Effort in Interactions with Imaginary Objects in Hindustani Vocal Music – Towards Enhancing Gesture-Controlled Virtual Instruments

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

[1] Department of Music Technology & Acoustics Eng., TEI of Crete, Greece
[2] Department of Music, Durham University, UK
[†] pashalidou@staff.teicrete.gr

ABSTRACT

In Hindustani Dhrupad vocal improvisation singers often appear to engage with melodic ideas by manipulating intangible, imaginary objects and materials with their hands, such as through stretching, pulling, and pushing. This observation suggests that some patterns of change in acoustic features of the voice allude to the interactions that objects through their physical properties can afford and they are defined by the sensation of effort required for fighting against or yielding in to the resistance involved in the interaction.

The present study reports on the exploration of relationships between the voice, the interaction possibilities of malleable (elasticity) vs. rigid (weight/friction) objects and the physical effort they appear to require as perceived by an observer in Dhrupad vocal improvisation.

1. INTRODUCTION

With the recent shift towards embodied approaches in music performance, physical effort has often been highlighted as an important aspect of expressivity. In fact, the lack of expressivity in designed gesture-sound interactions has been attributed by [1] to the rupture of the so-called 'efforted-input paradigm'. However, systematic approaches to the role of effort still remain limited. In this study we understand effort as a concept which reflects the active or passive attitude of the person in fighting against or giving in to the physical conditions that influence the movement while trying to achieve an intentional task [2].

The current work is ultimately intended to enhance the design of mappings in Virtual Musical Instruments [3], in which imaginary objects are sculpted to produce sounds. It proposes to achieve this by taking advantage of the physical effort induced by familiar gestures associated with handling real objects in our natural environment and thus by making sound sculpting feel more 'physically plausible' [4].

To meet this end, the current paper uses the ecologically valid setting of real performances, in order to gain a deeper understanding of motor-based metaphors in the specific cases where a Dhrupad singer seems to manually interact with imaginary objects and materials (MIIOMs). Dhrupad is an 'oral' music tradition (in the sense that it does not depend on music notation) and relies heavily on improvisation, which is rule-based and conforms to the raga (modal) system. More specifically, the paper aims to examine whether physical effort and gesture interaction types are arbitrarily related to the voice or if they could be predicted through a small number of audio and/or movement features. MIIOMs offer a unique opportunity to study profound gesture-sound links that are not just related to the mechanical couplings of instrumental playing.

2. METHODOLOGY

The work makes use of a mixed methodological approach, combining qualitative and quantitative methods on original recordings of interviews, audio-visual material and 3D-movement data of vocal improvisations. The present study reports on the first part of the vocal improvisations by two different Dhrupad vocalists of the same music lineage, namely Afzal Hussain (raga Jaunpuri, scale: 1,2,*b*3,4,5,*b*6,*b*7, 18 minutes) and Lakhan Lal Sahu (raga Malkauns, scale: 1,b3,4,*b*6,*b*7, 23 minutes).

Hussain and Sahu were selected out of 15 performers who were recorded performing the slow non-metered section of an improvisation called alap, which is sung to a repertoire of non-lexical syllables. Only the alap improvisation was used in this study in order to concentrate on melodic factors without having to take into account the metrical structure or the lyrical content of the later stages of a raga performance. Real performances rather than designed experiments were used, in order to insure ecological validity in the approach of the subject. Thus, all recordings were conducted in domestic spaces in India (2010). The recording sessions included video, audio and motion recordings (10-camera Optitrack passive marker mocap system) and were followed by interviews.

First, we applied thematic analysis to the interview material to identify action-based metaphors, which informed the annotation process of the video material that followed. For the video analysis we relied on third-person observations that aimed at identifying, labeling and later classifying the audio-visual material in terms of recurrent types (categorical descriptors) of MIIOMs, as well as perceived effort levels (numerical) that appear to be exerted by the performer in a range between 0 (lowest) and 10 (highest).

The annotations (cross-validated by two choreographers /dancers) were used in the quantitative part of the analysis as the response values of models that were fit to the measured movement and sonic features in order to (a) estimate the annotated effort levels (numerical, through linear regression) and (b) classify gestures (categorical, through logistic regression), specifically discriminating between interactions with elastic vs. rigid objects. The features that were used for estimating the responses were computed by first extracting continuous (time-varying) movement and audio features from the raw movement and audio data respectively and then computing representative statistical global measures (such as mean, SD, min, max). The movement and audio features

reported in [5] were first used as a starting point, but then a number of alternative features were also explored that were meant to raise the explained variance of the estimated responses.

3. RESULTS

Two variations of models were developed for each task (effort estimation and gesture classification):
(a) a model that best fits each individual performer, thus better reflecting the idiosyncratic aspect of the individual singer and
(b) a model that can better describe shared cross-performer behaviours.

