Music and movement share a dynamic structure that supports universal expressions of emotion

Beau Sievers[a,1], Larry Polansky[b], Michael Casey[b], and Thalia Wheatley[a,1]

[a]Department of Psychological and Brain Sciences and [b]Department of Music, Dartmouth College, Hanover, NH 03755

Music moves us. Its kinetic power is the foundation of human behaviors as diverse as dance, romance, lullabies, and the military march. Despite its significance, the music-movement relationship is poorly understood. We present an empirical method for testing whether music and movement share a common structure that affords equivalent and universal emotional expressions. Our method uses a computer program that can generate matching examples of music and movement from a single set of features: rate, jitter (regularity of rate), direction, step size, and dissonance/visual spikiness. We applied our method in two experiments, one in the United States and another in an isolated tribal village in Cambodia. These experiments revealed three things: (*i*) each emotion was represented by a unique combination of features, (*ii*) each combination expressed the same emotion in both music and movement, and (*iii*) this common structure between music and movement was evident within and across cultures.

cross-cultural | cross-modal

Music moves us, literally. All human cultures dance to music and music's kinetic faculty is exploited in everything from military marches and political rallies to social gatherings and romance. This cross-modal relationship is so fundamental that in many languages the words for music and dance are often interchangeable, if not the same (1). We speak of music "moving" us and we describe emotions themselves with music and movement words like "bouncy" and "upbeat" (2). Despite its centrality to human experience, an explanation for the music-movement link has been elusive. Here we offer empirical evidence that sheds new light on this ancient marriage: music and movement share a dynamic structure.

A shared structure is consistent with several findings from research with infants. It is now well established that very young infants—even neonates (3)—are predisposed to group metrically regular, auditory events similarly to adults (4, 5). Moreover, infants also infer meter from movement. In one study, 7-mo-old infants were bounced in duple or triple meter while listening to an ambiguous rhythm pattern (6). When hearing the same pattern later without movement, infants preferred the pattern with intensity (auditory) accents that matched the particular metric pattern at which they were previously bounced. Thus, the perception of a "beat," established by movement or by music, transfers across modalities. Infant preferences suggest that perceptual correspondences between music and movement, at least for beat perception, are predisposed and therefore likely universal. By definition, however, infant studies do not examine whether such predispositions survive into adulthood after protracted exposure to culture-specific influences. For this reason, adult cross-cultural research provides important complimentary evidence for universality.

Previous research suggests that several musical features are universal. Most of these features are low-level structural properties, such as the use of regular rhythms, preference for small-integer frequency ratios, hierarchical organization of pitches, and so on (7, 8). We suggest music's capacity to imitate biological dynamics including emotive movement is also universal, and that this capacity is subserved by the fundamental dynamic similarity of the domains of music and movement. Imitation of human physiological responses would help explain, for example, why "angry" music is faster and more dissonant than "peaceful" music. This capacity may also help us understand music's inductive effects: for example, the soothing power of lullabies and the stimulating, synchronizing force of military marching rhythms.

Here we present an empirical method for quantitatively comparing music and movement by leveraging the fact that both can express emotion. We used this method to test to what extent expressions of the same emotion in music and movement share the same structure; that is, whether they have the same dynamic features. We then tested whether this structure comes from biology or culture. That is, whether we are born with the predisposition to relate music and movement in particular ways, or whether these relationships are culturally transmitted. There is evidence that emotion expressed in music can be understood across cultures, despite dramatic cultural differences (9). There is also evidence that facial expressions and other emotional movements are cross-culturally universal (10–12), as Darwin theorized (13). A natural predisposition to relate emotional expression in music and movement would explain why music often appears to be cross-culturally intelligible when other fundamental cultural practices (such as verbal language) are not (14). To determine how music and movement are related, and whether that relationship is peculiar to Western culture, we ran two experiments. First, we tested our common structure hypothesis in the United States. Then we conducted a similar experiment in L'ak, a culturally isolated tribal village in northeastern Cambodia. We compared the results from both cultures to determine whether the connection between music and movement is universal. Because many musical practices are culturally transmitted, we did not expect both experiments to have precisely identical results. Rather, we hypothesized results from both cultures would differ in their details yet share core dynamic features enabling cross-cultural legibility.

