (d') would be the same, however, now, the hit rate and the false alarm rate would both increase, while the number of misses and correct rejections would decrease (the increase in hits is obscured in Figure 4.7's visualization by the false alarm segment from **B**'s distribution, but it too is increasing). The rates of these changes are not the same, though, with the number of false alarms increasing *more* than the number of hits.

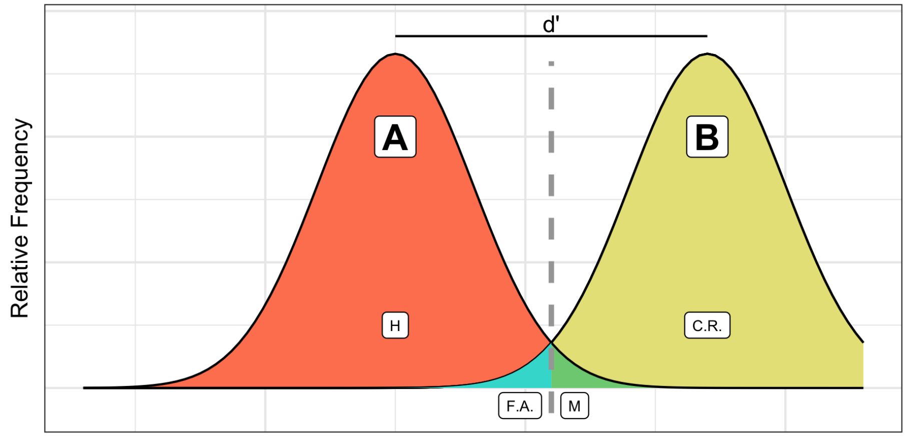


FIGURE 4.6: A second illustration of what d' is a measure of. Here, the value of d' would be much larger than in the previous illustration as the distance between the two distributions is greater. Here, the probability of "Hits" and "Correct Rejections" is far greater with only a small chance of "Misses" and "False Alarms" indicating that the two stimuli are much more easily discriminated between.

The measure of this bias, c, is the distance of this line from the central location, so a value of 0 indicates no bias while a value of ± 1 indicates complete bias (they only select **A** or only **B**).

d' was calculated per participant using the `dprime.ABX` function in the "psyphy" package (version 0.2-2; Knoblauch, 2020) in R. In analyzing ABX paradigms, it is better to assume that participants use a "differencing" strategy as opposed to an "independent-observations" strategy (Hautus & Meng, 2002) and so this was specified in the function's options. Using a difference decision strategy, the participant tries to find the stimulus—**A** or **B**—from which **X** is minimally different, whereas in an independent observations strategy, a participant categorizes **A**, then **B**, and, after having identified these two, categorizes **X**. The standard deviation for each value of d' was calculated according to equation 13.4 in Macmillan and Creelman (2004, p. 325) and the 95% confidence intervals are 1.96 standard deviations either side of the mean. If a participant's d' value and the 95% confidence interval surrounding this point are different from zero, this is evidence to support the argument that participants are indeed able to discriminate between the different drum performance timing profiles.

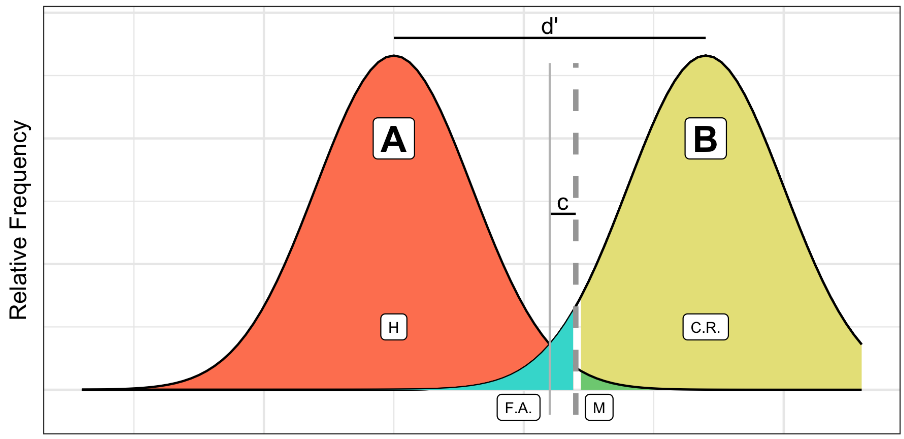


FIGURE 4.7: A third illustration of d', this time with the distributions exactly as in Figure 4.6, but with the decision criterion (dashed grey line) shifted right. The measure of bias, c is annotated, measuring the distance of the deflection from the midpoint.

It must be noted that some values in the following Results sections have had to be adjusted. This is because the calculation of both d' and c involves first evaluating the distance between the standardized distributions, which requires z scoring the data. Here, z scoring is a way of finding the x coordinate on a standard normal distribution (centered at 0 and with a standard deviation of 1) at which point the area under the curve equals the specified value. For example, $z(0.5)$ finds the x coordinate where 50% of the curve is to the left, i.e. the mean, which $= 0$ on the standard normal distribution. Regarding the challenge of calculating d' and c for some instances in the following analyses, $z(1)$ and $z(0)$ are both infinite as there is no x coordinate on the normal distribution that has 100% or 0%, respectively, of values beneath it (the normal distribution has no maximum or minimum). As such it impossible to calculate the bias-free measure of sensitivity d' or the response bias c when a participant's hit or false alarm rate is 100% or 0%. To avoid this issue, some hit and false alarm rates are adjusted up or down 0.05 (a hit/false alarm rate of 1 becomes 0.95, a rate of 0 becomes 0.05) to allow a value of d' and c to at least be approximated. Separately, the standard deviations of some participants' sensitivity, and therefore the associated confidence intervals, are incalculable as the hit rate or false alarm rate is 0.5 and the formula for calculating the SD involves dividing by z(hit rate) and/or z(false alarm rate), which would involve dividing by 0 (a mathematical impossibility). These cases are marked "NA" in the following tables.

Lastly, even though a general d' score, calculated using all data, will be presented at the start of the analyses to help provide an overview of responses, each participant introduces their own subjective criteria variable that cannot be combined with others, so they should be analyzed individually. This, however, results in a fundamental challenge of signal detection theory: it provides a way of estimating sensitivity and bias with far more clarity than a significance test using a binomial distribution (reporting that, e.g., participants got 70% of trials correct which differs significantly/non-significantly from chance performance), but there are no predefined performance indexes (i.e., what value of d' counts as "good" performance) and it is not possible to conduct inferential statistics.[8] As such, following individual d' analyses, mixed effects models are presented that explore which factors impact participant performance: Is, for example, one pair of contrasts (e.g., laid back loose vs. laid back tight) more successfully discerned than others? And what is the impact, if any, of musical experience on participant performance?[9] Mixed-effects analyses, performed using the "lme4" package (version 1.1-26; Bates et al., 2015) in R, allow the separation of "fixed effects" (which represent planned features of the experiment) and "random effects" (which represent unanticipated or unpredictable variation between individuals, stimuli of the same category, and other aspects of the experiment and data). The inclusion of random effects accounts for clustered, repeated measures data and enhances the power of the analysis, the

[8]See Macmillan and Creelman (2004, pp. 336–7) for a discussion of this problem. A partial solution is to pool participants together into groups that share similar biases (c) and sensitivities (d') (Macmillan & Creelman, 2004, pp. 331–6), though this introduces its own complications as the pooling criteria have to be subjectively decided upon.

[9]The hypothesis here is that participants who are more musically sophisticated will perform better (i.e., have finer temporal acuity) than those who are less musically sophisticated (Kraus & Chandrasekaran, 2010; Tzounopoulos & Kraus, 2009). However, this may well not be the case as a recent large meta-analysis of the cognitive benefits of musical training shows no overall benefits (Sala & Gobet, 2020). Danielsen et al. (in review), in an extension of the RITMO group's P-center work, find that expert training in a musical genre affects low-level perceptions of sounds, showing that folk, jazz, and EDM musicians respond to sounds by instruments in the genre they specialize in differently to instrumental sounds they are less experienced with. Additionally, EDM producers tended to locate the P-center sooner in the sounds than the other experts, and folk musicians had larger amounts of variability to where they locate the P-center.

statistical strength of analysis of fixed effects, and also allows the direct study of inter-individual variation.

Results

First, to get a general understanding of participant performance and an overview of the data, a one-sided exact binomial test was conducted to ascertain whether, overall, participants were able to discriminate between drummer performances with different timing profiles or whether their performance was indistinguishable from chance. As a reminder, Experiment 1 presented participants with all possible contrasts of performance timing profiles. Of the 719 individual trials that are included in the analysis (after screening out five participants and 25 trials that took too long), participants got 532 matches correct (74.0%) and failed to select the correct pairing in 187 cases, which is statistically different from chance performance ($p < .0001$). If we analyze all of the data using signal detection theory, this is also significantly different from chance: $d' = 1.962$, 95% CI $[1.858, 2.065]$.

Looking more precisely at the data, it is possible to evaluate the performance of each individual participant. Figure 4.8 shows each participants' hit rate against their false alarm rate, giving an indication of how successful each participant is at correctly discriminating between the two heard timing profiles. Annotated on this figure are dashed lines that show the boundaries for achieving each d' score. As can be seen, most participants have a sensitivity of at least $d' = 1$ or 1 normal standard deviation above chance performance.[10] Going further, Figure 4.9 plots each participant's d' value (indicated by the point in the plot) as well as the 95% confidence interval about this point (also see Table 4.7 at the end of this chapter for full information). Of the 29 responses with analyzable data,[11] all but one participant have d' values whose 95% confidence intervals

[10] The two values in the bottom right-hand corner of Figure 4.8 have incalculable d' scores as their hit rate is lower than their false alarm rate.

[11] Some participant data, even after adjustment, cannot be used for calculation of d'.

do not overlap with a score of 0 (an overlap that would indicate that there was no perceptible difference in the timing profiles). There are a further four participants with low *d′* scores and incalculable confidence intervals who may be presumed to fail to accurately distinguish between the performances. As such, it is possible to say that, for the vast majority of participants (24 out of 29 participants), they *are able to distinguish* between drum performances that have different timing profiles.

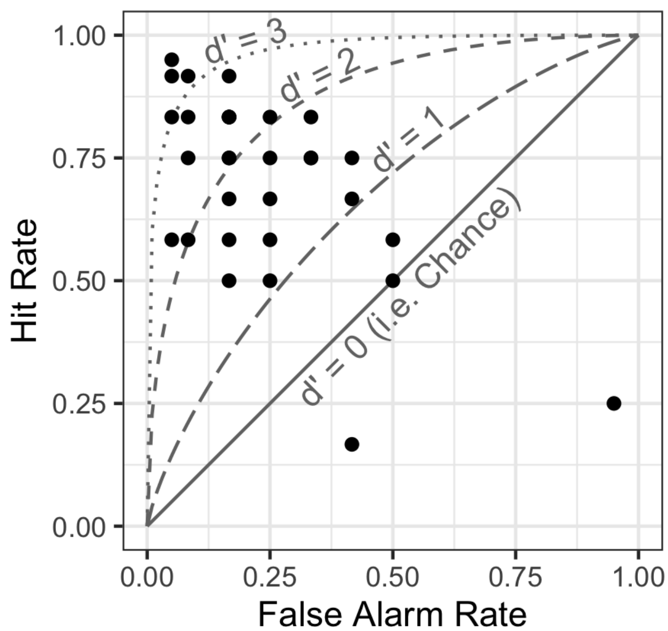


Figure 4.8: Each participant in Experiment 1's "ROC" (Receiver-Operating Characteristics), plotting their hit rate against their false alarm rate. Also annotated are lines representing the boundaries for specific *d′* values. The two values to the right of the diagonal center line do not have a calculable *d′* (see Table 4.7 at the end of the chapter).

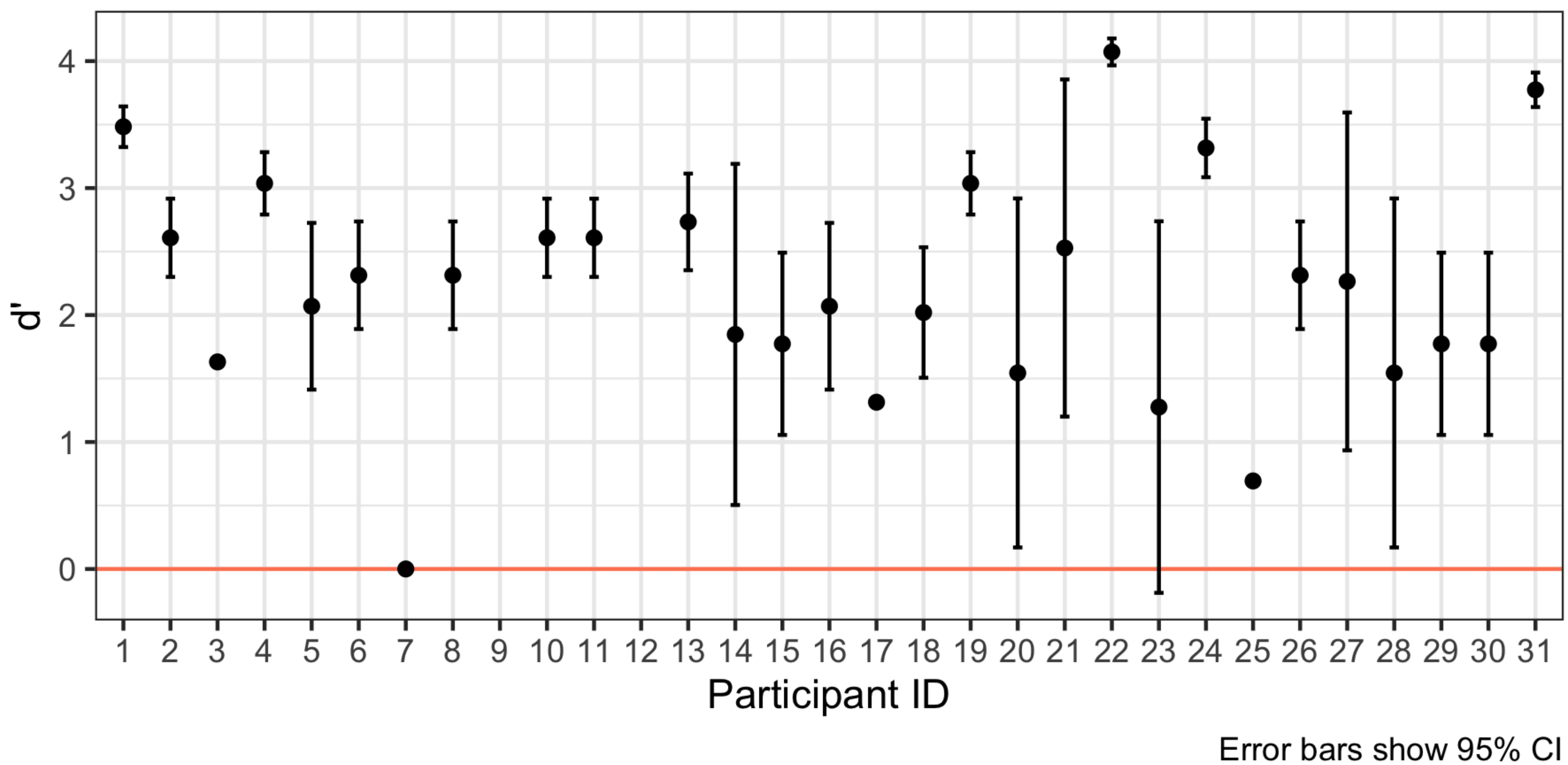


FIGURE 4.9: d' scores for each participant (Experiment 1). The horizontal coral/light orange line is to highlight where the threshold ($d' = 0$) is. Participants 9 and 12 are missing as their hit rate, false alarm rate, or both, even after adjustment, result in an incalculable d' (see Table 4.7). Some data points are lacking confidence intervals as they are incalculable, too (see explanation on page 134).

Analyzing this data set further, one can see whether there are any differences in each participant's ability to discern between timing profiles that are heard at 80 BPM and at 110 BPM. Figure 4.10 plots each participant's d' value, subdivided by tempo. As an example of what this figure reveals, participant 3 could not distinguish between performance timing profiles at 80 BPM (their d' value is 0), yet they performed very well at 110 BPM, having a d' well above zero, indicating that they can clearly distinguish between profiles. When breaking down the data in this way, because of the design of the experiment limiting the number of trials available for analysis at each tempo, there is a far greater number of incalculable statistics with some participants only having a d' value available at one tempo, plus some participants have no d' values whatsoever—see Table 4.8 (end of chapter). Analysis across all participants shows that there is a significant difference in the ability of participants to discern different timing profiles at different tempi. The d' values, on average, are higher for performances at 110 BPM ($M = 2.89, SD = 0.82$) than at 80 BPM ($M = 1.80, SD = 1.25$), paired t-test: $t(26) = 3.96, p < .0001, d = 0.762$. Since d' values are standardized, this indicates that participants are 61% better able to discriminate between different timing profiles at 110 BPM than at 80 BPM.

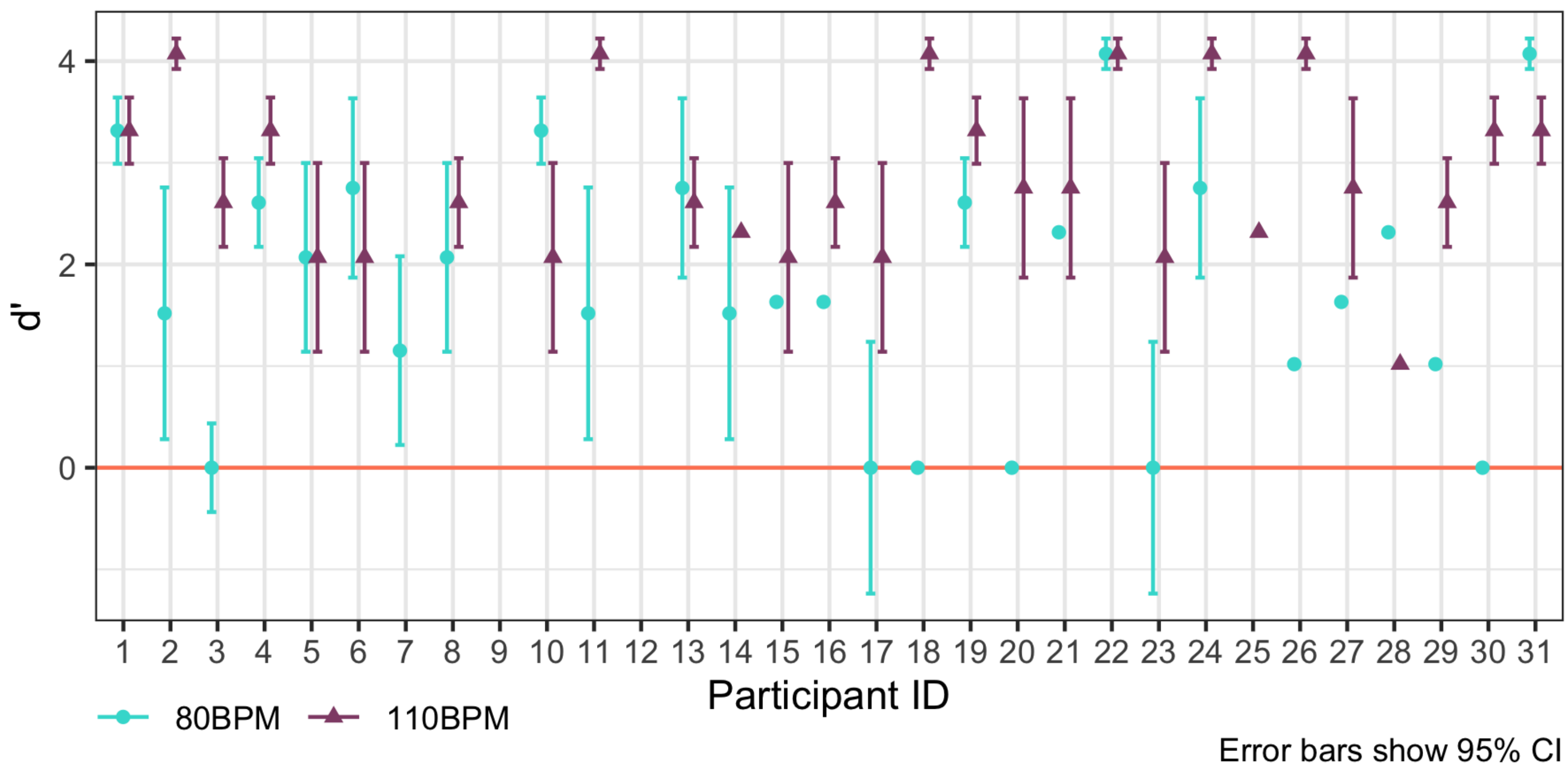


FIGURE 4.10: *d′* scores for each participant, broken down by tempo (Experiment 1). Several participants have no *d′* scores available, or only at one tempo, as their hit rates or false alarm rates limit available analyses (see Table 4.8).

Mixed Effects Model

A mixed effects model allows one to unpack which factors contribute to participant performance in the discrimination task. A logistic mixed model (estimated using the maximum likelihood method, confidence intervals using the Wald method) was fitted to predict whether participants can correctly match **X** with either **A** or **B** given the contrast that participants are presented with (e.g., Pushed Loose vs. Pushed Tight) and including participant ID and stimuli as random effects.[12] The model's total explanatory power is moderate (conditional $R^2 = 0.244$) and the marginal R^2, the part related to the fixed effects alone, has a small effect size = 0.065. The model's intercept, corresponding to Laidback Loose vs. Laidback Tight and when the Tempo = 110 BPM, is at 0.973 (95% CI [0.36, 1.59], $p = .002$). Within this model, there are several significant contrasts between the drum performance timing profiles. First, participants were significantly better, when compared with the intercept, at detecting the difference between Laid Back Loose vs. Pushed Tight profiles ($\beta = 0.829$, 95% CI [0.105, 1.553], $p = .025$). Participant success at discriminating between profiles was also significantly improved when they were presented with Laid Back Tight vs. Pushed Loose ($\beta = 1.120$, 95% CI [0.373, 1.867], $p = .003$). Of note is that the contrast pairs that are significant in the model are the maximally contrasting pairs, that is, both the location (Laid Back vs. Pushed) *and* the scale (Tight vs. Loose) of the pockets differ in these significant pairings. Lastly, there was a significant decrease in participant success when performances were at 80 BPM than when they were at 110 BPM ($\beta = -0.729$, 95% CI [-1.156, -0.303], $p < 0.001$). No other timing profiles or interactions had significant effects on participants' ability to correctly match the timing profile of **X** with one of **A** or **B** (see Table 4.4).

[12] Formula: `Match ~ Profile_Comparison + Tempo + (1 | ID) + (1 | Stimuli)`. Models with additional fixed effects of hours of listening and musical training were also compared, but were found to not be significantly more informative than this parsimonious model, $\chi^2(1) = 0.052, p = .820$ and $\chi^2(1) = 1.777, p = .183$, respectively. Additionally, both undesirably increased the Akaike Information Criterion (AIC—an estimator of prediction error) from 827.43 for the model without this information to 829.38 with listening hours, and 829.61 with musical training.

