Towards a Mappable Database of Emergent Gestural Meaning

Doug Van Nort
DisPerSion Lab
School of the Arts, Media, Performance & Design
York University, Toronto, ON
vannort@yorku.ca

Ian Jarvis
DisPerSion Lab
School of the Arts, Media, Performance & Design
York University, Toronto, ON
jarv314@yorku.ca

Michael Palumbo
DisPerSion Lab
School of the Arts, Media, Performance & Design
York University, Toronto, ON
info@palumbomichael.com

ABSTRACT

This paper presents our work towards a database of performance activity that is grounded in an embodied view on meaning creation that crosses sense modalities. Our system design is informed by the philosphical and aesthestic intentions of the laboratory context within which it is designed, focused on distribution of performance activity across temporal and spatial dimensions, and expanded notions of the instrumental system as environmental performative agent. We focus here on design decisions that result from this overarching worldview on digitally-mediated performance.

Author Keywords

gesture database, emergence, distributed performance, mapping, improvisation

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing

1. INTRODUCTION

As noted in [16], the term "gesture" has been used often, but with a variety of meanings and in a variety of applications, within the NIME community. The author claims that the term has primarily been used in one of three ways: to describe an act of communication, of control, or as a metaphor for action from the physical world. A typological approach to examining musical performance gestures in their varying interpretations, particularly in regards to communication and control, is presented in [7]. Meanwhile, the metaphorical concept of the "musical gesture" has seen a great deal of attention in the context of music theoretic analyses [14][13], with a focus on expressivity as uncovered within the structure of musical compositions. As evidenced by the collection edited by Leman and Godoy [12], researchers who examine gesture through an embodied music cognition lens have advanced the conversation on the linkages between these interpretations of gesture, as they manifest in physical action as well as in musical materials. We feel that perhaps most important to this embodied cognitive view is the recognition of the duality that exists between intention and reception as they are situated between performer and audience, between collaborating performers and indeed between the performer and themselves. This intentionality, which can cross modalities between movement and sound, can be hard to locate in the continuum between goal-oriented actions and the emotive substrate of what we might call an "expressive gesture". It is therefore understandable that Jensenius [15] has favoured action-related terms to gesture-related ones, and concepts such as "action-sound coupling" to "gesture mapping". We recognize and agree that there is indeed a gap that exists between the action that manifests in both kinetic and sonic realms in the context of musical performance and gestural meaning that arises in context, as a product of a specific work, and indeed in the minds of a specific audience member. We further recognize that being careful with such language is important in conceptual framing of a NIME-related system design or artwork.

Having said this, we focus here precisely on gesture as an emergent phenomenon that results from viewing and listening to a performed work in context, where something one might receive as "gesture" is reinforced by the unfolding of embodied activities and events over time, building what has been called an action-oriented ontology [18]. This work reports on our efforts to develop a framework whose goal is to segment, capture and build a database of temporal units of performance information that might be considered to have gestural meaning. These include sonic and kinetic actions, and go beyond this to any sensed temporal information that arises in performance. The goal is to scale upwards towards a large dataset of such units in a fashion that they may be analyzed, compared, retrieved and re-injected into other performance contexts wherein they might develop new gestural meaning, perhaps referencing some trace of their original intention. This work is intended to serve both the analysis of past performances as well as a platform for the creation of new works. The laboratory context within which this work is being developed fluidly moves between analysis-oriented research examination of embodied human action, perception and cognition in situations of distributed creativity (e.g. free improvisation), as well as the creation of new works that merge the performing and computational arts. To this end, our developing platform utilizes distributed and web-based modes of representation and interfacing, so that we may scale outwards to other laboratory contexts with the hopes that they might contribute to a growing body of gestural information, while being free to apply analyses related to their own research-creation questions to the emergent gestural database. We report here on our developing framework, and the laboratory context within which the concept has been framed and is currently being developed.

NIME'16, July 11-15, 2016, Griffith University, Brisbane, Australia.