General Method

We created a computer program capable of generating both music and movement; the former as simple, monophonic piano melodies, and the latter as an animated bouncing ball. Both were controlled by a single probabilistic model, ensuring there was an isomorphic relationship between the behavior of the music and the movement of the ball. This model represented both music and movement in terms of dynamic contour: how changes in the stimulus unfold over time. Our model for dynamic contour comprised five quantitative parameters controlled by on-screen slider bars. Stimuli were generated in real time, and manipulation of the slider bars resulted in immediate changes in the music being played or the animation being shown (Fig. 1).
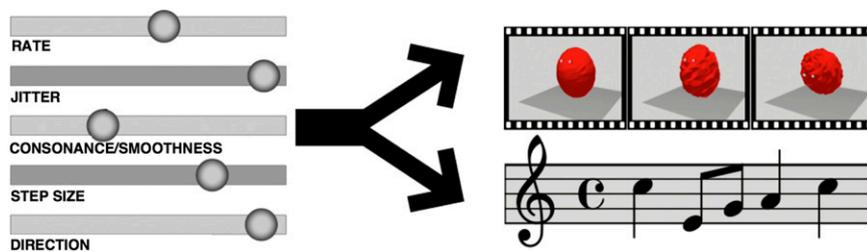
PSYCHOLOGICAL AND COGNITIVE SCIENCES

Fig. 1. Paradigm. Participants manipulated five slider bars corresponding to five dynamic features to create either animations or musical clips that expressed different emotions.

The five parameters corresponded to the following features: rate (as ball bounces or musical notes per minute, henceforth beats per minute or BPM), jitter (SD of interonset interval), direction of movement (ratio of downward to upward movements, controlling either pitch trajectory or ball tilt), step size (ratio of big to small movements, controlling pitch interval size or ball bounce height), and finally consonance/smoothness [quantified using Huron's (15) aggregate dyadic consonance measure and mapped to surface texture].

The settings for each of these parameters affected both the music and the movement such that certain musical features were guaranteed to correspond with certain movement features. The rate and jitter sliders controlled the rate and variation in interonset interval of events in both modalities. The overall contour of each melody or bounce sequence was determined by the combined positions of the direction of movement and step-size sliders. Absolute pitch position corresponded to the extent to which the ball was tilted forward or backward. Low pitches corresponded with the ball "leaning" forward, as though looking toward the ground, and high pitches corresponded with "leaning" backward, looking toward the sky. For music, the consonance slider controlled the selection of one of 38 possible 5-note scales, selected from the 12-note Western chromatic scale and sorted in order by their aggregate dyadic consonance (15) (*SI Text*). For movement, the consonance slider controlled the visual spikiness of the ball's surface. Dissonant intervals in the music corresponded to increases in the spikiness of the ball, and consonant intervals smoothed out its surface. Spikiness was dynamic in the sense that it was perpetually changing because of the probabilistic and continuously updating nature of the program; it did not influence the bouncing itself. Our choice of spikiness as a visual analog of auditory dissonance was inspired by the concept of auditory "roughness" described by Parncutt (16). We understand "roughness" as an apt metaphor for the experience of dissonance. We did not use Parncutt's method for calculating dissonance based on auditory beating (17, 18). To avoid imposing a particular physical definition of dissonance, we used values derived directly from aggregated empirical reports of listener judgments (15). Spikiness was also inspired by nonarbitrary mappings between pointed shapes and unrounded vowels (e.g., "kiki") and between rounded shapes and rounded sounds (e.g., "bouba"; refs. 19–21). The dissonance-spikiness mapping was achieved by calculating the dissonance of the melodic interval corresponding to each bounce, and dynamically scaling the spikiness of the surface of the ball proportionately.

Three of these parameters are basic dynamic properties: speed (BPM), direction, and step-size. Regularity (jitter) and smoothness were added because of psychological associations with emotion [namely, predictability and tension (22)] that were not already captured by speed, step-size, and direction. The number of features (five) was based on the intuition that this number created a large enough possibility space to provide a proof-of-concept test of the shared structure hypothesis without becoming unwieldy for participants. We do not claim that these five features optimally characterize the space.

These parameters were selected to accommodate the production of specific features previously identified with musical emotion (2). In the music domain, this set of features can be grouped as "timing" features (tempo, jitter) and "pitch" features (consonance, step size, and direction). Slider bars were presented with text labels indicating their function (*SI Text*). Each of the cross-modal (music-movement) mappings represented by these slider bars constituted a hypothesis about the relationship between music and movement. That is, based on their uses in emotional music and movement (2, 23), we hypothesized that rate, jitter, direction of movement, and step size have equivalent emotional function in both music and movement. Additionally, we hypothesized that both dissonance and spikiness would have negative valence, and that in equivalent cross-modal emotional expressions the magnitude of one would be positively correlated with the magnitude of the other.

United States

Methods. Our first experiment took place in the United States with a population of college students. Participants ($n = 50$) were divided into two groups, music ($n = 25$) and movement ($n = 25$). Each participant completed the experiment individually and without knowledge of the other group. That is, each participant was told about either the music or the movement capability of the program, but not both.

After the study was described to the participants, written informed consent was obtained. Participants were given a brief demonstration of the computer program, after which they were allowed unlimited time to get used to the program through undirected play. At the beginning of this session, the slider bars were automatically set to random positions. Participants ended the play session by telling the experimenter that they were ready to begin the experiment. The duration of play was not recorded, but the modal duration was ~5–10 min. To begin a melody or movement sequence, participants pressed the space bar on a computer keyboard. The music and movement output were continuously updated based on the slider bar positions such that participants could see (or hear) the results of their efforts as they moved the bars. Between music sequences, there was silence. Between movement sequences, the ball would hold still in its final position before resetting to a neutral position at the beginning of the next sequence.