3.1. Effort estimation

3.1.1. Idiosyncratic schemes:

Different idiosyncratic schemes of associating the perceived physical effort with acoustic and movement features were identified through linear regression, that are based on the pitch space organisation of the raga and the mechanical strain of voice production.
- In the case of Hussain, the use of 5 non-collinear audio and movement features yielded a good fit (R^{2adj}) of about 60%. According to the results, higher bodily effort levels are required for slower movements that involve moving the hands wider apart and exhibit a larger speed variation. They are accompanied by singing melodic glides that start from lower degrees (closer to the lower tonic) and ascend to higher degrees (further away from the lower tonic) of the scale within the boundaries of each individual octave, in other words for larger melodic movements that move higher up in the octave. Pitch values are calculated here on a relative logarithmic scale and thus they describe glides associated with particular degrees of the melodic scale, better describing characteristic qualities of the specific raga.
- In the case of Sahu, the use of 4 non-collinear audio and movement features yielded an adequately good fit (R^{2adj}) of about 44%. According to the results, higher bodily effort is required for hand movements that exhibit a larger variation of hand divergence (change of speed in moving the hands apart or closer to each other), with a strong mean acceleration in the beginning (onset). They are accompanied by larger melodic glides that reach up to higher maximum pitches. In this case, pitch is expressed in an absolute logarithmic scale, which better reflects the mechanics of voice production. As the alap is organised based on a gradual ascent towards the pitch climax, pitch is here also representative of the alap macro-structure.

3.1.2. Generic scheme:

A more generic cross-performer estimation of effort was achieved by almost identical linear models for the two singers, yielding a good fit (R^2adj) of about 53% (with 5 features) for Hussain and 42% (with 4 features) for Sahu respectively. According to these models, high bodily effort levels are required by both singers for melodic movements that start from a lower and reach up to a higher pitch (on an absolute logarithmic scale, reflecting the mechanics of voice production). They are accompanied by movements which are slow on average but exhibit a large variation of speed, and in the specific case of Hussain by movements during which the hands move further away.

3.2. Gesture (MIIOM) classification

3.2.1. Idiosyncratic schemes:

Different modes of gesture class association to acoustic and movement features were identified through logistic regression, that are based on regions of particular interest in the raga pitch space organization and analogous cross-domain morphologies.
- In the case of Hussain, the use of 5 non-collinear audio and movement features yielded a high classification rate (AUC) of about 95%. According to the results, it is more likely that interactions with elastic objects are performed by hand gestures that exhibit a low absolute mean acceleration value and a large variation in the hands' divergence (speed in moving the hands apart). They are associated with slower and larger melodic movements that ascend to a higher degree of the scale. The highest degree of the scale (7th) in raga Jaunpuri is the most unstable and therefore it dictates a change in pitch movement direction from ascent to descent (i.e. a double pitch glide, changing direction on the 7th degree), which is similar to the direction change of the hands when interacting with an elastic object. It could be therefore suggested that the MIIOM type is associated with the grammatical rules of the raga.
- In the case of Sahu, the use of 4 non-collinear audio and movement features yielded a high classification rate (AUC) of about 80%. According to the results for Sahu, interactions with elastic objects are more likely to be performed with pitch movements of a larger interval and larger duration and with the hands moving faster and remaining bound to each other.

3.2.2. Generic scheme:

A more generic cross-performer MIIOM classification was achieved by almost identical general logistic models for the two singers, yielding a good fit (AUC) of about 86% (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 for larger melodic movements (calculated on an absolute logarithmic scale), and with the hands moving further apart for Hussain and less apart but faster in the case of Sahu.

4. CONCLUSIONS

Based on the results, we rejected the null-hypothesis that gesture and effort are unrelated to the melody and found statistically significant movement and sound features that (a) best fit each individual performer and (b) describe the phenomena in the most generic way across performers in estimating effort and classifying MIIOMs.

Overall, models were more heavily dependent on acoustic rather than movement features and there was a bigger overlap of acoustic rather than movement features between models. 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 movement features as non-trivial.

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 their exerted effort levels with melody provides good evidence for non-arbitrariness; this may reflect the dual nature of mapping in being associated with the mental organisation of the melodic context, the mechanical strain of vocalization, the alap

macro-structure and shared cross-modal morphologies. This conclusion leads us to further suggest that in designing an electronic musical instrument, a more flexible mapping scheme should be aimed for, which would enable a performer to switch between mapping modes.

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

REFERENCES

[1] J. d'Escriván, "To sing the body electric: Instruments and effort in the performance of electronic music", *Contemporary Music Review*, vol. 25, no. 1-2, pp. 183-191, 2006.

[2] P. Hackney, *Making connections: Total body integration through Bartenieff fundamentals*. Routledge, New York, NY, 1998.

[3] A. Mulder, *Design of virtual three-dimensional instruments for sound control*, 1st ed. Ottawa: National Library of Canada = Bibliothèque nationale du Canada, 1999.

[4] N. Castagné, C. Cadoz, *A Goals-Based Review of Physical Modelling*, In Proceedings of the International Computer Music Conference, Barcelona, Spain, pp. 343-346, 2005.

[5] K. Nymoen, R. Godøy, A. Jensenius and J. Torresen, "Analyzing correspondence between sound objects and body motion", *ACM Transactions on Applied Perception*, vol. 10, no. 2, pp. 1-22, 2013.