Mapping physical effort to gesture and sound during interactions with imaginary objects in Hindustani vocal music

Stella Paschalidou[*1], Martin Clayton[#2], Tuomas Eerola[#3]

[*]*Dept. of Music Technology & Acoustics Eng., TEI of Crete, Greece*
[#] *Dept. of Music, Durham University, UK*
[1]pashalidou@staff.teicrete.gr, [2]martin.clayton@durham.ac.uk, [3]tuomas.eerola@durham.ac.uk

ABSTRACT

In the Dhrupad genre of Hindustani vocal improvisation singers often appear to engage with melodic ideas by manipulating intangible, imaginary objects with their hands. Despite the fact that no real object is involved, we observe vocalists executing repeated patterns of bi-manual effortful gestures, comprising gripping, exerting and releasing phases (Rahaim, 2009). They stretch, pull, push, throw and perform other movements, whereby they appear to be fighting against or yielding to some imaginary resistive force. This observation suggests that some patterns of change in acoustic features of the voice allude to the effortful interactions that such objects through their physical properties can afford, such as viscosity, elasticity, weight, friction etc. The present paper focuses exclusively on these types of gestures (abbreviated as 'MIIOs', i.e. Manual Interactions with Imaginary Objects) and it seeks to examine whether hand movements during MIIOs relate to the action of imagined forces, and if so, what type of forces and on what types of materials. We consider such voluntary imitations of interactions with real objects of particular interest to embodied music cognition research, as it has been previously argued that musical thinking is grounded in the ubiquitous patterns of actions we possess through our ecological knowledge of interacting with objects of the real world. Despite the specificity of the genre, the paper aims to address concerns in the study of gesture-sound relationships that are of interest to the wider research community and thus outcomes may be extended to other music lineages.

I. INTRODUCTION

Dhrupad singing offers a distinctive case for studying links between sound and physically-inspired concepts, as the deliberate and easily identifiable imitation of interactions we know from the real world makes it easier to take advantage of the ecologically valid conditions of real music performances. Recording music performances allows one to study more robust gesture-sound links that vocalists have established as active practitioners over years of practice rather than spontaneous responses to stimuli by listeners. Dhrupad is an interesting case of music-making for studying links to the non-discrete nature of gestures and phenomena of mechanics, as smooth melodic glides (mīṇḍ) in this genre abound and notes are approached through a strong sense of a pitch continuum. Additionally, it is suitable for gesture-sound analysis, since performances start at a strikingly low pace, they exhibit a slow melodic development, and performers are especially rigorous about the precision and subtleties of the intonation. Finally, Dhrupad is an 'oral' music tradition, which means that students do not rely on written notation and knowledge is transmitted through direct imitation from teacher to disciple. This imitation includes body movements too and it supports the deliberate choice (described in the methodology) of collecting material from a single musical lineage, namely students of Zia Fariduddin Dagar, allowing the investigation of similarities and differences between the gesturing habits of musicians sharing the same teacher.

Singing gestures have drawn little attention in the research community, although there is a growing body of published research on Hindustani music (e.g. Clayton, 2007; Leante, 2009; Rahaim 2009; Pearson 2013). None of the existing works has however studied the Dhrupad genre and none has dealt with MIIOs. MIIOs offer a special case where the hands—although free to move—are deliberately constrained by the conception of such an object. Familiar interactions from the real world that would not normally produce sound but are so spontaneously employed by Hindustani vocalists may offer a unique opportunity in the context of embodied music cognition. Due to the absence of a real mediator, they may allow significant cognitive processes to be revealed, that are associated with more fundamental concepts than unequivocal mechanical cross-modal couplings of a particular instrument.

Dhrupad relies heavily on improvisation, which is rule-based and conforms to the 'rāga' (modal) system. The notion of a rāga as a movement in a melodic pitch 'space' (Fatone et al., 2011) is also accompanied by smooth hand movements in the real space (deliberate or unconscious). According to Rahaim (2009) and personal observations, Hindustani singers seem to engage with the melodic content in two main ways, which are also supposed to reflect two different relationships between the body and the voice:

(a) by keeping the hand(s) open and tracing curves in an effortless way as trajectories in space, whereby the geometrical properties of manual movement provide a melographic representation of the sound and

(b) by closing the hands (forming a grip) and manipulating some imaginary object by effortfully acting against the imagined opposing force associated with changing its shape, size and position, whereby the gesture-voice link seems to reside in the consistent physicality of the object employed.

