Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music

Author(s): Vijay Iyer

Source: *Music Perception: An Interdisciplinary Journal* , Vol. 19, No. 3 (Spring 2002), pp. 387-414

Published by: University of California Press

Stable URL: https://www.jstor.org/stable/10.1525/mp.2002.19.3.387

Music Perception
Spring 2002, Vol. 19, No. 3, 387–414

Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music

VIJAY IYER
New York City

The dual theories of embodied mind and situated cognition, in which physical/temporal embodiment and physical/social/cultural environment contribute crucially to the structure of mind, are brought to bear on issues in music perception. It is argued that cognitive universals grounded in human bodily experience are tempered by the cultural specificity that constructs the role of the body in musical performance. Special focus is given to microrhythmic techniques in specific forms of African-American music, using audio examples created by the author or sampled from well-known jazz recordings.

A great majority of the research in music perception and cognition has focused on a rather narrow segment of human musical phenomena, namely the tonal concert music of pre-20th-century Western Europe, as filtered through contemporary European-derived performance practices. Hence we have an abundance of tonal-music-inspired models and representations for perceptual and cognitive phenomena, focusing mostly on pitch organization in the large-scale time domain. Some well-known examples are theories of recursive formal hierarchies (Lerdahl & Jackendoff, 1983) and of musical meaning from deferred melodic or harmonic expectations (Meyer, 1956; Narmour, 1990). Lerdahl and Jackendoff contend that in the way that information-processing stages are organized, musical cognition is fully analogous to linguistic cognition. Such models suppose that the cognition of music consists of the logical parsing of recursive tree structures to reveal greater and greater levels of hierarchical organization. However, because so much musical behavior is nonlinguistic in nature, music tends to challenge dominant linguistic paradigms, which reduce all cognition to rational thought processes such as problem solving, deductive reasoning, and inference. With its emotional and associative qualities and its connection

Address correspondence (including requests for audio examples) to Vijay Iyer, 606 W. 116th St., #2, New York, NY 10027. (e-mail: vijay@vijay-iyer.com)
ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

to dance and ritual, music seems to provide a counterexample to such theories of mind.

While quite far-reaching in the case of Western tonal music, 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. Although groove is a highly subjective quality, music that grooves can sustain interest or attention for long stretches of time to an acculturated listener, even if "nothing is happening" on the musical surface. A prime example is James Brown's (1991) music (audio ex. 1),[1] which frequently has precious little melodic or harmonic material and is highly repetitive, but would never be described as static. The fact that groove carries enough weight to override other musical factors in certain kinds of musical experience suggests that the traditional linguistics-based viewpoint does not suffice in describing the entirety of music cognition.

This mismatch between tonal-music grammars and most musics of the world cannot be ascribed to relative levels of musical sophistication or complexity. More likely, one must account for major cultural disparities in approaches to the organization, production, and cognition of music. In this article, I claim that an essential component of these disparities is the status of the body and physical movement in the act of making music. Although every music listener has a body (from which fact one could derive hypothetical universals of music), every culture "constructs" the human body differently (thereby tempering any claims to universality). The role of the body in various musics of the world becomes clearer when one observes the functions that music and dance assume in these cultures. Later, I address the roles of physical embodiment and sociocultural situatedness in music perception and cognition, developing flexible concepts that complement and supplement the linguistic theories just mentioned. After summarizing some key concepts from the theory of embodiment, I employ these concepts to address the issue of expressive microtiming in African-American music.

Embodied and Situated Cognition

Much can be said of the role of the body in cognition in general. Recent conceptual developments in cognitive science move toward the inclusion of the body in our understanding of the mind. In particular, cognitive scientists have begun to infer connections between the structure of mental processes and physical embodiment (e.g., Lakoff & Johnson, 1999). The viewpoint known as *embodied* or *situated* cognition treats cognition as an activity

1. Audio examples may be obtained by contacting the author.

that is structured by the body situated in its environment—that is, as embodied action. In this view, cognition depends upon experiences based in having a body with sensorimotor capacities; these capacities are embedded in an encompassing biological, psychological, and cultural context. Sensory processes (perception) and motor processes (action), having evolved together, are seen therefore as fundamentally inseparable, mutually informative, and structured so as to ground our conceptual systems. (Varela, Thompson, & Rosch, 1991, p. 173)

EMBODIMENT

The embodiment hypothesis suggests an alternative basis for cognitive processes. Perception is understood as *perceptually guided action*. We explore our environments with our bodies and our senses, learning to correlate multisensory input with our bodily experience. Such behavior is facilitated through elaborate feedback mechanisms among sensory and motor apparatus. The temporal information in our sensory input is matched to the motor image of the body in the sensorimotor loop (Todd, 1999). Hence, cognitive structures emerge from the recurrent sensorimotor patterns that enable the perceiver to guide his or her actions in the local situation; the emergent, learned neural connections between the senses and the motor system form the basis for cognition. The mind's embodiment provides natural biases for inductive models and representations, and thus automatically grounds cognitive processes that might normally be considered disembodied (Varela et al., 1991). This view provides a sharp contrast from the standard information-processing viewpoint, in which cognition is seen as a problem of recovering details of the pregiven outer world (e.g., Pylyshyn, 1984). In the embodied viewpoint, the mind is no longer seen as passively reflective of the outside world, but rather as an active constructor of its own reality. In particular, cognition and bodily activity intertwine to a high degree. In this perspective, the fundamental building blocks of cognitive processes are control schemata for motor patterns that arise from perceptual interaction with the body's environment. The drives for the cognitive system arise from within the system itself, in the form of needs and goals (Prem, 1996).

Neuroscientific evidence corroborates this viewpoint. We can make more sense of our brains and bodies if we view the nervous system as a system for producing motor output. The cerebellum is connected almost directly to all areas of the brain—sensory transmissions, reticular (arousal/attention) systems, hippocampus (episodic memories), limbic system (emotions, behavior). All areas of our brain seem geared to coping with their functions as they pertain to problems of motor control (Hardcastle, 1996, p. 7). Such evidence from neuroscience allows for postulating shared mechanisms for low-level control of embodied action and higher-level cognition; motor plans for limb movement could interact with goal-oriented abstract plans. The mind thereby becomes a *distributed* entity, an emergent characteristic of the whole sensory-central-motor neural system, existing in the elaborate network of interconnections that extend throughout the body.

SITUATEDNESS

The above characterization of the embodied mind covers half of the picture. If we grant that cognition is structured at least to some degree by bodily experience, then we must understand the body to be immersed in an environment that shapes its experience. Hence the cognitive theory of embodiment also stresses temporal, physical, and sociocultural situatedness. The body and its environment not only provide constraints but also *enable* cognition. Work in animal behavior has addressed the potential links between sensory and motor systems, as in the classic experiment by Held and Hein (1958). In this study, a group of kittens were raised in the dark and exposed to light only under controlled conditions. A first group of animals was allowed to move around normally, but each of them had to pull around a cart in which rode a member of the second group. The two groups thus shared nearly identical visual experience, but one group experienced the world actively and the other passively. Upon release after a few weeks, the first group of kittens behaved normally, but the second group behaved as if blind, bumping into objects and falling over edges. Hence objects in the world are apprehended not simply by visual extraction of features, but rather by the visual guidance of action (Varela et al., 1991, p. 175). In a more humane fashion, it has been observed that infants who can walk have qualitatively different reactions to certain stimuli, such as slopes and falloffs, than preambulatory infants (Thelen & Smith, 1994, pp. 217–220).

