

# EFFORT IN GESTURAL INTERACTIONS WITH IMAGINARY OBJECTS IN THE CONTEXT OF DHRUPAD VOCAL IMPROVISATION

**Stella Paschalidou**

Dept. of Music Technology &  
Acoustics Eng., TEI of Crete,  
Greece  
pashalidou@staff.  
teicrete.gr

**Martin Clayton**

Dept. of Music, Durham University,  
UK  
martin.clayton@  
durham.ac.uk

**Tuomas Eerola**

Dept. of Music, Durham University,  
UK  
tuomas.eerola@durham.ac.uk

## ABSTRACT

This paper examines relationships between the voice and the ecological knowledge of how the human body moves when interacting with objects of the environment in the context of Hindustani vocal music. In the Dhrupad genre of singing improvisation vocalists often appear to engage with melodic ideas by manipulating intangible, imaginary objects with their hands, such as through stretching, pulling, pushing etc. Our main focus is on how much effort each of such gestures is perceived to require and whether and how this is linked to its melodic counterpart.

The work makes use of both qualitative and quantitative methods in a sequential order on original recordings of interviews, audio-visual material and 3D-movement data of Dhrupad vocal improvisations. Findings indicate that there is a certain level of consistency in gesture-sound links despite flexibility caused by idiosyncratic movement patterns of individuals. Different schemes of cross-modal associations were revealed for the singers analysed, that depend on the pitch space organisation of melodic modes (rāgas), the mechanical requirements of voice production, the macro-structure of the (ālāp) improvisation and morphological cross-domain analogies. Results further suggest that a good part of the variance in both physical effort and gesture type can be explained through a small set of sound and movement features.

## 1. INTRODUCTION

### 1.1 Background

The current paper takes an embodied approach in the analysis of Dhrupad vocal improvisation. Dhrupad is a sub-genre of Hindustani music that 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' [1] is also accompanied by smooth hand movements in the real space (deliberate or unconscious).

*Copyright: © 2018 Stella Paschalidou et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

It is often the case that listeners and performers alike will report that they experience virtual worlds of forces in motion in relation to music and sound [2, 3]. In fact, in Hindustani music (and the Dhrupad genre in particular) singers make a frequent use of linguistic expressions for motor-based metaphors alluding to the sensation of a resistive force that the agent (the performer) needs to fight against. Additionally, during vocal performances singers seem to engage with the melody and its intricate qualities by employing and manipulating imaginary objects. They stretch, pull, push, collect, throw and execute other movements, whereby they appear to be fighting against or yielding to some imaginary resistive force. Although the object is not real, we observe them executing repeated patterns of bi-manual effortful gestures, comprising gripping (closing the hand), action and releasing phases [4].

We consider such voluntary imitations of interactions with real objects (that would not necessarily produce any sound) 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 [5, 6]. Such Manual Interactions with Imaginary Objects (that we will call MIIOs) offer a special case where the hands—although free to move (see previous studies on free-hand sound tracings [7, 8])—are deliberately constrained by the conception of an object. On the other hand, 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.

### 1.2 Aim

The current paper offers a systematic examination of whether and how the kinaesthetic sensation of effort that is conveyed through movement during MIIOs relates to the voice in Dhrupad vocal performances. In other words, the paper explores whether the link between gestures and sound seems to reside in the interaction possibilities and the levels of effort that these virtual objects afford [9] according to their physical properties, such as viscosity, elasticity, weight, friction etc. Effort is understood here in its common usage, which reflects our understanding of how hard a person must work, either mentally or physically, in order to achieve an intended goal. Although effort in music has often been regarded as an important

aspect of both listening and performing [10], systematic approaches to its role have been limited [11].

Singing gestures have drawn little attention in the field of embodied music cognition, although there is a growing body of work on Hindustani music (e.g. [12, 13, 4, 14, 15]). None of the existing works has however studied the Dhrupad genre and none has dealt with MIOs and their perceived levels of effort. Dhrupad singing offers a distinctive case for studying underlying links between sound and effort, as it exhibits easily identifiable imitations of real manual interactions with objects and a plethora of smooth melodic glides (*mīṇḍ*) where notes are approached with a sense of pitch continuum, which makes it a good case for studying the non-discrete nature of the mechanics of gestures. It is also a suitable case of music making due to the strikingly slow melodic progression and the rigour and precision of intonation by singers. Finally, the 'oral' way of music transmission means that students do not rely on written notation, which allows the transmission of embodied knowledge through direct visual engagement of disciples with their teachers during teaching sessions. However, physical metaphors in musical thinking are not a Dhrupad exclusivity [1] and therefore, 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 of embodied music cognition and thus outcomes may be extended to other music lineages.

