Gesture and Embodied Metaphor in Spatial Music Performance Systems Design

Ricky Graham
Stevens Institute of Technology
Hoboken, NJ 07030
USA
rgraham1@stevens.edu

Brian Bridges
University of Ulster
Magee campus, Derry~Londonderry
Northern Ireland, UK
bd.bridges@ulster.ac.uk

ABSTRACT
This paper describes the design, theoretical underpinnings and development of a hyperinstrumental performance system driven by gestural data obtained from an electric guitar. The system combines a multichannel audio feed (parsed for its pitch contour, spectral content and note inter–onset time data) with motion tracking of the performer's larger–scale bodily movements using a Microsoft Xbox Kinect sensor. These gestural materials provide the basis for the system's musical mapping strategies, informed by an integration of embodied cognitive models with electroacoustic/electronic music theory (specifically, Smalley's *spectromorphology*). The performance system's sound processing is further animated using the *boids* flocking algorithm by Reynolds. This provides an embodied/ecological base for connecting Lerdahl's spatial and syntactical models of tonal harmony with sound spatialization and textural processing. Through this work, we aim to advance broadly applicable *performance gesture ecologies*, providing typologies that facilitate creative (but still coherent) mappings from physical and *figurative* performance gestures to spatial and textural structures.

Keywords
Gesture, embodied, schemas, boids, flocking, spatialization, mapping

1. INTRODUCTION
Investigations concerning the form, role and meaning of 'gesture' preoccupy multiple academic domains, permeating technical design, philosophy and experimental investigations into human cognition, action and experience. In the music sphere, this interest is manifested in the use of embodied cognition to interrogate theoretical models within musicology [1,2]**.** Furthermore, researchers in the field of musical interaction design continue to explore the practical implications of gestural models, exploiting concepts of affordances and embodied models in systems design [3,4**]**. This paper will discuss the design features in an early iteration of a spatial music performance system whose constituent mappings and control structures adopt bodily bases, supporting the exploration of relationships between spatiotemporal performance gestures on both physical and *figurative* [3] (perceptual integration of individual sounded acts) planes. (See also Smalley's [5] discussion of *gestural surrogates* for a similar concept within the field of electroacoustic music theory.)

We propose that through the use of bodily–driven models in systems design, a computer music performer (or hyperinstrumentalist) may creatively explore compatible embodied bases for connections between pitch, spatialization and creative sound processing. We suggest that practice–led investigations of embodied concepts in music will allow us to advance *performance gesture ecologies* that provide unifying frameworks for connecting individual performance gestures and musical macrostructure, facilitating intuitive engagement with complex performance systems.

2. EMBODIED COGNITION AND GESTURAL TYPOLOGIES AS A BASIS FOR SPATIAL MUSIC PERFORMANCE SYSTEMS DESIGN

2.1 Familiar Gestural Affordances in Performance Systems Design
Our performance system consists of a hyperinstrumental expansion of an electric guitar. It uses a multichannel pickup on the guitar to facilitate (a) the tracking of musical event data from the guitar, and (b) the separate processing and spatialization of different voices from the guitar's (channel–per–string) audio feed (see Figure 1). The tracking and processing parts of the system are implemented in Pure Data (Pd).

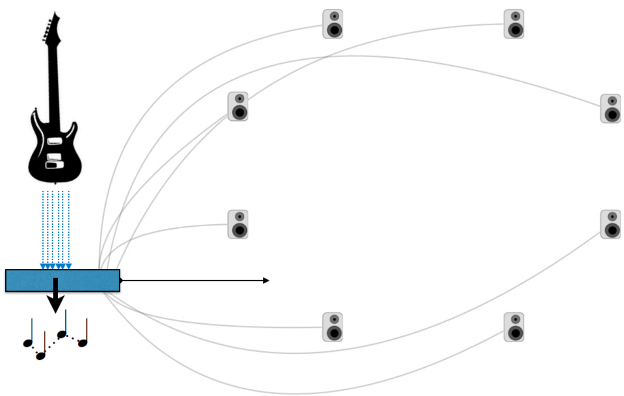

Figure 1: Basic system hardware configuration: multichannel audio pickup facilitates parsing and spatialization of voices

The central rationale behind the system's initial design was that it should provide the musician with the opportunity to easily derive control of spatialization from familiar performance gestures**.** Therefore, we based our system on an electric guitar whose physical affordances were essentially unmodified: there were no 'bolt–ons' of additional control surfaces or sensors [6]. Our perspective was that deviating from instrumental norms would impose additional cognitive loads when learning a complex spatial music performance system. In addition, they may undermine a performer's established technical strengths, due to emerging issues of control complexity and fragmentation.