There is reason to believe that such programmability of the brain extends far beyond childhood. Although it is commonly known that cognitive development proceeds rapidly along with brain growth and cortical myelinization during the crucial first years of life (Passingham, 1982, pp. 112ff), it is not often recognized that many networks of the brain retain a susceptibility to reprogramming throughout an individual's life (Laughlin, McManus, & D'Aquili, 1992, p. 41). The remarkable case studies of Ramachandran and Blakeslee (1998) and Sacks (1985) further attest to the adaptability and plasticity of the brain throughout adulthood. Hence we may discern a continuum of neural structures ranging from "hard-wired" evolutionary traits to highly flexible, environmentally adaptive features (Shore, 1996, p. 17). The existence of this continuum supports the embodied cognitive paradigm, which encompasses the body and its environment.

In addition to the universals of cognition based upon having a body, its sensorimotor systems, and its physical environment, researchers study the particular social and cultural factors that contribute to the development of mind. Cognition is seen in part as a social phenomenon, distributed over mind, body, activity and cultural context. We rely upon various attributes of our physical, social, and cultural environment to support or augment our mental capacities. Lave (1988) studied the arithmetic used by adults of various backgrounds while grocery shopping. Lave's results showed that in making purchase decisions, these situated agents employed a flexible real-time arithmetic in order to select better prices per unit weight, continually taking into account the constraints imposed by the layout of the stores, the capacities of their home refrigerators, and the dietary requirements of their family members. Such shopping prowess—skill at situated arithmetic—was rarely reflected in subjects' performance on grade-school math problems; cognition as demonstrated in *practice* was found to differ significantly from cognition in an abstract, desituated setting. In a related formulation, Clark (1997, pp. 213–216) discusses the ambiguity of the boundary between "mind" and "world," along the lines of a distinction between "user" and "tool." When a bird drops a nut from a great height to crack it open, does the ground become a tool (p. 214)? Rather, the bird is exploiting an aspect of its environment to extend its physical capabilities; the concept of a tool dissolves. Hence the mind may be viewed as symbiotically embedded not only within its body but also in that body's environment, and structured by its surroundings.

In sum, the theory of embodiment encompasses both neuropsychological and socioenvironmental views of cognition. Embodied cognition stresses physical, temporal, and functional situatedness, and enforces interaction between the agent's body and its environment. Such a holistic view prevents oversimplifications or unrealistic assumptions, because it provides specific grounding in reality (Mataric, 1996). And, quite significantly, the embodied view of cognitive science allows for direct cultural interaction, which is undeniably crucial for both language and music.

Embodied Music Perception

How might we connect the theoretical framework of embodied cognition with the study of music? First, we should examine the role of the body in music perception, cognition, and production, and attempt to take into account the realities of our perceptual systems. Let us address connections between aspects of musical time—rhythm, timing, meter, phrasing—and the body. One often speaks of a musical groove as something that induces motion. In describing his aesthetic criteria for rhythm tracks, a colleague involved in hip-hop music distinguished between a musical excerpt that "makes me bob my head" and one that doesn't (M. Bilal, personal communication, 1997). Many of us have witnessed motion induced in infants or toddlers via music, but this behavior is not universal, involuntary, or even reliable. This capacity to entrain to a regular aural pulse may be an evolutionary vestige of a previously useful ability that has more recently fallen into disuse. In any case, this phenomenon clearly involves regular, rhythmic bodily movement as a kind of sympathetic reaction to regular rhythmic sound—that is, as a kind of dance. The very ubiquity of dance and music in world cultures (Brown, 1991) attests to the intimate connection between dance and rhythm perception.

Recent neuropsychological studies of music perception have affirmed the cognitive role of body motion in music perception and production. From a meta-analysis of studies of brain-damaged patients with lesions localized in various regions of the brain, it was suggested that the "rhythmic component ... of an auditory image cannot be activated without recruiting neural systems known to be involved in motor activity, especially those involved in the planning of motor sequences" (Carroll-Phelan & Hampson, 1996; see also Peretz, 1993). Such neuropsychological data have allowed hypotheses about the induction of a sense of beat or pulse in terms of the so-called sensorimotor loop, which includes the posterior parietal lobe, premotor cortex, cerebrocerebellum, and basal ganglia. In the sensorimotor perspective, a perceived rhythm is literally an imagined movement: "[I]f the spatiotemporal form of certain [sensory] stimuli are matched to the dynamics of the motor system, then they may evoke a motion of an internal representation, or motor image, of the corresponding synergetic elements of the musculoskeletal system, *even if the musculoskeletal system itself does not move*" (Todd, 1999 [emphasis in original]; see also Todd, Lee, & O'Boyle, 1999). Hence, the act of listening to rhythmic music involves the same mental processes that generate bodily motion.

One might suppose that musical gestures might be more efficacious in eliciting such sympathetic behavior if they represent aspects of human motion somehow. For example, musical phenomena might evoke the dynamic swells associated with breathing, the steady pulse associated with walking, and the rapid rhythmic figurations associated with speech. Note that each of these three examples occurs at a different timescale; characteristic frequencies of the first regime might fall in the range of 0.1 to 1 Hz, the second 1 to 3 Hz, the third 3 to 10 Hz. In fact, it is interesting to observe the correspondences in frequency range in these groups of behaviors:

Versions of Table 1 were suggested by Fraisse (1982) and Todd (1994). Because of the adaptability and variability of the human body itself, these regimes should be regarded as fuzzy categories with some degree of overlap. (The last row of Table 1 is discussed in the section on expressive microtiming.) It is a plausible hypothesis that musical activity on these timescales or "channels" might exploit these correspondences. A variety of simple truisms support this view. Most wind-instrument phrase lengths are

TABLE 1
Musical Correlates to Bodily Motions

Body Motion	Musical Correlate	Approximate Frequency Range (Hz)
Breathing, moderate arm gesture, body sway	Musical phrase	0.1–1
Heartbeat, sucking, chewing, walking, sexual intercourse, head nod	Musical pulse (*tactus*)	1–3
Speech/lingual motion, hand gesture, digital motion	Smallest musically salient subdivisions of musical pulse; fast notated rhythms	3–10
Phoneme, rapid flam between fingers or limbs	Grace notes, deviations, asynchronies, microtiming	10–60

naturally constrained by lung capacity; tactus-heavy urban dance music often makes sonic references to foot-stomping and to sexually suggestive slapping of skin; blues guitarists, jazz pianists, and *quinto* players in Afro-Cuban *rumba* are said to "speak" with their hands and fingers. All such instances involve the embodiment of the musical performer and the listening audience.

ECOLOGICAL PERCEPTION

All too often, theorists and psychologists have treated musical motion in terms of abstract, time-varying auditory images, while ignoring the motions exerted by the performer. Musical motion is usually discussed as a structural abstraction in pitch space or as involving the play of forms against one another. A more grounded approach is taken by Friberg and Sundberg (1999) and Friberg, Sundberg, and Frydén (2000), who investigate the psychological associations of music with certain rhythmic gaits and other locomotive phenomena. But a review of the concept of musical motion by Shove and Repp (1995) highlights the important and overlooked fact that musical motion is, first and foremost, *audible human motion*. To amplify this view, Shove and Repp make use of Handel's (1990, p. 181) three levels of event awareness: the raw psychophysical perception of tones, the perception of abstract qualities of the tones apart from their source, and last the apprehension of environmental objects that give rise to the sound event. This last level is aligned with the "ecological level" of perception as suggested by Gibson (1979). At this level, "the listener does not merely hear the *sound* of a galloping horse or bowing violinist; rather, the listener hears a horse *galloping* and a violinist bowing" (Shove & Repp, 1995, p. 59). In this ecological framework, the source of perceived musical movement is the human performer, as is abundantly clear to the listener attending to music as a performance event (p. 60). We connect the perception of musical motion at the ecological level to human motion. This suggests that, at this level, musical perception involves an understanding of bodily motion—that is, a kind of *empathetic* embodied cognition. Table 2 suggests some distinctions and correspondences in this vein.