After indicating they were ready to begin the experiment, participants were instructed to take as much time as needed to use the program to express five emotions: "angry," "happy," "peaceful," "sad," and "scared." Following Hevner (24), each emotion word was presented at the top of a block of five words with roughly the same meaning (*SI Text*). These word clusters were present on the screen throughout the entire duration of the experiment. Participants could work on each of these emotions in any order, clicking on buttons to save or reload slider bar settings for any emotion at any time. Only the last example of any emotion saved by the participant was used in our analyses. Participants could use all five sliders throughout the duration of the experiment, and no restrictions were placed on the order in which the sliders were used. For example, participants were free to begin using the tempo slider, then switch to the dissonance slider at any time, then back to the tempo slider again, and so on. In practice, participants constantly switched between the sliders, listening or watching the aggregate effect of all slider positions on the melody or ball movement.

Results. The critical question was whether subjects who used music to express an emotion set the slider bars to the same positions as subjects who expressed the same emotion with the moving ball.

To answer this question, the positions of the sliders for each modality (music vs. movement) and each emotion were analyzed using multiway ANOVA. Emotion had the largest main effect on slider position [$F(2.97, 142.44) = 185.56$, $P < 0.001$, partial $\eta^2 = 0.79$]. Partial η^2 reflects how much of the overall variance (effect plus error) in the dependent variable is attributable to the factor in question. Thus, 79% of the overall variance in where participants placed the slider bars was attributable to the emotion they were attempting to convey. This main effect was qualified by an Emotion × Slider interaction indicating each emotion required different slider settings [$F(4.81, 230.73) = 112.90$, $P < 0.001$; partial $\eta^2 = 0.70$].

Although we did find a significant main effect of Modality (music vs. movement) [$F(1,48) = 4.66$, $P < 0.05$], it was small (partial $\eta^2 = 0.09$) and did not interact with Emotion [Emotion × Modality: $F(2.97, 142.44) = 0.97$, $P > 0.4$; partial $\eta^2 = 0.02$]. This finding indicates slider bar settings for music and movement were slightly different from each other, regardless of the emotion being represented. We also found a three-way interaction between Slider, Emotion, and Modality. This interaction was significant but modest [$F(4.81, 230.73) = 4.50$, $P < 0.001$; partial $\eta^2 = 0.09$], and can be interpreted as a measure of the extent to which music and movement express different emotions with different patterns of dynamic features.

To investigate the similarity of emotional expressions, we conducted a Euclidean distance-based clustering analysis. This analysis revealed a cross-modal, emotion-based structure (Fig. 2).

These results strongly suggest the presence of a common structure. That is, within this experiment, rate, jitter, step size, and direction of movement functioned the same way in emotional music and movement, and aggregate dyadic dissonance was functionally analogous to visual spikiness. For our United States population, music and movement shared a cross-modal expressive code.

Cambodia

Methods. We conducted our second experiment in L'ak, a rural village in Ratanakiri, a sparsely populated province in northeastern Cambodia. L'ak is a Kreung ethnic minority village that has maintained a high degree of cultural isolation. (For a discussion of the possible effects of modernization on L'ak, see *SI Text*.) In Kreung culture, music and dance occur primarily as a part of rituals, such as weddings, funerals, and animal sacrifices (25). Kreung music is formally dissimilar to Western music: it has no system of vertical pitch relations equivalent to Western tonal harmony, is constructed using different scales and tunings, and is performed on morphologically dissimilar instruments. For a brief discussion of the musical forms we observed during our visit, see *SI Text*.

Fig. 2. Music-movement similarity structure in the United States data. Clusters are fused based on the mean Euclidean distance between members. The data cluster into a cross-modal, emotion-based structure.

Fig. 3. (*A*) Kreung participants used a MIDI controller to manipulate the slider bar program. (*B*) L'ak village debriefing at the conclusion of the study.

The experiment we conducted in L'ak proceeded in the same manner as the United States experiment, except for a few modifications made after pilot testing. Initially, because most of the participants were illiterate, we simply removed the text labels from the sliders. However, in our pilot tests we found participants had difficulty remembering the function of each slider during the movement task. We compensated by replacing the slider labels for the movement task with pictures (*SI Text*). Instructions were conveyed verbally by a translator. (For a discussion of the translation of emotion words, see *SI Text*.) None of the participants had any experience with computers, so the saving/loading functionality of the program was removed. Whereas the United States participants were free to work on any of the five emotions throughout the experiment, the Kreung participants worked out each emotion one-by-one in a random order. There were no required repetitions of trials. However, when Kreung subjects requested to work on a different emotion than the one assigned, or to revise an emotion they had already worked on, that request was always granted. As with the United States experiment, we always used the last example of any emotion chosen by the participant. Rather than a mouse, participants used a hardware MIDI controller (Korg nanoKontrol) to manipulate the sliders on the screen (Fig. 3*A*).