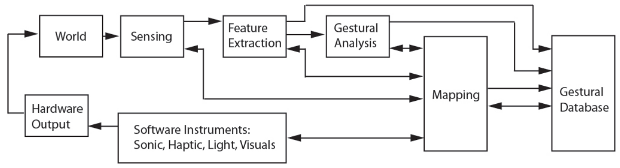

Figure 1: DisPerSion lab ecosystem

2. GESTURALITY AND DISTRIBUTED PERFORMANCE

The concept of the emergent gestural database arises from the first author's many years of practice in electro-acoustic free improvisation in contexts that integrate both instrumental systems and machine improvising systems [28], and often integrates movement and visual media. This has led to research into the understanding of how meaning is constructed in situations where there is no score or schema, and how awareness of such can be used as a framing concept in the design of interactive systems that augment and support these performance endeavours [27]. This work has evolved into the founding of the DisPerSion lab [1], where the research is expanding outward from the paradigm of the solo performer in a music-centric context, towards the notion of environments that possess performative agency in the interaction between collectives of improvisers, in contexts that integrate a variety of sound and movement based practices.

In this expanded view, the concept of distributed performance is applied to encompass both the notion of distributed creativity across performers as well as between human performers and computational agents. In both cases, this distribution of creativity is potentially across space (telematics) as well as across time (engagement between past and present performance contexts). Following an embodied cognitive view and the concept of an action-oriented ontology, the concept of the gesture – as a phenomenon that might manifest in movement, sound or other temporal forms – becomes the fundamental point of access for design of the larger laboratory environment-as-instrumental agent.

The design of the lab infrastructure has been constrained by this approach to digitally-mediated collective improvisation. The desire to engage students and visiting artist-researcher collaborators with varying levels of technical expertise, as well as the need to move fluidly between input/output media types has led to a modular design, as depicted in figure 1. Embedded within this, we see a classic signal flow of sensing, feature analysis, mapping and instrumental output via some sort of hardware display. However, the configuration here is expanded to accommodate a database of gestural information, which may be queried and whose values may be mapped across modalities to software "instruments" that generate sound, light, haptics, or visuals.

2.0.1 *Flexible Representation*

Much like concatenative systems such as cataRT [25], which develops a signal-driven timbre space that may be queried and retrieved as replacement for control input, we focus on a gesture space of temporal units which may be retrieved and re-injected into a performance event. However the purpose of a system like cataRT is to represent a fundamental concrete unit (i.e. the audio grain) by a vector of descriptors, as a means to retrieve this unit. In our case, we represent a series of parallel streams of time-tagged information. Each parallel stream is considered equivalent and regarded as a *unique window view into the gestural unit.* While we find inspiration with SDIF and even more so the GDIF project [17], we differ in our open approach to creating linkages of information about performance actions. It is not hierarchical structures of temporal streams from a common source we are examining, but rather the *concurrence and coarticulation of potentially gesture-relevant phenomena, possibly from geographically distributed sources.* Indeed, the key is that each one of these streams co-occurred in time, and were grouped together by virtue of a chosen configuration. For example, all data streams might be chunked separately every 1000 ms, or the onset and offset of an audio stream might trigger the segmentation of a large set of data streams from all performers. This framework is informed by Godoy's work on chunking, which examines the perceptual fusion of disparate action-sound trajectories by virtue of concurrence and proximity in both time and space [11]. This work also reaffirms the importance of meso-level events in the range of 0.5 to 5 seconds, which in practice has guided the scale of gestural events that we have examined thus far.

As with any analysis task, segmentation remains an open and important question. Following our agnostic approach, we currently apply a variety of segmentation and extraction algorithms. This is in light of our goal to discover kinetic and sonic trajectories that might be considered to be subsumed within one gestural profile (sub-chunks), that are connected by larger contexts at the level of sections and movements (supra-chunks), and that share similarity with activity from other performance contexts. In building towards this larger goal, currently all streams are tagged with an association of their source (i.e. IMU sensors, audio stream, etc.), their method of segmentation, the global mapping information from the performance situation, a timestamp relative to some beginning of performance time, and a global unix time stamp. This flexible approach to representation, and the desire to filter and run comparative measures between gestural units that may be located in disparate physical locations has informed our choice of database design.