Predictors	Estimate	95% CI	z	p
(Intercept)	0.973	[0.362, 1.585]	3.118	**0.002**
Laid Back Loose vs. Pushed Loose	0.457	[-0.250, 1.164]	1.268	0.205
Laid Back Loose vs. Pushed Tight	0.829	[0.105, 1.553]	2.243	**0.025**
Laid Back Tight vs. Pushed Loose	1.120	[0.373, 1.867]	2.940	**0.003**
Laid Back Tight vs. Pushed Tight	0.601	[-0.114, 1.315]	1.648	0.099
Pushed Loose vs. Pushed Tight	0.113	[-0.576, 0.803]	0.322	0.747
Tempo: 80 BPM	-0.729	[-1.156, -0.303]	-3.353	**<0.001**
Random Effects				
$\tau_{00\ \mathrm{ID}}$ [random intercept variance]	0.69			
$\tau_{00\ \mathrm{Stimuli}}$	0.09			
ICC [intraclass-correlation coefficient]	0.19			
N_{ID}	31			
N_{Stimuli}	24			
Observations	744			
Marginal R^2 / Conditional R^2	0.065 / 0.244			

TABLE 4.4: Summary of the logistic model for Experiment 1.

Discussion

This first study has demonstrated that listeners are able to discern between drum performances using specific pairings of tight/loose and laid back/pushed, but not between all such pairings. Success was highest when both tight/loose and pushed/laid back were contrastive between stimuli in a pair. Of the 31 participants who listened to sets of stimuli with different performance timing profiles, 24 of them could clearly distinguish between profiles. When the participant data are analyzed according to the two separate tempi heard in the study (80 and 110 BPM), it is once again apparent that these listeners can clearly hear the differences between the different timing profiles that are heard in performances lasting only 8 bars. What is noticeable at this level, however, is how the participant responses to performances at different tempi do appear to differ, with performances at 80 BPM generally being less easily distinguished than those at 110 BPM. One might assume that detecting differences at slower tempos would be easier since the variations in the drum performances would be larger (a delay of 0.01 of a beat at 80 BPM lasts longer in milliseconds than the same delay of 0.01 beats at 110 BPM), though this does not appear to be the case with the participants in this study. Perhaps the performances at 110 BPM are closer to the tempo of a typical pop song[13] and so participants are more experienced with processing performances at this speed, or perhaps the attentional windows required by performances at 80 BPM are just too large for these listeners to hold information about the performances in their working memory when comparing the differences between the various ones that are heard.

Some participants provided optional written comments on the study. Participant 5, who could consistently correctly identify the performance timing profiles with a sensitivity of $d' = 2.069$, wrote "They all start to blend together, but I did my best." Similarly, participant 29 ($d' = 1.773$) commented "That was actually incredibly hard! Fun, but hard." No participants commented to say the task was easy. This suggests that listeners might not be able to identify clearly what the differences or specific qualities of a performance are, yet they are evidently still able to operationalize abstract knowledge about a performance to correctly make comparisons.

Participants also filled out a survey that asked various questions about their music listening, musical activity, and training. As was mentioned in the summary of all of the experiments, the goal here is not to provide a rigorous investigation of participants' musical experience and behaviors (such a survey would have extended the length of the study significantly). Instead, the goal was to get a broad impression of how participants engage with music in their daily lives and to see if any preliminary connections between, for example, musical training and performance in the discrimination task may be found. A regression of d' values by the number of hours that participants reported listening to music per week was not significant ($F(1, 27) = 0.033, p = .857$, adj. $R^2 = -.0358$), meaning that no substantive comments may be made about the relationship of the amount of music that participants listen to and their ability to differentiate between

[13]Measures of average BPM for pop songs vary widely depending on the genres, artists, and timeframe that are being considered, but 120 BPM is a reasonable approximation of the average tempo of contemporary Billboard pop songs.

different performance styles. Additionally, no significant relationship was found between whether participants had musical training and their d' scores ($F(2, 26) = 3.296, p = .053$, adj. $R^2 = .141$).

Participants were also surveyed about which musical genres they listened to in order to get a sense for how familiar they were with popular music drumming and nuanced performance timings. Focusing on one data point of interest, participant 30 specifically mentioned that they listen to the "lo-fi" and "chillhop" sub-genres of hip hop, a hallmark of which is drum beats that are often quite significantly "off grid" and unquantized (Winston & Saywood, 2019). Their sensitivity ($d' = 1.773$) is approximately half that of participants 1 ($d' = 3.483$), 24 ($d' = 3.316$), and 31 ($d' = 3.774$), participants who do not mention hip hop whatsoever (they listen to jazz, blues, pop, non-Western, and reggae music). This is but one data point so may be nothing more than a coincidence, but it encourages future investigations of the connection between the types of music people listen to and their ability to hear performance timing nuances. For example, it might be possible to suggest that an affinity for, and experience with music that uses highly noticeable drum timing profiles does *not* necessarily lead to a higher acuity as a result of enculturation or familiarity.

This experiment did present a few challenges to providing a comprehensive analysis of participant performance. First, due to a concern for not overburdening participants with too many trials, this experiment was limited in the amount of data that could be collected from each participant. Ideally, each stimulus pairing would be presented numerous times to obtain more reliable data about the participants' hit and false alarm rates. Secondly, there were problems with calculating d' and c for two participants (9 & 12) whose performance, as evaluated in the first analysis (Figure 4.9 and Table 4.7), was too poor and so could not be evaluated, even after adjustment. These two participants had very high false alarm rates and comparatively lower hit rates, perhaps suggesting that they misunderstood the experimental task and, in reality, had relatively good sensitivity, but selected the *different* performance rather than the *matching* performance. If this was the case, participant 9 would have a d' value of 3.005 [2.634, 3.376] and participant 12 would have a d' of 1.405 [0.0561, 2.754].

The mixed effects model can help, in some ways, to clarify whether participants are sensitive to differences in performance timing profiles even when d' cannot provide the full picture. The model, summarized in Table 4.4, shows that the specific types of timing profile heard *do* have an impact on participants' ability to correctly match **X** with either **A** or **B**, suggesting that some timing profiles might be more easily heard and/or utilized by listeners. In this first experiment, participants appear to be more successful at matching the performance timing profiles when comparing Laid Back with Pushed profiles than when comparing two types—Tight and Loose—of Laid Back or two types of Pushed. This is not the full picture, however, as not all comparisons of Laid Back with Pushed were significant in the mixed effects model. What the model highlights is that the vast majority of the variance in successful matches lies within participants and within the individual stimuli (this is captured by the ICC, which shows that 81% of variance (1 – ICC) is within the grouping structure).

Experiment 2: Pushed vs. Laid Back

The aim of Experiment 2 is to investigate whether listeners are able to distinguish between drum performances that have pushed or laid back timing profiles (i.e., the snare and bass drum onsets tend to land slightly before (pushed) or after (laid back) the metronomic hi-hat onset). This study refines Experiment 1 by homing in on these two particular timing profile contrasts to isolate one facet of performance timing that a drummer may manipulate to engender a particular feel.

Method

The design of Experiment 2 is identical to that of Experiment 1 (summarized earlier in Figure 4.2) with only the stimuli sets changed to include only pushed vs. laid back profile comparisons.

Participants

36 participants (17 female), aged between 26 and 66 ($M = 39.6$, $SD = 10.8$) were recruited through Amazon MTurk in early 2021. One participant was screened out for self-reported hearing difficulties and a further two were screened out for failing the attention checks. Participants could not participate if they had taken part in Experiment 1.

Procedure & Stimuli

As in Experiment 1, 24 trials (12 × 80 BPM, 12 × 110 BPM) were presented online to participants, each consisting of three performances **A**, **B**, and the target performance **X**. Here, **A** and **B** are only ever comparisons of pushed and laid back profiles, with the amount of variance in the performance (described here by the contrast between loose and tight) held steady for the comparisons. So, participants are asked to discriminate between "pushed loose" and "laid back loose" or "pushed tight" and "laid back tight," but never between one profile that is "tight" and another that is "loose." The same attention checks and screening criteria as Experiment 1 were utilized.

Results

Of the 773 pushed vs. laid back trials included in the analysis (19 individual trials were screened out for having durations over 100 seconds), 437 were correctly matched by participants (56.5%), which is significantly different from chance performance ($p < .001$). Calculating d' across all participant data, this is also significantly different from chance performance: $d' = 0.881$, 95% CI [0.160, 1.602]. Per participant sensitivity and bias is reported in Figure 4.11 and Table 4.9 (at end of chapter). The first thing to note is that there is a large amount of missing information about individual participant d' scores and confidence intervals. Out of the 33 participants who passed the attention checks and data screening, 17 have incalculable d' and/or confidence intervals. This is because several participants' false alarm rates were higher than their hit rate (possibly suggesting misunderstanding of the task) or, for those with a calculable d' but no confidence interval, one of the hit or false alarm rates was exactly 0.5. For the 16 participants that can be evaluated using signal detection theory methods, there appear to be three groups of participants: Nine participants have confidence intervals that do not overlap with 0, meaning that only these nine may confidently be said to successfully discern between pushed and laid back performances. A further four participants' (1, 9, 17, and 29) sensitivity and confidence intervals overlap with 0 by a small amount, while the remaining three (2, 11, and 19) have sizable overlaps. Analysis across all participants shows no significant difference in d' scores for when the laid back vs. pushed timing profiles are loose ($M = 1.53, SD = 0.66$) or tight ($M = 1.11, SD = 0.63$), paired t-test: $t(20) = 1.75, p = .095$. Additionally, there was no significant difference in participant d' scores for stimuli heard at 80 BPM ($M = 1.09, SD = 0.72$) than at 110 BPM ($M = 1.40, SD = 0.63$), paired t-test: $t(22) = -1.54, p = .137$.

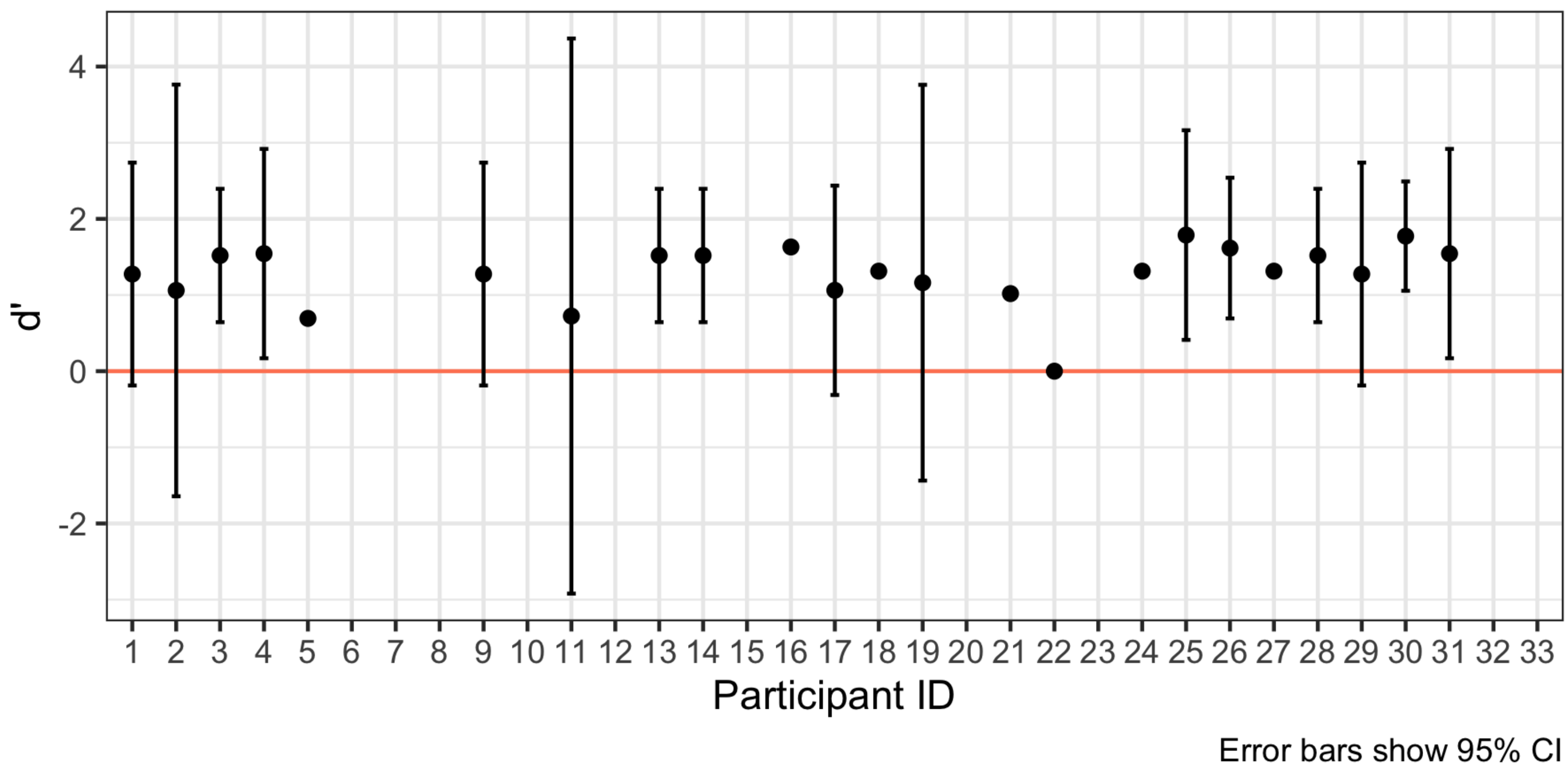


FIGURE 4.11: *d′* scores for each participant (Experiment 2). Several participants are missing information on this graph as their hit rate, false alarm rate, or both, result in an incalculable *d′* (see Table 4.9).

Mixed Effects Model

Signal detection theory is limited in what it can reveal about these participants' abilities to discriminate between pushed and laid back performances and so a mixed effects model is especially valuable. A logistic mixed model was fitted as in Experiment 1, with the contrast pair and tempo as fixed effects, and participant ID and stimuli as random effects.[14] Since Experiment 2 only looks at Laid Back vs. Pushed profiles, the contrast pairs are Laid Back *Loose* vs. Pushed *Loose* and Laid Back *Tight* vs. Pushed *Tight*. The model's total explanatory power is large (conditional $R^2 = 0.418$), though the marginal R^2 (fixed effects alone) $= 0.009$, as a large amount of the variance in the model is associated with the random intercepts (see $\tau_{00\ \text{Stimuli}}$—the random intercept variance—in Table 4.5). The near-zero size of the fixed effects in the model, as well as the non-significant findings for both whether the contrast is performed Tight or Loose and whether the contrast is heard at 80 or 110 BPM (see Table 4.5), suggests that the ability to successfully hear these subtle timing nuances is highly dependent on the individual participant and the specific musical performance that is heard.

Discussion

Participants in Experiment 2, who listened to stimuli sets that contrasted performances with laid back and pushed timing profiles, appear to have a difficult time successfully recognizing the timing profile of performance **X** and selecting a performance (**A** or **B**) that had a similar profile. While they did get nearly 60% of trials correct—a success rate that is significantly above chance—individual participant performance varied widely. This can be seen both in the range of confidence interval magnitudes (compare, for example, the size of the error bars for participant

[14]Formula: `Match ~ Profile_Comparison + Tempo + (1 | ID) + (1 | Stimuli)`. Models with additional fixed effects of hours of listening and musical training were also compared, but were found to not be significantly more informative than this parsimonious model, $\chi^2(1) = 0.571, p = .450$ and $\chi^2(1) = 0.505, p = .477$, respectively. Additionally, both undesirably increased the Akaike Information Criterion: from 873.67 for the model without this information to 875.09 with listening hours, and 876.59 with musical training.

Predictors	Estimate	95% CI	*z*	*p*
(Intercept)	0.666	[-0.408, 1.740]	1.216	0.224
Laid Back Tight vs. Pushed Tight	-0.285	[-1.518, 0.949]	-0.452	0.651
Tempo: 80 BPM	-0.351	[-1.585, 0.883]	-0.558	0.577
Random Effects				
$\tau_{00\ \mathrm{ID}}$	0.14			
$\tau_{00\ \mathrm{Stimuli}}$	2.18			
ICC	0.41			
N_{ID}	33			
N_{Stimuli}	24			
Observations	792			
Marginal R^2 / Conditional R^2	0.009 / 0.418			

Table 4.5: Summary of the logistic model for Experiment 2.

13 with participant 11 in Figure 4.11) and also the intraclass-correlation coefficient for the mixed effects model, which estimates that 59% of the variance in successful matches is attributable to the individuals' performances and the effect of the different stimuli. Confirming the challenge this experiment's task presented to participants, there were a few comments submitted by participants such as "It was really difficult" and "I honestly co[u]ld not hear the difference in most of the clips, this was hard."

The mixed effects logistic model does not have much explanatory power, however it does inform us that the vast majority of variance (59%) in participant performance is attributable to individual participants and the specific stimuli. It also provides additional information beyond what the d' analysis could show, suggesting that participant success at discriminating between laid back and pushed timing profiles does not change significantly whether those laid back and pushed profiles are loose or tight. Lastly, it supports the finding from the signal detection analysis that there is no significant effect of tempo on participant success.

In the course of constructing the logistic mixed model, it was made apparent that musical training and number of hours listened do not have a significant relation to participants' ability to match the timing profiles. This is confirmed by linear regressions of d' values with musical training ($F(2, 20) = 0.632, p = .521$, adj. $R^2 = -.0035$) and the self-reported number of listening hours ($F(1, 21) = 1.842, p = .189$, adj. $R^2 = .037$).

Overall, the data from Experiment 2 give some support to the hypothesis that listeners can discriminate between drum performances that are laid back and those that are pushed; however, it is important to consider the finer details of this conclusion. Participants' overall performance is better than chance and several participants' d' scores are significantly different from 0 (9 out of 16 with fully analyzable data, or up to 13 out of 16 if we are generous to those participants whose CIs overlap with 0 marginally), but the prevalence of wide confidence intervals and high false alarm rates amongst the participants does leave the question somewhat still open as to whether participants can successfully discriminate between laid back and pushed timing profiles.

Experiment 3: Tight vs. Loose

The aim of Experiment 3 is to investigate whether listeners are able to distinguish between drum performances that have tight or loose timing profiles (i.e., there is less or more variance, respectively, to the location in time of snare and bass drum onsets in the basic drum groove). This study refines Experiment 1, and complements Experiment 2, by concentrating on the remaining two timing profile contrasts so that the perceptibility of the feel-engendering performance timing may be evaluated.

Method

The design of Experiment 3 is identical to that of Experiments 1 and 2 (summarized earlier in Figure 4.2) with the stimuli sets changed to use only tight vs. loose profile comparisons.

Participants

36 participants (14 female), aged between 20 and 65 ($M = 38.4$, $SD = 10.7$) were recruited through Amazon MTurk in early 2021. Two participants were screened out for self-reported hearing difficulties and no further participants were screened out for failing the attention checks. Participants could not have taken part in either of the previous studies.

Procedure & Stimuli

As in Experiments 1 and 2, 24 trials (12 × 80 BPM, 12 × 110 BPM) were presented online to participants, each consisting of three performances **A**, **B**, and the target performance **X**. Here, **A** and **B** are only ever comparisons of tight and loose profiles. The degree to which the center of the timing distribution comes before or after the metronomic hi-hat in the performance (described here by the contrast between pushed or laid back) is held steady for the comparisons. Therefore, participants are asked to discriminate between "pushed loose" and "pushed tight" or "laid back loose" and "laid back tight," but never between one profile that is "pushed" and another that is "laid back." The same attention checks and screening criteria as in Experiments 1 and 2 were utilized.

Results

Of the 797 tight vs. loose trials included in the analysis (19 individual trials were screened out for having durations over 100 seconds), 464 were correctly matched by participants (58.2%), which is significantly above chance performance ($p < .0001$). However, when all of the participant data is analyzed using signal detection theory methods, participant sensitivity is *not* significantly different from chance: $d' = 1.010$, 95% CI$[-0.952, 2.971]$. Per participant sensitivity and bias is reported in Figure 4.12 and Table 4.10 (at the end of the chapter). Just as in the previous experiments, there is a large amount of missing information about individual participant d'

scores and confidence intervals. 16 out of the 34 participants who passed the attention checks and data screening have incalculable d' and/or confidence intervals. For the 18 participants that can be evaluated using signal detection theory methods, 12 have confidence intervals that do not overlap with 0 meaning that only these 12 may confidently be said to successfully discern between tight and loose performances. A further two participants' (2 and 24) sensitivity and confidence intervals overlap with 0 by just a small amount. Analysis across all participants shows no significant difference in d' scores for when the tight vs. loose timing profiles are pushed ($M = 1.28, SD = 0.88$) or laid back ($M = 1.63, SD = 0.60$), paired t-test: $t(19) = 1.47, p = .158$. Additionally, there was no significant difference in participant d' scores for stimuli heard at 80 BPM ($M = 1.47, SD = 0.73$) than at 110 BPM ($M = 1.24, SD = 0.78$), paired t-test: $t(22) = -1.15, p = .264$.