## 2. METHODOLOGY

The work combines qualitative and quantitative methods applied on original recordings of interviews, audio-visual material and 3D-movement data (using a 10-camera Optitrack passive marker system) of vocal improvisations by fourteen vocalists of the same music lineage (all students of Zia Fariduddin Dagar, including the maestro himself). Real recordings of performances were chosen rather than designed laboratory experiments of subjects responding to stimuli, in order to reveal robust gesture-sound links that vocalists may have established over years of practice rather than spontaneous responses to stimuli by listeners. These performances were collected for the specific study in domestic spaces in India (2010-2011). In order to avoid the metrical structure and the lyrical content of the later stages of the performance and only concentrate on melodic factors, only the slow non-metered section of the improvisation (*ālāp*) was used, which is sung to a repertoire of non-lexical syllables. The qualitative part of the analysis includes a thematic analysis of interview material by six vocalists and a video observation analysis of four vocalists. The quantitative part of the analysis involves the development of mathematical descriptions for the classification of MIOs and the inference of effort levels by fitting (linear) models to a small set of movement and sonic features for two of the performers.

The thematic analysis of interview material aimed to reveal cross-modal concepts embedded in explicit and implicit knowledge by the interviewees and also to

ground the coding scheme that was later developed during the video analysis stage. The analysis also intended to highlight the importance of visual and motor imagery in the conception of music as movement by musicians and to concentrate in particular on how the extension of the body through imaginary objects and materials may be facilitating engagement with melodic ideas. The interview material was therefore collected and thematic analysis was applied to recurrent sensorial and pictorial descriptors (adjectives, verbs and nouns) of motor-based metaphors and performer-object interactions when talking about sound and music.

The video observation analysis was performed in the ANVIL annotation environment and it aimed to use third-person observations in order to identify and classify types of MIOs and melodic motives, to find recurrent gesture-sound associations and to draw some first conclusions about the level of consistency in their co-appearance for a single or across performances. The coding scheme was informed by findings of the previously conducted thematic analysis but mainly emerged in ANVIL progressively during multiple viewings of the video footage for each individual performer/performance. This was done by visually identifying, segmenting, labeling and classifying repeated patterns of (90) manual gesture events that allude to MIOs. These gestures were also annotated in terms of melodic movement type, pitch interval, octave range, sung syllable and melodic context (intention of the melody in moving towards the tonic immediately after the annotated phrase) of their melodic counterpart, as well as in terms of amount of effort (on a scale between 0-10, 0 being the highest) that each gesture was perceived to require. The most prominent gesture types varied between stretching an elastic object and pulling or pushing away a rigid object. An inter-coder validation of gesture class annotations was also carried out by two professional dancers/choreographers for one of the performances, in order to assess whether the annotations of the main coder could be considered 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 quantitative part of the analysis aimed at exploring whether it would be possible to computationally infer the visually annotated effort levels and gesture classes (in terms of interactions with elastic (stretching, compressing) versus rigid (pulling, pushing-away, collecting, throwing) objects) based on a small set of movement and sound features extracted from the captured material. This stage of the analysis 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 sets of linear models were devised, those which best fit to the data of each individual performer (describing rather idiosyncratic aspects of gesture-sound links for each vocalist) and those that—despite their lower goodness of fit—overlapped to a greater extent across performers (displaying a more generic power over performers). Models considered as best fitting were those that displayed the best trade-off between model accuracy, compactness and simplicity in interpretation. The models were developed for the per-

formance of Afzal Hussain in *rāga Jaunpurī* (scale: 1,2,b3,4,5,b6,b7, 18 minutes) and that of Lakhan Lal Sahu in *rāga Mālkaunś* (scale: 1,b3,4,b6,b7, 23 minutes). The explorative character of this process meant that a considerable part of the process involved finding a set of most important descriptors to fit the models to. Initially, the (15) most relevant gesture and audio features reported in [16] on sound-tracing experiments were used, but then a number of alternative motion/acoustic variables was progressively added as to raise the goodness of fit while retaining the compactness of the model. These were high-level statistical global measures (such as mean, SD, min, max) computed from a number of time-varying audio and movement features that derived from the raw data. Although far more sophisticated features may be probed, derivatives of position coordinates were chosen as they have proven to be robust and pertinent in other music performance contexts [17] and in sound tracing experiments [18, 19, 20].