3. THE GESTURE DATABASE

The database portion of the lab infrastructure are conditioned by the above goals, and the creative practices of the DisPerSion Lab related to scalable distributed collaboration – from small to large ensembles across local networks and the internet. Beyond scalability, this includes an agnostic UI (in regards to systems/instruments/sensors being mapped to the database); online use in performance (querying both immediate data that is captured as well as all historical data in the database) and computationally intensive offline analysis. Our design is towards the end of a *performable database*, which we conceive as an active agent within the performance ecosystem that extends the embodied knowledge available for both human and computational agents in the lab environment.

We design the database so as to facilitate web-based collaboration and performance in recognition of current practices such as those found in [20] [24] [29], and scalability in anticipation of large-scale web-based collaboration such as that found in [10]. Our specific focus on gestural action is informed by current gestural database frameworks found in the work of [23] and [9], yet our expanded notion of gesture as a dynamic semantic unit drives our desire for open-ended ontologies and scalability. This includes scalability in regards to the potential amount of data that is captured, analyzed, compared, archived and re-injected into current and future performance contexts, as well as openness in regards to the number and type of devices that are connected to the database.

In an extreme case, such a "device" can include a live coding system [29], with such an example depicted in figure 2, which demonstrates a situation in which live coding is used as a control of the sensor-sound mapping and control conditioning. In this example, we commit both live coding keystrokes (which effectively represent meta-knowledge about mapping) as well as gestural output from the sensor device into the gesture database. This limit case problematizes the conceptualization of hierarchical structures of temporal streams as each stream becomes a node of potential infinite regress into nested computational abstractions networked across numerous, potentially dynamic, geographically dispersed locations; all of which becomes potentially relevant gestural information. That said, intention and meaning is found in the execution and co-articulation: keystrokes invoke action-sound potential, and knowledge of this as well as the concurrence of the sensor-based and textural actions is preserved for future analysis and querying. Meanwhile, we describe a more classic data capture situation in section 5, while the concept of committing meta-knowledge about the mapping state is discussed in the next section.

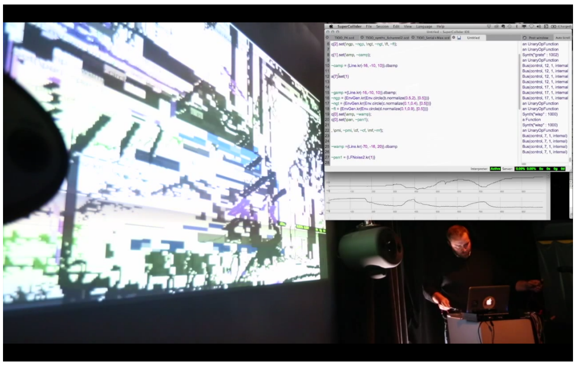

Figure 2: Integration of physical action and live coding as gestural units for the database

In regards to the platform for an interactive and web-accessible performable database, one finds two competing frameworks currently in use that are divided by requirements for how the data is structured. In traditional relational databases it is necessary to know how the data (or schema) is structured before you can add data to the database, which can increase query times and code complexity [5]. This has led to the development of a highly refined Structured Query Language (SQL) for managing database tasks. By contrast, on the rise is the use of noSQL, or Not Only Structured Query Language to account for the vast increase in non-structured, and semi-structured data. In our initial design of the performable database we have implemented the noSQL database MongoDB [8], using node.js [26] for the server architecture. This model has been successfully implemented in a project focused on large-audience performances [10] as well as one which required an increasingly large physiological sensor database [6].

This is also highly relevant in our application, given our desire to scale outward to telematic engagement with data from partner sites that collectively contribute to (and query from) the larger database. Rather than using tables (columns and rows) as in SQL databases, MongoDB commits data to the database in documents based on JSON (JavaScript Object Notation) for flexible and schemaless data models. Documents and sub-documents, can be inserted, created, or modified at different levels of hierarchy, by several different users, without imposing a data structure. As noted, scalability in both the number of users and the amount of data is of particular importance to our design of the performable database, and MongoDB avoids scaling issues through horizontal scaling (adding more machines), which can occur without limit. By contrast, SQL databases scale vertically (increase in the CPU and RAM of a single machine) and thus are often limited to the power of a single machine [5]. Meanwhile, node.js [26] is based on Chrome's JavaScript runtime, and is designed to efficiently build scalable networked applications. We take advantage of MongoDB's efficient integration of Javascript to enhance performing database requests in order to reduce the initial overhead of resources and streamline connection between networked databases. In practice, we run search queries in order to reduce the set of relevant data, which is then brought into Max/MSP for analysis and similarity comparisons. The flexibility in design of our approach is oriented towards future linkage with other software for analysis such as Python for machine learning and R for temporal and cross-correlation analysis.