In other words, Dhrupad vocalists do not merely project the melographic content of a phrase by the absolute position of the hands and their trajectories in space; the dynamic aspect of the movement is also employed to convey various types of related sonic information. The current paper is only concerned with the second type of gestures. For example, one may notice gestures that look as if the singer is stretching something like a rubber band, moving a heavy object around space or bouncing something like a ball on the floor (Fatone et al., 2011; Rahaim, 2009). At the core of such events lies a sensation of resistance that the agent (the performer) needs to defy in order to finally move the object. When these gestures appear, the reference is moved from the hands themselves to the imagined object, with which the performer remains in

contact. It looks like the body is temporarily extended through an interface to the real world, where individual notes can be gripped as objects. This contact lasts as long as the object serves the purpose of the vocalization; vocalists seem to unconsciously drop it thereafter. In each of these cases, by grasping, holding, extending and then releasing them, notes can be manipulated as smooth pitch glides (mīṇḍ). Despite the absence of a real object, the link to the sound seems to be mediated by the imagined material. There is an observable match between the voice and the apparent manipulative gestures in terms of synchronization and temporal congruency of various features, such as melody, dynamics and timbre. Grasping, stretching and releasing an elastic material are examples of such fundamental embodied experiences. However, the link between hand gestures and voice on the occasions of such interactions has not been studied before in a systematic way.

Therefore, by drawing on theories by O'Regan & Noë (2001), Thompson & Varela (2001) and Varela et al. (1991) that consider human cognition as founded on repeated patterns of interactions with the real world and by extending Gibson's (1977) ecological theory of affordances to cover the imaginary objects of MIIOs, we aim to demonstrate that prior sensorimotor experiences from familiar interactions with the environment are linked with the melodic activity while singing. In other words, the current work examines the assumption that when Dhrupad singers appear to interact with an imaginary object, the link to the sound resides in the interaction possibilities the object affords. Central to this is an exploration of actions that draw on our familiarity in interacting with the real world through the constraints that the materials involved impose on the movement.

II. METHODOLOGY

The work makes use of a combined methodological approach, consisting of qualitative and quantitative methods on original recordings of interviews, audio-visual material and 3D-movement data (using an Optitrack passive marker system) of vocal improvisations, which were collected by conducting recordings by fourteen vocalists of the same music lineage (all students of Zia Fariduddin Dagar, including the maestro himself) in domestic spaces in India (2010-2011) for the specific study. The qualitative part of the analysis includes a thematic analysis of interview material (by six Dhrupad vocalists) and a video observation analysis of performance recordings (by two performers), while the quantitative part of the analysis involves the development of formalized description for the classification of MIIOs by fitting (linear) general logistic models to a small set of movement and sonic features (for two performers).

The thematic analysis of interview material aimed to best unfold common underlying cross-modal concepts embedded in explicit and implicit knowledge by the interviewees and also to ground the later annotations of video observations and inform the coding scheme that would be developed. The analysis focused on the conception of melody as movement by the participants and the importance of visual and motor imagery in the way musicians conceive sound. Most importantly, this part concentrated on the melodic function of imagined objects with which the performer may interact, in an attempt to offer a better understanding of how the extension of the body through imaginary objects and materials may be facilitating engagement with melodic ideas. Thematic analysis

was therefore applied to sensorial descriptors (adjectives, verbs and nouns) of motor-based metaphors or pictorial terms of performer-object interactions that are used recurrently when talking about sound and music.