KINESTHETICS

The psychology of bodily feedback—referred to variously as *kinesthetic*, *haptic*, or *proprioceptive*—involves the sensation of bodily position, presence, or movement resulting from tactile sensation and from vestibular input. We rely on such awareness whenever we engage in any physical activity; it helps us hold objects in our hands, walk upright, lean against walls, guide food into our mouths, and swallow it. In these cases, there is a strong interaction between kinesthetic and visual input. Similarly, in the playing of musical instruments, we must treat sonic and kinesthetic dimen-

TABLE 2
Embodied Correlates to Abstract Musical Perception

Abstract Music Perception	Ecological Music Perception
Sounds	Sound sources
Perception	Recognition
Abstraction	Embodiment
Music	Sonic trace of organized human activity
Rhythm	Human motion
Tempo	Speed of human motion
Meter	Regularity of human motion; an *invariant* of the musical environment
Expressive timing	Deviations from invariance
Polyrhythm	Coordinated contrasting human motions
Timbre	Specific instrument/voice/sound source
Loudness	Degree of effort, exertion; number of individuals in unison
Melody	Sustained vocalization, vocal cord use, lung exertion, control
Harmony	Polyphony, interacting sound sources
Form, recurrence, organization	Events, situational/environmental factors
Unison	Synchronized action
Compositional time, musical time	Real time
Piece of music/composition/score	Performance/event

sions as interacting parameters; we must bear in mind the spatiomotor mode of musical performance (Baily, 1985).

For musicians, a major part of musical competence involves the bodily coordination of limbs, digits, and for wind instruments, breathing. Such bodily awareness is most demanding on polyphonic instruments, where multiple sonic streams are generated simultaneously. (In this way, the modern drum set and keyboard instruments are the paradigm for body-centered polyphony. The drum set and the organ are among the only Western four-limb instruments; piano should be considered a three-limb instrument including the use of the pedals, which are often coordinated strongly with the sounds generated by the hands.) For musical performers, the difference between musical and human motion collapses to some degree; the rhythmic motions of the performer and of the musical object overlap (Shove & Repp, 1995, p. 60).

TIMING

In groove-based contexts, rhythmic expression occurs at an extremely fine timescale—rapid enough to rule out a simple auditory-feedback mechanism for its implementation (Fraisse, 1982). This relates to an age-old question in neuroscience, known as the problem of serial order in behavior (Lashley, 1951). The question is how to explain our assimilation and production of very fast sequences of events in time, given that human reflexes and neural transmission speeds are too slow to account for them. Lashley cites the common experience of mistakes in serial order of rapid sequences, such as typing, as evidence of hierarchical organization in this kind of behavior.

There is evidence that temporal, rhythmic, and grouping judgments and productions employ different modes of processing for times under roughly one half second than they do for longer times (Fraisse, 1956, pp. 29–30, cited in Clarke, 1999; Preusser, 1972; Michon, 1975). These short-time processes are described variously as "precognitive," "sensory," or "immediate"—as a kind of sensation, recognition, or gestalt perception, rather than a kind of analytical or counting process. (Other researchers have failed to detect such a perceptual boundary; see Friberg & Sundberg, 1995; Wittman & Poppel, 1999.) Some suggest that this cutoff, if valid, corresponds to the transition between so-called echoic and short-term memory, as indicated by the timescales involved and by other experiments (Brower, 1993; Michon, 1975). These different regimes of memory should distinguish musical rhythms above and below this approximate cutoff as qualitatively different phenomena. For pulse-based music, this cutoff lies in the middle of the tactus range, at about 300 to 800 ms; rhythmic material below this is perceived categorically as combinations of subdivisions of a main regulating pulse, and durations above it are considered to be on the level of metric grouping of pulses. By this division, echoic memory covers the immediate timescale of rhythmic activity, whereas short-term or working memory covers meter and phrases. These different types of memory involve different kinds of processing. We *entrain* to a pulse based on the echoic storage of the previous pulse and some matched internal oscillator periodicity; we *feel* the relationships among strong and weak beats (accentual meter); we *count* times between phrases or bars (metric grouping); and we *recognize* subpulse rhythms qualitatively (Brower, 1993). An embodied account of rhythm perception and cognition—one that takes into account the inherent limitations and abilities of the situated mind-body system—would need to factor in these inherent distinctions of human memory.

The role of different kinds of memory points to the need for different models to explain rhythmic expertise at such a fine scale. In bat and owl echolocation, neural delay-line architectures serve to give the creatures much higher temporal resolution than neural transmission would seemingly allow (Jeffress, 1948). One could say that the animals' temporal acuity exists "in" these long neural pathways—in the physical structure of the perceptual apparatus. A working hypothesis, inspired by the existence of such structures, is that precisely timed rhythmic activity involves the entire body in a complex, holistic fashion, combining audio, visual, and somatosensory channels.

According to the embodiment hypothesis, cognitive structures emerge from reinforced intermodal sensorimotor coupling. In this view, short-time rhythm cognition might include physical sensation, visual entrainment, and sonic reinforcement, unmediated by a symbolic representation. Cognition on the part of musicians apparently involves the physical act of making music as a primary ingredient. Consider the components of the sensory-motor image associated with rhythm perception that are rooted in echoic memory: a phrasal/body-sway component (breath-based), a tactus/foot-tap component (limb-based) (Todd, 1994), and a rapid multiple-finger-tap component (speech- or digit-based). According to the embodied-cognition viewpoint, what we have called our internal representations may consist of precisely these sensorimotor couplings.

Expressive Microtiming in African-American Music

Gradually during the 1990s, in studying West African dance-drumming and in performing on keyboard instruments in jazz, hip-hop, and funk contexts, I experienced an interesting revelation: the simplest repetitive rhythmic patterns could be imbued with a universe of expression. I often witnessed the Ghanaian percussionist and teacher C. K. Ladzekpo stopping the drum ensemble to chide his students for playing their parts with no emotion. One might wonder how much emotion one can convey on a single drum whose pitch range, timbral range, and discrete rhythmic delineations are so narrow that the only two salient elements at one's disposal are intensity and timing. Yet it became clear over time that a great deal can be conveyed with just those two elements. Miniscule timing deviations in performed music are often misleadingly described as "discrepancies" (Keil & Feld, 1994), "motor noise," or "inaccuracies" (Rasch, 1988). Nonetheless, a vast amount of research has been dedicated to the uncovering of structure in these so-called inaccuracies. (For a review of such research in jazz contexts, see Collier & Collier, 1996). It turns out that these deviations from strict metronomicity both convey information about musical structure and provide a window onto internal cognitive representations of music.

Some of my arguments in this section draw upon cultural aspects of music listening. Working from the documented historical lineage between West African and African-American cultures, Wilson (1974, 1981) has identified a constellation of conceptual tendencies that exist in the musics of that vast diversity of cultures. Among the musical preferences and principles enumerated were the following: a connection between music and physical body movement; a principle of rhythmic contrast; a tendency towards rhythmic stratification; the concept of antiphony (i.e. "call and response"); a high degree of percussivity; a continuum between speech and song; a heterogeneous sound ideal, in which musical space is filled up with a variety of contrasting timbres; and a concept of music as meaningful "in motion," that is, as part of everyday life. These and other concepts can serve as the beginnings of a pan-African musical aesthetic, because so many of these notions appear so often in so many different kinds of West African and African-American music.

It is worth reminding ourselves that the study of African-American culture has been plagued by racist mythologies surrounding the idea of the body. Historically, African-American cultural practice has been seen by mainstream Western culture as the realm of the physical, the sensual, and the intuitive, in diametric opposition to the intellectual, the formal, and the logical. As outlined in McClary and Walser (1994, p. 80), I must stress that "to discuss the bodily aspects of cultural texts or performances is not to *reduce* them," but rather to elevate the crucial role of embodiment in all aspects of cultural and perceptual activity. An enlightened treatment of black music that draws from theories of embodiment can get beyond the old mind-body binary, particularly in its racialist manifestations. Such an approach affirms the African diasporic aesthetics that engendered this powerful body of music, while illuminating aspects of all world music.