4. DATABASE MAPPING IN THE NETWORK

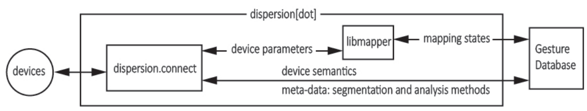

Figure 3: dispersion[dot]: integrating device and mapping meta-data into database framework

As a means of integrating performative laboratory activity and the database, we have created a suite of software-based machine learning and automation tools to facilitate bidirectional dataflow throughout database and mapping states. Named *dispersion[dot]*, this toolkit is attentive to how the disparate elements of the lab's ecosystem are connected, including when and how mapping structures have changed, the original context of the data being stored, and in what contexts it has since been used. At present, *dispersion[dot]* is built around *libmapper* [19], an open-source mapping toolkit whose primary components, the *map.device*, *map.out*, and *map.in*, can be embedded in a program in order to represent any device and its respective input and output parameters. We currently utilize libmapper because it enables us to rapidly connect the outputs of any map.device to inputs of any other map.device, operates stably, is open-source and can be implemented in projects written in Max/MSP, Python, C, C++, Java, and Pd.

While libmapper is adept at facilitating the cross-coupling of parameter data, it has limited use in situations which require the introduction of new devices at the real-time speed of improvisation, as well as the use of data types apart from floats and integers. Such constraints are understandable given that libmapper began as a toolkit for flexible parameter mapping during the prototyping phase of new instrument design. We detail two components of *dispersion[dot]* which address these limitations.

Operating alongside a running instance of libmapper, *dispersion.connect* actively listens on a given port for any external device that is already capable of sending OSC, and establishes a connection to libmapper by scripting a *map.device* that is recursively optimized based on learned semantic information, metadata, and usage patterns specific to its agent source. This is particularly useful towards the inclusion of a wireless OSC-enabled device whose source code may be inaccessible, thus preventing a user from embedding a *map.device*. Furthermore, incoming address patterns are checked against others stored in the current session, and if found to be unique, are scripted as a new map.out. *Dispersion.connect* includes a subsystem named *dispersion.bridge* that connects machines that do not already have an OSC implementation, nor have an installation of libmapper. It is presently available in Max/MSP using only vanilla objects. *Dispersion.bridge* takes parameter names as arguments, passing them over multicast to a subsystem of the *dispersion.connect* instance in order to script a respective *map.device*.

Drawing on DisPerSion Lab's objective of creating open and flexible topologies, *dispersion[dot]* has been conceived as agnostic to data structures that are idiosyncratic to the environments of our remote collaborators in distributed performance contexts. To facilitate this we integrate a range of datatypes, such as trees or strings, towards the end of committing to the database sessional meta-knowledge of devices, mapping states, relevant namespace [4] content, changes to code, and how connected devices influenced each other. We do this in recognition of the fact that any meaning arising from a gestural language developed within mediated performance is not only due to temporal reinforcement of an action-oriented ontology, but further is due to the spatial and semantic cross-coupling of all agents of sensing and display within the interactive performance ecosystem that gives rise to them. This is in awareness of what Paine has recently referred to as the techno-somatic dimension [22] of a mediated performance environment. With this in mind, we therefore integrate mapping as meta-information about the performance state, and commit this to the database in linkages with gestural-temporal semantic units.

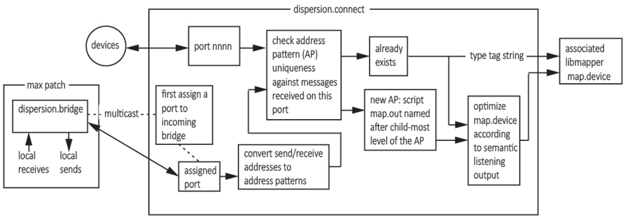

Figure 4: Semantic listening for devices via dispersion.connect and dispersion.bridge