When presented with continuous sliders as in the United States experiment, many participants indicated they were experiencing decision paralysis and could not complete the task. To make the task comfortable and tractable we discretized the sliders, limiting each to three positions: low, medium, and high (*SI Text*). As with the United States experiment, participants were split into separate music ($n = 42$) and movement ($n = 43$) groups.

Results: Universal Structure in Music and Movement

There were two critical questions for the cross-cultural analysis: (*i*) Are emotional expressions universally cross-modal? and (*ii*) Are emotional expressions similar across cultures? The first question asks whether participants who used music to express an emotion set the slider bars to the same positions as participants who expressed the same emotion with the moving ball. This question does not examine directly whether particular emotional expressions are universal. A cross-modal result could be achieved even if different cultures have different conceptions of the same emotion (e.g., "happy" could be upward and regular in music and movement for the United States, but downward and irregular in music and movement for the Kreung). The second question asks whether each emotion (e.g., "happy"), is expressed similarly across cultures in music, movement or both.

To compare the similarity of the Kreung results to the United States results, we conducted three analyses. All three analyses required the United States and Kreung data to be in a comparable format; this was accomplished by making the United States data discrete. Each slider setting was assigned a value of low, medium, or high in accordance with the nearest value used in the Kreung experiment. The following sections detail these three analyses. See *SI Text* for additional analyses, including a linear discriminant

analysis examining the importance of each feature (slider) in distinguishing any given emotion from the other emotions.

ANOVA. We z-scored the data for each parameter (slider) separately within each population. We then combined all z-scored data into a single, repeated-measures ANOVA with Emotion and Sliders as within-subjects factors and Modality and Population as between-subjects factors. Emotion had the largest main effect on slider position [$F(3.76, 492.23) = 40.60, P < 0.001$, partial $\eta^2 = 0.24$], accounting for 24% of the overall (effect plus error) variance. There were no significant main effects of Modality (music vs. movement) [$F(1,131) = 0.004, P = 0.95$] or Population (United States, Kreung) [$F(1,131) < 0.001, P = 0.99$] and no interaction between the two [$F(1,131) = 1.15, P = 0.29$].

This main effect of Emotion was qualified by an Emotion × Sliders interaction, indicating each emotion was expressed by different slider settings [$F(13.68, 1791.80) = 38.22, P < 0.001$; partial $\eta^2 = 0.23$]. Emotion also interacted with Population, albeit more modestly [$F(3.76, 492.23) = 11.53, P < 0.001$, partial $\eta^2 = 0.08$] and both were qualified by the three-way Emotion × Sliders × Population interaction, accounting for 7% of the overall variance in slider bar settings [$F(13.68, 1791.80) = 10.13, P < 0.001$, partial $\eta^2 = 0.07$]. This three-way interaction can be understood as how much participants' different emotion configurations could be predicted by their population identity. See *SI Text* for the z-scored means.

Emotion also interacted with Modality [$F(3.76, 492.23) = 2.84, P = 0.02$; partial $\eta^2 = 0.02$], which was qualified by the three-way Emotion × Modality × Sliders [$F(13.68, 1791.80) = 4.92, P < 0.001$; partial $\eta^2 = 0.04$] and the four-way Emotion × Modality × Sliders × Population interactions [$F(13.68, 1791.80) = 2.8, P < 0.001$; partial $\eta^2 = 0.02$]. All of these Modality interactions were modest, accounting for between 2% and 4% of the overall variance.

In summary, the ANOVA revealed that the slider bar configurations depended most strongly on the emotion being conveyed (Emotion × Slider interaction, partial $\eta^2 = 0.23$), with significant but small influences of modality and population (partial η^2's < 0.08).

Monte Carlo Simulation. Traditional ANOVA is well-suited to detecting mean differences given a null hypothesis that the means are the same. However, this test cannot capture the similarity between populations given the size of the possibility space. One critical advance of the present paradigm is that it allowed participants to create different emotional expressions within a large possibility space. Analogously, an ANOVA on distance would show that Boston and New York City do not share geographic coordinates, thereby rejecting the null hypothesis that these cities occupy the same space. Such a comparison would not test how close Boston and New York City are compared with distances between either city and every other city across the globe (i.e., relative to the entire possibility space). To determine the similarity between Kreung and United States data given the entire possibility space afforded by the five sliders, we ran a Monte Carlo simulation. The null hypothesis of this simulation was that there are no universal perceptual constraints on music-movement-emotion association, and that individual cultures may create music and movement anywhere within the possibility space. Showing instead that differences between cultures are small relative to the size of the possibility space strongly suggests music-movement-emotion associations are subject to biological constraints.

We represented the mean results of each experiment as a 25-dimensional vector (five emotions × five sliders), where each dimension has a range from 0.0 to 2.0. The goal of the Monte Carlo simulation was to see how close these two vectors were to each other relative to the size of the space they both occupy. To do this, we sampled the space uniformly at random, generating one-million pairs of 25-dimensional vectors. Each of these vectors represented a possible outcome of the experiment. We measured the Euclidean distance between each pair to generate a distribution of intervector distances (mean = 8.11, SD = 0.97).