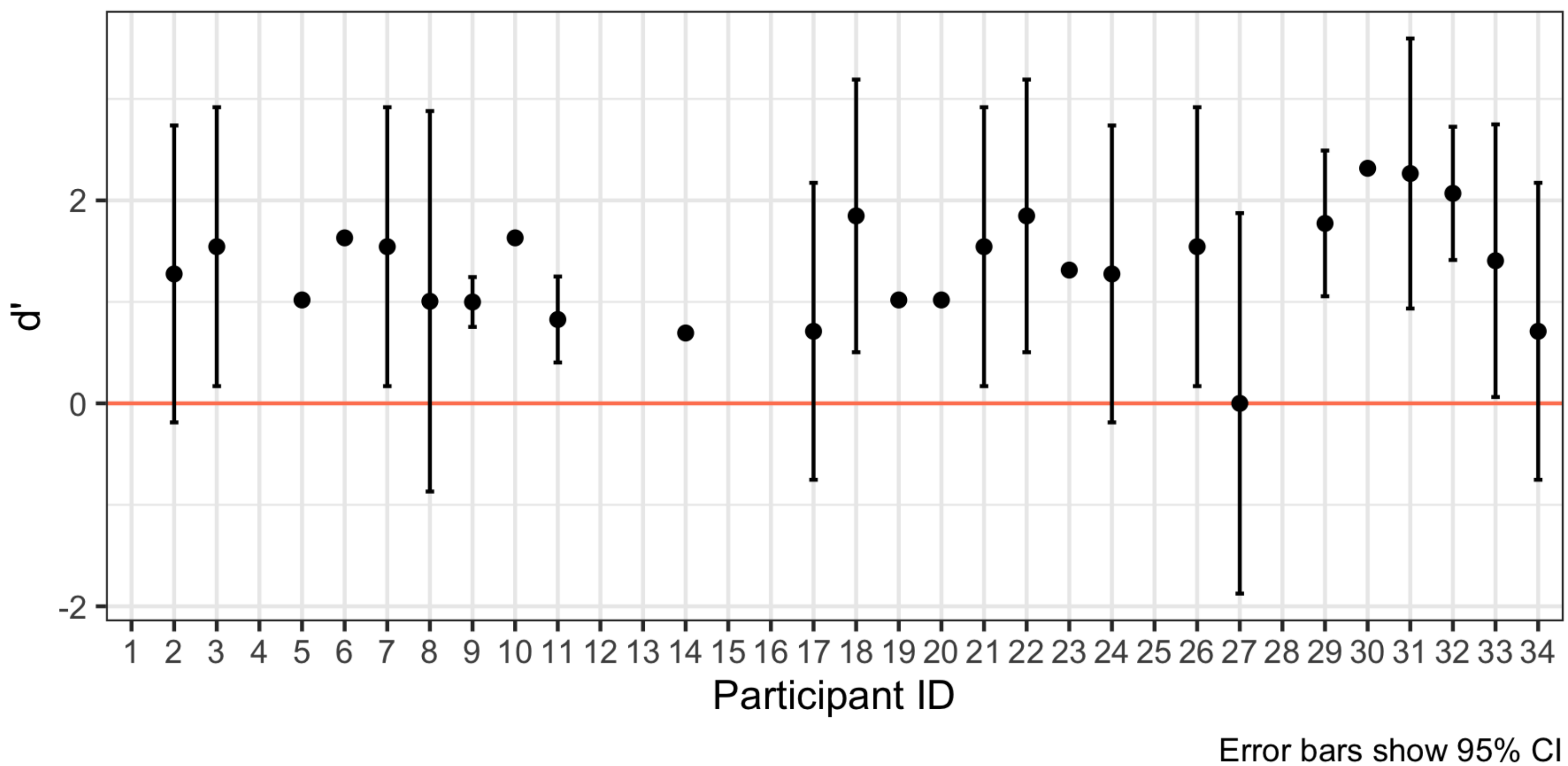


Figure 4.12: *d′* scores for each participant (Experiment 3). Several participants are missing information on this graph as their hit rate, false alarm rate, or both, result in an incalculable *d′* (see Table 4.10).

Predictors	Estimate	95% CI	*z*	*p*
(Intercept)	0.441	[-0.268, 1.151]	1.218	0.223
Pushed Loose vs. Pushed Tight	-0.195	[-1.003, 0.614]	-0.471	0.637
Tempo: 80 BPM	0.069	[-0.739, 0.878]	0.168	0.867
Random Effects				
$\tau_{00\ \mathrm{ID}}$	0.11			
$\tau_{00\ \mathrm{Stimuli}}$	0.87			
ICC	0.23			
N_{ID}	34			
$\mathrm{N}_{\mathrm{Stimuli}}$	24			
Observations	797			
Marginal R^2 / Conditional R^2	0.002 / 0.231			

Table 4.6: Summary of the logistic model for Experiment 3.

Mixed Effects Model

Lastly, as in the analysis of Experiments 1 and 2, a logistic mixed model was fitted to predict whether participants can correctly differentiate between the tight vs. loose timing profiles given the contrast pair and tempo as fixed effects, and including the unique participants and stimuli as random effects.[15] The model's total explanatory power is moderate (conditional $R^2 = 0.231$), though the marginal R^2 (fixed effects alone) = 0.002, which suggests that nearly all of the variance in the model is associated with the random intercepts. Within this model there are no significant effects, neither for when the contrasted pairs of Tight vs. Loose performance profiles were Pushed, nor Laid Back, nor for when performances were at 80 or 110 BPM, on the ability to match performance **X** with the correct timing profile (summarized in Table 4.6).

[15]Formula: `Match ~ Profile_Comparison + Tempo + (1 | ID) + (1 | Stimuli)`. Models with additional fixed effects of hours of listening and musical training were also compared, but were found to not be significantly more informative than this parsimonious model, $\chi^2(1) = 0.003, p = .959$ and $\chi^2(1) = 2.556, p = .110$, respectively. Additionally, both undesirably increased the Akaike Information Criterion: from 1001.1 for the model without this information to 1003.0 with listening hours, and 1002.5 with musical training.

Discussion

Experiment 3 looked at whether participants could successfully recognize and act upon perceived differences in drum performances that had tight and loose timing profiles. The results of this experiment are mixed and do not uniformly support the hypothesis that participants can successfully do this. While, overall, they achieved a success rate of 58.2%, which is significantly above chance, the sensitivity as calculated using signal detection theory is not significantly different from chance $d' = 1.010$, 95% $\mathrm{CI}[-0.952, 2.971]$. Evidently, they did not find this task easy. As can be seen in Figure 4.12 and Table 4.10 (at the end of this chapter), just 12 of the participants' sensitivity was significantly above chance performance. It is important to note that this is not exactly the same as saying that only 12 out of 34 participants could perceive the differences in tight vs. loose performances as some participants had high sensitivity, but it was not possible to calculate a 95% confidence interval owing to the hit rate being 0.5 (for example, participant 30's $d' = 2.316$ and they got 86.4% of trials correct, but their confidence intervals are incalculable).

The mixed effects model, as in Experiment 2, does not have much explanatory power with the vast majority (77%) of the variance in the ability to successfully differentiate between the performances' timing profiles attributable to the individual participants and the specific stimuli they heard. The model did provide information that the difference between tight and loose profiles is not significantly more or less perceptible when the contrasting pair is pushed or laid back, as well as confirming that there is no effect of tempo. Both the mixed effects model and linear regressions of d' scores show that there is no meaningful connection between participants' success and their musical training ($F(2, 23) = 0.176, p = .839$, adj. $R^2 = -.071$) and the self-reported number of listening hours ($F(1, 24) = 0.184, p = .672$, adj. $R^2 = -.034$).

To summarize Experiment 3, the findings suggest that listeners are *not* able to successfully discriminate between these particular tight and loose performance profiles, and it is clear that they find this task challenging. Individual participants vary widely in their abilities, as seen by the range of d' and confidence interval values.

General Discussion

This exploratory set of studies is a first step towards investigating whether or not listeners are able to perceive some types of subtle nuances and differences between drummers' timing profiles. In both the omnibus comparison of all performance profile types and the controlled comparisons of pushed vs. laid back, success rates and individual participant d' scores are significantly above chance performance for a majority of participants who participated in the studies. However, in the case of tight vs. loose profiles, profiles that have less or more amounts of variation to the microtiming nuances of the onsets, it is not clear beyond reasonable doubt that listeners are able to consistently tell the difference between a tight performance and a loose one. I would argue that this does not prohibit analyses of music that consider the tightness or looseness of a performer, but Experiment 3's findings show that many listeners may not be sensitive—or their sensitivity may vary widely—to these quite subtle differences. As such, any conclusions about the impact of tightness or looseness on the listener experience would need to be accompanied by a caveat about how individual listeners vary in their awareness of this subtle performance detail.

For all three experiments, there was no connection between participant success at the discrimination task and their musical training, listening habits, or genres that they tend to listen to, as assessed by their self-reported answers to a simple survey. We may interpret this as giving confidence to the idea that listeners, no matter their background, experience, or training, are able to hear subtle differences in performances (again, with the caveat that tight vs. loose is far more challenging). One might expect that participants in the study who are more musically sophisticated would perform better (i.e. have finer temporal acuity) than those who are less musically sophisticated (Kraus & Chandrasekaran, 2010; Matthews et al., 2016; Tzounopoulos & Kraus, 2009), though a large meta-analysis of the cognitive benefits of musical training shows no overall benefits (Sala & Gobet, 2020).[16] In the present study, recognizing different time feels is not an elite or specialized skill. That being said, recruitment for these three studies was limited to people residing in the USA; it would be very interesting to see whether people whose rhythmic prototypes may vary as a result of embeddedness in different global music-cultural environments (Jakubowski et al., 2018; Polak et al., 2018), or whether hyper-specialized, US-based listeners such as audio engineers and producers, vary from the present studies' sample populations in their ability to successfully hear differences in performance timing.[17]

This collection of experiments was not without its challenges. First, there is the challenge, throughout, of missing or incalculable d' scores and confidence intervals. The absence of a score or confidence interval, as noted in Experiment 3's discussion (page 157), is not necessarily an indicator that the participant was not sensitive to differences between performance timing profiles, but it limits the ability to present a comprehensive analysis. In an ideal world, these studies would involve far more trials than the 24 that were presented to participants, which would provide greater reliability and would also afford finer-grained hit and false alarm rates, and most likely would fill in some of the gaps in d' and 95% confidence intervals. 24 trials was chosen as it was the fewest number of trials that would completely cover all combinations of performance types under consideration at two different tempi (Experiment 1). ABX paradigms tend to use hundreds of short trials (e.g., one second-long recordings of single phonemes), but that is not feasible here both because there are limits to how many trials may be presented to an unsupervised participant and expect meaningful, quality data, and also because of the nature of the musical phenomena

[16] The meta-analysis presented in Sala and Gobet (2020) pertains to the cognitive benefits of music training with *children*, which is separate from the population sampled here. That being said, Kraus and Chandrasekaran (2010) and Tzounopoulos and Kraus (2009) also look at younger populations to argue that musical training, even just fifteen months, can result in structural changes in, for example, the primary auditory and primary motor areas. Matthews et al. (2016) contrasts highly specialized musicians (10+ years on their primary instrument) with "non-musicians" (<3 years of training and at least a gap of 5 years since any training). I raise these details to highlight how there is an underdeveloped area of research on the impact of casual music making or long-past musical training on rhythm perception.

[17] Also see Danielsen et al. (in review), which finds sizable differences in where professional musicians from different genres hear the perceptual center of musical sounds, suggesting that training and experience impact foundational perceptual processes.

being presented. Audio stimuli could, perhaps, be shortened from eight bars to one or two bars, but then the variance that is captured by the qualitative descriptors "tight" and "loose" would all but vanish since only a few onsets are heard.

The possibility of using shorter performances as stimuli also connects to another consideration in this experiment: is the task too challenging for participants? Holding two or three drum performances in mind and then making decisions based off this information most definitely tests the limits of working memory. If the performances were shorter, this might ease some of the cognitive load of the task and thereby enable participants to better compare performances. Phonological working memory is estimated to have a span of only 1.5 to 2 seconds (Baddeley, 1986; Baddeley & Hitch, 1974) and this limit may also apply beyond verbal material and be a limit to the processing of rhythms (summarized in Repp and Doggett, 2007, p. 368; original sources include Grube, 1996, 1998; Larsen and Baddeley, 2003; Saito, 2001; Saito and Ishio, 1998; Wilson, 2001). That means that, unless all three performances can somehow occur in this frame, some kinds of memory storage and retrieval processes must be utilized in making the judgements that this protocol requires. Of note is that this storage and retrieval process might lead to different findings if this set of experiments were to be repeated with a harmonic instrument rather than the (largely) inharmonic sounds of a drum kit as there are differences in how these types of sounds are encoded in memory (McPherson & McDermott, 2020). Additionally, I acknowledge that some may contest that the ABX paradigm is not an ecologically valid way of evaluating listeners' ability to hear these subtle performance differences as the protocol requires listeners to hold several different performances in their working memory, a scenario that is not encountered in typical day-to-day listening.

Going forward, this study has operationalized the qualitative labels pushed, laid back, tight, and loose, which is language drawn from the community of performers in numerous American and European popular genres. It would be illuminating to assess what descriptive words lay listeners might use to describe these profiles in free response trials, potentially illuminating other facets of the concept of feel that have not yet been considered (in much the same way as my phenomenological study into the qualitative dimensions of groove—Hosken, 2020).

Overall, the results of the study add to the analyses presented in other chapters of this dissertation, suggesting at least partial support for the claims that these subtle changes in timing are perceivable and not below a perceptual threshold, and therefore they may have meaningful aesthetic and/or embodied consequences for listeners. The findings of this set of experiments establish the "meaningfulness of resolution" that Pressing questioned, and provide context and credence for the analyses of "feel" in musical performances.

Summary Tables of Participant Scores for All Experiments

ID	Hit Rate	FA Rate	% Correct	d'	SD	95% CI	c
1	0.917	0.083	92.3	3.483	0.082	[3.323, 3.643]	0
2	0.833	0.167	84.6	2.609	0.157	[2.3, 2.917]	0
3	0.5	0.167	69.2	1.631	NA	NA	0.484
4	0.917	0.167	90.9	3.037	0.125	[2.792, 3.283]	-0.208
5	0.833	0.333	76.9	2.069	0.335	[1.413, 2.726]	-0.268
6	0.75	0.167	80	2.313	0.216	[1.89, 2.737]	0.146
7	0.5	0.5	53.8	0	NA	NA	0
8	0.833	0.25	80.8	2.313	0.216	[1.89, 2.737]	-0.146
9	0.25	0.95†	20.8	NA	NA	NA	NA
10	0.833	0.167	84	2.609	0.157	[2.3, 2.917]	0
11	0.833	0.167	84.6	2.609	0.157	[2.3, 2.917]	0
12	0.167	0.417	45.8	NA	NA	NA	NA
13	0.75	0.083	84	2.734	0.194	[2.353, 3.114]	0.354
14	0.583	0.167	73.1	1.847	0.685	[0.504, 3.191]	0.378
15	0.75	0.333	73.1	1.773	0.366	[1.055, 2.491]	-0.122
16	0.667	0.167	76.9	2.069	0.335	[1.413, 2.726]	0.268
17	0.5	0.25	65.4	1.313	NA	NA	0.337
18	0.75	0.25	76.9	2.02	0.262	[1.506, 2.534]	0
19	0.833	0.083	88.5	3.037	0.125	[2.792, 3.283]	0.208
20	0.75	0.417	72	1.544	0.701	[0.169, 2.918]	-0.232
21	0.583	0.05†	90	2.528	0.677	[1.2, 3.856]	0.717
22	0.95†	0.05†	100	4.072	0.05†4	[3.966, 4.178]	0
23	0.667	0.417	65.4	1.275	0.746	[-0.188, 2.738]	-0.11
24	0.833	0.05†	92	3.316	0.118	[3.086, 3.547]	0.339
25	0.583	0.5	58.3	0.693	NA	NA	-0.105
26	0.75	0.167	80.8	2.313	0.216	[1.89, 2.737]	0.146
27	0.583	0.083	76.9	2.265	0.679	[0.934, 3.595]	0.586
28	0.583	0.25	70.8	1.544	0.701	[0.169, 2.918]	0.232
29	0.75	0.333	79.2	1.773	0.366	[1.055, 2.491]	-0.122
30	0.667	0.25	73.1	1.773	0.366	[1.055, 2.491]	0.122
31	0.917	0.05†	96.2	3.774	0.069	[3.638, 3.91]	0.131

TABLE 4.7: d' Statistics per Participant (Experiment 1). NB: Hit rate and False Alarm rate values that have been adjusted either up or down 0.05 are indicated with †.

Table 4.8: d' Statistics per Participant, broken down by tempo (Experiment 1). NB: Hit rate and False Alarm rate values that have been adjusted either up or down 0.05 are indicated with †. Even with this adjustment, d' for some participants still cannot be calculated.

ID	Tempo	Hit Rate	FA Rate	% Correct	d'	SD	95% CI	c
1	80BPM	0.95†	0.167	91.7	3.316	0.166	[2.99, 3.642]	-0.339
	110BPM	0.833	0.05†	91.7	3.316	0.166	[2.99, 3.642]	0.339
2	80BPM	0.667	0.333	66.7	1.519	0.632	[0.28, 2.757]	0
	110BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
3	80BPM	0.167	0.167	50	0	0.222	[-0.436, 0.436]	0.967
	110BPM	0.833	0.167	83.3	2.609	0.222	[2.173, 3.044]	0
4	80BPM	0.833	0.167	87.5	2.609	0.222	[2.173, 3.044]	0
	110BPM	0.95†	0.167	91.7	3.316	0.166	[2.99, 3.642]	-0.339
5	80BPM	0.833	0.333	75	2.069	0.474	[1.141, 2.998]	-0.268
	110BPM	0.833	0.333	75	2.069	0.474	[1.141, 2.998]	-0.268
6	80BPM	0.667	0.05†	81.8	2.752	0.45	[1.87, 3.634]	0.607
	110BPM	0.833	0.333	75	2.069	0.474	[1.141, 2.998]	-0.268
7	80BPM	0.333	0.167	58.3	1.152	0.474	[0.224, 2.08]	0.699
	110BPM	0.667	0.833	41.7	NA	NA	NA	NA
8	80BPM	0.833	0.333	75	2.069	0.474	[1.141, 2.998]	-0.268
	110BPM	0.833	0.167	83.3	2.609	0.222	[2.173, 3.044]	0
9	80BPM	0.5	0.95†	27.3	NA	NA	NA	NA
	110BPM	0.05†	0.95†	0	NA	NA	NA	NA
10	80BPM	0.95†	0.167	90.9	3.316	0.166	[2.99, 3.642]	-0.339
	110BPM	0.667	0.167	75	2.069	0.474	[1.141, 2.998]	0.268
11	80BPM	0.667	0.333	66.7	1.519	0.632	[0.28, 2.757]	0
	110BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
12	80BPM	0.167	0.5	36.4	NA	NA	NA	NA
	110BPM	0.167	0.333	45.5	NA	NA	NA	NA
13	80BPM	0.667	0.05†	81.8	2.752	0.45	[1.87, 3.634]	0.607
	110BPM	0.833	0.167	83.3	2.609	0.222	[2.173, 3.044]	0
14	80BPM	0.667	0.333	66.7	1.519	0.632	[0.28, 2.757]	0
	110BPM	0.5	0.05†	75	2.316	NA	NA	0.822
15	80BPM	0.833	0.5	66.7	1.631	NA	NA	-0.484
	110BPM	0.667	0.167	75	2.069	0.474	[1.141, 2.998]	0.268
16	80BPM	0.5	0.167	66.7	1.631	NA	NA	0.484
	110BPM	0.833	0.167	83.3	2.609	0.222	[2.173, 3.044]	0
17	80BPM	0.333	0.333	50	0	0.632	[-1.238, 1.238]	0.431

Table 4.8: d' Statistics per Participant, broken down by tempo (cont.).

ID	Tempo	Hit Rate	FA Rate	% Correct	d'	SD	95% CI	c
	110BPM	0.667	0.167	75	2.069	0.474	[1.141, 2.998]	0.268
18	80BPM	0.5	0.5	50	0	NA	NA	0
	110BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
19	80BPM	0.833	0.167	83.3	2.609	0.222	[2.173, 3.044]	0
	110BPM	0.833	0.05†	91.7	3.316	0.166	[2.99, 3.642]	0.339
20	80BPM	0.5	0.5	54.5	0	NA	NA	0
	110BPM	0.95†	0.333	83.3	2.752	0.45	[1.87, 3.634]	-0.607
21	80BPM	0.5	0.05†	88.9	2.316	NA	NA	0.822
	110BPM	0.667	0.05†	88.9	2.752	0.45	[1.87, 3.634]	0.607
22	80BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
	110BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
23	80BPM	0.667	0.667	50	0	0.632	[-1.238, 1.238]	-0.431
	110BPM	0.667	0.167	75	2.069	0.474	[1.141, 2.998]	0.268
24	80BPM	0.667	0.05†	83.3	2.752	0.45	[1.87, 3.634]	0.607
	110BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
25	80BPM	0.667	0.95†	33.3	NA	NA	NA	NA
	110BPM	0.5	0.05†	80	2.316	NA	NA	0.822
26	80BPM	0.5	0.333	58.3	1.019	NA	NA	0.215
	110BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
27	80BPM	0.5	0.167	66.7	1.631	NA	NA	0.484
	110BPM	0.667	0.05†	83.3	2.752	0.45	[1.87, 3.634]	0.607
28	80BPM	0.5	0.05†	80	2.316	NA	NA	0.822
	110BPM	0.667	0.5	58.3	1.019	NA	NA	-0.215
29	80BPM	0.667	0.5	63.6	1.019	NA	NA	-0.215
	110BPM	0.833	0.167	90.9	2.609	0.222	[2.173, 3.044]	0
30	80BPM	0.5	0.5	50	0	NA	NA	0
	110BPM	0.833	0.05†	91.7	3.316	0.166	[2.99, 3.642]	0.339
31	80BPM	0.95†	0.05†	100	4.072	0.076	[3.922, 4.222]	0
	110BPM	0.833	0.05†	91.7	3.316	0.166	[2.99, 3.642]	0.339

ID	Hit Rate	FA Rate	% Correct	d'	SD	95% CI	c
1	0.667	0.417	65.4	1.275	0.746	[-0.188, 2.738]	-0.11
2	0.545	0.364	62.5	1.06	1.379	[-1.643, 3.763]	0.117
3	0.667	0.333	69.2	1.519	0.447	[0.643, 2.395]	0
4	0.583	0.25	69.2	1.544	0.701	[0.169, 2.918]	0.232
5	0.5	0.417	57.7	0.693	NA	NA	0.105
6	0.333	0.417	50	NA	NA	NA	NA
7	0.545	0.75	44	NA	NA	NA	NA
8	0.333	0.417	50	NA	NA	NA	NA
9	0.583	0.333	65.4	1.275	0.746	[-0.188, 2.738]	0.11
10	0.167	0.583	34.6	NA	NA	NA	NA
11	0.545	0.455	58.3	0.724	1.859	[-2.921, 4.368]	0
12	0.444	0.556	50	NA	NA	NA	NA
13	0.667	0.333	69.2	1.519	0.447	[0.643, 2.395]	0
14	0.667	0.333	69.2	1.519	0.447	[0.643, 2.395]	0
15	0.5	0.583	50	NA	NA	NA	NA
16	0.5	0.167	69.2	1.631	NA	NA	0.484
17	0.417	0.25	61.5	1.062	0.701	[-0.313, 2.436]	0.442
18	0.5	0.25	65.4	1.313	NA	NA	0.337
19	0.545	0.333	64	1.162	1.326	[-1.437, 3.76]	0.158
20	0.333	0.5	46.2	NA	NA	NA	NA
21	0.5	0.333	61.5	1.019	NA	NA	0.215
22	0.5	0.5	53.8	0	NA	NA	0
23	0.417	0.583	46.2	NA	NA	NA	NA
24	0.75	0.5	65.4	1.313	NA	NA	-0.337
25	0.818	0.417	72	1.787	0.702	[0.411, 3.163]	-0.349
26	0.636	0.273	69.6	1.616	0.471	[0.692, 2.54]	0.128
27	0.5	0.25	65.4	1.313	NA	NA	0.337
28	0.667	0.333	69.2	1.519	0.447	[0.643, 2.395]	0
29	0.667	0.417	65.4	1.275	0.746	[-0.188, 2.738]	-0.11
30	0.667	0.25	73.1	1.773	0.366	[1.055, 2.491]	0.122
31	0.583	0.25	68	1.544	0.701	[0.169, 2.918]	0.232
32	0.333	0.455	48	NA	NA	NA	NA
33	0.364	0.6	43.5	NA	NA	NA	NA

TABLE 4.9: d' Statistics per Participant (Experiment 2).