The video observation analysis aimed to identify and classify MIIOs and melodic movements, to find recurrent gesture-sound associations in terms of gesture classes, melodic phrases and the melodic context and to draw some first conclusions about the level of consistency in gesture-sound associations both within a single performance and across performances/performers. It was performed in the ANVIL annotation environment by first visually identifying, segmenting, labeling and classifying repeated patterns of manual gestures that allude to MIIOs for each individual performer/performance, and also annotating in terms of amount of effort level that each gesture was perceived to require. The used coding scheme was informed by the analysis of the interview material, but mainly emerged progressively during multiple viewings of the video footage. Despite the lack of a real object between the hands, we still expected that a third-person would be able to make judgments on the kind of interaction and imagined object that a vocalist seemed to be employing through visual observation of the dynamic and kinetic properties of the motion. For each of the gestures, the associated audio content was also coded and classified in terms of recurrently used melodic movement types, pitch interval, octave range, sung syllable, as well as the larger phrase in which the melodic movement was embedded and the melodic context (the intention of the melody in moving towards the tonic immediately after the annotated phrase). The choice for the features used in the coding scheme of each performer was based on indications coming from simple visual observation about the possibility that these acoustic features might be associated with the gestures. Therefore, they differed slightly between performers. An inter-coder validation of the visual material in terms of gesture classes (MIIO types) was then carried out by two professional dancers/choreographers for the performance of Afzal Hussain, in order to assess whether the annotations of the main coder could be considered as reliable. Finally, a gesture-sound association analysis was conducted in order to identify recurrent associations between categorical aspects of the hand movements and the voice.

The logistic regression models that were developed aimed to validate and augment the findings from the qualitative methods by using exact measurements in order to classify gestures in terms of interactions with elastic (stretching, compressing) versus rigid (pulling, pushing-away, collecting, throwing) objects based on a small number of acoustic and movement features. The responses (gesture classes) that were used in the models as ground-truth—or in other words the correct output values—were drawn from the manual annotations of the video analysis. The models also aimed to examine to what extent models might mostly reflect idiosyncratic elements of gesturing or more generic gesture-sound links shared across vocalists. Therefore, two variations of models were presented: (a) one that best fits each individual performer and (b) one that largely overlaps between performers and thus can best describe more generic cross-performer behaviors. The best fitting models were chosen as a trade-off between model accuracy, compactness and simplicity in interpretation and aimed at expressing rather

idiosyncratic aspects of gesture-sound links for each performer.

For the video analysis and the general logistic models only the slow non-metered section of the improvisation—called ālāp—was used, which is sung to a repertoire of non-lexical syllables, in order to avoid the metrical structure and the lyrical content of the later stages of a raga performance and only concentrate on melodic factors. These performances include raga Jaunpuri (scale: 1,2,*b*3,4,5,*b*6,*b*7, 18 minutes) by Afzal Hussain and rāgā Malkauns (scale: 1,*b*3,4,*b*6,*b*7, 23 minutes) by Lakhan Lal Sahu.

III. RESULTS

During interviews performers displayed a high visual element in the conceptualization of music and they made a frequent use of physically inspired linguistic descriptors in talking about music. Some vocalists made in fact explicit and concrete statements about categorical gesture-sound associations, such as the following by Sahu:
- mīṇḍ – stretching a rubber band,
- gamak – applying pressure,
- hudak – throwing.
Additionally, some musicians acknowledged the role of body transmission and similarities between themselves and their teacher's gesturing habits, not always as the exact replica of movements, but in terms of some shared movement qualities. Most importantly, there was a heavy use of motor-based metaphors alluding to the sensation of a resistive force that the performer needs to fight against, such as elasticity/stretching, pressure/pushing, weight/lifting and friction/scratching. Such recurrent descriptors were extracted, coded and organized in meaningful ways in classes of overarching themes. A final table was produced which was intended to inform the coding scheme of the performance analysis in later stages.

Through the video observation analysis the following recurrent gesture classes were identified for each of the performers: stretching or compressing an elastic object, pulling, collecting pushing away or throwing a rigid object. An association analysis between categorical aspects of gesture and voice resulted in a consistent link that is shared between performers and mainly reflects shared cross-modal morphologies. For instance, double-sloped pitch glides are associated with gestures comprising two opposing spatial directions, such as in stretching or compressing an elastic object, while monotonic ascending melodic glides are associated with gestures of a single direction in space, as in pushing a way an object of considerable mass. Additionally, results for Hussain displayed a strong link between specific types of gestures with specific pitch areas of the raga and with the intention of the melody to ascend to or towards the tonic, where the melody typically resolves. Finally, interactions with elastic objects tend to require higher levels of effort than interactions with rigid objects.