A great majority of this music falls in the category of groove-based music that I have mentioned, meaning that it features a steady, virtually isochronous pulse that is established collectively by an interlocking composite of rhythmic entities and is either intended for or derived from dance. This somewhat inadequate description should not be viewed as a *definition* of the concept of groove; indeed, to some degree, that definition is what we are searching for with this work. One could say that, among other functions, groove *gives rise to the perception of a steady pulse* in a musical performance—in other words, it engages the "walk" (locomotor) channel of the listener's sensorimotor system, giving rise to *entrainment*.

In groove-based music, this steady pulse is the chief structural element, and it may be articulated in a complex, indirect fashion. In groove contexts, musicians display a heightened, seemingly microscopic sensitivity to musical timing (on the order of a few milliseconds). They are able to evoke a variety of rhythmic qualities, accents, or emotional moods by playing notes slightly late or early relative to a theoretical metric time point. Numerous studies have dissected the nuances of expressive *ritardandi* and other tempo-modulating rhythmic phenomena (Desain & Honing, 1996; Friberg & Sundberg, 1999; Repp, 1990; Todd, 1989), whereas fewer careful quantitative studies have focused on expressive timing with respect to an isochronous pulse (Bilmes, 1993; Collier & Collier, 1996). In groove-based contexts, even as the tempo remains constant, fine-scale rhythmic delivery becomes just as important a parameter as, say, tone, pitch, or loudness. All these musical quantities combine dynamically and holistically to form what some would call a musician's "feel." Individual players have their own feel, that is, their own ways of relating to an isochronous pulse. Musical messages can be passed at this level. A musician can pop out of a polyphonic, rhythmically regular texture by a "deviation" from strict metricality, or a set of such deviations. As I discuss later, these kinds of performance variations create an attentional give-and-take to emphasize different moments interactively. Such techniques are manipulated with great skill by experienced musicians playing together, as a kind of communication at the "feel" level. This variety of expressive timing against an isochronous pulse contains important information about the inner structure of groove.

Analyzing microtiming in European classical music performance, Drake and Palmer (1993) investigated three types of accent structure: rhythmic grouping, melodic grouping, and metric accents. In studies of timing of numerous skilled classical pianists, they found systematic deviations from strict regularity that correlated with these accent structures; small performance variations in timing, intensity, and duration are used to enhance aspects of musical structure. Drake and Palmer concluded that these performance variations facilitate listeners' segmentation of musical sequences, since the accent structures serve to break up a musical sequence into smaller, more tractable chunks. Microrhythmic expression occurs in ensemble contexts as well; in his studies of microtiming variation in small chamber groups, Rasch (1988) conducted statistical analyses of intermusician differences in note onsets from recorded ensemble performances. Although he averaged out all musical structure, including metric accents and tempo variation, he found that on average, in a string trio, the violin's lead voice tends to play ahead by 5 to 10 ms, the cello tends to follow, and the viola's middle voice showed a net lag of another 5 to 10 ms. Rasch's valuable discovery of systematic variation in ensemble performance supplements the wealth of research on expressive timing of solo performers.

The studies just cited focus on aspects of European classical-music performance that would not fall into the realm of groove-based music because of the former's reliance on tempo variation for expressive purposes. Indeed, these aforementioned results indicate that beats are frequently lengthened or shortened by the performer. Also, as discussed elsewhere (Iyer, 1998), the treatment of metric organization as implying a series of weak and strong beats does not apply particularly well to West African or African-American musics. In these contexts there is not always such a thing as a metric accent, from the point of view of performance variation. It would be instructive to conduct a similar microtiming analysis for a percussion ensemble, particularly in instances of groove-based music, which is much less forgiving in the realm of tempo variation and *rubati* than a string trio might be. Bilmes (1993) conducted a timing analysis of a recorded performance of Los Muñequitos de Matanzas (1990), an Afro-Cuban *rumba* group (audio ex. 2). In a performance averaging 110 beats per minute (such that what would be a notated sixteenth note lasts around 135 ms), both the *quinto* and the *segundo* (lead and middle conga drum, respectively) tend to play about 30 ms ahead, or "on top." On the other hand, the *tumbao* (low conga drum) had a much broader distribution, nearly as often late as early. It should be noted that here the position of the theoretical beat was *not* determined by the norm set by these three instruments themselves, as it was in the case of Rasch's string trio. Rather, the beat was established by a reference instrument, in this case a *clave* or a *guagua*. Hence it was possible for all three instruments to be ahead of the nominal beat, which was not the case for the string trio. In Bilmes's work, the *average* interdrum asynchrony was not calculated; indeed, such a measure would ignore any relationship between timing and musical structure. But a frequency analysis of the microtiming variations revealed systematic structure. For example, the rhythmic deviations of the repetitive *segundo* part displayed a strong peak corresponding to the frequency of repetition, suggesting that the microtiming variations were correlated to the rhythmic cycle and were not entirely random.

Given this apparent systematicity of fine-scale rhythmic expression in groove contexts, we can take cues from the results discussed earlier, as well as from our expanded view of cognition involving the theory of embodiment, to make guesses about the function of such rhythmic expression. Thus we hypothesize that microtiming variations in groove music play any of the following roles: (1) highlight structural aspects of the musical material, (2) reflect specific temporal constraints imposed by physical embodiment, and/or (3) fulfill some aesthetic function. Later, I shall address all of these possibilities via a few examples. To make these examples clear, we make use of a representation for rhythmic expression (Bilmes, 1993; Iyer, Bilmes, Wright, & Wessel, 1997) in which, like pitch, loudness, and timbre, we treat microtiming as an *attribute* of a musical note in a metered context. The microtiming of a note is measured as a positive or negative temporal deviation from the note's ideal (i.e., notated) metric position, as measured by some reference instrument in the polyphonic texture.

ASYNCHRONY

The asynchronous unison attacks (Rasch, 1988, discussed earlier) support a psychoacoustically meaningful explanation. Rasch reports an earlier study of asynchrony: a 1977 experiment investigating the effect of onset difference times on the perception of quasi-simultaneous tones. The threshold for perceiving the upper of two quasi-synchronous tones could be decreased from between zero and -20 dB to about -60 dB by introducing an onset difference time of 30 ms. Above the audible threshold, this slight asynchrony was shown to contribute to the perceived transparency of a compound tone. So a function of these quasi-synchronous attacks, or *flams*, could be to aid in the perception of the timbral constituents of the unison attack.

The accompanying audio example demonstrates this tactic (Figure 1, audio ex. 3 & 4). In the first version, the double stroke on the first beat is essentially synchronous; in the second version, a delay of 30 ms is introduced between the two notes, resulting in a small flam. This serves to enhance the perception of the two timbral constituents.

In some musical situations in which blending is preferred, this kind of multitimbral asynchrony may be undesirable, but it is often a valued musical trait in groove-based music. On Wilson's list of African and African-American aesthetic concepts is the notion of a heterogeneous sound ideal, a tendency to value the presence of a variety of contrasting timbres. Another important cultural characteristic is a *collectivist* ideal, in which music is construed as a communal activity among groups of people. The rhythmic asynchronies described earlier aid in the perception of a multiplicity of timbres, as well as, in the ecological view of music perception (Gibson, 1979; Shove & Repp, 1995), the multiplicity of human bodies behind those

Fig. 1. Quasi-synchronous unison (audio ex. 3 and 4).

timbres. That is, rhythmic asynchrony contributes both to the heterogeneous sound ideal and to the sense of collective participation. To be sure, exact synchrony is impossible with groups of people anyway; but this principle applies even when different sounds are played by one individual, as on the modern drum set. So here is an instance in which a kind of subtle rhythmic expression fulfills both a perceptual and an aesthetic function.