ID	Hit Rate	FA Rate	% Correct	d'	SD	95% CI	c
1	0.333	0.5	41.7	NA	NA	NA	NA
2	0.667	0.417	65.4	1.275	0.746	[-0.188, 2.738]	-0.11
3	0.583	0.25	66.7	1.544	0.701	[0.169, 2.918]	0.232
4	0.333	0.417	50	NA	NA	NA	NA
5	0.5	0.333	61.5	1.019	NA	NA	0.215
6	0.5	0.167	66.7	1.631	NA	NA	0.484
7	0.583	0.25	69.2	1.544	0.701	[0.169, 2.918]	0.232
8	0.583	0.417	61.5	1.006	0.956	[-0.869, 2.88]	0
9	0.167	0.083	57.7	0.999	0.125	[0.753, 1.244]	1.175
10	0.5	0.167	69.2	1.631	NA	NA	0.484
11	0.25	0.167	56	0.826	0.216	[0.403, 1.25]	0.821
12	0.25	0.5	42.3	NA	NA	NA	NA
13	0.333	0.583	41.7	NA	NA	NA	NA
14	0.5	0.417	57.7	0.693	NA	NA	0.105
15	0.5	0.583	50	NA	NA	NA	NA
16	0.333	0.417	50	NA	NA	NA	NA
17	0.417	0.333	57.7	0.71	0.746	[-0.753, 2.173]	0.321
18	0.583	0.167	73.1	1.847	0.685	[0.504, 3.191]	0.378
19	0.5	0.333	61.5	1.019	NA	NA	0.215
20	0.667	0.5	61.5	1.019	NA	NA	-0.215
21	0.583	0.25	69.2	1.544	0.701	[0.169, 2.918]	0.232
22	0.833	0.417	73.1	1.847	0.685	[0.504, 3.191]	-0.378
23	0.5	0.25	65.4	1.313	NA	NA	0.337
24	0.583	0.333	65.4	1.275	0.746	[-0.188, 2.738]	0.11
25	0.417	0.5	52	NA	NA	NA	NA
26	0.583	0.25	69.2	1.544	0.701	[0.169, 2.918]	0.232
27	0.583	0.583	53.8	0	0.956	[-1.875, 1.875]	-0.21
28	0.25	0.5	45.5	NA	NA	NA	NA
29	0.75	0.333	73.1	1.773	0.366	[1.055, 2.491]	-0.122
30	0.5	0.05†	86.4	2.316	NA	NA	0.822
31	0.583	0.083	76	2.265	0.679	[0.934, 3.595]	0.586
32	0.667	0.167	76.9	2.069	0.335	[1.413, 2.726]	0.268
33	0.417	0.167	65.4	1.406	0.685	[0.062, 2.749]	0.589
34	0.667	0.583	57.7	0.71	0.746	[-0.753, 2.173]	-0.321

TABLE 4.10: d' Statistics per Participant (Experiment 3).

CHAPTER 5

Dynamically Changing Pockets and Form in "Superstition"

So far in this dissertation, pockets have been viewed either locally (a single window of time in which a few different instruments' onsets coalesce within one pocket) or cumulatively (for instance, the delineation over the course of a few bars of the probabilistic domain associated with each individual drummer in Chapter 3), but always in isolation and removed from the broader musical context. As yet unexplored is how pockets manifest and operate in the context of a full piece of music and whether any variations in the shape of a pocket interact with other musical elements, such as form or hypermeter. This chapter changes the scale on which pockets are analyzed, zooming out to ask if microtemporal details might interact with the macrostructural design of famous pop songs to explore the artistic and aesthetic effects of ever-changing pockets. Through a case study of Stevie Wonder's "Superstition" (1972), I explore how the shape of the pocket changes across the different formal sections of the song, working alongside other compositional features of the track to amplify the contrast between each part of the song's structure.

This chapter introduces an analytic methodology that may be used to understand pockets in full ensemble recordings and provides musical context for the theory of beats as domains. Quantitative analyses of micromusical phenomena may unearth statistically significant findings, but these findings need to be deeply embedded into musical analyses and connected with what musicians do in their art. We may analyze a piece through the lens of the theory of pockets and find, for example, that a melody instrument is consistently earlier in the pocket than a rhythm or harmony part, or that one performer in an ensemble has a notably different pocket shape to the rest of the ensemble. These details must then be integrated into a consideration of what it means for the experience of the music, what it means for the performer and/or listener, and perhaps even raise questions about why a performer may have made such a decision (consciously or unconsciously).

"Superstition"

"Superstition" has not only achieved long-lasting commercial success and recognition, regularly featuring in "top songs of all time"-style lists (see, for example Time Magazine's "All-time 100 Songs" [Wolk, 2011] and Rolling Stone's "500 Greatest Songs of All Time" [Rolling Stone, 2003]), but it has also become a common point of reference for groove research after Janata et al. (2012) found it to have the "highest mean groove rating" of the various tracks they asked experimental participants to evaluate. Many other researchers have since used the Janata et al. tracks, often just "Superstition," in their own investigations of groove phenomena (e.g., Christensen & Bemman, 2017; Hove et al., 2020; Ross et al., 2016; Senn et al., 2020; Stupacher et al., 2016; Stupacher et al., 2013; Stupacher et al., 2017). "Superstition," evidently, is a significant song both culturally and academically that warrants further investigation in this chapter's case study.

Hughes (2003) provides an analytic overview of the track, describing its genesis and overall construction, as well as presenting some quantitative analyses of the timing of the opening drum pattern's swing that demonstrate the inconsistency of sixteenth-note events in Wonder's performance (see also an analysis of the microtemporal details of the core clavinet part in Ashley, 2014, pp. 159–61). Hughes draws these timing measurements together with a discussion of harmonic and instrumentation choices by Wonder to argue that:

> In this song, there is no question of confusing the location of the beat, which is steadily sounded by the drums, bass, and clavinet. Instead, Wonder is playing with the flow created by that beat, pushing our ears forward by tantalizing, satisfying, or surprising us while simultaneously creating the sense of metric informality usually associated with live recordings of musical groups.
>
> (Hughes, 2003, p. 162)

Hughes's analysis emphasizes the fact that the sense of beat is dynamic throughout the track and that this dynamism has consequences for the listener's qualitative experience. Metric locations are never ambiguous, but Wonder is artistically shaping the individual beats.

One last motivation for focusing on "Superstition" in this study is that Stevie Wonder multitrack recorded all the parts himself (apart from the saxophone and trumpet parts). That is to say that he recorded the drums, rewound the tapes and recorded the Moog bass synth on a new track of the same tape while listening to the drum performance, then the same for the two Hohner clavinet parts, and then finally sang (for more on the recording process, see Deane, 1995b, 11:34ff.). As the performance was recorded with no external, metronomic reference (no click track was used),[1] the rhythmic frameworks that guide "Superstition" arise wholly from Wonder's internal clock. I argue that the expression of musical time is therefore completely under his control as composer–performer and so, through analyzing the details of this performed time, we can understand how Wonder uses different pockets in tandem with other performative and compositional techniques to shape the experience of the song.

[1]Click tracks have existed for synchronizing sound and film since at least the 1920s, but only really came into the music recording studio in the 1980s—a decade after "Superstition" (Brennan, 2020, p. 295; also see Théberge, 2016).

Analytic Method

To analyze the pockets in "Superstition," I first need to know the exact location in time of every single sound event, which can be difficult with the complex sound of full-band performances that we hear on commercial records. Fortunately, the original, individual recorded tracks ("stems") for "Superstition," as well as an eclectic array of other famous songs, are available through online communities of fans who share the individual instrument parts so that they may make remixes (e.g., they can use just the bass synthesizer part in their own song) or simply because they love the songs/performances and want to listen more closely to individual elements. The source of these stems is hard to ascertain due to online anonymity and the legal quagmire of sharing fragmented copyrighted materials that were not intended to be released to the public. Stems arrive on these fan websites from multiple possible sources. They might get leaked by someone working at the recording studio. They might appear through remix projects that are officially endorsed (an artist uploading their raw materials to invite others to remix them[2]) or that are not so official (someone who is doing an official, sanctioned remix leaks the materials without permission). I believe this kind of route—a leak of the official, original material either from the studio or from a remix—is the origin of the "Superstition" stems. Another sizable source of stems, though not for "Superstition" as it does not feature in these, is unencrypted data from the video games "Guitar Hero" and "Rock Band." In order to allow players to perform the guitar, bass, or drum part of the famous songs, the game uses the original stems so that it can subtract that instrument out and allow the game player to fulfill that role. Lastly, there are sophisticated and impressively successful artificial intelligence/neural network methods available for separating individual musical sources out of a track, most typically separating the vocal part from the rest of the ensemble, for example

[2]For instance, Radiohead ran a few remix contests in 2008, one for "Nude" and another for "Reckoner" that some argue was as a novel marketing ploy to boost iTunes rankings through double sales (Buskirk, 2008). Likewise, in 2011, R.E.M. uploaded the raw files for their song "It Happened Today" to their fan website (https://remhq.com/news/remix-it-today-5/) with a Creative Commons non-commercial license and invited people to "remix, reimagine, remake."

Lalal.ai,[3] Spleeter (Hennequin et al., 2020), CatNet (Song et al., 2021), Open-Unmix (Stöter et al., 2019), and the improved version CrossNet-Open-Unmix (Sawata et al., 2021).

The novel analytic method I used to understand pockets in complex, ensemble performances is summarized in Figure 5.1. It is important to note that the computational tools and processes I have selected have a direct impact on the results I obtain, and that different tools and methods may yield different results; such matters fuel ongoing research agendas in music perception and computational musicology.

First, using the individual recording stems, I collected information about the onset times for each instrumental track in "Superstition" using MiningSuite 0.10 (Lartillot, 2019).[4] This method is used to detect the onsets for the bass synth, bass drum, and the two clavinets. The snare and hi-hat parts, however, were both recorded using overhead microphones and so there is not a unique stem containing only snare or hi-hat onsets. As such, I used ADTLib (Southall et al., 2017), an automatic drum transcription tool that employs neural networks, to distinguish snare and hi-hat events (for a review of various drum transcription tools, see Wu et al., 2018). To make sense of the various event timings gathered with the MiningSuite and ADTLib, and to consider them in relation to one another, we need to approximate where metric locations—i.e., beat 1, 2, etc.—are so that onset clusters around these locations may be analyzed as pockets. In order to do this, I made use of the beat tracking algorithm from the madmom audio signal processing library (Böck et al., 2016), which uses a large number of trained Recurrent Neural Networks that have Long-Short Term Memory cells that are then filtered by a Dynamic Bayesian

[3] https://www.lalal.ai/

[4] This is a beta development of a new Matlab package that is framed (in part) as version 2.0 of the MIR-toolbox (Lartillot et al., 2008a). The process and parameters for extracting the onset times are as follows: `a = aud.events('sound_file.wav', 'Detect', 'Peaks'); b = sig.peaks(a, 'Threshold', 0.1); get(b, 'PeakPrecisePos');`. In this sequence of functions, events are detected with a peak-picking algorithm, all peaks below 10% of the maximum peak are screened out (to eliminate bleed-through from other instruments), then the precise time code for each peak is recorded. A subset of onsets for each instrument was also manually inspected, looking at the waveform for each instrument stem in Logic Pro X and verifying that the output of the `aud.events` function corresponded to musically meaningful peaks in the sound signal.

Network to arrive at the final estimations of the beat locations.[5] This provides time codes for locations that approximate where a metronome click or listener tap may be. These approximate beat locations were qualitatively validated by creating an audio file that overlaid a beep at the locations madmom suggested onto the original audio. The resultant output of original audio with an overlaid reference structure matched my own listening experience of the beat locations. All events within a window of time around this location—defined here as 10% of the inter-beat interval—get labelled with that metric position.[6] All other events falling outside these windows, for example eighth and sixteenth notes, are excluded from the present analyses. These filtering processes work with the onset times as located by the MiningSuite and ADTLib functions and don't compensate for, for example, the different attack durations of the individual instruments. Working directly with the onset-detection output is the best choice available as any manual post-processing would reduce the replicability of this analytic method and its findings. Finally, the earliest onset associated with each metric position is defined as the start of the pocket and the time codes of all other onsets associated with that metric position are compared with that first onset.

[5] It is possible to use the MiningSuite for this function, too, but madmom's method is one of the highest performing models in the MIREX Beat Tracking competition.

[6] The boundaries of the pocket were set as ±10% of the inter-beat interval (IBI) as a result of careful consideration of the rhythmic elements of "Superstition." The foundation of the track's groove is a sixteenth-note shuffle. The swing ratios in "Superstition" vary widely (see Ashley, 2014; Hughes, 2003), so it is hard to prescribe a fixed value for the pocket boundaries that would ensure that the swung sixteenth-notes are not accidentally captured in the analysis of the following metric position. Perfectly isochronous sixteenth-notes would be 25% of the IBI. The most heavily swung sixteenth notes in the first measures (according to Hughes, 2003, Table 3) near a 3:1 ratio, which would make the short note last 12.5% of the IBI. Therefore, 10% of the IBI is a reasonable boundary to set either side of the metric position found by madmom such that swung sixteenths are not accidentally included in the analysis. Lastly, this ±10% was reconsidered after the full set of data were obtained in case a significant number of outliers were captured. This was not the case, as will be seen in the density plots presented below. The process of determining these boundaries illustrates the highly contextual features which contribute to the probabilistic nature of the pocket boundaries, with this being my informed but personal view of the pocket.

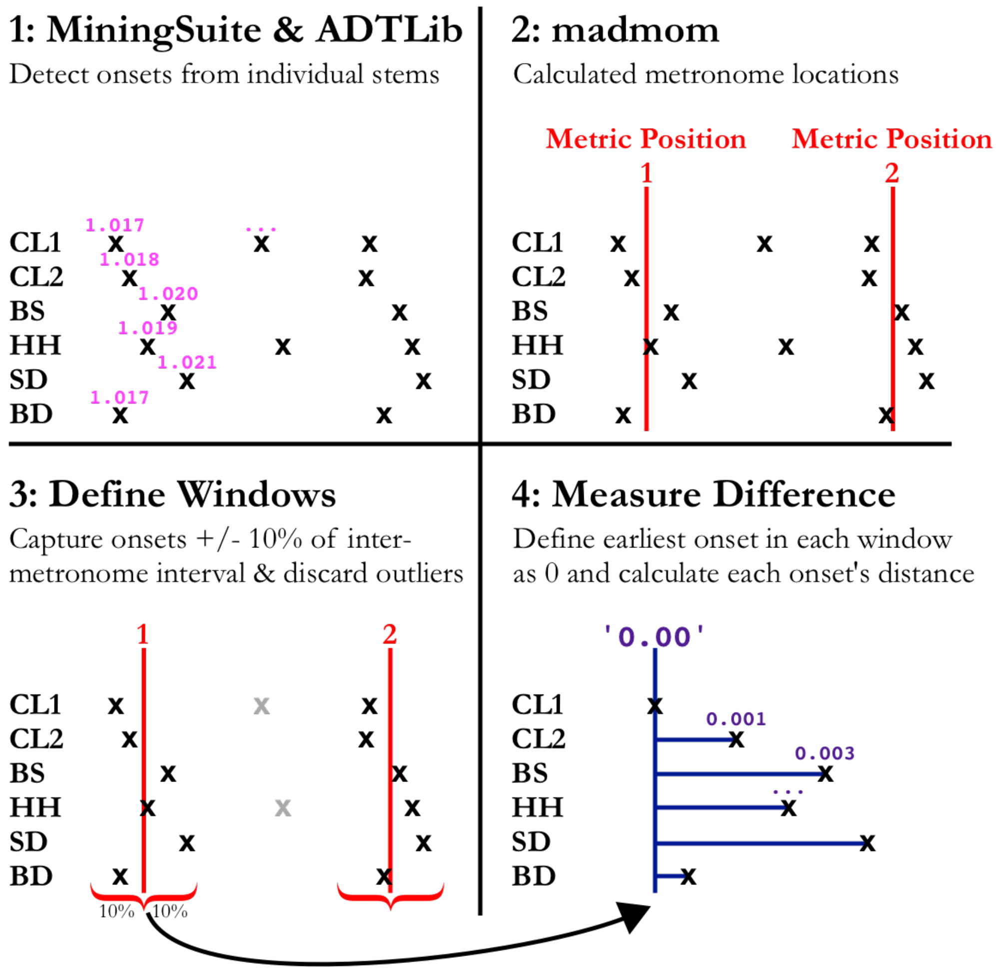


Figure 5.1: An illustration of the analytic steps used to evaluate the pockets in "Superstition." CL1 = clavinet 1, CL2 = clavinet 2, BS = bass synthesizer, HH = hi-hat, SD = snare drum, BD = bass drum.

Analysis

The majority of the following analysis focuses on pockets whose profiles are defined by aggregating onsets over numerous metric positions (cumulative pockets) across 8- and 16-bar formal sections of "Superstition," but, first, it is interesting to progress step-by-step through the opening bars to see how individual local pockets are integrated into the cumulative ones. The track opens with drums only, playing a shuffle groove (see Figure 5.2 for a transcription of the opening bars of "Superstition" in staff notation and Figure 5.3 for the precise timing details of each instrument in bar 1). The top half of Figure 5.4 shows an exaggerated illustration of each of the four main local pockets afforded by this first bar.[7] Starting with the downbeat of bar 1, we hear a hi-hat and a bass drum. The onset-detection analysis shows that the hi-hat sounds before the bass drum, so we have two distinct events associated with metric position 1. With so little other information available, we might hypothesize that the pocket is constructed in such a way that it enfolds both onsets and that the location most likely to be defined as "the beat" is halfway between the two onsets. The second metric position features hi-hat, snare, and bass drum onsets, again with a small amount of non-alignment between them. Taking this moment in isolation, we can hypothesize once more about what kind of probabilistic domain could contain all three events, skewing the pocket slightly. Repeating this process for the remaining two locations, we end up with four distinct local pockets. These four pockets can be amalgamated (lower half of Figure 5.4) into a cumulative pocket that starts to define what the shape of the beat domain is in the introduction to "Superstition."[8]

[7] The anacrusis to the track is on solo snare drum and therefore doesn't provide much information about the contours of the pocket (though these two beats do initiate metric processes and start to define the metric reference structure for the track).

[8] It is important to stress that the methodology presented in this chapter weights the contributions of each instrument to the cumulative pocket equally. In reality, the contribution of each instrument to the sounding pocket is likely to vary a lot due to factors such as the instrument's prominence in the sound box of the final recording/how the individual stems are mixed together and also the role each instrument is performing in the ensemble.

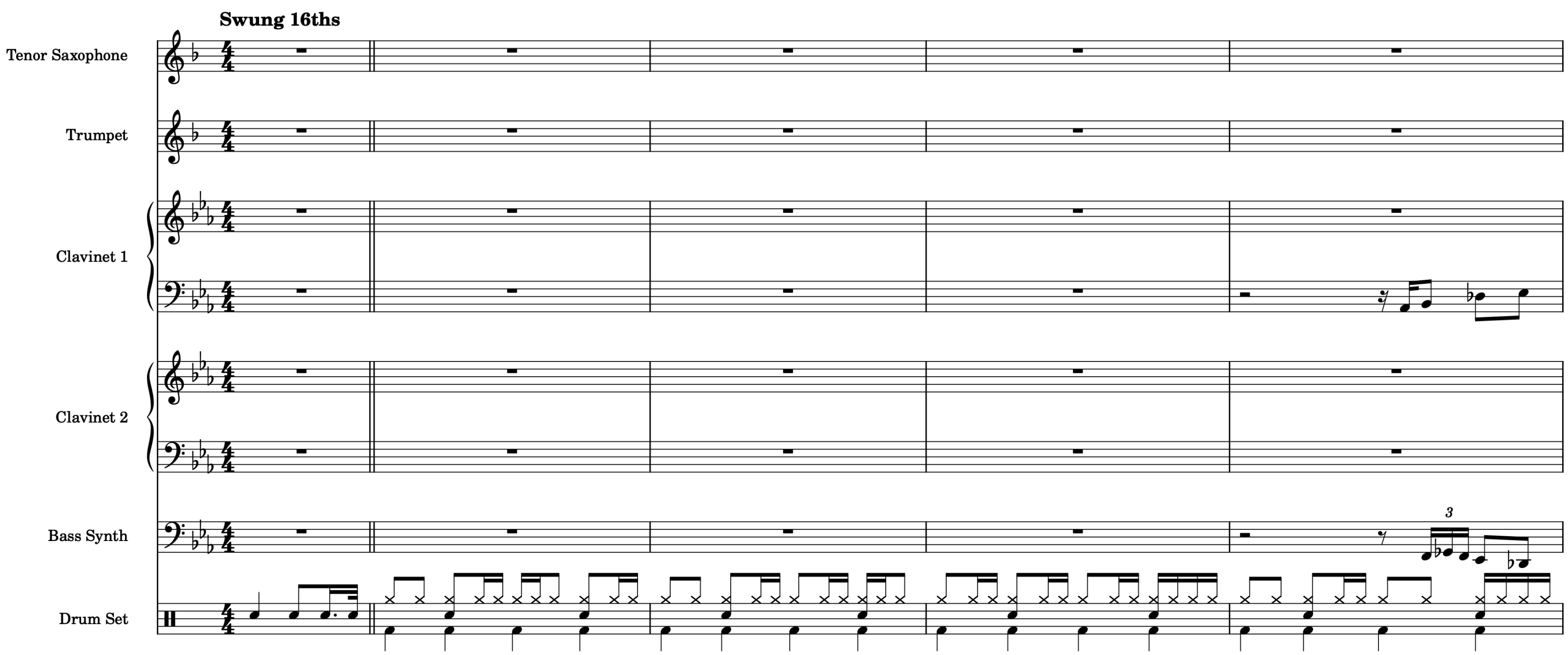

Swung 16ths
Tenor Saxophone
Trumpet
Clavinet 1
Clavinet 2
Bass Synth
Drum Set
3

Figure 5.2: Transcription of the opening bars of "Superstition." NB: 16th notes are swung throughout.

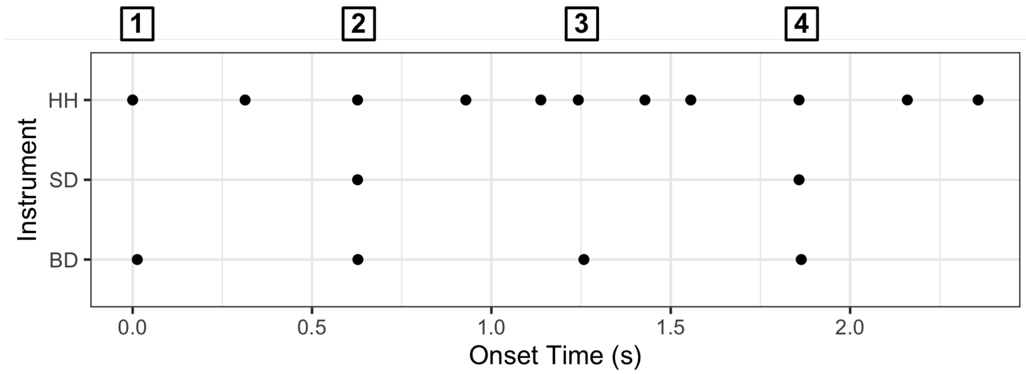


Figure 5.3: All onsets in bar 1 of "Superstition." Above the figure, approximate metric locations ("beat 1," "beat 2," etc.) are indicated to aid with tracking the performance.