As a result, the multichannel audio feed becomes our primary source of control data through the extraction of note–gesture (event) information, including pitch–class, spectral content and note inter–onset data for each voice. This atomistic data is integrated to provide gestural macrostructures (i.e. figurative gestures [3]) such as pitch contours and average note inter–onset time that can then be directly applied to musical mappings.

2.2 Mapping, Macrostructure and Coherence

2.2.1 Pitch contour as gestural data

The system's core design philosophy is that coherence and accessibility is facilitated through the use of broadly isomorphic embodied models at its various levels, from input event parsing to its output audio processing. *Figurative* instrumental data is extracted in Pure Data (Pd), including pitch (frequency Hz), note (attack) inter-onset time (ms), amplitude (dB), macro–melodic contours (and computation of syntactical relationships such as melodic tension). Initial developments concentrated on applying pitch–contour data to the control of first–order ambisonic spatialization azimuth direction and distance parameters. The system's pitch–tracking (via the monophonic channel–per–string audio feeds) was implemented using the time-domain, 'Specially Normalized AutoCorrelation' or SNAC–based [helmholtz~] object for Pd [7]. This was then parsed into pitch–class data (for twelve pitch–class divisions, *pc0–pc11*, in relation to a user–defined tonal center, *pc0*).

In our first version, *pc0* to *pc11* were mapped to a 360–degree rotation on the azimuth plane (clockwise from front center). Its spatial dynamics were reinforced by our distance mapping, which is derived from cognitive models of tonal hierarchies (see Figure 2).

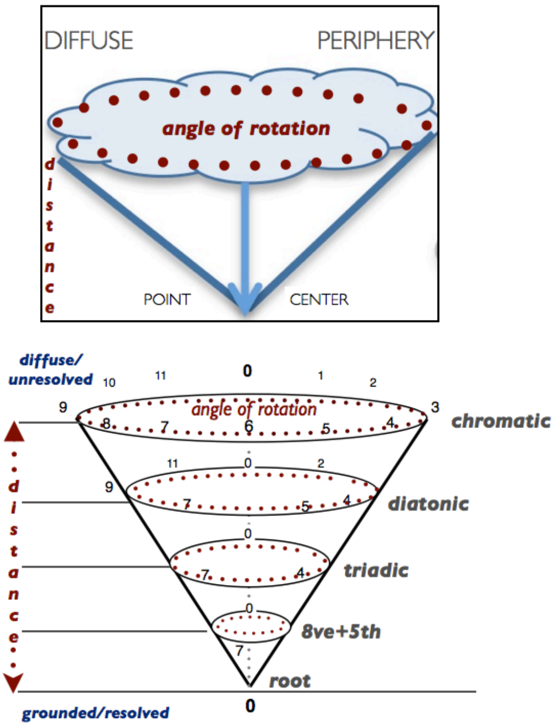

Figure 2: the initial system's mapping from a spatial model of tonality—after Lerdahl [9]—to spatialization (center–periphery and angle of rotation)

Music cognition research [8] and cognitively–informed music theory [9] suggests that different intervals within tonal structures may occupy distinct levels for tonal 'distance' within a hierarchically–divided cone structure (a *basic space* for tonal relations). These distance factors are thus available for mapping to the ambisonic distance parameter, connecting tonal center–periphery with spatial dynamics.

In terms of the embodied associations of this configuration, *pc0* is treated as stable: centered/grounded/resolved. Deviations from this pitch–class are conceptualized as increasing hierarchical (and functional) distance, with associated greater instability. As a result, pitch materials closer to tonal center (octaves, fifths, triads with respect to *pc0*) are spatialized towards the center; materials which are more tonally distant (chromatic level) are spatialized towards the periphery. The tonally distant materials (commonly judged to be more 'dissonant') are 'contained' at the periphery. The materials closer to pc0 reinforce a sense of a grounding center. Thus, spatial ideas implicit within pitch structures are reified in our spatialization strategy. This may facilitate the performer's exploration of *gestural narratives* between pitch contours and the system's embodied–ecological response.