Interestingly, the word "flam" is an onomatopoeia. What distinguishes it from a syllable like "tam" is the two-consonant combination "fl," which evokes the very sound of a quasi-synchronous attack. In speech perception, the timescale of a phoneme is around 30 ms, which is the same rhythmic range as in the asynchrony example just given. This correspondence suggests an interesting correlation between speech processing and music perception at the auditory level. Indeed, we can add it to our list of correspondences between body motion and music (Table 1): the mouth movements related to the *syllable* correspond to rapid rhythmic digital motion, whereas the quicker lingual motion associated with phonemes correspond to the microrhythmic asynchronies of groove-based music.

STREAMING

It is well established that auditory stream segregation is a function of both pitch and timbre (Bregman, 1990). From our work, it appears that microtiming can also contribute to streaming. This claim builds on the role of asynchrony in facilitating the perception of multiple tones. Audio examples 5 and 6 consist of a steady stream of triplets on tom-toms, together with a series of sparse tom-tom strokes at a lower volume (Figure 2).

In the first audio example, the unison strokes are as simultaneous as is allowed by the MIDI protocol (i.e., within a couple of milliseconds of each other). In the second example, the second stream is delayed by 30 ms with respect to the first, as indicated by the arrows in the figure, but kept at its same low volume. In the former case, the different timbres fuse into one stream, whereas in the latter case, the second stream is clearly audible as a separate entity. This example shows clearly how such miniscule timing variations can contribute to streaming. This technique is especially important in a context where the aesthetic tendency is to "fill up the musical space"

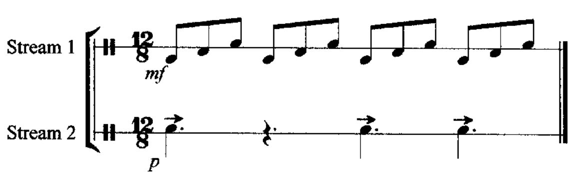

Fig. 2. Auditory streaming caused by quasi-synchronous unison (audio ex. 5 and 6).

(Wilson, 1974). Timing variations can allow an instrument that is sonically buried to draw attention to itself in the auditory scene. Thus the presence of multiple instruments of similar timbres, as in a West African drum ensemble or a large jazz ensemble, need not be viewed as enforcing the subordination of individual identity. Individual musicians can improvise at this microrhythmic level to create an attentional give-and-take. This streaming effect also serves an aesthetic function, in that it enhances the perception of different rhythmic groups as separate animate entities with distinct "personalities" as Ladzekpo (1995) stresses.

SPREADING

It was not until the advent of automated machinery that human ears were ever treated to inhuman rhythmic precision. Perhaps by virtue of their familiarity, the sonic traces of temporal constraints imposed by the body are in certain contexts perceived as aesthetically pleasing, while inhuman rhythmic regularity often is not (except insofar as it elicits human motive response, as in much contemporary urban dance music). Audio examples 7 and 8 (Figure 3) consist of two versions of the same rhythm.

The first rendition is executed as close to the theoretical ideal as the computer allows—that is, mathematically accurate to MIDI precision. The second features timing inflections designed to imitate an aspect of human performance. The difference is not simply the injection of random temporal slop. Rather, it involves the spreading apart of consecutive attacks played by the same hypothetical limb or digit. An individual effector such as a limb, hand, or digit has a time constant associated with its motion; the nerves and muscles have a brief down time as they prepare for the next action. Try tapping a single finger as rapidly as possible, and you find that you cannot surpass a frequency of about 7 taps per second. This is a universally familiar characteristic of the human body.

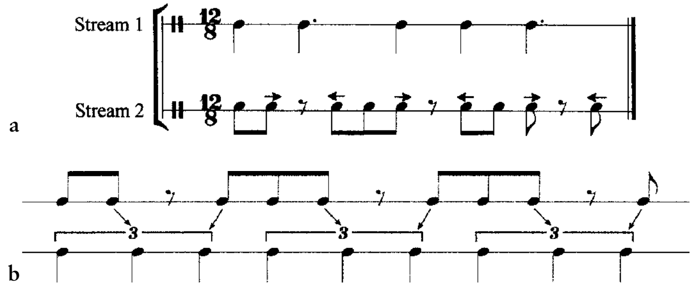

Fig. 3. (a) Microrhythmic spreading as a sonic trace of embodiment (audio ex. 7 and 8). (b) "Tripletizing" a duple rhythm.

The rhythmic expression added in Figure 3a is systematic; in stream 2, the first of each group of three taps is about 30 ms early, and the last is about 30 ms late, as indicated by the arrows. In addition to enhancing perceived separation, this example depicts the encoding of bodily movement in musical material. Nearly all listeners are familiar with the kind of motion suggested by these synthetic tapped rhythms, but that motion is strongly implied only by the second, "imperfect" version. Again, this description recalls the embodied, ecological view of musical perception, in which the listener perceives the *source* of the sounds, rather than the sounds themselves. In a music that embraces physical body motion (Wilson, 1974) and that is contiguous with everyday experience, this sonic trace of the body is a valued aesthetic.

An interesting and perhaps contrasting interpretation of this example has to do with *categorical ambiguity.* The deviations in stream 2 seem to stretch the rhythm away from the cross-rhythmic duple construct (two eighth notes, an eighth rest, and an eighth note) and toward the cross-rhythmic triple construct (three quarter-note triplets) (Figure 3b). In collaborations with Afro-Cuban musicians, I have occasionally been asked to perform this transformation—to partially (but not fully) "tripletize" a duple rhythm. It is an open question whether such categorical ambiguity is a cause or an effect of the sonic trace of the body introduced in this example.

Coding for Invariance

The preceding three examples demonstrate the notion of invariance. At the most basic level, expressive microtiming represents a departure from regularity, so it is likely to be noticed in relief against the more regular background. Gibson (1975) claimed that our perceptual systems are attuned to variants and invariants in the environment; they code for change. As an example, consider the way that vibrato or a trill can facilitate auditory scene analysis by drawing our attention to a particular instrument in an otherwise blended orchestral texture (audio ex. 9). The microvariation of a single pitch is enough to make that voice pop out in the auditory scene.

We can make a similar generalization with rhythm. That which is regular, or invariant, in an isochronous-pulse context is the norm set by the regularity of pulsation, along with its salient multiples and subdivisions; that which is irregular comprises the variable rhythmic material along with its continuous expressive variation. Microrhythmic expression signals a departure from the implied norm, hence marking a particular sound or group of sounds as worthy of attention or analysis by our perceptual systems. This argument contributes to an ecological view of rhythm perception, in which we are attuned to variations in an otherwise regular environment characterized by invariants.

SWING

A kind of rhythmic expression that seems to be indigenous to African-American culture developed in jazz of the first half of the 20th century. Known as swing, this structure can be thought of as modified duple subdivisions of the main pulse, or as modified triplet subdivisions, or both concurrently. As duple subdivisions, they divide the interval of a pulse into two unequal portions, of which the first is slightly longer. They are occasionally rendered in triplet notation as a quarter note followed by an eighth note; however, this exaggerates the typical swing ratio, which tends to fall in the gray area between duple and triple and is strongly tempo dependent, typically lower for fast tempi and higher for slow ones (see Collier & Collier, 1996 and references therein). An individual musician has a particular range of preferred ratios and particular ways of manipulating them, which together form crucial dimensions of that individual's sound, rhythmic feel, and musical personality.

In a related experiment on rhythm, Fraisse (1982) has studied the ability of musically trained and untrained subjects to reproduce rhythmic patterns of various degrees of complexity. "Arrhythmic" sequences with arbitrary relationships between time intervals caused the most difficulty. In more regular rhythmic cases, subjects tended to simplify the ratios between intervals, almost always settling on exactly two classes of time interval: long (400-800 ms) and short (200-400 ms). People tend to understand rhythms to feature two and only two interval lengths, roughly in the ratio of 2:1. This drive towards rhythmic simplicity recalls some of the classical perceptual laws, namely the principle of economy in organization; but usually as performed or as "preferred," the ratio is lower—in fact closer to swing, about 1.75:1, about 57% (Fraisse, 1982).