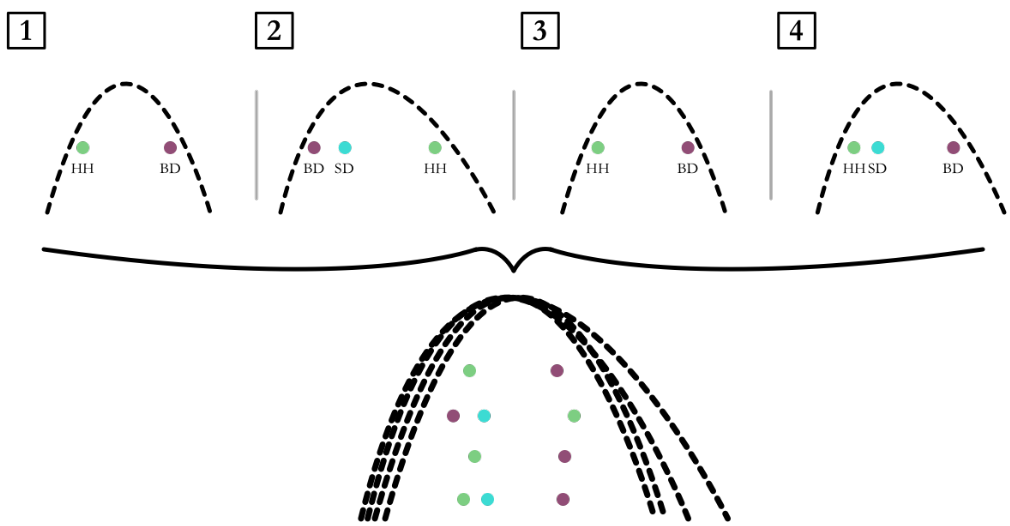


Figure 5.4: A sketch of the four local pockets and the amalgamation into a cumulative pocket in bar 1 of "Superstition."

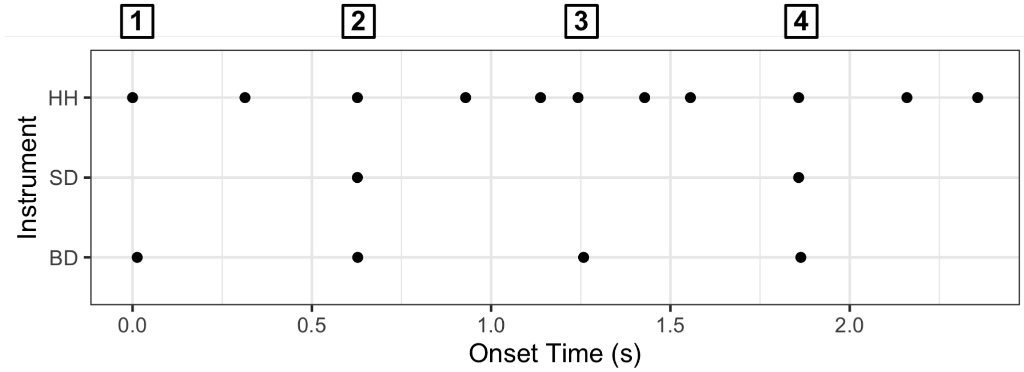
Hughes (2003, pp. 154–6) analyzes the variation in the amount of swing in the introduction's hi-hat part, having relatively consistent eighth notes, but a large amount of irregularity to the sixteenths, and argues that:

> The beat is smooth and predictable overall, but convoluted and highly unpredictable in detail. In other words it is loosely organized, by design. This feeling of loose

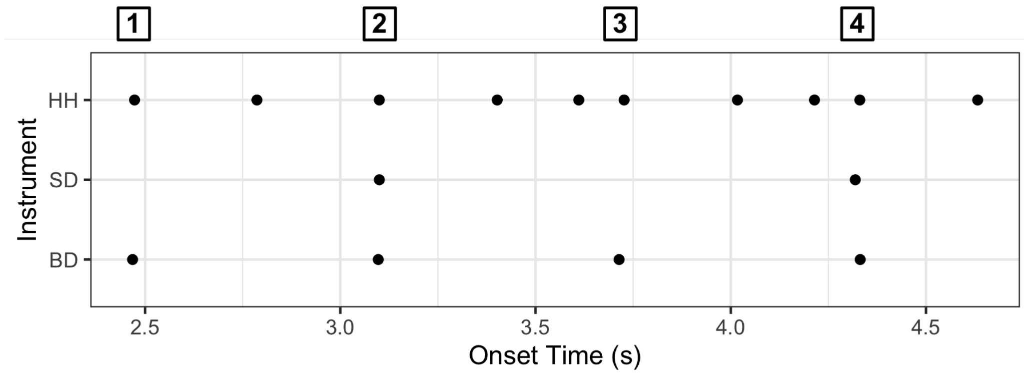
> organization, of "standing on shaky ground" (to paraphrase a Temptations song with a similar groove) is clearly established in the opening drum part.
>
> (Hughes, 2003, p. 156)

The visualization of the precise details of ensemble timing present in Figure 5.3 extends Hughes's observations, showing how the detail of the beat is indeed loosely organized not just as a result of the swung shuffle pattern on the hi-hat, but also because of the microtemporal details of the local and cumulative pockets.

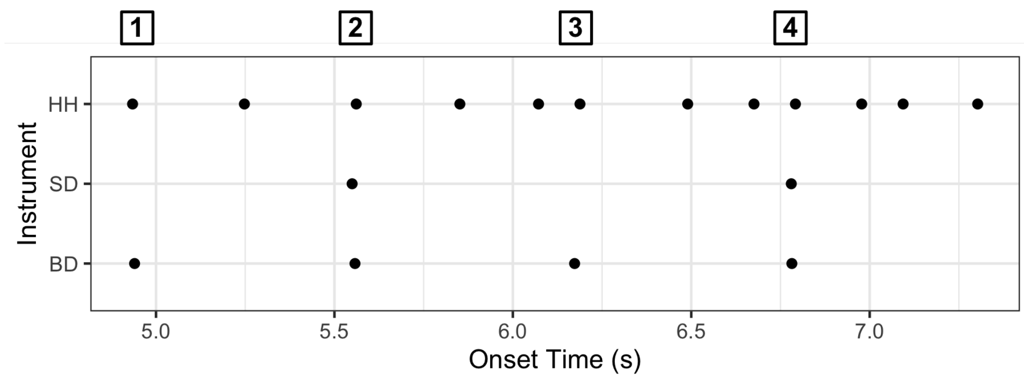
Stepping through this process bar-by-bar for the opening four bars (Figure 5.5), this "loose organization" of the beats is seen repeatedly. For instance, the non-alignment of hi-hat, snare drum, and bass drum onsets can be seen across all the of subfigures. The hi-hat appears to lag behind the other instruments, occurring later in the pocket. The composite sound of all of these individual events does not necessarily reach our ears as numerous distinct onsets due to the very small size of the timing differences; however, the lag of the hi-hat in the pocket, while perhaps not aurally distinct, might be said to add some aural "thickness" in these opening bars. While not necessarily clearly separate, the width of the pocket in these opening bars (and elsewhere in the track) can impact, for example, the timbre. This idea of "thickness" is potentially an additional qualitative property beyond tight/loose and pushed/laid back that the theory of pockets can offer to our understanding of performances. Lastly, if we were to sketch the local pockets for metric locations seen in this opening, as in Figure 5.4, we would observe how the pocket is dynamic and ever-changing in response to new information.



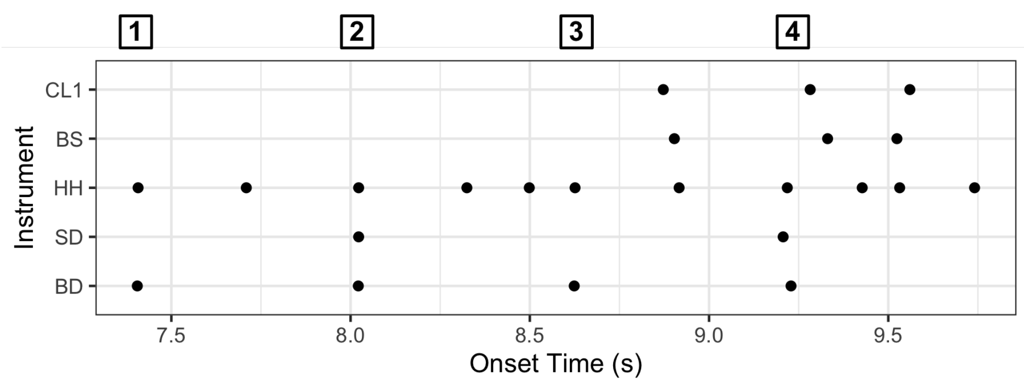
(5.5.1) Bar 1.



(5.5.2) Bar 2.



(5.5.3) Bar 3.



(5.5.4) Bar 4.

Figure 5.5: All onsets in the opening four bars of "Superstition."

When more instruments enter the musical texture, the amount of information available to define each local pocket increases, accelerating the pace of the development of the cumulative pocket and more clearly defining its shape. In bar 5, the first clavinet enters along with the bass synth (see Figure 5.6, the pickup to bar 5 is shown at the end of Figure 5.5.4) and the way that the beat domains are pulled in either direction, as well as having the core, peak position of the pocket reinforced, becomes more evident. The four main metric positions of bar 5 are well defined, but the clavinet in particular seems to play against the rest of the ensemble, laying back on the first entry, and swinging so much on "3 +" that this onset almost coincides with the fourth beat of the bar (see just after 11.5 seconds in Figure 5.6).[9] The clavinet's tendency to lay far back in the pocket can be seen throughout the performance. As one more example of this, see Figure 5.7, which shows the instrumental parts behind the first vocal entry of "Superstition" in bar 9, the start of the first verse. Note also, here, that the second clavinet part, which plays an off-beat dotted-eighth figure that rhythmically interlocks with the first clavinet part (see Figure 5.2), also plays after the drum part in each beat's pocket (for example, observe the slight delay on the fourth beat of the bar, around 21.4 seconds, when compared with the drums at the bottom of the figure).

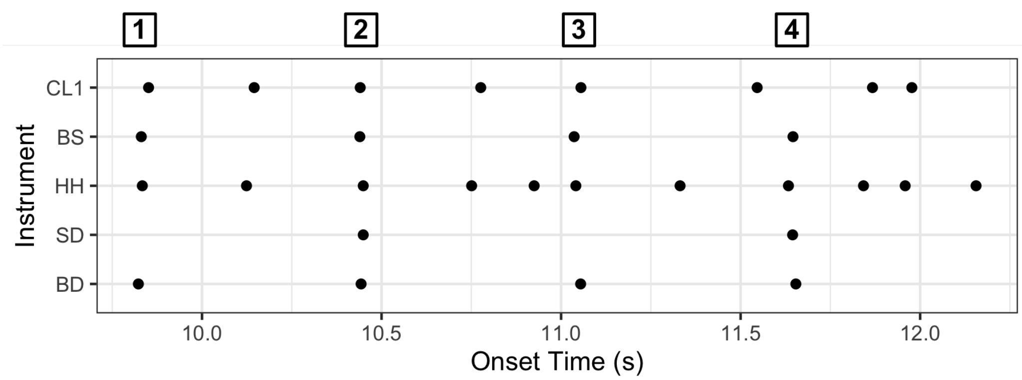


FIGURE 5.6: All onsets in bar 5 of "Superstition."

[9] Although the analysis of pockets presented in this chapter does not capture this laid back swing (as it falls outside the ±10% bracket defined in the analytic methodology outlined in Figure5.1), it may well be the case that the heavy swing has some kind of gravitational pull on the pocket from the outside, widening beat 4's pocket further (a kind of "beat 4 in anticipation," after Danielsen, 2006, Chapter 5).

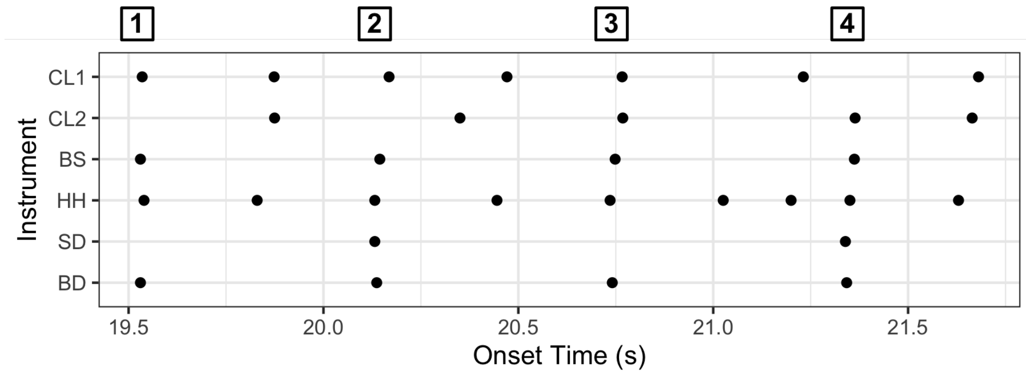


FIGURE 5.7: All onsets in bar 9 of "Superstition."

Pockets & Formal Sections

Moving beyond the individual pockets in the introduction, Figure 5.8 uses density plots to provide an overview of the scale and shape of the pockets across each of the main formal sections of "Superstition" and to gain a general sense of any variations. The *x* axis is measured in seconds, starting from the earliest onset associated with each pocket (as defined in Figure 5.1). The height of the curve on the *y* axis represents how dense events are at that moment in the pocket. As an example of how to interpret this plot, there are very few sound events within most pockets from 0.04 seconds onwards so the curve is low, but there are a large concentration of events earlier on in each pocket and so the curve is higher there. This plot breaks the form of the song down into subsections, so each larger building block (e.g., verse 1) gets split in half (A and B), allowing more detailed insight into how the pocket dynamically evolves over the course of the song. From this first look, it can be observed that there are indeed differences in the shape and spread of the pockets for each section, suggesting that Stevie Wonder changes the way that he expresses beats over the course of the performance of a song. The ways in which the pocket changes between sections will now be explored further.

Focusing on the verses and choruses, there appears to be a pattern whereby verses (coral/light orange in Figure 5.8) and choruses (turquoise) have different pocket shapes. The verses' pockets

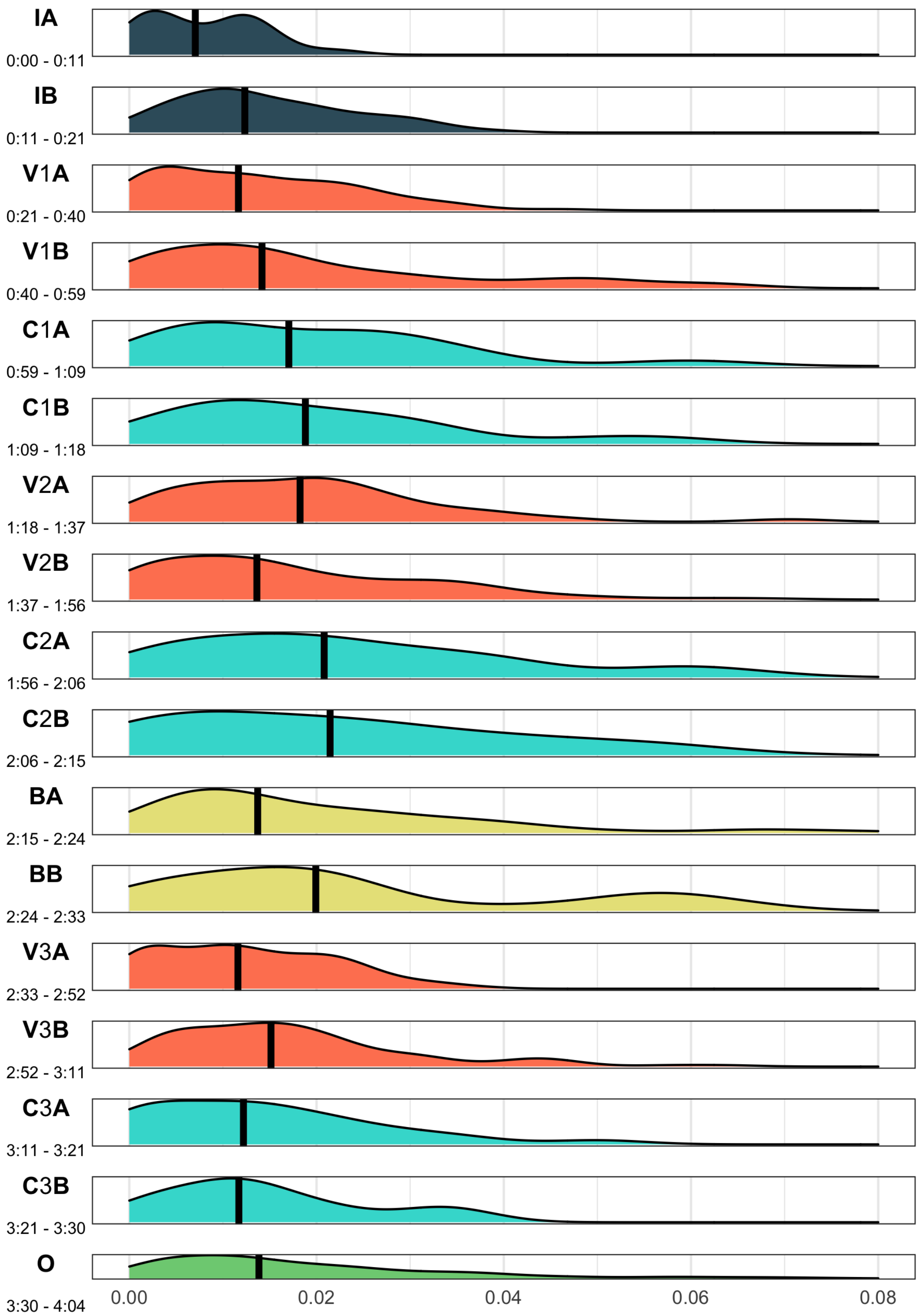


FIGURE 5.8: An overview of pockets (measured in seconds on the *x* axis) across the form of "Superstition." Black vertical lines indicate where the median of each pocket is. IA = first half of the introduction, V1B = second half of the first verse, C2A = first half of the second chorus, BA = first half of the bridge, O = outro.

are not as spread out as the choruses' (represented by horizontal spread in the density plots), being more bunched up to the left-hand side than the choruses, which have long right-hand tails. Additionally, the median amount of spread in how wide (in seconds) the pockets are for each section is further right for choruses than verses, at least for the first two verse–chorus pairs (see the black vertical lines). This would suggest that the onsets in the verses are more closely synchronized, with fewer/smaller discrepancies between the instruments. The third verse–chorus pair appears to change this pattern somewhat, with these two sections of the form having rather similar profiles. This is more easily visible in Figure 5.9, which summarizes the onsets only from the verses and choruses (for simplicity and to facilitate comparison, the sections are taken as wholes, rather than in halves as previously). Statistically, there are some differences between the size of the cumulative pockets that are performed in verses and chorus.[10] Chorus 1 has significantly more spread to its pocket that verse 1, $t(148.31) = 2.23, p = .027$; however, chorus 2 does not have significantly more spread than verse 2, $t(110.32) = 1.771, p = .079$, nor does chorus 3 have more spread than verse 3, $t(118.17) = 0.481, p = .632$.

Interpreting the descriptive overview of the verse and chorus pockets presented in Figure 5.8 as well as the above statistical findings in more musical ways and using the language of "feel," we might say that the first two verses have more consistently *tight* pockets than the first two choruses, which have more spread and so may be described as being more *loose*. The musical materials that make up the verses and choruses differ quite noticeably: properties of the verses include harmonic stasis, interlocking independent rhythmic ostinati from the clavinets, rhythmic interplay between

[10] Because of the way that the pockets are measured (always from the earliest point), there is a large amount of positive skew and so models that presume normally distributed data have an increased risk of Type I Error. As such, onset times are Box–Cox transformed (Box & Cox, 1964) before being analyzed for all analyses presented in this chapter. The Box–Cox transformation is an exponential transformation that takes the original, non-normal data and finds the optimal value of the function's exponent, λ, that will result in normally distributed data, allowing for the application of a wider variety of statistical tools. In this approach, in contrast to simply applying a logarithmic transformation (which is actually the same as a Box–Cox transformation with $\lambda = 0$), a transformation parameter that is dependent on the distribution of the data under investigation is estimated prior to transforming the data. This makes the Box–Cox transformation more flexible than other methods because the conversion into the transformed scale is specifically a function of the distribution of the data.

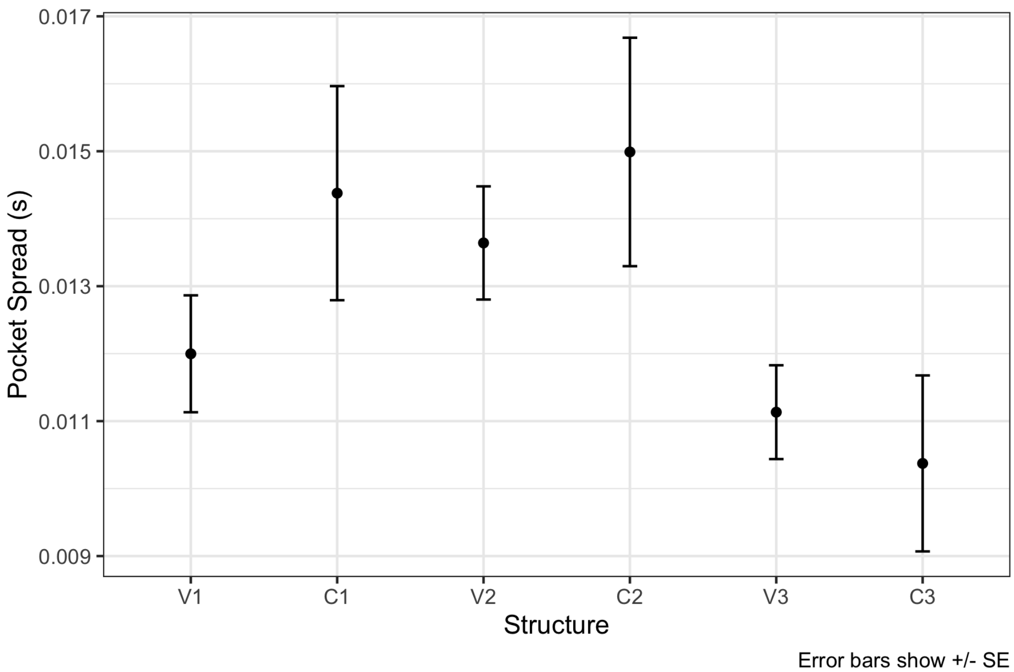


Figure 5.9: Mean pocket sizes for verses and choruses in "Superstition."

the instrumental parts and Stevie Wonder's singing, and a relatively static/non-teleological sense of jamming on the groove without an urgent need to develop or lead somewhere. The choruses, on the other hand, feature chromatic harmonic movement, relatively homorhythmic texture with more sustained notes, a swirling effect caused by pitch bends and rising and falling dynamics on the sustained notes, and a clear trajectory towards the climactic V^{+} followed by a dramatic break in instrumental activity. It would seem that the shape of the pocket for each section is another parameter of the performance that Stevie Wonder is modifying to enhance the contrast between the verses and choruses. Whether this is intentional or not is unknowable, as is whether the pockets are more loose in the choruses independent of the other changes or whether, perhaps, the larger physical gestures required to play louder and the move from the close-to-the-body hi-hat in the verses to the arm's length away ride cymbal in the choruses, to shift harmonically around the instrument, or to play the drum fills in the chorus may cause an increase in the size of the pocket as a byproduct of other musical actions. The third verse and chorus might, however, provide counter-evidence to the argument that any change in pockets is simply a side effect of an intensifying of physical gestures as they have a higher level of energy than the first two verse–chorus pairs yet their pockets, as seen in Figure 5.9, are smaller. This third verse–chorus pair is not as aurally contrasting as the first two, which perhaps also explains why the statistical analysis showed no difference in pocket shape. Musically, this may be a decision to vary the performance on the third repeat, a familiar compositional logic seen in a range of musical styles (e.g., the "rule of three" as seen in AAB song structures de Clercq and Margulis, 2018, p. 151 or a Schoenbergian musical sentence, which breaks the pattern that had been established through repetition on the third instance).