2.2.2 Dynamic mappings: embodied forces and performative frames

In addition to their application to basic contour mappings, pitch–class profiles form the basis of more sophisticated mapping strategies based on Lerdahl's dynamic models of tonal relations [9,10]. Although Lerdahl's primary influence is that of 'traditional' cognitive psychology [8], his theories can also be viewed as implicitly embodied. His dynamic models treat musical syntax dynamics as forces, including an explicit invocation of gravitation via an inverse–square law [9,10]. This use of embodied concepts provides the potential for integration with more explicitly embodied approaches. Following this lead, dynamic melodic syntax data (such as attraction and inertia values) from Lerdahl's models were mapped to in–kind parameters (namely, *attraction* and *inertia*) of a *boids* flocking algorithm, controlling dynamic spatialization [11,12,13,14] relative to a specified central point within two–dimensional Cartesian space.

Although this algorithm has previously been applied to spatialization [13], our innovation [14] lies in its integration with embodied force–dynamic metaphors which we identify as common to Lerdahl's [9,10] model and the *boids* algorithm. Resulting sound structures are *animated* spatially using *boids* [11] as an embodied metaphor for melodic syntax. Each *boid* controls the movement of a single voice from the multichannel instrument output, with strength of tonal attraction reflected in the overall flock's centricity or peripherality (see Figure 3).

Spatial dynamics are thus subject to control via an embodied/environmental metaphor, with individual spatialization parameters consolidated within a holistic model connecting tonality with embodied spatial associations. These parametric mappings could therefore be viewed as a practical implementation of Johnson's embodied cognitive theory of common practice musical structures: his *music–as–moving–force* metaphor [15]. The mapping of tonal attraction values to flocking centricity is illustrated in our first video example: [video example 1: 11'05–16'21].[1] In addition, Lerdahl's *implicative denial* (denial of potentially more attractive pitch–class candidates) [9,10] can be observed to be associated with more vigorous flocking behaviors, such as greater acceleration and avoidance values for the *boids* (and, hence, the spatialized voices).

[1] http://www.youtube.com/watch?v=XO1n_GS9I5w
Note that all video examples are stereo reductions: ambisonic B-format decoded to standard stereo.

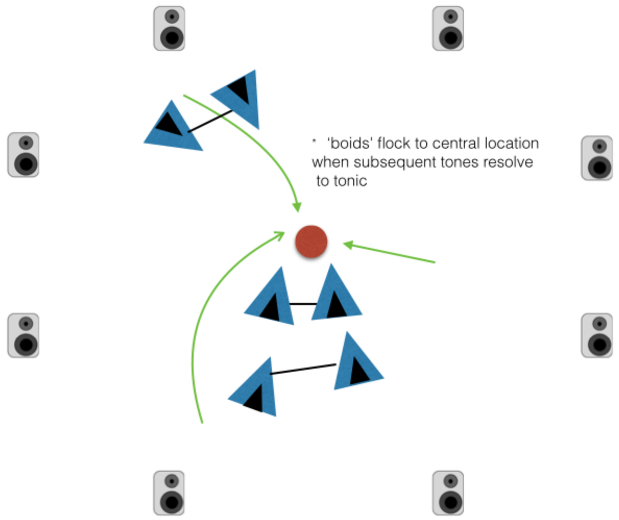

Figure 3: *boids* mappings resolve tonal center to spatial center (providing 'spatial closure')

3. GESTURAL QUALITIES AND BODILY BASES FOR EXTENDED MAPPING STRATEGIES

3.1 Ancillary/Accompaniment Gestures and Embodied Functional Associations

The first iteration of the system is thus based on facilitating accessible connections between figurative gestures and spatial metaphors via tonal hierarchy models**.** Connections between musical structures and embodied spatial domains may be further explored and strengthened through the extraction and mapping of additional gesture data, specifically indirect *ancillary/accompaniment* gestures [3]. These are bodily movements that a performer may consciously or unconsciously execute alongside sounded performance gestures. They may be conceptualized as embodied accompaniments to musical structures. Although such usage entails the presence of additional input controls, we consider the application of such gestures to mappings to be broadly compatible with our initial aim of maintaining clear connections with familiar performance gestures. In addition, such gestures are already broadly accessible as by–products of established performance practice. They are therefore less likely to be experienced by the performer as contributing to fragmented control and impeding their musical execution.