The distance between the Kreung and United States mean result vectors was 4.24, which was 3.98 SDs away from the mean. Out of one-million vector pairs, fewer than 30 pairs were this close together, suggesting it is highly unlikely that the similarities between the Kreung and United States results were because of chance (Fig. 4).

Taken together, the ANOVA and Monte Carlo simulation revealed that the Kreung and United States data were remarkably similar given the possibility space, and that the combined data were best predicted by the emotions being conveyed and least predicted by the modality used. The final analysis examined Euclidean distances between Kreung and United States data for each emotion separately.

Cross-Cultural Similarity by Emotion: Euclidean Distance. For this analysis, we derived emotional "prototypes" from the results of the United States experiment. This derivation was accomplished by selecting the median value for each slider for each emotion (for music and movement combined) from the United States results and mapping those values to the closest Kreung setting of "low," "medium," and "high." For example, the median rate for "sad" was 46 BPM in the United States sample. This BPM was closest to the "low" setting used in the Kreung paradigm (55 BPM). Using this method for all five sliders, the "sad" United States prototype was: low rate, low jitter, medium consonance, low ratio of big to small movements, and high ratio of downward to upward movements. We measured the similarity of each of the Kreung datapoints to the corresponding United States prototype by calculating the Euclidean distance between them.

For every emotion except "angry," this distance analysis revealed that Kreung results (for music and movement combined) were closer to the matching United States emotional prototypes than they were to any of the other emotional prototypes. In other words, the Kreung participants' idea of "sad" was more similar to the United States "sad" prototype than to any other emotional prototype, and this cross-cultural congruence was observed for all emotions except "angry." This pattern also held for the movement results when considered separately from music. When the music results were evaluated alone, three of the five emotions (happy, sad, and scared) were closer to the matching United States prototype than any nonmatching prototypes.

The three Kreung emotional expressions that were not closest to their matching United States prototypes were "angry" movement,
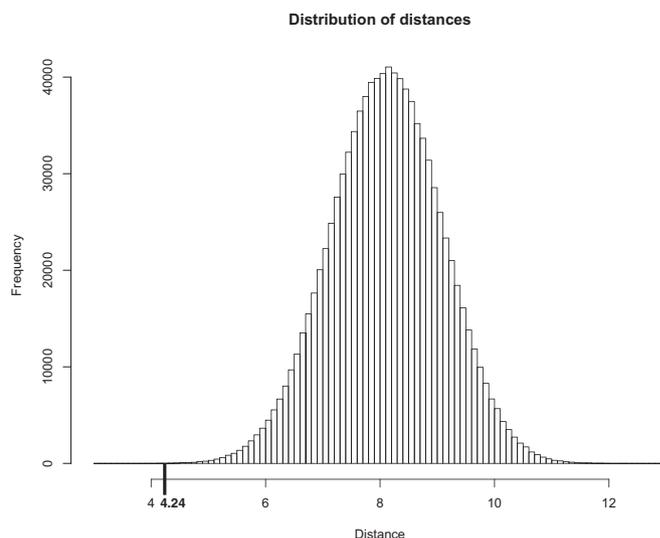

Fig. 4. Distribution of distances in the Monte Carlo simulation. Bold black line indicates where the similarity of United States and Kreung datasets falls in this distribution.

"angry" music, and "peaceful" music; however, these had several matching parameters. For both cultures, "angry" music and "angry" movement were fast and downward. Although it was closer to the United States "scared" prototype, Kreung "angry" music matched the United States "angry" prototype in four of five parameters. Kreung "peaceful" music was closest to the United States "happy" prototype, and second closest to the United States "peaceful" prototype. In both cultures, "happy" music was faster than "peaceful" music, and "happy" movement was faster than "peaceful" movement.

Discussion

These data suggest two things. First, the dynamic features of emotion expression are cross-culturally universal, at least for the five emotions tested here. Second, these expressions have similar dynamic contours in both music and movement. That is, music and movement can be understood in terms of a single dynamic model that shares features common to both modalities. This ability is made possible not only by the existence of prototypical emotion-specific dynamic contours, but also by isomorphic structural relationships between music and movement.

The natural coupling of music and movement has been suggested by a number of behavioral experiments with adults. Friberg and Sundberg observed that the deceleration dynamics of a runner coming to a stop accurately characterize the final slowing at the end of a musical performance (26). People also prefer to tap to music at tempos associated with natural types of human movement (27) and common musical tempi appear to be close to some biological rhythms of the human body, such as the heartbeat and normal gait. Indeed, people synchronize the tempo of their walking with the tempo of the music they hear (but not to a similarly paced metronome), with optimal synchronization occurring around 120 BPM, a common tempo in music and walking. This finding led the authors to suggest that the "perception of musical pulse is due to an internalization of the locomotion system" (28), consistent more generally with the concept of embodied music cognition (29).