However, it is not apparent why the interval would be divided unequally in the first place. It would seem even simpler and more economical if there were no such difference in duration between the first and second of two consecutive swung notes. But the point is that this difference facilitates the perception of higher level rhythmic structure. An immediate consequence of the swing feel is that it suggests the next level of hierarchical organization. In conventional terms, the swung eighth-note pairs are perceptually grouped into the larger regular interval, that is, the quarter note. If all subdivisions were performed with exactly the same duration, it would be more difficult to perceive the main beat. The lengthening of the first of two swung notes in a pair amounts to a durational accentuation of the beat. (Often in practice, the second note of the swung pair is given a slight accent in intensity, as if to compensate for its shorter duration.) Hence swing enhances the perception of the main pulse, as the examples (Figure 4, audio ex.10 & 11) demonstrate.

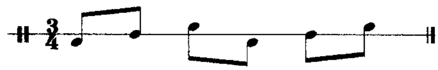

Fig. 4. Swing enhances perception of main pulse (audio ex. 10 and 11).

The first version plays all eighth notes exactly equivalently and is therefore metrically indistinct, whereas the second version introduces a slight swing, which immediately marks the pulse. The observation that swing enhances the perception of the tactus is no surprise, given its primary function in dance contexts. Common jazz expressions referring to a "walking" bass or a "four-on-the-floor" drumbeat affirm that the rhythmic content of "swing" is grounded in the locomotive channel of human motion.

IN THE POCKET: BACKBEAT DELAY

The notion of a backbeat is indigenous to the modern drum kit, an instrument pioneered by African Americans in this century. It consists of a strongly accented snare drum stroke or handclap on beats two and four of a four-beat metric cycle, where the beat is typically a moderate tactus rate (Figure 5, audio ex. 12).

The backbeat appears to have arisen in the middle of the 20th century, as the popular swing rhythm yielded to the even more popular, more bombastic rock and roll rhythms of artists such as Little Richard and Chuck Berry.

In his musical interpretation of Stuckey's (1987) study of the culture of enslaved Africans and its influence on modern African-American culture, Floyd (1995) discusses the important African diasporic cultural ritual known as the *ring shout* as a distinctive space in which, among other things, music and dance were fused. This activity "helped preserve ... what we have come to know as the characterizing and foundational elements of African-American music," including "constant repetition of rhythmic and melodic figures and phrases," "hand clapping, foot patting, and approximations thereof," and "the metronomic pulse that underlies all music." (Floyd, 1995, p. 6) As a cultural model, the ring shout serves for Stuckey as a hermeneutical point of departure in the study of African-American art forms. It provides an alternative lens through which to view these later practices, a lens

Fig. 5. Backbeat (audio ex. 7 and 8).

that is grounded on African, rather than European, concepts and aesthetics. (See Rosenbaum, 1998 for more documentation of the ring shout.)

The backbeat that is so prevalent in postwar African-American popular music seems to reference the role of the body in the ring shout—the bass drum (struck with a mallet attached to a foot pedal, in the modern drumset) and snare drum (struck manually with a stick) replacing the stomp and clap, respectively. In fact, a real or synthetic handclap sound is often superimposed on the backbeat's snare drum sound in popular urban dance music. The edginess and repetition of the backbeat embodies the cyclic, earthy atmosphere of the ring-shout ritual. The backbeat taps into the hypnotic, functional role of repetition in such rituals, in which steady, moderate tempo, rhythmic ostinati, and physical body motion (stomping and clapping) were combined in a collective setting to create a shared multisensory experience. It is plausible that the earliest musical activities of humankind possessed many of these qualities. The backbeat is best understood as a contemporary, popular remnant of what is probably some very ancient human musical behavior, filtered through a sophisticated, stylized African ritual and through centuries of African-American musical development.

The curious point about the backbeat in practice is that when performed by the most esteemed drummers, it frequently displays a microscopic lopsidedness. If we consider the downbeat to be exactly when the bass drum is struck, then the snare drum is very often played ever so slightly *later* than the midpoint between two consecutive pulses (audio ex. 13). Often musicians are aware of this to some degree, and they have a term for it: the drummer is said to play "in the pocket." Although perhaps unaware of the exact temporal details of this effect, a skilled musician or listener in this genre hears this kind of expressive microdelay as "relaxed" or "laid back" as opposed to "stiff" or "on top." This effect is much subtler than the salient rhythmic categorization of the long and short durations of swing. It is a miniscule adjustment at the level of the tactus, rather than the substantial fractional shift of rhythmic subdivisions in swing. The delay functions possibly as a kind of accent, because it involves the postponement of an expected consequent (Meyer, 1956). The optimum snare-drum offset that we call the "pocket" may well be that precise rhythmic position that maximizes the accentual effect of a delay without upsetting the ongoing sense of pulse. This involves the balance of two opposing forces: the force of regularity that resists delay, and the backbeat accentuation that demands delay.

Note that the concept of a backbeat, and the slight delay associated with it, does not pertain if a single sound is used for both the downbeat and the backbeat. (As an example, the urban dance-music genre known as "house" features an isochronous bass drum on all four beats, with the snare-drum backbeat occasionally dropping out.) The effect seems tied to the difference between the two sounds, and perhaps also to the actual sounds themselves and the imagined bodily activity that gives rise to it. In a related study, Fraisse (1982) reports that when asked to synchronize a finger tap or a foot tap with a periodic sound, the finger tap anticipates the sound by roughly 30 ms, and the foot tap anticipates even more. This systematic difference between timing of the hand and the foot suggests that the subject understands synchronization to be the coincidence of auditory and kinesthetic information in the brain. For precise cortical synchronization, the tap must slightly precede the sound to allow for the length of neural transmission, and the temporal offset scales with the distance. This delay architecture amounts to the subject's hand coming *later* than the foot for perceived synchronization, since the anticipatory "error" is greater for the foot. This seems to predict that a regularly alternating stomp-clap pattern would contain a microscopic asymmetry similar to that found in the modern backbeat. (However, it suggests that we frame it as a downbeat anticipation, rather than a backbeat delay; this distinction is a matter of perspective.) Given that the bass drum both references and is played by the foot, and similarly the snare drum both points to and involves the hand, it is possible that this resultant delay structure was transferred to the drumset. Though these arguments are quite speculative, it is plausible that there is an important relationship between the backbeat and the body, informed by the African-American cultural model of the ring shout.

Rhythmic Expression: Situated Musical Examples

THELONIOUS MONK PLAYS "I'M CONFESSIN'"

One of the most fascinating skills displayed by Monk and many other pianists of the genre is a high degree of independence between the two hands, to the degree that one hand can appear to perform rhythms that are ambiguously if at all related to those performed by the other. Often, as in stride piano, this takes the form of a steady pulse or repetitive bass rhythm in the left hand (the "ground"), and rhythmically expressive melodies in the right hand (the "figure"). A classic example is Monk's 1963 solo recording of "I'm Confessin' (That I Love You)" (Monk, 1998; audio ex. 14). In this piece, after carrying on in this expressive stride fashion for some time, the last two bars of the first chorus give rise to an improvised, rapid melodic stream that tumbles rhythmically into the next bar (Figure 6, audio ex. 15).