### 3. RESULTS

Interview analysis revealed a high visual element in the conceptualisation of music and a heavy use of motor-based metaphors expressed through physically inspired linguistic descriptors that allude to the sensation of resistive forces a performer needs to fight against, such as elasticity/stretching, pressure/pushing, weight/lifting and friction/scratching. These recurrent descriptors were organised in meaningful ways in order to inform the coding scheme of the video analysis in later stages. The transmission of musical knowledge through visual engagement and imitation of gestural habits between teachers and disciples was also acknowledged, not perhaps as exact replicas but in terms of shared movement qualities. Some vocalists made even explicit statements about categorical gesture-sound associations.

The association analysis carried out between categorical aspects of gesture and voice during the video analysis stage resulted in a consistent link that is shared between performers and mainly reflects shared cross-modal morphologies. However, a few other modes of association were also revealed, such as a strong link between interaction type and effort level, with the stretching gesture of an elastic object requiring higher levels of effort than pulling/pushing away a rigid object, as well as a strong link between specific types of gestures with specific pitch areas of the *rāga* and with the melodic context (the intention of the melody to ascend to or towards the tonic) in the case of Afzal Hussain.

By using the manual annotations of the preceding video analysis as ground-truth responses—or other words the correct output values—a number of compact linear models were devised in the quantitative part of the analysis, according to which the null-hypothesis that voice and gesture are unrelated was rejected. By using a small set (of four or five) non-collinear movement and sound features it was possible to estimate a good part of the variance in gesture classification and effort inference.

### 3.1 Idiosyncratic schemes

#### 3.1.1 Effort level inference

Differing idiosyncratic schemes of associating effort with acoustic and movement features were identified, that are based on the pitch space organisation of the *rāga* as well as the mechanical requirement for producing the voice.

*Hussain*: The use of 5 non-collinear audio and movement features yielded a good fit of about 60% ( $R^2_{adj}$ ). According to this model, higher effort levels are required when the hands move slower and further apart and exhibit a larger speed variation. These are accompanied by melodic glides that start from lower degrees and ascend to higher degrees of the *rāga* scale in the range of an individual octave, thus they are considered to be associated with characteristic qualities of the specific *rāga*.

*Sahu*: The use of 4 non-collinear audio and movement features yielded an adequately good fit of about 44% ( $R^2_{adj}$ ). According to this model, higher bodily effort is required for hand movements that exhibit a larger variation of hand divergence (speed in moving the hands further apart) and a strong onset acceleration. They are accompanied by larger melodic glides that ascend to higher maximum pitches, thus reflecting the increased mechanical requirement of voice production. This may also be reflecting the gradual ascent towards the pitch climax of the *ālāp* macro-structure.

#### 3.1.2 Gesture classification

Different modes of gesture class association with acoustic and movement features were revealed, especially with pitch regions of particular interest for the specific *rāga* and analogous cross-domain morphologies.

*Hussain*: The use of 5 non-collinear audio and movement features in the logistic models that were developed yielded a high classification rate of about 95% (AUC). According to this, it is more likely that interactions with elastic objects (rather than rigid) are performed by hand gestures that exhibit a low absolute mean acceleration and a large variation in hands' divergence. They are accompanied by slower and larger melodic movements that ascend to a higher degree of the scale.

*Sahu*: The use of 4 non-collinear audio and movement features in the logistic model yielded an adequately good fit of about 80% (AUC). According to this model, interactions with elastic objects are more likely performed with pitch movements of a larger interval and larger duration and with the hands moving faster and remaining bound to each other.

### 3.2 Generic scheme

#### 3.2.1 Effort level inference

Two almost identical linear models were developed, yielding a good fit of about 53% ( $R^2_{adj}$ ) with 5 features for Hussain and 42% with 4 features for Sahu respectively. According to these models, higher bodily effort levels are required by both singers for melodic movements that start from lower and reach up to higher pitches, reflecting the mechanical strain of voice production. They are accompanied by movements which are slow on average but

with a large variation of speed, and in the specific case of Hussain with the hands moving further apart.

### 3.2.2 Gesture classification

Two almost identical general logistic models were developed, yielding a good fit of about 86% (AUC) with 3 features for Hussain and 78% with 4 features for Sahu respectively. According to these models, interactions with elastic objects are more likely to be performed at lower pitches during larger melodic movements, and with the hands moving further apart for Hussain and less apart but faster in the case of Sahu.