5. EXAMPLE: SONIC-BIOPHYSICAL CONCURRENCE IN EAO

In section 3 we discussed the limit-case of live coding keystrokes as a form of textual "gesture" that modifies information of action-sound relationships. In a more classic case, we currently are active in capturing sonic and biophysical activity produced during a variety of movement and sounding contexts such as Deep Listening [21] sessions, where there exists a mixture between structured collective action and individual agency for each participant, leading to larger emergent forms. Another important analysis context is the Electro-Acoustic Orchestra (EAO) [2], where the focus has been on capturing audio streams from each player (ranging from 5-15) as well as MYO data (EMG, acceleration and orientation) from a subset of 5 players at a time, with a focus on the dominant-hand forearm of a variety of performers (including Wacom tablet electronics, flute, guitar, saxophone, and violin). Such a session is depicted in figure 5. These sessions have been freely improvised, and our focus has been on discovering emergent patterns and temporal concurrences in

Figure 5: Electro-Acoustic Orchestra

regards to time of onset and gestural similarity (over time and between users).

As noted above, our analysis thus far has been conducted in Max/MSP. In particular, using the MUBU and PIPO libraries [3]. Segmentation has been defined by examining short-term threshold changes to energy profiles: RMS, loudness, YIN fundamental frequency or the first MFCC coefficient. This segmentation defines the gestural units that are aggregated in the database, and we have explored using segmentations in MYO as well as in audio to act as the determinant of onset/offset times for both audio streams. Therefore, each gestural unit contains knowledge about the session, session time as well as UTC time, segmentation method, knowledge about devices and mapping, as well as the extracted features associated with each gesture. As a first pass of analysis, we have used search functions (e.g. all units within a given average range of fundamental frequency) in order to load JSON files back into Max/MSP, in order to run further similarity analysis, such as using the XMM probabilistic modelling library accompanying MUBU in order to examine similarities between units of the same type (kinetic action or sound) from different moments in time, and also to compare concurrent action-sound models at different points in time. More computationally simple yet no less interesting in its result, we also examine the temporal proximity of parallel segmentations, as in figure 6. In this instance, we ran segmentation on audio MFCCs and MYO y-axis accelerometer in parallel, and gestural units with high-correlation (defined as having nearly time-aligned segments) were returned. This information is then annotated and the database entry updated. We also annotate the similarity measure results and update this with each gestural unit, thereby adding linkages between data and creating a topology that could potentially be queried and "traversed" in a real-time interactive performance context. This example highlights the importance of the openness and scalability that underpins our system design choices, and the future goal of interacting with the database in a performative fashion.

6. CONCLUSION

We present here our initial works towards a design that is informed by a distributed view on performance across space as well as time, with the emergent concept of the gesture as the fundamental point of inquiry. The DisPerSion Lab, regarded in this context as an environmental instrument having performative agency, remains the primary site of development and real-world challenge to this work. Our initial

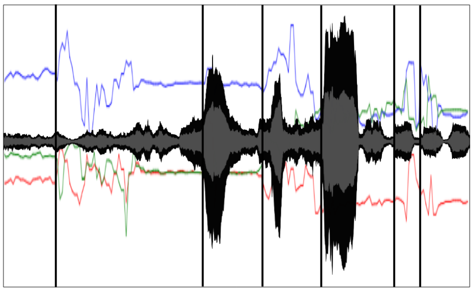

Figure 6: Moments of correlation between audio and MYO acceleration data (x=red, y=blue, z=green) from performer during EAO Session.

gestural database and mapping framework has proven useful thus far in facilitating complex performance events and in post-hoc analysis. Going forward, future challenges will include integrating more complex query functions on the database side, that might allow for deeper analysis, learning and refinement of the data to happen outside of environments such as Max/MSP, where bottlenecks occur in regards to efficiently passing large amounts of data. While fairly unstructured and improvisational music and movement contexts remain a primary interest in this research, we further hope to refine our approach to segmentation, analysis and temporal similarity measures in a way that is informed by contemporary knowledge of embodied human perceptual and cognitive factors.