The embodiment of musical meter presumably recruits the putative mirror system comprised of regions that coactivate for perceiving and performing action (30). Consistent with this hypothesis, studies have demonstrated neural entrainment to beat (31, 32) indexed by beat-synchronous β-oscillations across auditory and motor cortices (31). This basic sensorimotor coupling has been described as creating a pleasurable feeling of being "in the groove" that links music to emotion (33, 34).

The capacity to imitate biological dynamics may also be expressed in nonverbal emotional vocalizations (prosody). Several studies have demonstrated better than chance cross-cultural recognition of several emotions from prosodic stimuli (35–39). Furthermore, musical expertise improves discrimination of tonal variations in languages such as Mandarin Chinese, suggesting common perceptual processing of pitch variations across music and language (40). It is thus possible, albeit to our knowledge not tested, that prosody shares the dynamic structure evinced here by music and movement. However, cross-modal fluency between music and movement may be particularly strong because of the more readily identifiable pitch contours and metric structure in music compared with speech (4, 41).

The close relationship between music and movement has attracted significant speculative attention from composers, musicologists, and philosophers (42–46). Only relatively recently have scientists begun studying the music-movement relationship empirically (47–50). This article addresses several limitations in this literature. First, using the same statistical model to generate music and movement stimuli afforded direct comparisons previously impossible because of different methods of stimulus creation. Second, modeling lower-level dynamic parameters (e.g., consonance), rather than higher-level constructs decreased the potential for cultural bias (e.g., major/minor). Finally, by creating emotional expressions directly rather than rating a limited set of stimuli prepared in advance, participants could explore the full breadth of the possibility space.

Fitch (51) describes the human musical drive as an "instinct to learn," which is shaped by universal proclivities and constraints. Within the range of these constraints "music is free to 'evolve' as a cultural entity, together with the social practices and contexts of any given culture." We theorize that part of the "instinct to learn" is a proclivity to imitate. Although the present study focuses on emotive movement, music across the world imitates many other phenomena, including human vocalizations, birdsong, the sounds of insects, and the operation of tools and machinery (52–54).

We do not claim that the dynamic features chosen here describe the emotional space optimally; there are likely to be other useful features as well as higher-level factors that aggregate across features (37, 55–56). We urge future research to test other universal, cross-modal correspondences. To this end, labanotation—a symbolic language for notating dance—may be a particularly fruitful source. Based on general principles of human kinetics (57), labanotation scripts speed (rate), regularity, size, and direction of movement, as well as "shape forms" consistent with smoothness/spikiness. Other Laban features not represented here, but potentially useful for emotion recognition, include weight and symmetry. It may also be fruitful to test whether perceptual tendencies documented in one domain extend across domains. For example, innate (and thus likely universal) auditory preferences include: seven or fewer pitches per octave, consonant intervals, scales with unequal spacing between pitches (facilitating hierarchical pitch organization), and binary timing structures (see refs. 14 and 58 for reviews). Infants are also sensitive to hierarchical pitch organization (5) and melodic transpositions (see ref. 59 for a review). These auditory sensitivities may be the result of universal proclivities and constraints with implications extending beyond music to other dynamic domains, such as movement. Our goal was simply to use a small set of dynamic features that describe the space well enough to provide a test of cross-modal and cross-cultural similarity. Furthermore, although these dynamic features describe the space of emotional expression for music and movement, the present study does not address whether these features describe the space of emotional experience (60, 61).

Our model should not be understood as circumscribing the limits of emotional expression in music. Imitation of movement is just one way among many in which music may express emotion, as cultural conventions may develop independently of evolved proclivities. This explanation allows for cross-cultural consistency yet preserves the tremendous diversity of musical traditions around the world. Additionally, we speculate that, across cultures, musical forms will vary in terms of how emotions and their related physical movements are treated differentially within each cultural context. Similarly, the forms of musical instruments and the substance of musical traditions may in turn influence differential cultural treatment of emotions and their related physical movements. This interesting direction for further research will require close collaboration with ethnomusicologists and anthropologists.

By studying universal features of music we can begin to map its evolutionary history (14). Specifically, understanding the cross-modal nature of musical expression may in turn help us understand why and how music came to exist. That is, if music and movement have a deeply interwoven, shared structure, what does that shared structure afford and how has it affected our evolutionary path? For example, *Homo sapiens* is the only species that can follow precise rhythmic patterns that afford synchronized group behaviors, such as singing, drumming, and dancing (14). *Homo sapiens* is also the only species that forms cooperative alliances between groups that extend beyond consanguineal ties (62). One way to form and strengthen these social bonds may be through music: specifically the kind of temporal and affective entrainment that music evokes from infancy (63). In turn, these musical entrainment-based bonds may be the basis for *Homo sapiens*' uniquely flexible sociality (64). If this is the case, then our evolutionary understanding of music is not simply reducible to the capacity for entrainment. Rather, music is the arena in which this and other capacities participate in determining evolutionary fitness.