Structures based on accompaniment gestures are obtained using a combination of the Xbox Kinect and the third-party application, *Synapse* [16]. This provides skeletal point–based data, including values for velocity and acceleration. The result of this extension is that we are now able to access a combination of gestural types—(1) physically small–scale sounding gestures (either as individual note articulations and composite figurative gestures) alongside (2) more expansive bodily *accompaniment/ancillary* gestures—facilitating more holistic interpretations and mappings via force metaphors (see Figure 4).

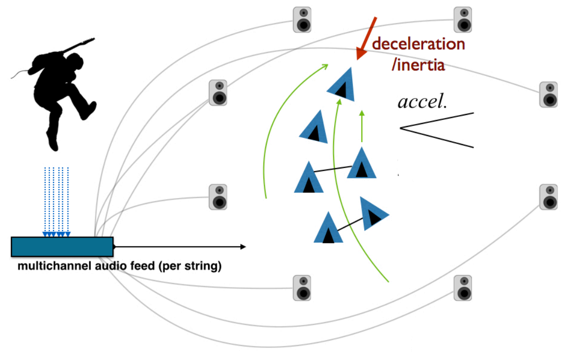

Figure 4: physical motion connected with musical motion (integration of figurative and ancillary gesture)

An example application of ancillary bodily movements may be seen in [video example 2].[2] In this example, the velocity of larger–scale movements are applied to various parameters of granular processing corresponding to a *rate–effort–to–density* dynamic. Movements are mapped as follows:

(1) The velocity of the left hand (at the head stock of the guitar) controls the flocking speed

(2) The velocity of the right hand (picking hand) is mapped to the following granular parameters: feedback, buffer position, grain reveal and time variation

(3) Other bodily accompaniment gestures (higher–amplitude 'waving' gestures with the picking hand, tracked via skeletal velocity) are used to instantiate more dynamic spatial movements around the central attraction flocking point, that is determined by the torso position of the performer.

In addition, figurative gestural materials also articulate the *rate–effort–to–density* approach via the mapping of average note–inter–onset to granular density and grain size.

Fundamentally, this mode of interaction can be seen as relational: functional dynamics of engagement with the environment are embodied within performance gestures. Johnson [15] has proposed a typology of *qualitative* dimensions of movement: *tension, projection* and *linearity*. These dimensions deal with the connection between the manner of the movement's initiation and the form of the resulting gesture. We believe it is significant that this typology bears a striking resemblance to Smalley's [5] account of *energy–motion profiles* in electroacoustic music see table 1, below.

Table 1. Comparison of Johnson's *dimensions of movement* with Smalley's *energy–motion profiles* and embodied associations

Johnson	Embodied Association	Smalley
Tension	Rate–effort=>overcoming inertia	*Motion rootedness*
Projection	Sudden rate-change / transient movement	*Motion launching*
Linearity	Coherence of path	*Contour energy/inflection*

Tension and *motion–rootedness* are correlated with an embodied expectation (*force–dynamic*) of the effort required to overcome inertia: the persistence of a system's grounded/stable state (in our case, primarily associated with tonal and spatial

[2] http://www.youtube.com/watch?v=5wEgpVuB9-w

center). *Projection/motion–launching* implies the significant rate–effort may instantiate a larger–scale dynamic movement, the degree of which may dictate the form that the continuing gesture's *linearity/contour–energy* takes (more coherent or incoherent path away from or back to its initial resting state). Given the broad correspondence between these two theories, they may be seen as contributing a shared *performance gesture ecology* for mappings connecting physical gesture with musical macrostructures ranging from common practice tonal syntaxes [15] to textural and electroacoustic approaches [5].

3.2 Ancillary/Accompaniment Gestures and Integrated Mappings

[Video example 3][3] adopts the presented typology to integrate input modalities and output domains. In part I, the presence/absence of attack transients may be conceptualized as mapping via *projection* (see Figure 5) to spatial trajectory.