The first and second halves of each formal section, when looking at Figure 5.8 and especially the median values as indicated by the vertical black lines, also appear to have potentially different pocket profiles. The second half of most sections appear to have wider, looser pockets than the first half (apart from verse 2 and perhaps chorus 3). When analyzing this statistically, however, this difference between first and second halves is not significant, $t(980.7) = 1.857, p = .064$. The second halves of verses feature horn riffs (not included in the analysis as they are not played by Stevie Wonder) that are paralleled in the bass synth while all other voices remain the same as in the first half. The second halves of choruses and the bridge feature a return of the verse's clavinet ostinati and more sustained horn parts with the previously described atmosphere of non-teleological jamming out. All of these second halves of sections, though in different ways, have a sense of expanding or developing previous material, an expansion that might be said to be reflected in the expansion of the size of the pocket. That being said, the fact that verse 2 is musically very similar to verse 1, but does not mirror the pattern of expanding of the pocket for a looser second half complicates the suggestion that there are systematic changes in pocket shape associated with specific formal sections of "Superstition." One possible explanation for, or interpretation of, this disconfirming evidence involves reframing how we contextualize the pockets in verse 2: It is not that the second half is less loose, rather it is that the first half is remarkably loose to start and so there is limited scope to loosen the pocket in the second half. Verse 2's clavinet parts begin to develop the now well-established ostinati and add a few flourishes, which could be why the first half is comparatively loose—Stevie Wonder may be relaxing into the patterns and embodying a more loose aesthetic, and/or these flourishes are breaks from the repetitive physical motion that has been locked in for the previous minute and a half, and so there may well be some more looseness here as an artifact of the physical variation of the patterns.

Instrument-Specific Pockets

Each instrument that Stevie Wonder plays in "Superstition" contributes differently to the cumulative ensemble pocket. Figure 5.10 presents cumulative pockets for the entire track, divided by instrument, and it is very clear that the drums have very differently shaped pockets to the keyboard instruments. The keyboards have wider pockets both in terms of where their median values are and also in how substantial their right-hand tails are. The drum kit instruments, on the other hand, are far tighter. There also appears to be some bimodality in the hi-hats and snare drums with twin peaks that roughly align and mirror each other, perhaps suggesting that the hi-hat sometimes leads the snare drum within the pocket, but most of the time it is the snare drum which leads the hi-hat.[11] Statistical investigations of the impact of different instruments on the shape of the pocket show that the observed differences from Figure 5.10 are indeed significant and that each instrument makes a sizable impact on the shape of the pocket. A regression on Box–Cox transformed data (summarized in Table 5.1; refer back to footnote 10 for an explanation of the transform) shows that each instrument contributes uniquely to the overall, ensemble pocket. When read in conjunction with (unadjusted) summary statistics for the instruments (Table 5.2), it

[11] This could, potentially, simply be an artifact of the automatic drum transcription method required to obtain onset times for the snare and hi-hat parts. Many of the onset times it assigns to these two drums are identical because of the challenge of separating the two when they occur so close together in time.

can be observed that the keyboard instruments have significantly more spread (more "looseness") to their pockets than do the drum kit pockets and occur, on average, later in the pocket (more "laid back").

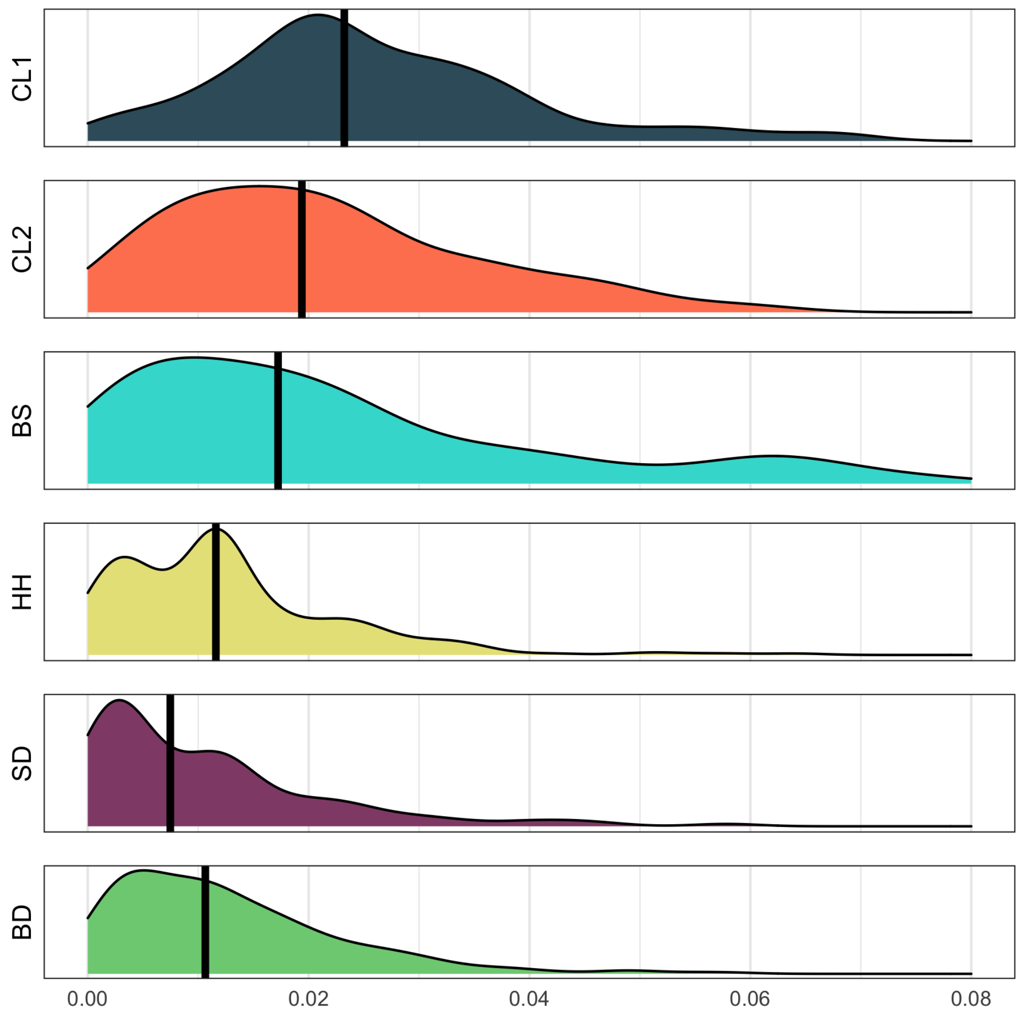


FIGURE 5.10: An overview of each instrument's pocket in "Superstition." As previously, black vertical lines indicate the median for each pocket.

Are these differences in pocket shapes for each instrument due to the role of each instrument in the ensemble, with keyboards perhaps having the ability to lay back behind the drums that have to delineate time by defining the pocket earlier on? Or could it be because of the way in which "Superstition" was recorded, with the drums being performed first and the other parts

Predictor	Estimate	SE	t	p
Intercept	-1.991	0.011	-188.135	<.0001
CL2	-0.051	0.018	-2.865	.004
BS	-0.064	0.017	-3.833	.0001
HH	-0.167	0.015	-11.241	<.0001
SD	-0.207	0.018	-11.256	<.0001
BD	-0.166	0.014	-11.645	<.0001

Observations = 1217
R^2 / Adj. R^2 = .167 / .164
$p < .0001$

TABLE 5.1: Summary of regression analysis for instruments predicting metric spread/pocket size. NB: the estimates are for Box–Cox transformed data ($\lambda = 0.384$). They can be read relatively, but the positive or negative coefficients cannot be interpreted literally, nor is it possible to directly get a sense of scale (e.g., a β of 0.5 is not twice as large an effect as a β of 0.25).

Instrument	N	Mean	SD	Median	MAD
CL1	245	0.0257	0.0138	0.0232	0.0115
CL2	134	0.0218	0.015	0.0194	0.0135
BS	162	0.0228	0.0198	0.0172	0.0159
HH	250	0.0126	0.0103	0.0116	0.00953
SD	122	0.0113	0.0132	0.00747	0.00811
BD	304	0.0128	0.0103	0.0106	0.00931

TABLE 5.2: Summary statistics for the location and scale (in seconds) of each instrument's pocket in "Superstition" (untransformed data with zero values—i.e., onsets that are defined as the start of each pocket—filtered out).

being recorded while listening back to the drums (and without the abstract, external time keeping of a click track)? Within the confines of an analysis of "Superstition," these questions may not be answered, though a more broad analysis (discussed below) may well be able to disentangle the roles of each instrument within an ensemble.

One alternative justification for why the keyboard instruments have such different pocket shapes to the drums may be that a hallmark of drum sounds is that they all have very sharp "attacks" (roughly defined as the portion of a sound where the sound begins and increases in volume to its peak). When a drum or cymbal (though not very large cymbals like gongs) is struck, we immediately hear its onset—the attack stage of the sound is almost instantaneous. The keyboard parts in "Superstition," however, present a variety of sound profiles whose attack times may potentially be linked to the pocket sizes they afford: The clavinets have a bright, staccato sounds while the Moog bass synthesizer has a much more muddy, legato sound with large amounts of portamento between pitches. Looking at the envelopes for each sound (Figure 5.11), we can observe how each of the instruments used in "Superstition" has a unique profile for how the sound is expressed through time.[12] For instance, the snare drums clearly have a single, sharp attack with immediate decay, while the bass synthesizer rises quickly, then sustains before decaying (the oscillator that gives the Moog its unique timbre is also visible in the rapid wiggles in the envelope). Therefore, in order to take steps to understand the differences in pocket sizes by instruments, it is important to analyze the attack times of the instrumental sounds in "Superstition."

For each instrument, a representative sample of 10 onsets was edited out of the individual stems and each of their envelopes were analyzed using the MiningSuite's attack time function.[13] This function defines the start of the attack as the time point when the amplitude is first above a noise threshold (40 dB below the maximum amplitude for the given onset) and then finds the time from this point to local maxima/peaks in the signal. It considers multiple maxima, not just the first peak, in case the sound rises, has a small dip, then rises again to the true peak (for an in-depth discussion and illustration of the method, see Lartillot et al., 2021, p. 726ff.). Summary statistics for the length of the attack phase of each instrument's onsets are shown in Table 5.3 and a comparison is shown in Figure 5.12. From this, we can see that there is some variation in attack times between each instrument, and very little variation within each instrument (seen in the very small error bars). An ANOVA of attack times (NB: not Box–Cox transformed) showed significant variation between instruments, $F(5, 24) = 11.99, p < .0001$. A post hoc Tukey test

[12] The differences in height on the y axis are simply due to differences in amplitude (/loudness) on the raw recording files—a consequence of microphone placement or recording gain settings—and are not indicative of the loudness in the final, complete mix.

[13] `aud.attacktime('sound_file.wav');`

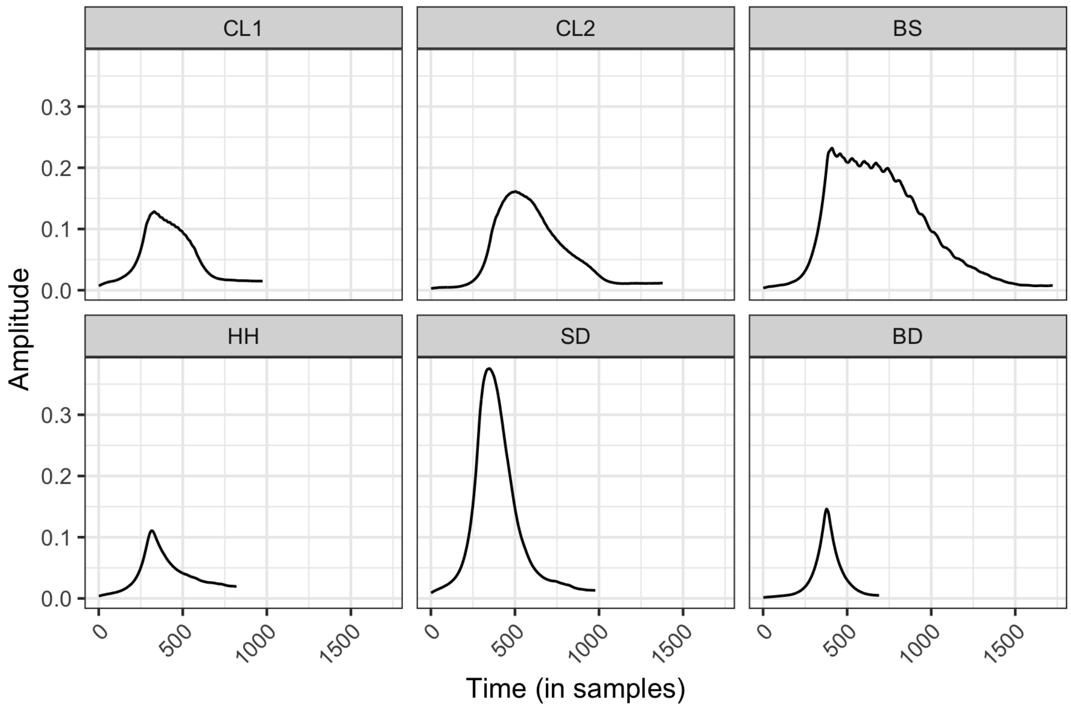


FIGURE 5.11: An illustration of each instrument's envelope. The snippet of audio used to determine the envelope varies in length, hence why, for example, the analysis of clavinet 1 ends around 1000 samples, while the bass synthesizer part fills the frame. The sampling frequency of the recordings is 44.1 kHz, so 1000 samples ≈ 0.0227 seconds.

showed that clavinet 2's attack time differs significantly to all other instruments apart from the snare drum ($p < .001$), and bass drum attack times and snare drum attack times also differed significantly ($p < .05$). No other instruments' attack times differed.

These analyses of the instrument attack profiles in "Superstition" help clarify whether the previously described differences in how the individual instruments contribute to the cumulative, ensemble pocket are *musical* in their nature or a technical consequence of the sounds used in the performance. Although there are indeed some differences in the instruments' attack times, these are almost entirely in clavinet 2. Looking back to the whole track analyses in Figure 5.10 and Tables 5.1 & 5.2, this attack time finding does not convincingly explain why there were differences in how the instruments contribute to the full-band pocket since, if the attack times were noteworthy, then all of the keyboard sounds would have noticeably slower attack phases to their sounds, which is not the case.

Instrument	**Mean**	**SD**	**SE**	**95% CI**
CL1	0.0143	0.00149	0.000472	[0.0132, 0.0153]
CL2	0.0184	0.00488	0.00154	[0.0149, 0.0218]
BS	0.0117	0.000762	0.000241	[0.0112, 0.0122]
HH	0.0142	0.000645	0.000204	[0.0137, 0.0146]
SD	0.0151	0.00121	0.000384	[0.0142, 0.0159]
BD	0.0121	0.000379	0.00012	[0.0118, 0.0124]

TABLE 5.3: Summary statistics for each instrument's attack time (measured in seconds).

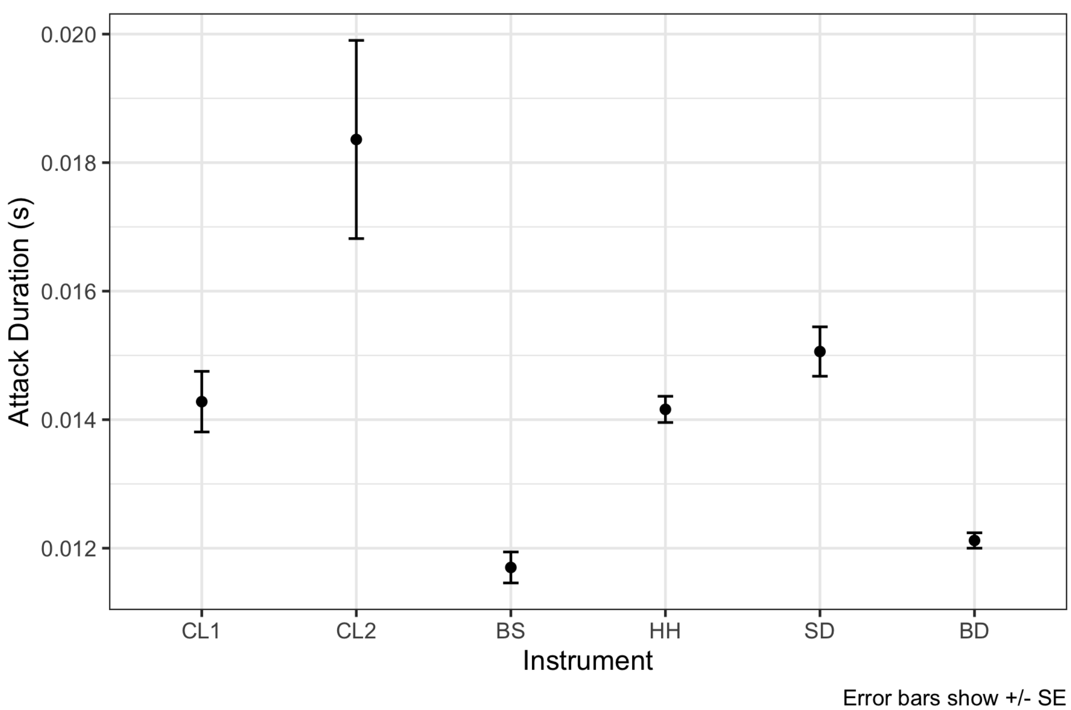


FIGURE 5.12: Mean attack times for each instrument's sound envelope.

Researchers at the RITMO Centre for Interdisciplinary Studies in Rhythm, Time, and Motion have developed the computational methods used above that analyze sonic profiles, and have related this to the heard experience of a sound (Danielsen et al., 2019; Lartillot et al., 2021;

Nymoen et al., 2017).[14] They have demonstrated that attack times are associated with "P-centers"—a term borrowed from phonology to indicate the perceptual center of the sound, understood as the specific moment at which it is perceived to occur (Morton et al., 1976). While it is incorrect to perfectly equate perceived attack regions and P-centers, it has long been established that "The temporal localization of notes largely depends on the characteristics of the attack region (and, more generally, the rising region), [and so] it has been called the 'perceptual attack time' (Gordon, 1987)" (Lartillot et al., 2021, p. 723; also see the discussion on p. 735). While this chapter's focus is on the performance of pockets and not the perceptual beat bins that P-centers are associated with, this investigation of the differences in how the individual instrument onsets sound is an important piece to consider when zooming out from analyzing single onsets and considering how the performance may be perceived.

Conclusion

This chapter draws together the theoretical argument for conceiving of beats as domains introduced in Chapter 2 along with the empirical analyses of drum performances from Chapter 3 to present an analysis of Stevie Wonder's "Superstition" that uses a novel methodology to explore how artists employ dynamically changing pockets across their musical performances. I have shown how different formal sections and subsections of the song have different pocket shapes and how these differences in microtemporal spread feed, alongside compositional and other performance choices like loudness and register, into the overall experience of these sections of a piece. The language of "feel" is especially important here, allowing simple explanation in the vernacular language of pop musicians of the relatively "tight" first two verses and the contrasting "looseness" of the choruses. The empirical quanta of the performance are connected with the qualitative

[14] This work builds on previous research including Bechtold and Senn (2018), Villing (2010), and Wright (2008) (not associated with the RITMO group).

details of the performance.

Alongside these music-analytic findings, I have illustrated a methodology by which full-band recordings may be parsed using a range of music information retrieval tools to extract information about the dimensions and contents of pockets for each metric position of a performance. Using this information, the pockets for the whole track, for each formal section (e.g., verse 1), for each subsection (e.g., the first half of verse 1), for each instrument, and, though not presented here because it is not especially illuminating in this case, for each metric position within a bar may be analyzed.

The experimental work presented in Chapter 4, which asked whether these microtemporal phenomena are even perceivable to listeners, helps contextualize the analysis presented here. Though the performance nuances may be slight, and though Chapter 4 focused exclusively on drum performances so the transferability of these findings to full-band contexts is not fully known, listeners do appear to be sensitive to subtle changes in the shape of pockets. Further, participants in those listening experiments only heard eight bars of music, not the over-100 bars of "Superstition," so they would benefit from acquiring a lot more information about the structure of the pockets in the full song. Even if some of the differences in pocket shape in "Superstition" did blur for some listeners, it is also important to remember how pockets may interact with other musical and perceptual phenomena. For example, the exact differences between onsets within a pocket may not always be precisely discernible, but the differences are likely to impact the resultant timbre and, if all onsets are exactly aligned, it is very hard to perceptually separate the instruments.

Taking the theory of pockets and the analytic methodology developed here forward, there are numerous possible avenues for discovering details about musical beats in performances. For instance, the present analysis of "Superstition" could be further developed by situating it within a corpus of other famous tracks. Such a corpus study would enable the investigation of a more generalizable relationship between the evolution of the shape of a performer or ensemble's pocket and the felt experience of the sections of a song. Additionally, the role of individual instruments or sets of instruments within full-band contexts could be investigated, perhaps contrasting the rhythm section pockets and lead instruments' pockets to explore how a soloist weaves their own sense of time into the sonic fabric of the rest of the band, or considering how a player can be known for "pushing" time within a band or "playing on the edge of the pocket," yet there is no global tempo change.