The shared structure of emotional music and movement must be reflected in the organization of the brain. Consistent with this view, music and movement appear to engage shared neural substrates, such as those recruited by time-keeping and sequence learning (31, 65, 66). Dehaene and Cohen (67) offer the term "neuronal recycling" to describe how late-developing cultural abilities, such as reading and arithmetic, come into existence by repurposing brain areas evolved for older tasks. Dehaene and Cohen suggest music "recycles" or makes use of premusical representations of pitch, rhythm, and timbre. We hypothesize that this explanation can be pushed a level deeper: neural representations of pitch, rhythm, and timbre likely recycle brain areas evolved to represent and engage with spatiotemporal perception and action (movement, speech). Following this line of thinking, music's expressivity may ultimately be derived from the evolutionary link between emotion and human dynamics (12).

1. Baily J (1985) *Musical Structure and Cognition*, eds Howell P, Cross I, West R (Academic, London).
2. Juslin PN, Laukka P (2004) Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *J New Music Res* 33(3): 217–238.
3. Winkler I, Háden GP, Ladinig O, Sziller I, Honing H (2009) Newborn infants detect the beat in music. *Proc Natl Acad Sci USA* 106(7):2468–2471.
4. Zentner MR, Eerola T (2010) Rhythmic engagement with music in infancy. *Proc Natl Acad Sci USA* 107(13):5768–5773.
5. Bergeson TR, Trehub SE (2006) Infants' perception of rhythmic patterns. *Music Percept* 23(4):345–360.
6. Phillips-Silver J, Trainor LJ (2005) Feeling the beat: Movement influences infant rhythm perception. *Science* 308(5727):1430.
7. Trehub S (2000) *The Origins of Music.* , Chapter 23, eds Wallin NL, Merker B, Brown S (MIT Press, Cambridge, MA).
8. Higgins KM (2006) The cognitive and appreciative import of musical universals. *Rev Int Philos* 2006/4(238):487–503.
9. Fritz T, et al. (2009) Universal recognition of three basic emotions in music. *Curr Biol* 19:1–4.
10. Ekman P (1993) Facial expression and emotion. *Am Psychol* 48(4):384–392.
11. Izard CE (1994) Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychol Bull* 115(2):288–299.
12. Scherer KR, Banse R, Wallbott HG (2001) Emotion inferences from vocal expression correlate across languages and cultures. *J Cross Cult Psychol* 32(1):76–92.
13. Darwin C (2009) *The Expression of the Emotions in Man and Animals* (Oxford Univ Press, New York).
14. Brown S, Jordania J (2011) Universals in the world's musics. *Psychol Music*, 10.1177/0305735611425896.
15. Huron D (1994) Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Percept* 11(3):289–305.
16. Parncutt R (1989) *Harmony: A Psychoacoustical Approach* (Springer, Berlin).
17. McDermott JH, Lehr AJ, Oxenham AJ (2010) Individual differences reveal the basis of consonance. *Curr Biol* 20(11):1035–1041.
18. Bidelman GM, Heinz MG (2011) Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *J Acoust Soc Am* 130(3):1488–1502.
19. Köhler W (1929) *Gestalt Psychology* (Liveright, New York).
20. Ramachandran VS, Hubbard EM (2001) Synaesthesia: A window into perception, thought and language. *J Conscious Stud* 8(12):3–34.
21. Maurer D, Pathman T, Mondloch CJ (2006) The shape of boubas: Sound-shape correspondences in toddlers and adults. *Dev Sci* 9(3):316–322.
22. Krumhansl CL (2002) Music: A link between cognition and emotion. *Curr Dir Psychol Sci* 11(2):45–50.
23. Bernhardt D, Robinson P (2007) *Affective Computing and Intelligent Interaction*, eds Paiva A, Prada R, Picard RW (Springer, Berlin), pp 59–70.
24. Hevner K (1936) Experimental studies of the elements of expression in music. *Am J Psychol* 48(2):246–268.
25. United Nations Development Programme Cambodia (2010) *Kreung Ethnicity: Documentation of Customary Rules* (UNDP Cambodia, Phnom Penh, Cambodia). Available at www.un.org.kh/undp/media/files/Kreung-indigenous-people-customary-rules-Eng.pdf. Accessed June 29, 2011.
26. Friberg A, Sundberg J (1999) Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. I. *Acoust Soc Am* 105(3):1469–1484.
27. Moelents D, Van Noorden L (1999) Resonance in the perception of musical pulse. *J New Music Res* 28(1):43–66.
28. Styns F, van Noorden L, Moelants D, Leman M (2007) Walking on music. *Hum Mov Sci* 26(5):769–785.
29. Leman M (2007) *Embodied Music Cognition and Mediation Technology* (MIT Press, Cambridge, MA).
30. Molnar-Szakacs I, Overy K (2006) Music and mirror neurons: From motion to 'e'motion. *Soc Cogn Affect Neurosci* 1(3):235–241.
31. Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized timing of isochronous sounds is represented in neuromagnetic β oscillations. *J Neurosci* 32(5):1791–1802.
32. Nozaradan S, Peretz I, Missal M, Mouraux A (2011) Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31(28):10234–10240.
33. Janata P, Tomic ST, Haberman JM (2012) Sensorimotor coupling in music and the psychology of the groove. *J Exp Psychol Gen* 141(1):54–75.
34. Koelsch S, Siebel WA (2005) Towards a neural basis of music perception. *Trends Cogn Sci* 9(12):578–584.
35. Bryant GA, Barrett HC (2008) Vocal emotion recognition across disparate cultures. *J Cogn Cult* 8(1-2):135–148.
36. Sauter DA, Eisner F, Ekman P, Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc Natl Acad Sci USA* 107(6):2408–2412.
37. Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 70(3):614–636.
38. Thompson WF, Balkwill LL (2006) Decoding speech prosody in five languages. *Semiotica* 2006(158):407–424.
39. Elfenbein HA, Ambady N (2002) On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol Bull* 128(2):203–235.
40. Marie C, Delogu F, Lampis G, Belardinelli MO, Besson M (2011) Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *J Cogn Neurosci* 23(10):2701–2715.
41. Zatorre RJ, Baum SR (2012) Musical melody and speech intonation: Singing a different tune. *PLoS Biol* 10(7):e1001372.
42. Smalley D (1996) The listening imagination: Listening in the electroacoustic era. *Contemp Music Rev* 13(2):77–107.
43. Susemihl F, Hicks RD (1894) *The Politics of Aristotle* (Macmillan, London), p 594.
44. Meyer LB (1956) *Emotion and Meaning in Music* (Univ of Chicago Press, Chicago, IL).
45. Truslit A (1938) *Gestaltung und Bewegung in der Musik. [Shape and Movement in Music]* (Chr Friedrich Vieweg, Berlin-Lichterfelde). German.
46. Iyer V (2002) Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Percept* 19(3):387–414.
47. Gagnon L, Peretz I (2003) Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cogn Emotion* 17(1):25–40.
48. Eitan Z, Granot RY (2006) How music moves: Musical parameters and listeners' images of motion. *Music Percept* 23(3):221–247.
49. Juslin PN, Lindström E (2010) Musical expression of emotions: Modelling listeners' judgments of composed and performed features. *Music Anal* 29(1-3):334–364.
50. Phillips-Silver J, Trainor LJ (2007) Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition* 105(3):533–546.
51. Fitch WT (2006) On the biology and evolution of music. *Music Percept* 24(1):85–88.
52. Fleming W (1946) The element of motion in baroque art and music. *J Aesthet Art Crit* 5(2):121–128.
53. Roseman M (1984) The social structuring of sound: The temiar of peninsular malaysia. *Ethnomusicology* 28(3):411–445.
54. Ames DW (1971) Taaken sàmàarii: A drum language of hausa youth. *Africa* 41(1):12–31.
55. Vines BW, Krumhansl CL, Wanderley MM, Dalca IM, Levitin DJ (2005) Dimensions of emotion in expressive musical performance. *Ann N Y Acad Sci* 1060:462–466.
56. Russell JA (1980) A circumplex model of affect. *J Pers Soc Psychol* 39(6):1161–1178.
57. Laban R (1975) *Laban's Principles of Dance and Movement Notation*. 2nd edition, ed Lange R (MacDonald and Evans, London).
58. Stalinski SM, Schellenberg EG (2012) Music cognition: A developmental perspective. *Top Cogn Sci* 4(4):485–497.
59. Trehub SE, Hannon EE (2006) Infant music perception: Domain-general or domain-specific mechanisms? *Cognition* 100(1):73–99.
60. Gabrielsson A (2002) Emotion perceived and emotion felt: Same or different? *Music Sci* Special Issue 2001–2002:123–147.
61. Juslin PN, Västfjäll D (2008) Emotional responses to music: The need to consider underlying mechanisms. *Behav Brain Sci* 31(5):559–575, discussion 575–621.
62. Hagen EH, Bryant GA (2003) Music and dance as a coalition signaling system. *Hum Nat* 14(1):21–51.
63. Phillips-Silver J, Keller PE (2012) Searching for roots of entrainment and joint action in early musical interactions. *Front Hum Neurosci*, 10.3389/fnhum.2012.00026.
64. Wheatley T, Kang O, Parkinson C, Looser CE (2012) From mind perception to mental connection: Synchrony as a mechanism for social understanding. *Social Psychology and Personality Compass* 6(8):589–606.
65. Janata P, Grafton ST (2003) Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. *Nat Neurosci* 6(7):682–687.
66. Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: Auditory-motor interactions in music perception and production. *Nat Rev Neurosci* 8(7):547–558.
67. Dehaene S, Cohen L (2007) Cultural recycling of cortical maps. *Neuron* 56(2):384–398.