7. REFERENCES

[1] DisPerSion Lab. `http://www.dispersionlab.org`. *Last Accessed April 17, 2016.*

[2] Electro-Acoustic Orchestra. `http://www.dispersionlab.org/eao`. *Last Accessed April 17, 2016.*

[3] MUBU and PIPO. `http://ismm.ircam.fr/category/software/`. *Last Accessed April 17, 2016.*

[4] I. Bergstrom and J. Llobera. OSC-Namespace and OSC-State: Schemata for Describing the Namespace and State of OSC-Enabled Systems. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Goldsmiths, University of London, 2014.

[5] C. Birgen, H. Preisig, and J. Morud. Sql vs. nosql. 2014.

[6] B. Bortz, J. Jaimovich, and R. B. Knapp. Emotion in motion: A reimagined framework for biomusical/emotional interaction. *Proceedings of the New Interfaces for Musical Expression*, 2:10, 2015.

[7] C. Cadoz and M. M. Wanderley. Gesture-Music. In M. M. Wanderley and M. Battier, editors, *Trends in Gestural Control of Music*, pages 315–334. IRCAM – Centre Pompidou, 2000.

[8] K. Chodorow. *MongoDB: the definitive guide.* O'Reilly Media, Inc., 2013.

[9] W. Forsynthe. Motion bank. `http://motionbank.org/`, 2016.

[10] J. Freeman, S. Xie, T. Tsuchiya, W. Shen, Y.-L. Chen, and N. Weitzner. Using massmobile, a flexible, scalable, rapid prototyping audience participation framework, in large-scale live musical performances. *Digital Creativity*, 26(3-4):228?244, Oct 2015.

[11] R. I. Godøy, A. R. Jensenius, and K. Nymoen. Chunking in music by coarticulation. *Acta Acustica united with Acustica*, 96(4):690–700, 2010.

[12] R. I. Godøy and M. Leman. *Musical gestures: Sound, movement, and meaning.* Routledge, 2010.

[13] A. Gritten and E. King. *Music and gesture.* Ashgate Publishing, Ltd., 2006.

[14] R. S. Hatten. *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert.* Indiana University Press, 2004.

[15] A. R. Jensenius. Action-sound: Developing methods and tools to study music-related body movement. 2007.

[16] A. R. Jensenius. To gesture or not? an analysis of terminology in nime proceedings 2001-2013. 2014.

[17] A. R. Jensenius, T. Kvifte, and R. I. Godøy. Towards a gesture description interchange format. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 176–179. IRCAM Centre Pompidou, 2006.

[18] M. Leman. *Embodied music cognition and mediation technology.* MIT Press, 2008.

[19] J. Malloch, S. Sinclair, and M. M. Wanderley. Libmapper: (a library for connecting things). In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 3087–3090, New York, NY, USA, 2013. ACM.

[20] D. Ogborn, E. Tsabary, I. Jarvis, A. Cárdenas, and A. McLean. extramuros: making music in a browser-based, language-neutral collaborative live coding environment. *Proceedings of the International Conference on Live Coding*, 2015.

[21] P. Oliveros. *Deep listening: a composer's sound practice.* IUniverse, 2005.

[22] G. Paine. Interaction as material: The techno-somatic dimension. *Organised Sound*, 20(01):82–89, 2015.

[23] P. Pasquier, T. Schiphorst, and S. DiPaola. Movement and meaning middleware. `http://www.mplusm.ca/`, 2016.

[24] C. Roberts, M. Wright, and J. Kuchera-Morin. Music programming in gibber. *Proceedings of the International Conference on Live Coding*, 2015.

[25] D. Schwarz. Corpus-based concatenative synthesis. *Signal Processing Magazine, IEEE*, 24(2):92 –104, 2007.

[26] S. Tilkov and S. Vinoski. Node. js: Using javascript to build high-performance network programs. *IEEE Internet Computing*, 14(6):80, 2010.

[27] D. Van Nort, P. Oliveros, and J. Braasch. Developing Systems for Improvisation based on Listening. In *Proceedings of the International Computer Music Conference*, New York, 2010.

[28] D. Van Nort, P. Oliveros, and J. Braasch. Electro/acoustic improvisation and deeply listening machines. *Journal of New Music Research*, 42(4):303–324, 2013.

[29] S. Wilson, N. Lorway, R. Coull, K. Vasilakos, and T. Moyers. Free as in beer: Some explorations into structured improvisation using networked live-coding systems. *Computer Music Journal*, 38(1):54?64, 2014.