## 4. CONCLUSIONS

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 a better understanding of the role of voluntary imitations of interactions with objects of the real world by using original material of interviews and recordings of real performances captured in the field in India and by proposing a novel sequential methodological approach that comprises both qualitative and quantitative methods. MIIOs offer a particular case where motor imagery is “materialised” through physical actions directed towards an imagined object.

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 and generic associations that are not necessarily performer-specific or stylistic. Such association may be grounded in some shared cross-modal morphologies, but in some cases also in the melodic organization of the ālāp improvisation. Regardless of the idiosyncratic element in gesturing noticed by individual performers, 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.

Combining ethnographic data with exact measurements of real performances has brought about the advantages of ecological validity, however it has also posed a number of important challenges and limitations, such as the small dataset. Collecting a larger dataset of multiple performers, performances and rāgas for each performer will allow a more systematic comparison between performers, performances and rāgas.

## 5. REFERENCES

- [1] 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.
- [2] Stern, D. (2010). *Forms of vitality : Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford ; New York: Oxford University Press.
- [3] Eitan, Z., & Granot, R. Y. (2006). How Music Moves. *Music Perception*, 23, 3, pp. 221-248.
- [4] Rahaim, M. (2009). *Gesture, Melody, and the Param-paric Body in Hindustani Vocal Music* (Doctoral thesis, University of California, Berkeley, USA). Retrieved from <http://gradworks.umi.com/33/69/3369126.html> (last accessed December 15, 2016).
- [5] Noë, A. (2004). *Action in perception* (Representation and mind). Massachusetts ; London: MIT Press.
- [6] Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind : Cognitive science and human experience*. Cambridge, Mass.: MIT Press.
- [7] Godøy, R. I., Song, M., Nymoen, K., Haugen, M. R., & Jensenius, A. R. (2016). Exploring Sound-Motion Similarity in Musical Experience. *Journal of New Music Research*, 45, 3, pp. 210-222.
- [8] Caramiaux, B., Bevilacqua, F., & Schnell, N. (2010b). *Mimicking sound with gesture as interaction paradigm*. Tech. rep., IRCAM-Centre Pompidou.
- [9] Gibson, J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Lawrence Erlbaum.
- [10] Krefeld, V., & Waisvisz, M. (1990). The Hand in the Web: An Interview with Michel Waisvisz. *Computer Music Journal*, 14, 2, pp. 28-33.
- [11] Bennett, P., Ward, N., O'Modhrain, S., & Rebelo, P. (2007). DAMPER: a platform for effortful interface development. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression* (pp. 273-276), New York City, USA.
- [12] Clayton, M. (2007). Time, gesture and attention in a Khyāl performance. *Asian Music*, 38(2), 71-96.
- [13] 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.
- [14] Pearson, L. (2013). Gesture and the sonic event in Karnatak music. *Empirical Musicology Review*, 8(1), 2-14.
- [15] Kelkar, T., & Jensenius, A. R. (2018). Analyzing Free-Hand Sound-Tracings of Melodic Phrases. *Applied Sciences*, 8(1), 135.
- [16] Nymoen, K., Godøy, R. I., Jensenius, A. R., & Torresen, J. (2013). Analyzing correspondence between

sound objects and body motion. *Acm Transactions on Applied Perception*, 10, 2, pp. 1-22.

- [17] Rasamimanana, N. H., Flety, E., & Bevilacqua, F. (2006). Gesture Analysis of Violin Bow Strokes. *Lecture Notes in Computer Science*, 3881, pp. 145-155.
- [18] Nymoen, K., Torresen, J., Godoy, R. I., & Jensenius, A. R. (2012). A Statistical Approach to Analyzing Sound Tracings. *Lecture Notes in Computer Science*, 7172, pp. 120-145.
- [19] Nymoen, K., Caramiaux, B., Kozak, M., & Torresen, J. (2011). Analyzing sound tracings: a multimodal approach to music information retrieval. In *Proceedings of the 1<sup>st</sup> International ACM workshop on Music information retrieval with user-centered and multimodal strategies* (pp. 39-44), Scottsdale, AZ, USA.
- [20] Glette, K. H., Jensenius, A. R., & Godøy R. I. (2010). *Extracting action-sound features from a sound-tracing study*. Tech report, University of Oslo, pp. 63-66.