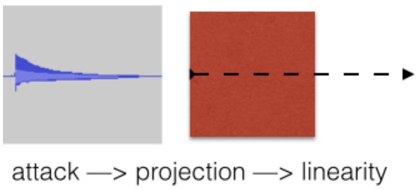

Figure 5: Guitar pluck (attack profile) causes sound object to be projected spatially

Rapid rates of change (clearly detected attack transients) overcome the implied 'grounding' (gravitational) dynamics of current states/positions. In this example, the clear articulation of a note will reset the voices to the central location, whilst the sustain phase will instantiate spatial diffusion processes (along with some additional tonal–textural processing). In Part II, we see a *boids* mapping in which the movement of the guitar head stock controls the *speed* of flocking behaviors and the picking hand's ancillary motion controls *inertia* and *avoidance* parameters. In such a mapping, the detection of a new note–onset (increasing *gesture–energy*) will center the instrument's output in the spatial array and present it with tonal/textural clarity. The continuant phase of each note event will see the note–gesture(s) being moved progressively towards a peripheral location, based on a metaphor of lower–energy articulations being assigned to the periphery. Thus, a synthesis of Johnson's and Smalley's theories may be seen as informing the use of audio analysis techniques (such as note–attack detection, envelope following and spectral analysis tools) as source structures for integrated embodied mappings alongside ancillary/accompaniment gestures.

4. CONCLUSION

By advancing a unifying framework (*performance gesture ecology*) that links together figurative, accompaniment and physical sounded gesture, we provide an intuitive means of integrating familiar performance gestures and metaphors into the design of complex spatial music performance system. We recognize that there are some inherent limitations in creating such schemas as fixed models. For example, our choice of reflecting tonal *center–periphery* directly in the spatial structure is, from one perspective, an aesthetic decision, and alternate mappings are certainly conceivable. However, the present version of our *center–consonant and peripheral/distant–dissonant* mapping is consistent with cognitive studies carried out by Krumhansl [8], in addition to Lerdahl's theoretical extension [9,10]. Furthermore, the correspondence between Johnson's typology of qualitative dimensions of movement and Smalley's energy–motion profiles may fruitfully inform the investigation of embodied perspectives on gestural mapping strategies. Future work will see the continued exploration of how embodied schemas and dynamics may provide coherent mappings via the exploitation of in–kind and isomorphic relationships between a performance's physical and figurative domains.

5. REFERENCES

[1] L. M. Zbikowski. *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford University Press USA, New York, 2002.

[2] C. Brower. A Cognitive Theory of Musical Meaning. *Journal of Music Theory*, 44, 2 (Autumn 2000), 323–379.

[3] C. Cadoz and M. Wanderley. Gesture–Music. In: M. Wanderley, M. Battier (eds*) Trends in Gestural Control of Music*. IRCAM, Paris, 2000, 77–8.

[4] K. Wilkie, S. Holland and P. Mulholland. What Can the Language of Musicians Tell Us about Music Interaction Design? *Computer Music Journal*, 34, 40 (Winter 2010).

[5] D. Smalley. Spectromorphology: explaining sound–shapes. *Organised Sound*, 2, 2 (August 1997), 107–126.

[6] R. Graham and B. Bridges. Mapping and Meaning: embodied metaphors and non–localised structures in performance system design. In: *Re–New 2013 Conference Proceedings* (ISSN: 2245-7801): available online at: http://www.l--l.dk/downloadables/re-new_2013_conference_proceeding.pdf

[7] K. Vetter. Helmholtz~ [software library], 2011. Available at: http://www.katjaas.nl/home/home.html

[8] C. Krumhansl. *Cognitive Foundations of Musical Pitch.* Oxford University Press, Oxford, 1990.

[9] F. Lerdahl. *Tonal Pitch–space*. Oxford University Press, Oxford, 2001.

[10] F. Lerdahl and C. Krumhansl. Modeling Tonal Tension. *Music Perception*, 24, 4 (2007), 329–366.

[11] C. Reynolds. Flocks, Herds and Schools: A distributed behavioral model. *SIGGRAPH Comput. Graph*., 21, 4, (1987), 25–34.

[12] E. Bates. *The Composition and Performance of Spatial Music.* Ph.D. Thesis. University of Dublin, Trinity College, Dublin, 2010.

[13] E. Singer. Boids for Max. [software library], 1997. Available at: http://ericsinger.com/cyclopsmax.html

[14] R. Graham. *Expansion of Electronic Guitar Performance Practice through the Application and Development of Interactive Digital Music Systems*. Ph.D. Thesis. University of Ulster, Northern Ireland, 2012.

[15] M. Johnson. *The Meaning of the Body: Aesthetics of Human Understanding*. University of Chicago Press, Chicago, 2007.

[16] R. Challinor. Synapse [software application], 2011. Available at: http://synapsekinect.tumblr.com/Synapse

[3] http://www.youtube.com/watch?v=jxWBDaSSX2w