In this excerpt, the melodic structure in the right hand temporarily overrides and upsets the underlying rhythmic structure, only to be righted again. We can interpret Monk's unquestionably gripping display here as a struggle with constraints; the norm of established pulse regularity, set by what has come before, is threatened by Monk's exploratory right-hand musical gestures. It seems to offer an example of a case in which such regularity is sacrificed briefly to allow for a case of extreme rhythmic expression. Nearly

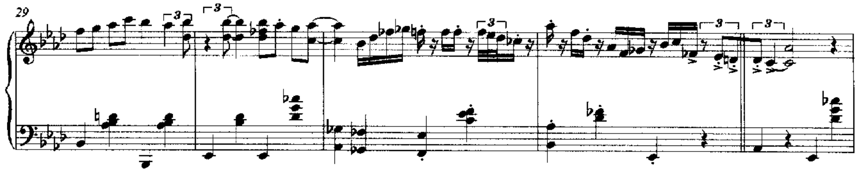

Fig. 6. Author's approximate transcription of bars 29–33 of "I'm Confessin'" as performed by Thelonious Monk (audio ex. 14 and 15).

upsetting the regular pulse, Monk takes a chance and chooses to follow through on a melodic idea that momentarily takes him rhythmically far afield. But note that the pulse is never lost; Monk leaves out a couple of quarter-note chords in the left hand (possibly due to pitch range overlap with the right hand), but otherwise provides strong and accurate pulse reinforcement in the stride style. The rhythmic underpinning of the left hand compensates for the apparent deviation from regularity.

The question of whether Monk "intended" to play this in exactly this way is a pejorative one, akin to reifying the role of "mistakes" in jazz (as in Walser, 1995). It is never clear what is "supposed" to happen in improvised music, so it makes little sense to talk about mistakes. From the perspective of an improvisor, the notion of a mistake is supplanted by the concept of interaction with the structure suggested by the sonic, physical, and temporal environment; in other words, improvisation privileges embodied cognition. This improvisation-friendly framework allows for the possibility of musical exploration and experimentation within constraints, including impromptu rhythmic variation of the sort described here.

It is also interesting to note, in this and the following example, that these extreme instances of microrhythmic expression occurs at a formal boundary—the transition from one 32-bar song-form chorus to the next. The often implied characterization of the symbolic as high-level and the embodied as low-level is misleading, for these functions may interact with each other bilaterally. In particular, one should *not* claim that the high-level processes "direct" the low-level, for in some cases it is not clear that there is any such hierarchical organization. Indeed, the tendency to posit such a hierarchy stems from our prejudice of mental processes as more "elevated" than physical ones. Further consideration of the example at hand suggests a heterarchical interconnectivity of body and mind. For instance, in the midst of an improvisation, the temporally situated pianist is always making choices. These choices are informed not simply by which note, phrase, or gesture is "correct," but rather by which activities are executable at the time that a given choice is made. (Similar observations have been made by Sudnow, 1978.) That is, a skilled improvisor is always attuned to the constraints imposed by the musical moment. This requires an awareness of the palette of musical acts available in general, and particularly of the dynamically evolving subset of this palette that is *physically* possible at any given moment. In this way, for example, the improvising pianist is more likely to choose piano keys that lie under her current hand position than keys that do not. Such weak constraints (which may be overridden, with physical and melodic repercussions) combine holistically with formal directives such as melody and harmony (which may also be overridden). Indeed, improvisation—musical and otherwise—may be understood partially as a dialectic between formal/symbolic and situational/embodied constraints.

The sense of embodiment provides a useful interpretive framework that can occasionally transcend genre; when the recording of this piece was played for a roomful of cognitive science undergraduates, most of whom had little familiarity with jazz or piano music, this excerpt elicited a burst of spontaneous laughter.

AHMAD JAMAL PLAYS "BUT NOT FOR ME"

A wonderfully extemporaneous, playful spirit is captured masterfully in pianist Ahmad Jamal's 1952 trio version of the standard tune "But Not for Me" (audio ex. 16; Jamal, 1980). In this piece, Jamal manipulates his relationship to the pulse actively and voluntarily through the skillful use of microtiming variation. Nearly every single phrase in Jamal's rendition contains some interesting microrhythmic manipulations, but here I will focus on one fragment, namely the end of the first chorus into the beginning of the second chorus. In measure 31 (Figure 7), Jamal initiates a repeating three-beat figure in the four-beat metric context. This additive rhythmic technique is a common one in African-American music, and Jamal carries it out to a humorous extreme, letting the blues-inflected figure cycle 12 full times (nine measures). The first four measures of this passage are displayed in Figure 7. I have adhered to the convention of representing swung rhythms with regular eighth notes, but it should be understood that there is much more to this passage than meets the eye. In particular, Jamal plays this figure extremely behind the beat, so much so as to enhance the humorous

Fig. 7. Author's approximate transcription of bars 31–34 of "But Not for Me" as performed by Ahmad Jamal (audio ex. 16 and 17).

effect of the repeating melodic figure by casting it in starker relief against the more ordinary rhythmic background (audio ex. 17).

In these four measures, the quarter note averages 469 ms (128 beats per minute). The note events in the piano that are displayed as occurring *on* the beat tend to begin actually around 40% of a beat *later* than the drummer's rimshots, which are indicated with x's in the figure. This places him consistently more than a triplet behind the beat; but idiomatically, there is no question of the transcribed metric interpretation as eighth notes. Furthermore, Jamal's second eighth note in each swung pair tends to occur about 85% of the way through the beat. This means that the swing ratio here is effectively inverted; the *first* eighth note in a delayed pair lasts about 45% of a beat (less than half), and the *second* lasts about 55% (more than half). It would appear that the perception of swing arises from complex variations in timing, intensity, or articulation; in this case, it is not merely a matter of achieving the "correct" microrhythmic ratio.

How does Jamal pull off this apparent rhythmic violation of an inverted swing? The answer seems to lie in his 40% *phase shift* relative to the beat established by the accompanying instruments. If, while maintaining this phase relationship, he were to adhere to the usual swing ratio of around 57%, then the second note in a swung pair would be close enough to the onset of the next beat (only a few percent early) that it would be heard as on-the-beat. By employing a relative anticipation of the second eighth note in each pair, Jamal avoids this problem, instead sounding squarely "between" the beats. The 40% delay also affords him enough rhythmic ambiguity so that the inverted swing does not sound jarring. Also, Jamal enhances the sense of swing by accenting the second of each pair (a common technique, as mentioned earlier). So here is a case in which one kind of rhythmic expression interacts with another; the usual long-short relationship of swing is altered in order to accommodate the "laid-back" quality of the melodic figure.

What is accomplished by playing in this laid-back, behind-the-beat fashion? One might expect the same simple perceptual effects (such as enhancing stream segregation) if he instead played *ahead* of the beat, for example. Playing behind the beat is definitely a cultural aesthetic in African-American music, especially jazz; a purposely "late" note seems to carry a certain emphasis, a kind of microrhythmic accent. From the ecological point of view, playing behind the beat might be normally associated with a physical or mental state of relaxation. The fact that the phase shift is sustained consistently is striking, though idiomatically quite familiar. My own introspection as a pianist suggests a conscious physical lag introduced in the execution, exploiting the time taken by the neural signals to reach the fingers. Playing squarely *on* the beat naturally requires a form of anticipation, in the sense that the neural signals must originate well before the beat in order for a note to sound at the proper moment. One can imagine relaxing this anticipatory poise, in a way that is literally conveyed by the ensuing temporal lag in execution. The expressive content of this gesture may have to do with the association with the physical experience of relaxation. Meanwhile, for a skilled practitioner, pulse entrainment provides an internally generated periodic function that protects against metric slippage despite such extreme microrhythmic deviation. Such hypotheses would demand further investigation.

Conclusion

In this article I have employed the dual frameworks of embodied and situated cognition to address some aspects of rhythmic expression that are quite distinct from the commonly discussed European classical musical performance techniques. Instead of (or in addition to) expressive concepts like *rubato*, *ritardando*, and *accelerando*, we have seen deliberately asynchronous unisons, subtle separation of rapid consecutive notes, asymmetric subdivisions of a pulse, and microscopic delays. As further illustration, we have seen extremely deft manipulation of fine-scale rhythmic material in examples from the jazz idiom. I have chosen to focus on African and African-American musics because they often feature these concepts in isolation from the possible interference of tempo variation, and because they tend to involve percussive timbres that facilitate precise microrhythmic analysis.