The results of the quantitative analysis rejected the null-hypothesis that voice and gesture are unrelated. By using a small set of non-collinear movement and sound features it was possible to estimate a good part of the variance in gesture classification. In the case of Hussain, the use of 5 audio and movement features yielded a high classification rate of about 95%, while in the case of Sahu, the use of 4 audio and movement features gave a slightly lower but still considerable classification rate of about 80%. A more generic cross-performer gesture classification was also achieved by almost identical general logistic models for the two singers, yielding a good fit of about 86% (with 3 features) for Hussain and 78% (with 4 features) for Sahu respectively. Results indicate that the inference of gesture classes is more heavily dependent on acoustic rather than movement features and that there is a bigger overlap of acoustic rather than movement features between performers. This observation possibly highlights a stronger idiosyncratic factor in how MIIOMs are manually performed, which renders the task of describing them through a small number of low-level movement features as non-trivial. However, in the case of Hussain gesture classes tend to be also associated with particular areas of the rāga pitch space, which is not the case with Sahu. For instance, for Hussain interactions with elastic objects are more likely associated with melodic movements that ascend to a higher degree of the scale; the opposite applies for interactions with rigid objects. As the 7th is the highest and also most unstable degree of the scale in rāga Jaunpuri, it naturally forces a change in pitch movement direction from ascent to descent, which is similar to the direction change of the hands when deforming an elastic object. It could be therefore suggested that the MIIO type is associated with the grammatical rules of the rāga.

IV. CONCLUSIONS

Findings indicate that despite the flexibility in the way a Dhrupad vocalist might use his hands while singing, the high degree of association between classes of virtual interactions and the voice provides good evidence for non-arbitrariness. This association may be grounded in shared cross-modal morphologies, but also in some cases in the melodic organization of the ālāp improvisation. Although there is higher consistency within an individual performance rather than across performers and performances, it could be still argued that the type of imagined object and the nature of the imagined opposing forces reflect qualities describing melodic aspects of the voice. Certainly a bigger dataset would be needed in order to make a strong and more generic claim about MIIOs in Dhrupad vocal music. However, the current paper has reported on the first study of gesture-sound associations during MIIOs in the Dhrupad genre of Hindustani vocal music. It has contributed to better understanding the role of voluntary imitations of interactions with real objects by using material from interviews and recordings of real performances in the field in India and by proposing a combined methodological approach that comprises both qualitative and quantitative methods.

By taking an embodied approach and mapping effort to a combination of features from both domains (auditory and movement), this work can also contribute to the enhancement of mapping strategies in empty-handed artificial interactions by making gesture-sound links more physical plausibility; novel interaction paradigms can be developed which are inspired by our interaction with the real world.

REFERENCES

Clayton, M. (2007). Time, gesture and attention in a Khyāl performance. *Asian Music*, *38*(2), 71-96.

Fatone, G. A., Clayton, M., Leante, L., & Rahaim, M. (2011). Imagery, melody and gesture in cross-cultural perspective. In Gritten, A. & King, E. F. (Eds.), *New Perspectives on Music and Gesture* (pp. 203-220), Surrey, Ashgate.

Gibson, J. J. (1977). *The theory of affordances*. Hilldale, USA.

Leante, L. (2009, November). The lotus and the king: imagery, gesture and meaning in a Hindustani rāg. In *Ethnomusicology Forum* (Vol. 18, No. 2, pp. 185-206). Taylor & Francis Group.

O'Regan, J. K., & Noë, A. (2001). Acting out our sensory experience. *Behavioral and Brain Sciences*, *24*(05), 1011-1021.

Pearson, L. (2013). Gesture and the sonic event in Karnatak music. *Empirical Musicology Review*, *8*(1), 2-14.

Rahaim, M. J. (2009). *Gesture, melody, and the paramparic body in Hindustani vocal music*. University of California, Berkeley.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences*, *5*(10), 418-425.

Varela, F. T., & Rosch, E. (1992). E.(1991) The Embodied Mind: Cognitive Science and Human Experience.