Moving beyond analytic and academic applications of the pocket, the language of "feel" that arises from and is substantiated by the theory of pockets could be applied practically in the studio (both the recording/production studio and the teaching studio) to communicate and achieve desired performance styles. The analyses and analytic methods presented in this dissertation may be utilized in pedagogical settings to dig deep into the fine details of what performers are doing and, by revealing precisely how, for example, Matt Chamberlain achieves a laid back feel or Stevie Wonder amplifies the felt intensity of a chorus section, facilitating emulation of sought-after performance styles. Or, growing out of this suggestion, the musically sensitive approach to microtiming analysis that the theory of pockets advances could potentially feed into the creation of algorithms that apply microtiming profiles to quantized performances, enhancing "humanizing" algorithms and other plug-ins for Digital Audio Workstations.

Bibliography

Abel, M. (2015). *Groove: An aesthetic of measured time*. Haymarket Books.

Adamo, M. (2013a). In the pocket – Funk drumming training camp part 1: Paradiddles and 16th-note subdivisions. *Modern Drummer*, (402), 68.

Adamo, M. (2013b). In the pocket – Funk drumming training camp part 2: More paradiddles. *Modern Drummer*, (403), 60.

Adamo, M. (2013c). In the pocket – Funk drumming training camp part 3: Originality. *Modern Drummer*, (404), 70–1.

Adamo, M. (2013d). In the pocket – Funk drumming training camp part 4. *Modern Drummer*, (407).

Adamo, M. (2013e). In the pocket – Funk drumming training camp part 5: Paradiddle-based grooves. *Modern Drummer*, (408).

Adamo, M. (2013f). In the pocket – Funk drumming training camp part 6: Changing subdivisions. *Modern Drummer*, (409), 70.

Anon. (2016). How to push or pull drum tracks to increase or reduce urgency. *Computer Music* https://www.musicradar.com/tuition/tech/how-to-push-or-pull-drum-tracks-to-increase-or-reduce-urgency-639436.

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, *48*(1), 66–79. https://doi.org/10.1006/brcg.2001.1304

Ashley, R. (2014). Expressiveness in funk. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 154–169). Oxford University Press.

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 647–667). Academic Press.

Bartneck, C., Duenser, A., Moltchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PLOS ONE*, *10*(4), e0121595. https://doi.org/10.1371/journal.pone.0121595

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

bdrum50. (2008). Yamaha groove night allstars 2. *YouTube* https://www.youtube.com/watch?v=0KizemOpuug.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217–239. https://doi.org/10.1016/0010-0277(95)00669-X

Bechtold, T. A., & Senn, O. (2018). Articulation and dynamics influence the perceptual attack time of saxophone sounds. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01692

Benadon, F. (2006). Slicing the beat: Jazz eighth-notes as expressive microrhythm. *Ethnomusicology*, *50*(1), 73–98.

Bengtsson, I. (1987). Notation, motion and perception: Some aspects of musical rhythm. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 69–80). Royal Swedish Academy of Music.

Bengtsson, I., & Gabrielsson, A. (1974). Empirische Rhythmusforschung in Uppsala [Empirical rhythm research in Uppsala]. *Hamburger Jahrbuch für Musikwissenschaft*, *1*, 195–220.

Bengtsson, I., & Gabrielsson, A. (1977). Rhythm research in Uppsala. *Music, room, and acoustics* (pp. 19–56). Royal Swedish Academy of Music.

Bengtsson, I., & Gabrielsson, A. (1980). Methods for analyzing performance of musical rhythm. *Scandinavian Journal of Psychology*, *21*(1), 257–268. https://doi.org/10.1111/j.1467-9450.1980.tb00369.x

Bengtsson, I., & Gabrielsson, A. (1983). Analysis and synthesis of musical rhythm. In J. Sundberg (Ed.), *Papers given at a seminar organized by the Music Acoustics Committee of the Royal Swedish Academy of Music* (pp. 27–60).

Bengtsson, I., Gabrielsson, A., & Gabrielsson, B. (1978). RHYTHMSYVARD: A computer program for analysis of rhythmic performance. *Swedish Journal of Musicology*, *60*, 15–24.

Bengtsson, I., Gabrielsson, A., & Thorsén, S.-M. (1969). Empirisk Rytmforskning (Empirical rhythm research). *Swedish Journal of Musicology*, *51*, 49–118.

Berliner, P. (1994). *Thinking in jazz: The infinite art of improvisation*. University of Chicago Press.

Bernard "Pretty" Purdie official website. (2021) https://www.bernardprettypurdie.com.

Berry, W. (1989). *Musical structure and performance*. Yale University Press.

Biamonte, N. (2014). Formal functions of metric dissonance in rock music. *Music Theory Online*, *20*(2).

Bilmes, J. (1993). *Timing is of the essence: Perceptual and computational techniques for representing, learning, and reproducing expressive timing in percussive rhythm* (Masters Thesis). Massachusetts Institute of Technology.

Binet, A., & Courtier, J. (1895). Recherches graphiques sur la musique. *L'Année Psychologique*, *2*(1), 201–222. https://doi.org/10.3406/psy.1895.1535

Blom, J.-P., & Kvifte, T. (1986). On the problem of inferential ambivalence in musical meter. *Ethnomusicology*, *30*(3), 491–517. https://doi.org/10.2307/851591

Böck, S., Korzeniowski, F., Schlüter, J., Krebs, F., & Widmer, G. (2016). Madmom: A new python audio and music signal processing library. *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, 1174–1178. https://doi.org/10.1145/2964284.2973795

Boley, J., & Lester, M. (2009). Statistical analysis of ABX results using signal detection theory. *Audio Engineering Society Convention 127*.

Boone, G. M. (2000). Marking mensural time. *Music Theory Spectrum*, *22*(1), 1–43. https://doi.org/10.2307/745851

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252.

Brennan, M. (2020). *Kick it: A social history of the drum kit*. Oxford University Press.

Brughera, A., Dunai, L., & Hartmann, W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. *The Journal of the Acoustical Society of America*, *133*(5), 2839–2855. https://doi.org/10.1121/1.4795778

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Buskirk, E. V. (2008). Radiohead's "Nude" charted due to remix tracks on iTunes. *Wired* https://www.wired.com/2008/04/did-radioheads/.

Butler, M. (2006). *Unlocking the groove: Rhythm, meter, and musical design in electronic dance music*. Indiana University Press.

Butterfield, M. (2006). The power of anacrusis: Engendered feeling in groove-based musics. *Music Theory Online*, *12*(4).

Butterfield, M. (2010). Participatory discrepancies and the perception of beats in jazz. *Music Perception: An Interdisciplinary Journal*, *27*(3), 157–176.

Butterfield, M. (2011). Why do jazz musicians swing their eighth notes? *Music Theory Spectrum*, *33*(1), 3–26.

Câmara, G. S., Nymoen, K., Lartillot, O., & Danielsen, A. (2020a). Effects of instructed timing on electric guitar and bass sound in groove performance. *The Journal of the Acoustical Society of America*, *147*(2), 1028–1041. https://doi.org/10.1121/10.0000724

Câmara, G. S., Nymoen, K., Lartillot, O., & Danielsen, A. (2020b). Timing is everything…or is it? Effects of instructed timing style, reference, and pattern on drum kit sound in groove-based performance. *Music Perception*, *38*(1), 1–26. https://doi.org/10.1525/mp.2020.38.1.1

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Choi, J. E., Moon, I. J., Baek, S.-Y., Kim, S. W., & Cho, Y.-S. (2019). Discrepancies between self-reported hearing difficulty and hearing loss diagnosed by audiometry: Prevalence and associated factors in a national survey. *BMJ Open*, *9*(4), e022440. https://doi.org/10.1136/bmjopen-2018-022440

Chor, I., & Ashley, R. (2006). Synchronization of perceptual onsets of performed bass notes. *Proceedings of the 9th International Conference for Music Perception and Cognition*, 18–21.

Christensen, J., & Bemman, B. (2017). Finding an ideal level of syncopation to elicit a groove response [Poster]. *2017 Biennial Meeting of the Society for Music Perception and Cognition, San Diego*.

Clarke, E. F. (1987a). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 19–33). Royal Swedish Academy of Music.

Clarke, E. F. (1987b). Levels of structure in the organization of musical time. *Contemporary Music Review*, *2*(1), 211–238.

Clarke, E. F. (1995). Expression in performance: Generativity, perception and semiosis. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 21–54). Cambridge University Press.

Clarke, E. F. (2000). Categorical rhythm perception and event perception. In C. Woods, G. Luck, R. Brochard, F. Seddon, & J. Sloboda (Eds.), *Proceedings of the 6th International Conference on Music Perception and Cognition*.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of Music Performance* (pp. 76–181). Royal Swedish Academy of Music.

Clynes, M. (1986). Generative principles of musical thought integration of microstructure with structure. *Journal for the Integrated Study of Artificial Intelligence, Cognitive Science and Applied Epistemology*, *3*(3), 185–223.

Cohn, R. (2021). Prefatory note: How music theorists model time. *Journal of Music Theory*, *65*(1), 11–16. https://doi.org/10.1215/00222909-9124702

Collier, G. L., & Collier, J. L. (1996). Microrhythms in jazz: A review of papers. *Annual Review of Jazz Studies; Newark, NJ*, 117–139.

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, *21*(4), 843–860. https://doi.org/10.3758/s13423-013-0564-3

Cook, N. (2012). Introduction: Refocusing theory. *Music Theory Online*, *18*(1).

Cook, N. (2013). *Beyond the score: Music as performance*. Oxford University Press.

Cook, N., & Pettengill, R. (Eds.). (2013). *Taking it to the bridge: Music as performance*. University of Michigan Press.

Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. University of Chicago Press.

Danielsen, A. (2006). *Presence and pleasure: The funk grooves of James Brown and Parliament*. Wesleyan University Press.

Danielsen, A. (2010a). Here, there and everywhere: Three accounts of pulse in D'Angelo's 'Left and Right'. In A. Danielsen (Ed.), *Musical rhythm in the age of digital reproduction* (pp. 19–35). Ashgate.

Danielsen, A. (2010b). Introduction – Rhythm in the age of digital reproduction. In A. Danielsen (Ed.), *Musical rhythm in the age of digital reproduction* (pp. 1–16). Ashgate.

Danielsen, A. (Ed.). (2010c). *Musical rhythm in the age of digital reproduction*. Ashgate.

Danielsen, A. (2015). Metrical ambiguity or microrhythmic flexibility? Analysing groove in 'Nasty Girl' by Destiny's Child. In R. von Appen, A. Doehring, D. Helms, & A. Moore (Eds.), *Song interpretation in 21st-century pop music* (pp. 53–72). Routledge.

Danielsen, A. (2018). Pulse as dynamic attending: Analysing beat bin metre in neo soul grooves. In C. Scotto, K. M. Smith, & J. Brackett (Eds.), *The Routledge companion to popular music analysis: Expanding approaches* (pp. 179–189). Routledge.

Danielsen, A., Haugen, M. R., & Jensenius, A. R. (2015). Moving to the beat: Studying entrainment to micro-rhythmic changes in pulse by motion capture. *Timing & Time Perception*, *3*(1-2), 133–154. https://doi.org/10.1163/22134468-00002043

Danielsen, A., Nymoen, K., Anderson, E., Câmara, G. S., Langerød, M. T., Thompson, M. R., & London, J. (2019). Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 402–418. https://doi.org/10.1037/xhp0000611

Danielsen, A., Nymoen, K., Langerød, M. T., Jacobsen, E., Johansson, M., & London, J. (in review). Sounds familiar(?) expertise with specific musical genres modulates timing perception and micro-level synchronization to auditory stimuli. *Attention, Perception, & Psychophysics*.

Danielsen, A., Waadeland, C. H., Sundt, H. G., & Witek, M. A. G. (2015). Effects of instructed timing and tempo on snare drum sound in drum kit performance. *The Journal of the Acoustical Society of America*, *138*(4), 2301–2316. https://doi.org/10.1121/1.4930950

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*(2), 103–113.

Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies*, *6*, 279–301.

Davidson, J. W. (1995). What does the visual information contained in music performances offer the observer? Some preliminary thoughts. In R. Steinberg (Ed.), *Music and the Mind Machine* (pp. 105–113). Springer. https://doi.org/10.1007/978-3-642-79327-1_11

Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance:Aa case study approach: *Psychology of Music*, *35*(3), 381–401. https://doi.org/10.1177/0305735607072652

Davies, M., Madison, G., Silva, P., & Gouyon, F. (2013). The effect of microtiming deviations on the perception of groove in short rhythms. *Music Perception: An Interdisciplinary Journal*, *30*(5), 497–510. https://doi.org/10.1525/mp.2013.30.5.497

Davies, S. (2001). *Musical works and performances: A philosophical exploration*. Clarendon Press.

de Clercq, T. (2016). Measuring a measure: Absolute time as a factor for determining bar lengths and meter in pop/rock music. *Music Theory Online*, *22*(3).

de Clercq, T., & Margulis, E. H. (2018). A psychological perspective on repetition in popular music. In Olivier Julien & Christophe Levaux (Eds.), *Over and over: Exploring repetition in popular music* (pp. 147–62). Bloomsbury Academic.

Deane, E. (1995a). Make it funky: Interview with Jabo Starks [Part 1 of 2]. *Rock and Roll [WGBH]* http://openvault.wgbh.org/catalog/V_41C9C2391DD8471794563DE82AAA3A51.

Deane, E. (1995b). Make it funky: Interview with Malcolm Cecil and Robert Margouleff [Part 1 of 2]. *Rock and Roll [WGBH]* https://openvault.wgbh.org/catalog/V_D5482985134149A5927DED9849248EF7.

DeCaro, D. (2017). Groove theory. *Canadian Musician*, *39*(1), 25.

DeFord, R. I. (2015). *Tactus, mensuration and rhythm in Renaissance music*. Cambridge University Press.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, *32*(3), 341–365. https://doi.org/10.1068/p3370

Deveaux, S. (1997). *The birth of bebop*. University of California Press.

Doll, C. (2018). Some practical issues in the aesthetic analysis of popular music. In C. Scotto, K. M. Smith, & J. Brackett (Eds.), *The Routledge companion to popular music analysis: Expanding approaches* (pp. 3–14). Routledge.

Drake, C., & Bertrand, D. (2001). The quest for universals in temporal processing in music. *Annals of the New York Academy of Sciences*, *930*(1), 17–27. https://doi.org/10.1111/j.1749-6632.2001.tb05722.x

Eigeldinger, J.-J. (1986). *Chopin: Pianist and teacher as seen by his pupils* (N. Shohet, Trans.). Cambridge University Press.

Emond, A. (2018). Get Lucky but beats 2 and 4 are swapped. *YouTube* https://www.youtube.com/watch?v=dSvvlu5zTDQ.

Flans, R. (1981). Roger Hawkins: The drums of Muscle Shoals. *Modern Drummer*, (23), 14–17, 58–60, 69, 75, 85, 87.

Flans, R. (1982). Triumvirate: Earth, Wind & Fire's Ralph Johnson, Freddie White, and Philip Bailey. *Modern Drummer*, (30), 22–4, 56–7, 60, 62, 64–6.

Follmer, D. J., Sperling, R. A., & Suen, H. K. (2017). The role of MTurk in education research: Advantages, issues, and future directions. *Educational Researcher*, *46*(6), 329–334. https://doi.org/10.3102/0013189X17725519

Fraisse, P. (1947). De l'assimilation et de la distinction comme processus fondamentaux de la connaissance. *Miscellanea psychologica albert michotte* (pp. 181–195). Institut Supérieur de Philosophie.

Fraisse, P. (1956). *Les structures rythmiques: Étude psychologique*. Editions Unversitaires.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (1st Edition, pp. 149–180). Academic Press.

Friberg, A., & Sundberg, J. (1993). Perception of just-noticeable time displacement of a tone presented in a metrical sequence at different tempos. *STL-QPSR (Speech Transmission Laboratory. Quarterly Progress and Status Reports)*, *34*(2-3), 49–56. https://doi.org/10.1121/1.407650

Friberg, A., & Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception: An Interdisciplinary Journal*, *19*(3), 333–349. https://doi.org/10.1525/mp.2002.19.3.333

Frühauf, J., Kopiez, R., & Platz, F. (2013). Music on the timing grid: The influence of microtiming on the perceived groove quality of a simple drum pattern performance. *Musicae Scientiae*, *17*(2), 246–260. https://doi.org/10.1177/1029864913486793

Fujii, S., Hirashima, M., Kudo, K., Ohtsuki, T., Nakamura, Y., & Oda, S. (2011). Synchronization error of drum kit playing with a metronome at different tempi by professional drummers. *Music Perception*, *28*(5), 491–503. https://doi.org/10.1525/mp.2011.28.5.491

Fujii, S., & Schlaug, G. (2013). The Harvard Beat Assessment Test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00771

Gabrielsson, A. (1982). Perception and performance of musical rhythm. In M. Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music* (pp. 159–169). Springer.

Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (2nd Edition, pp. 501–602). Academic Press.

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology*, *24*(1), 193–213. https://doi.org/10.1111/j.1467-9450.1983.tb00491.x

Ganis, F., Dahl, S., Câmara, G. S., & Danielsen, A. (2021). Beat precision and perceived danceability in drum grooves. *Rhythm Production and Perception Workshop*.

Gardner, J. (2019). "Everybody wants to be a lead player, but if you're a working musician, 95 per cent of it is rhythm" – Cory Wong on the power of rhythm guitar. *Guitar.com | All Things Guitar*.

Garza, J. M., Jr. (2021). Transcending time (feels): Riff types, timekeeping cymbals, and time feels in contemporary metal music. *Music Theory Online*, *27*(1).

Gates, H. L. (1988). *The signifying monkey: A theory of African-American literary criticism*. Oxford University Press.

Gerischer, C. (2006). O suingue baiano: Rhythmic feeling and microrhythmic phenomena in Brazilian percussion. *Ethnomusicology*, *50*(1), 99–119.

Gillick, J., Roberts, A., Engel, J., Eck, D., & Bamman, D. (2019). Learning to groove with inverse sequence transformations. *Proceedings of the 36th International Conference on Machine Learning*.

Gjerdingen, R. O. (1989). Meter as a mode of attending: A network simulation of attentional rhythmicity in music. *Intégral*, *3*, 67–91.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research*, 237–288.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *The Journal of the Acoustical Society of America*, *110*(1), 563–572. https://doi.org/10.1121/1.1376133

Goebl, W., Flossmann, S., & Widmer, G. (2010). Investigations of between-hand synchronization in Magaloff's Chopin. *Computer Music Journal*, *34*(3), 35–44.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Gordon, J. W. (1987). The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*, *82*(1), 88–105. https://doi.org/10.1121/1.395441

Gouyon, F. (2007). Microtiming in "Samba de Roda" —Preliminary experiments with polyphonic audio. *XII Simpósio Da Sociedade Brasileira de Computação Musical (SBCM), São Paulo.*

Gouyon, F., Herrera, P., & Cano, P. (2002). Pulse-dependent analyses of percussive music. *Proceedings of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, 1–6. https://doi.org/10.1109/ICASSP.2002.5745626

Gracyk, T. (1996). *Rhythm and noise: An aesthetics of rock*. Duke University Press Books.

Graff, G. (1992). As the Crowes fly. *Detroit Free Press*, *Detroit*, Section C.

Grant, R. M. (2014). *Beating time and measuring music in the early modern era*. Oxford University Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley & Sons.

Grimes, W. (2011). Marshall Grant, 83, bassist behind Johnny Cash's sound. *The New York Times*.

Groven, E. (1971). Musikkstudiar - ikkje utgjevne før. 1. Rytmestudiar. [Studies in music – previously unpublished. 1. studies in rhythm]. In O. Fjalestad (Ed.), *Eivind Groven. Heiderskrift til 70-årsdagen 8. oktober 1971* (pp. 93–102). Noregs Boklag.

Grube, D. (1996). Verarbeitung akustisch dargebotener Zeitintervalle im Sekundenbereich: Eine Leistung der phonologischen Schleife des Arbeitsgedächtnisses? [Processing of acoustically presented temporal intervals in the second range: An achievement of the phonological loop of working memory?] *Zeitschrift für Experimentelle Psychologie*, *43*, 527–46.

Grube, D. (1998). Die Kapazität des phonetischen Speichers des Arbeitsgedächtnisses als 'auditive Präsenzzeit' und ihr Einfluß auf die Reproduktion von Zeitmustern [The capacity of the phonetic store of working memory as the 'auditory present' and its influence on the reproduction of temporal patterns]. In U. Kotkamp & W. Krause (Eds.), *Intelligente Informationsverarbeitung* (pp. 223–31). Deutscher Universitätsverlag.

Haid, M. (2007a). It's all about time! *Modern Drummer*, *31*(9), 76–80, 82, 85, 87, 89.

Haid, M. (2007b). Remembering Jeff Porcaro and his incredible time feel. *Modern Drummer*, *31*(9), 90–92, 94, 96–98, 100–101, 103–104, 106.

Harnad, S. R. (Ed.). (1987). *Categorical perception: The groundwork of cognition*. Cambridge University Press.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics*, *17*(4), 315–339. https://doi.org/10.2307/1165127

Hasty, C. (1997). *Meter as rhythm*. Oxford University Press.

Haugen, M. R. (2014). Studying rhythmical structures in Norwegian folk music and dance using motion capture technology: A case study of Norwegian Telespringar. *Musikk og tradisjon*, (28), 27–52.

Haugen, M. R., & Danielsen, A. (2020). Effect of tempo on relative note durations in a performed samba groove. *Journal of New Music Research*, *49*(4), 349–361. https://doi.org/10.1080/09298215.2020.1767655

Haugen, M. R., & Godøy, R. (2014). Rhythmical structures in music and body movement in samba performance. In M. Song (Ed.), *Proceedings of the ICMPC-APSCOM 2014 Joint Conference: 13th Biennial International Conference for music perception and Cognition and 5th Triennial Conference of the Asia Pacific Society for the Cognitive Sciences of music* (pp. 46–52).

Hautus, M. J., & Meng, X. (2002). Decision strategies in the ABX (matching-to-sample) psychophysical task. *Perception & Psychophysics*, *64*(1), 89–106. https://doi.org/10.3758/BF03194559

Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: A fast and state-of-the art music source separation tool with pre-trained models. *Journal of Open Source Software*, *5*(50), 2154. https://doi.org/10.21105/joss.02154

Hobbs, J. (2020). Delay & echo: A guide to using time-based audio effects. *LedgerNote*.

Hofmann, A., Wesolowski, B. C., & Goebl, W. (2017). The tight-interlocked rhythm section: Production and perception of synchronisation in jazz trio performance. *Journal of New Music Research*, *46*(4), 329–341. https://doi.org/10.1080/09298215.2017.1355394

Holden, C. J., Dennie, T., & Hicks, A. D. (2013). Assessing the reliability of the M5-120 on Amazon's Mechanical Turk. *Computers in Human Behavior*, *29*(4), 1749–1754. https://doi.org/10.1016/j.chb.2013.02.020

Honing, H. (2001). From time to time: The representation of timing and tempo. *Computer Music Journal*, *25*(3), 50–61.

Honing, H., & Bas de Haas, W. (2008). Swing once more: Relating timing and tempo in expert jazz drumming. *Music Perception*, *25*(5), 471–476.