Analysis of all of these phenomena is enhanced by developing a detailed, extensive understanding of the role of the human body in music perception and cognition. I have argued that African and African-American musics value these kinds of microrhythmic expression in part because of a cultural aesthetic that foregrounds the body. However, I believe that these techniques and traits are found to varying degrees in *all* world music—a claim that surely merits further study.

References

Baily, J. (1985). Music structure and human movement. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 237–258). London: Academic Press.

Bilmes, J. A. (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, Cambridge.

Bregman, A. (1990). *Auditory scene analysis: the perceptual organization of sound.* Cambridge, MA: MIT Press.

Brower, C. (1993). Memory and the perception of rhythm. *Music Theory Spectrum, 15,* 19–35.

Brown, D. E. (1991). *Human universals.* New York: McGraw-Hill.
Brown, J. (1991). *Star time.* Compact disc compilation of original releases 1956–1984. New York: PolyGram Records.
Carroll-Phelan, B., & Hampson, P. J. (1996). Multiple components of the perception of musical sequences: A cognitive neuroscience analysis and some implications for auditory imagery. *Music Perception*, *13,* 517–561.
Clark, A. (1997). *Being there: Putting brain, body, and world together again.* Cambridge, MA: Bradford/MIT Press.
Clarke, E. (1999). Rhythm and timing in music. In D. Deutsch (Ed.), *The psychology of music, 2nd ed.* (pp. 473–500). New York: Academic Press.
Collier, G. L., & Collier, J. L (1996). Microrhythms in jazz: a review of papers. *Annual Review of Jazz Studies*, *3,* 117–139.
Desain, P., & Honing, H. (1996). Physical motion as a metaphor for timing in music: The final ritard. *Proceedings of the 1996 International Computer Music Conference* (pp. 458–460). Hong Kong: ICMPC.
Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, *10,* 343–378.
Floyd, S. (1995). *The power of Black music.* New York: Oxford University Press.
Fraisse, P. (1956). *Les structures rhythmiques.* Louvain, Paris: Publications Universitaires de Louvain.
Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.
Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America, 98,* 2524–2531.
Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, *105,* 1469–1484.
Friberg, A., Sundberg, J., & Frydén, L. (2000). Music from motion: Sound level envelopes of tones expressing human locomotion. *Journal of New Music Research, 29,* 199–210.
Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser & N. Lawrence (Eds.), *The study of time II* (pp. 295–301). New York: Springer-Verlag.
Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.
Handel, S. (1990). *Listening.* Cambridge, MA: MIT Press.
Hardcastle, V. (1996). *How to build a theory in cognitive science.* Albany, NY: SUNY Press.
Held, R., & Hein, A. (1958). Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual-Motor Skills, 8,* 87–90.
Iyer, V. (1998). *Microstructures of feel, macrostructures of sound: Embodied cognition in West African and African American musics.* Unpublished doctoral dissertation, University of California, Berkeley.
Iyer, V., Bilmes, J., Wright, M., & Wessel, D. (1997). A novel representation for rhythmic structure. *Proceedings of the 1997 International Computer Music Conference* (pp. 97–100). Thessaloniki, Greece: ICMPC.
Jamal, A. (1980). *What's New.* Compact disc compilation of original 1952 & 1958 releases. Woodland Hills, CA: Telstar Records.
Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, *41,* 35–39.
Keil, C., & Feld, S. (1994). *Music grooves.* Chicago: University of Chicago Press.
Ladzekpo, C. K. (1995). Foundation course in African dance drumming (web document). Available at http://cnmat.CNMAT.Berkeley.EDU/~ladzekpo/Foundation.html
Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to Western thought. Chicago: University of Chicago Press.
Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffries (Ed.), *Cerebral mechanisms in behavior: The Hixon Symposium* (pp. 112–146). New York: John Wiley & Sons.
Laughlin, C., McManus, J., & D'Aquili, E. (1990). *Brain, symbol, & experience: Towards a neurophenomenology of human consciousness.* New York: Columbia University Press.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life.* New York: Cambridge University Press.

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

Los Muñequitos de Matanzas (1990). *Cantar Maravilloso.* London: GlobeStyle Records.

Mataric, M. (Ed.) (1996). *Embodied cognition & action: Papers from the 1996 AAAI Fall Symposium.* Menlo Park, CA: AAAI Press.

McClary, S., & Walser, R. (1994). Theorizing the body in African-American music. *Black Music Research Journal, 14,* 75–84.

Michon, J. A. (1975). Time experience and memory processes. In J. T. Fraser & N. Lawrence (Eds.), *The study of time II* (pp. 302–313). New York: Springer-Verlag.

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

Monk, T. (1998). *Monk Alone: The Complete Columbia Solo Studio Recordings: 1962-1968.* Compact disc compilation. New York: Columbia Records/Sony Music.

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

Passingham, R. (1982). *The human primate.* Oxford: Freeman.

Peretz, I. (1993). Auditory agnosia: a functional analysis. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 199–230). Oxford: Clarendon Press.

Prem, E. (1996). Elements of an epistemology of embodied AI. In M. Mataric (Ed.), *Embodied cognition & action: Papers from the 1996 AAAI Fall Symposium* (pp. 97–101). Menlo Park, CA: AAAI Press.

Preusser, D. (1972). The effect of structure and rate on the recognition and description of auditory temporal patterns. *Perception & Psychophysics, 11*, 233–240.

Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press/Bradford Books.

Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind.* New York: William Morrow & Co.

Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 70–90). Oxford: Clarendon Press.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America, 88,* 622–641.

Rosenbaum, A. (1998). *Shout because you're free: The African American ring shout tradition in coastal Georgia.* Athens, GA: University of Georgia Press.

Sacks, O. (1985). *The man who mistook his wife for a hat and other clinical tales.* New York: Summit Books.

Shore, B. (1996). *Culture in mind: Cognition, culture, and the problem of meaning.* New York: Oxford University Press.

Shove, P., & Repp, B. (1995). Musical motion and performance: theoretical and empirical perspectives. In J. Rink (Ed.), *The practice of performance* (pp. 55–83). Cambridge, U.K.: Cambridge University Press.

Stuckey, S. (1987). *Slave culture: Nationalist theory and the foundations of Black America.* New York: Oxford University Press.

Sudnow, D. (1978). *Ways of the hand.* Cambridge, MA: Harvard University Press.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.

Todd, N. P. M. (1989.) Towards a cognitive theory of expression: the performance and perception of rubato. *Contemporary Music Review, 4,* 405–416.

Todd, N. P. M. (1994). The auditory "primal sketch": a multiscale model of rhythmic grouping. *Journal of New Music Research, 23,* 25–70.

Todd, N. P. M. (1999). Motion in music: A neurobiological perspective. *Music Perception, 17,* 115–126.

Todd, N. P. M., Lee, C. S., & O'Boyle, D. J. (1999). A sensory-motor theory of rhythm, time perception and beat induction. *Journal of New Music Research, 28*, 5–29.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Walser, R. (1995). "Out of notes": Signification, interpretation, and the problem of Miles Davis. In K. Gabbard (Ed.), *Jazz among the discourses* (pp. 165–188). Durham, NC: Duke University Press.

Wilson, O. (1974). The significance of the relationship between Afro-American music and West-African music. *The Black Perspective in Music, 2,* 3–22.

Wilson, O. (1981). The association of movement and music as a manifestation of a black conceptual approach to music making. In D. Heartz & B. Wade (Eds.), *International Musicological Society Report of the Twelfth Congress* (pp. 98–105). Basel: Bärenreiter Kassel.

Wittman, M., & Pöppel, E. (1999). Temporal mechanisms of the brain as fundamentals of communication—with special reference to music perception and performance. *Musicae Scientiae [special issue] 1999-2000,* 13–28.