Hopner, J. (1792). Royal Academy: The annual dinner. *The Observer*, 3.

Hosken, F. (2020). The subjective, human experience of groove: A phenomenological investigation. *Psychology of Music*, *48*(2), 182–198. https://doi.org/10.1177/0305735618792440

Hosken, F., Bechtold, T., Hoesl, F., Kilchenmann, L., & Senn, O. (in press). Drum groove corpora. *Empirical Musicology Review*, *16*(1).

Hove, M. J., Martinez, S. A., & Stupacher, J. (2020). Feel the bass: Music presented to tactile and auditory modalities increases aesthetic appreciation and body movement. *Journal of Experimental Psychology: General*, *149*(6), 1137–1147. https://doi.org/10.1037/xge0000708

Hudson, R. (1994). *Stolen time: The history of tempo rubato*. Oxford University Press.

Hudson, S. S. (2019). *Feeling beats and experiencing motion: A construction-based theory of meter* (Doctoral Dissertation). Northwestern University.

Hudson, S. S. (2022). Bang your head: Construing beat through familiar drum patterns in metal music. *Music Theory Spectrum*, *44*(2).

Hughes, T. (2003). *Groove and flow: Six analytical essays on the music of Stevie Wonder* (Doctoral Dissertation). University of Washington.

Imbrie, A. (1973). 'Extra' measures and metrical ambiguity in Beethoven. In A. Tyson (Ed.), *Beethoven Studies I* (pp. 45–66). Norton.

Ito, J. P. (2004). *Impulse structure in tonal music: A theory of the metrical coordination of motor behavior in performers* (Doctoral Dissertation). Columbia University. New York.

Ito, J. P. (2012). Focal impulses and expressive performance. *From Sounds to Music and Emotions. Proceedings of the International Symposium on Computer Music Modeling and Retrieval*, 480–489.

Ito, J. P. (2021). *Focal impulse theory: Musical expression, meter, and the body*. Indiana University Press.

Iyer, V. (1998). *Microstructures of feel, macrostructures of sound: Embodied cognition in West African and African-American musics* (Doctoral Dissertation). University of California, Berkeley. California.

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception: An Interdisciplinary Journal*, *19*(3), 387–414.

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, *27*(3), 359–370. https://doi.org/10.1016/j.cub.2016.12.031

Jakubowski, K., Polak, R., & Jacoby, N. (2018). Aesthetic preferences and perceptual discrimination of microrhythmic variations in music: A comparative study across three cultures. *ICMPC15/ESCOM10*.

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of experimental psychology: General*, *141*(1), 54–75. https://doi.org/10.1037/a0024208

Jehan, T. (2007). *Creating music by listening* (Doctoral Dissertation). Massachusetts Institute of Technology. Cambridge, Mass.

Johansson, M. (2010). The concept of rhythmic tolerance: Examining flexible grooves in Scandinavian folk fiddling. In A. Danielsen (Ed.), *Musical rhythm in the age of digital reproduction* (pp. 69–83). Ashgate.

Jones, M. R. (1981). Only time can tell: On the topology of mental space and time. *Critical Inquiry*, *7*(3), 557–576.

Jones, M. R. (1986). Attentional rhythmicity in human perception. In J. R. Evans & M. Clynes (Eds.), *Rhythm in psychological, linguistic and musical processes* (pp. 13–40). Charles C. Thomas.

Jones, M. R. (2019). *Time will tell: A theory of dynamic attending*. Oxford University Press.

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

Kane, B. (2018). Jazz, mediation, ontology. *Contemporary Music Review*, *37*(5-6), 507–528. https://doi.org/10.1080/07494467.2017.1402466

Keightley, K. (2012). Tin pan allegory. *Modernism/modernity*, *19*(4), 717–736. https://doi.org/10.1353/mod.2012.0081

Keil, C. (1966). Motion and feeling through music. *The Journal of Aesthetics and Art Criticism*, *24*(3), 337–349.

Keil, C. (1987). Participatory discrepancies and the power of music. *Cultural Anthropology*, *2*(3), 275–283.

Kilchenmann, L., & Senn, O. (2011). "Play in time, but don't play time": Analyzing timing profiles in drum performances. *Proceedings of the International Symposium on Performance Science*, 593–598.

King, J. (2013). The time is out of joint [Excerpts from liner notes to Light in the Attic's 2012 vinyl reissue of Voodoo]. *Slate* http://www.slate.com/articles/arts/music_box/2013/02/behind_the_scenes_with_ questlove_and_d_angelo_on_voodoo.html.

Knoblauch, K. (2020). Psyphy v0.2-2 [https://CRAN.R-project.org/package=psyphy].

Komar, A. J. (1971). *Theory of suspensions: A study of metrical and pitch relations in tonal music*. Princeton University Press.

Kozak, M. (2020). *Enacting musical time: The bodily experience of new music*. Oxford University Press.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Krauss, R. (1979). Grids. *October*, *9*, 51–64. https://doi.org/10.2307/778321

Krebs, H. (1999). *Fantasy pieces: Metrical dissonance in the music of Robert Schumann*. Oxford University Press.

Kvifte, T. (1989). *Instruments and the electronic age*. Solum Forlag.

Kvifte, T. (2004). Description of grooves and syntax/process dialectics. *Studia Musicologica Norvegica*, *30*, 554–77.

Kvifte, T. (2007). Categories and timing: On the perception of meter. *Ethnomusicology*, *51*(1), 64–84.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159.

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal stm by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1249–1268. https://doi.org/10.1080/02724980244000765

Lartillot, O. (2019). MiningSuite: A comprehensive Matlab framework for signal, audio and music analysis, articulating audio and symbolic approaches. *Proceedings of the Society for Sound and Music Computing Conference*, 489–489 http://olivierlar.github.io/miningsuite.

Lartillot, O., Nymoen, K., Câmara, G. S., & Danielsen, A. (2021). Computational localization of attack regions through a direct observation of the audio waveform. *The Journal of the Acoustical Society of America*, *149*(1), 723–736. https://doi.org/10.1121/10.0003374

Lartillot, O., Toiviainen, P., & Eerola, T. (2008a). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Springer Berlin Heidelberg.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008b). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications: Proceedings of the 31st Annual Conference* *of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7-9, 2007* (pp. 261–8). Springer-Verlag.

Leech-Wilkinson, D. (2006). Portamento and musical meaning. *Journal of Musicological Research*, *25*(3-4), 233–261. https://doi.org/10.1080/01411890600859412

Leech-Wilkinson, D. (2007). Sound and meaning in recordings of Schubert's "Die Junge Nonne". *Musicae Scientiae*, *11*(2), 209–236. https://doi.org/10.1177/102986490701100204

Leech-Wilkinson, D. (2009). *The changing sound of music: Approaches to studying recorded musical performances*. CHARM.

Leech-Wilkinson, D. (2012). Compositions, scores, performances, meanings. *Music Theory Online*, *18*(1).

Lent, C. (1983). Funk. *Rock School, Ep. 6 [Educational Broadcasting Corporation]*.

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

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lindsay, K., & Nordquist, P. (2007). Pulse and swing: Quantitative analysis of hierarchical structure in swing rhythm. *The Journal of the Acoustical Society of America*, *122*(5), 2945–2945. https://doi.org/10.1121/1.2942485

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *24*(3), 732–753. https://doi.org/10.1037//0278-7393.24.3.732

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "F" test. *Review of Educational Research*, *66*(4), 579–619. https://doi.org/10.2307/1170654

London, J. (2012). *Hearing in time: Psychological aspects of musical meter* (2nd Edition). Oxford University Press.

London, J., Nymoen, K., Langerød, M. T., Thompson, M. R., Code, D. L., & Danielsen, A. (2019). A comparison of methods for investigating the perceptual center of musical sounds. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-019-01747-y

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd Edition). Psychology Press.

Madison, G. (2001). Different kinds of groove in jazz and dance music as indicated by listeners' ratings. *Proceedings of the VII International Symposium on Systematic and Comparative Musicology and III International Conference on Cognitive Musicology*, 108–112.

Madison, G. (2004). Detection of linear temporal drift in sound sequences: Empirical data and modelling principles. *Acta Psychologica*, *117*(1), 95–118. https://doi.org/10.1016/j.actpsy.2004.05.004

Madison, G. (2006). Experiencing groove induced by music: Consistency and phenomenology. *Music Perception*, *24*(2), 201–208.

Matsushita, S., & Nomura, S. (2016). The asymmetrical influence of timing asynchrony of bass guitar and drum sounds on groove. *Music Perception: An Interdisciplinary Journal*, *34*(2), 123–131. https://doi.org/10.1525/mp.2016.34.2.123

Matthews, T. E., Thibodeau, J. N. L., Gunther, B. P., & Penhune, V. B. (2016). The impact of instrument-specific musical training on rhythm perception and production. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00069

McCredie, M. N., & Morey, L. C. (2019). Who are the Turkers? A characterization of MTurk workers using the personality assessment inventory. *Assessment*, *26*(5), 759–766. https://doi.org/10.1177/1073191118760709

McPherson, M. J., & McDermott, J. H. (2020). Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proceedings of the National Academy of Sciences*, *117*(50), 32169–32180. https://doi.org/10.1073/pnas.2008956117

Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.

Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. University of Chicago Press.

Moelants, D. (2011). The performance of notes inégales: The influence of tempo, musical structure, and individual performance style on expressive timing. *Music Perception*, *28*(5), 449–460. https://doi.org/10.1525/mp.2011.28.5.449

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*(5), 405–408. https://doi.org/10.1037/0033-295X.83.5.405

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089642

Naveda, L., Gouyon, F., Guedes, C., & Leman, M. (2011). Microtiming patterns and interactions with musical properties in samba music. *Journal of New Music Research*, *40*(3), 225–238. https://doi.org/10.1080/09298215.2011.603833

Naveda, L., & Leman, M. (2009). Accessing structure of samba rhythms through cultural practices of vocal percussions. *Proceedings of the 6th Sound and Music Computing Conference*, 259–264.

Nixon, B. (1981). Charles Collins: Up close. *Modern Drummer*, (26), 48–50.

Nymoen, K., Danielsen, A., & London, J. (2017). Validating attack phase descriptors obtained by the timbre toolbox and mirtoolbox. In T. Lokki, J. Pätynen, & V. Välimäki (Eds.), *Proceedings of the 14th Sound and Music Computing Conference* (pp. 214–9).

Ohriner, M. S. (2019). *Flow: The rhythmic voice in rap music*. Oxford University Press.

Osborn, B. (2010). *Beyond verse and chorus: Experimental formal structures in post-millennial rock music* (Doctoral Dissertation).

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception & Performance*, *15*(12), 331–346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception: An Interdisciplinary Journal*, *14*(1), 23–56. https://doi.org/10.2307/40285708

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Pearson, D. M. (2019). Extreme hardcore punk and the analytical challenges of rhythm, riffs, and timbre in punk music. *Music Theory Online*, *25*(1).

Pfleiderer, M. (2010). Dimensionen der Groove-Erfahrung. Eine empirische Studie. *PopScriptum, 11 – The Groove Issue*.

Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic prototypes across cultures: A comparative study of tapping synchronization. *Music Perception: An Interdisciplinary Journal*, *36*(1), 1–23. https://doi.org/10.1525/mp.2018.36.1.1

Potter, J. (2012). Bernard Purdie: The hitmaker. *Modern Drummer*, *36*(10), 56–57.

Pressing, J. (1987). The micro- and macrostructural design of improvised music. *Music Perception: An Interdisciplinary Journal*, *5*(2), 133–172. https://doi.org/10.2307/40285390

Prögler, J. (1995). Searching for swing: Participatory discrepancies in the jazz rhythm section. *Ethnomusicology*, *39*(1), 21–54.

Ramsey, G. P. (2003). *Race music: Black cultures from bebop to hip-hop*. University of California Press.

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica*, *43*(2), 121–131.

Reinholdsson, P. (1987). Approaching jazz performances empirically: Some reflections on methods and problems. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 105–25). Royal Swedish Academy of Music.

Repp, B. H. (1989). Expressive microstructure in music: A preliminary perceptual assessment of four composers' "pulses". *Music Perception: An Interdisciplinary Journal*, *6*(3), 243–273. https://doi.org/10.2307/40285589

Repp, B. H. (1990a). Composers' pulses: Science or art? *Music Perception: An Interdisciplinary Journal*, *7*(4), 423–434. https://doi.org/10.2307/40285477

Repp, B. H. (1990b). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists and one computer. *The Journal of the Acoustical Society of America*, *83*, 622–41. https://doi.org/10.1121/1.2025195

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

Repp, B. H. (1995). Quantitative effects of global tempo on expressive timing in music performance: Some perceptual evidence. *Music Perception*, *13*(1), 39–57. https://doi.org/10.2307/40285684

Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *The Journal of the Acoustical Society of America*, *100*(6), 3917–3932. https://doi.org/10.1121/1.417245

Repp, B. H. (2000). The timing implications of musical structures. In D. Greer, I. Rumbold, & J. King (Eds.), *Musicology and Sister Disciplines: Past, Present, Future – Proceedings of the 16th International Congress of the International Musicological Society, London 1997* (pp. 60–70).

Repp, B. H. (2004). On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences. *Human Movement Science*, *23*(3-4), 389–413. https://doi.org/10.1016/j.humov.2004.08.014

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. https://doi.org/10.3758/BF03206433

Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: A comparison of musicians and nonmusicians. *Music Perception*, *24*(4), 367–376. https://doi.org/10.1525/mp.2007.24.4.367

Repp, B. H., Luke Windsor, W., & Desain, P. (2002). Effects of tempo on the timing of simple musical rhythms. *Music Perception*, *19*(4), 565–593. https://doi.org/10.1525/mp.2002.19.4.565

Rice, W. (1967). Trumpeter meets match at blues alley: Cellar door. *The Washington Post, Times Herald (1959-1973)*, C11.

Rink, J. (2001). The line of argument in Chopin's E Minor prelude. *Early Music*, *29*(3), 435–444.

Rink, J. (Ed.). (2002). *Musical performance: A guide to understanding*. Cambridge University Press.

Rolling Stone. (2003). 500 greatest songs of all time. *Rolling Stone*.

Ross, J. M., Warlaumont, A. S., Abney, D. H., Rigoli, L. M., & Balasubramaniam, R. (2016). Influence of musical groove on postural sway. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 308–319. https://doi.org/10.1037/xhp0000198

Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior*, *43*, 304–307. https://doi.org/10.1016/j.chb.2014.11.004

Saito, S. (2001). The phonological loop and memory for rhythms: An individual differences approach. *Memory*, *9*(4-6), 313–322. https://doi.org/10.1080/09658210143000164

Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: Effects of concurrent articulation on reproduction of rhythms. *Japanese Psychological Research*, *40*(1), 10–18. https://doi.org/10.1111/1468-5884.00070

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, *48*(8), 1429–1441. https://doi.org/10.3758/s13421-020-01060-2

Sapp, C. (2006). Mazurka Project. *CHARM Mazurka Project*
http://mazurka.org.uk/.

Sawata, R., Uhlich, S., Takahashi, S., & Mitsufuji, Y. (2021). All for one and one for all: Improving music separation by bridging networks. *arXiv:2010.04228 [cs, eess]*.

Schulze, H.-H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, *51*(1), 10–15. https://doi.org/10.1007/BF00309270

Schwartz, C. (1985). The beating heart. *Electronic Soundmaker & Computer Music*, (May 1985), 11–12.

Schwartz, D. (2019). The timeless appeal of J Dilla: Hip-hop's biggest innovator. *Highsnobiety*.

Seashore, C. E. (1938). *Psychology of music*. McGraw-Hill.

Seashore, H. G. (1936). An objective analysis of artistic singing. In C. E. Seashore (Ed.), *Objective analysis of musical performance* (pp. 12–157). University of Iowa Press.

Senn, O., Bechtold, T., Rose, D., Câmara, G. S., Düvel, N., Jerjen, R., Kilchenmann, L., Hoesl, F., Baldassarre, A., & Alessandri, E. (2020). Experience of groove questionnaire: Instrument development and initial validation. *Music Perception*, *38*(1), 46–65. https://doi.org/10.1525/mp.2020.38.1.46

Senn, O., Kilchenmann, L., Bechtold, T., & Hoesl, F. (2018). Groove in drum patterns as a function of both rhythmic properties and listeners' attitudes. *PLOS ONE*, *13*(6), e0199604. https://doi.org/10.1371/journal.pone.0199604

Senn, O., Kilchenmann, L., & Camp, M.-A. (2009). Expressive timing: Martha Argerich plays Chopin's Prelude Op. 28/4 in E Minor. *Proceedings of the International Symposium on Performance Science*, 107–12.

Senn, O., Kilchenmann, L., von Georgi, R., & Bullerjahn, C. (2016). The effect of expert performance microtiming on listeners' experience of groove in swing or funk music. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01487

Seppanen, J. (2001). Tatum grid analysis of musical signals. *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 131–134. https://doi.org/10.1109/ASPAA.2001.969560

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology*, *13*(3), 326–376. https://doi.org/10.1016/0010-0285(81)90013-X

Sheila E. (2020). Playing drums in a band. *MasterClass* https://www.masterclass.com/classes/sheila-e-teaches-drumming-and-percussion/.

Sioros, G., Câmara, G. S., & Danielsen, A. (2019). Mapping timing strategies in drum performance. *Proceedings of the 20th International Society for Music Information Retrieval Conference*, 776–83. https://doi.org/10.5281/zenodo.3527926

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception: An Interdisciplinary Journal*, *19*(1), 87–120. https://doi.org/10.1525/mp.2001.19.1.87

Smith, R. J. (2012). *The One: The life and music of James Brown*. Gotham Books.

Smith, Y., & Walsh, K. (1996). Make it funky. *Dancing in the Street, Ep 9 [BBC]*.

Song, X., Kong, Q., Du, X., & Wang, Y. (2021). Catnet: Music source separation system with mix-audio augmentation. *arXiv:2102.09966 [cs, eess]*.

Southall, C., Stables, R., & Hockman, J. (2017). Automatic drum transcription for polyphonic recordings using soft attention mechanisms and convolutional neural networks. *18th International Society for Music Information Retrieval Conference, Suzhou, China*, 7.

Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics*, *65*(2), 318–328. https://doi.org/10.3758/BF03194803

Stewart, M. (1987). The feel factor: Music with soul. *Electronic Musician*, *3*(1), 57–65.

Stöter, F.-R., Uhlich, S., Liutkus, A., & Mitsufuji, Y. (2019). Open-Unmix: A reference implementation for music source separation. *Journal of Open Source Software*, *4*(41), 1667. https://doi.org/10.21105/joss.01667

Stupacher, J., Hove, M. J., & Janata, P. (2016). Audio features underlying perceived groove and sensorimotor synchronization in music. *Music Perception: An Interdisciplinary Journal*, *33*(5), 571–589. https://doi.org/10.1525/mp.2016.33.5.571

Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., & Keller, P. E. (2013). Musical groove modulates motor cortex excitability: A TMS investigation. *Brain and Cognition*, *82*(2), 127–136. https://doi.org/10.1016/j.bandc.2013.03.003

Stupacher, J., Wood, G., & Witte, M. (2017). Synchrony and sympathy: Social entrainment with music compared to a metronome. *Psychomusicology: Music, Mind, and Brain*, *27*(3), 158–166. https://doi.org/10.1037/pmu0000181

Sturgis, M. (2021). Groove, pocket, and feel. *Rhythm*, (300), 94–6.

Sullivan, J. (1984). Neil Young: So long, rock; hello, country. *Boston Globe*, A11–12.

Sum, M. (2012). *Music of the Gnawa of Morocco: Evolving spaces and times* (Doctoral Dissertation). University of British Columbia. Vancouver. https://doi.org/10.14288/1.0058249

Théberge, P. (2016). Click/beat/body: Thoughts on the materiality of time and tempo. In M. Bull & L. Back (Eds.), *The auditory culture reader* (2nd edition). Routledge.

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, *16*(1), 19–40. https://doi.org/10.1177/1029864911423457

Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron*, *62*(4), 463–469. https://doi.org/10.1016/j.neuron.2009.05.002

Underwood, B. (1978). Profile: Bobby Lyle. *DownBeat Magazine*, *45*(1), 40.

Vernon, L. N. (1936). Synchronization of chords in artistic piano music. In H. G. Seashore (Ed.), *Objective analysis of musical performance* (pp. 306–345). University of Iowa Press.

Villing, R. (2010). *Hearing the moment: Measures and models of the perceptual centre* (Ph.D.). National University of Ireland Maynooth.

Vuoskoski, J. K., Thompson, M. R., Clarke, E., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics*, *76*(2), 591–604. https://doi.org/10.3758/s13414-013-0582-2

Waadeland, C. H. (2001). "It Don't Mean a Thing If It Ain't Got That Swing": Simulating expressive timing by modulated movements. *Journal of New Music Research*, *30*(1), 23–37.

Willman, C. (1990). Pop music review: Stansfield versatile as a vocalist. *Los Angeles Times*, F1.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review*, *8*(1), 44–57. https://doi.org/10.3758/BF03196138

Winston, E., & Saywood, L. (2019). Beats to relax/study to: Contradiction and paradox in lofi hip hop. *IASPM Journal*, *9*(2), 40–54.

Witek, M. A. G., Liu, J., Kuubertzie, J., Yankyera, A. P., Adzei, S., & Vuust, P. (2020). A critical cross-cultural study of sensorimotor and groove responses to syncopation among Ghanaian and American university students and staff. *Music Perception*, *37*(4), 278–297. https://doi.org/10.1525/mp.2020.37.4.278

Wolk, D. (2011). Is 'Superstition' one of the All-TIME 100 Best Songs? *Time*.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *The Journal of the Acoustical Society of America*, *60*(6), 1381–1389. https://doi.org/10.1121/1.381231

Wright, M. J. (2008). *The shape of an instant: Measuring and modeling perceptual attack time with probability density functions (If a tree falls in the forest, when did 57 people hear it make a sound?)* (Doctoral Dissertation). Stanford University. Stanford, CA.

Wu, C.-W., Dittmar, C., Southall, C., Vogl, R., Widmer, G., Hockman, J., Muller, M., & Lerch, A. (2018). A review of automatic drum transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *26*(9), 1457–1483. https://doi.org/10.1109/TASLP.2018.2830113

Yeston, M. (1974). *The stratification of musical rhythm*. Yale University Press.

Zampini, M., Brown, T., Shore, D. I., Maravita, A., Röder, B., & Spence, C. (2005). Audiotactile temporal order judgments. *Acta Psychologica*, *118*(3), 277–291. https://doi.org/10.1016/j.actpsy.2004.10.017

Zampini, M., Shore, D. I., & Spence, C. (2003). Multisensory temporal order judgments: The role of hemispheric redundancy. *International Journal of Psychophysiology*, *50*(1), 165–180. https://doi.org/10.1016/S0167-8760(03)00132-6

Zbikowski, L. (2004). Modelling the groove: Conceptual structure and popular music. *Journal of the Royal Musical Association*, *129*(2), 272–297. https://doi.org/10.1093/jrma/129.2.272