Effort in gestural interactions with imaginary objects in Hindustani Dhrupad vocal music

PASCHALIDOU, PANAGIOTA-STYLIANI

Effort in gestural interactions with imaginary objects in Hindustani Dhrupad vocal music

Panagiota-Styliani Paschalidou

ABSTRACT

Physical effort has often been regarded as a key factor of expressivity in music performance. Nevertheless, systematic experimental approaches to the subject have been rare. In North Indian classical (Hindustani) vocal music, singers often engage with melodic ideas during improvisation by manipulating intangible, imaginary objects with their hands, such as through stretching, pulling, pushing, throwing etc. The above observation suggests that some patterns of change in acoustic features allude to interactions that real objects through their physical properties can afford. The present study reports on the exploration of the relationships between movement and sound by accounting for the physical effort that such interactions require in the Dhrupad genre of Hindustani vocal improvisation.

The work follows a mixed methodological approach, combining qualitative and quantitative methods to analyse interviews, audio-visual material and movement data. Findings indicate that despite the flexibility in the way a Dhrupad vocalist might use his/her hands while singing, there is a certain degree of consistency by which performers associate effort levels with melody and types of gestural interactions with imaginary objects. However, different schemes of cross-modal associations are revealed for the vocalists analysed, that depend on the pitch space organisation of each particular melodic mode (rāga), 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. Based on the findings, I argue that gesturing in Dhrupad singing is guided by: the know-how of humans in interacting with and exerting effort on real objects of the environment, the movement–sound relationships transmitted from teacher to student in the oral music training context and the mechanical demands of vocalisation.

EFFORT
IN GESTURAL INTERACTIONS
WITH IMAGINARY OBJECTS
IN HINDUSTANI DHRUPAD VOCAL MUSIC

PANAGIOTA-STYLIANI PASCHALIDOU

THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF MUSIC

DURHAM UNIVERSITY

2017

Table of contents

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

Sections from chapters 4, 5, 6 and 8 of this thesis have been presented or are accepted for presentation in conferences.

Paschalidou, S., & Clayton, M. (2015, March). Towards a sound-gesture analysis in Hindustani Dhrupad vocal music: effort and raga space. Paper presented at *the (ICMEM) International Conference on the Multimodal Experience of Music*, Sheffield, UK, 23 - 25 March, 2015.

Paschalidou, S., Eerola, T., & Clayton, M. (2016, July). Voice and movement as predictors of gesture types and physical effort in virtual object interactions of classical Indian singing. In *Proceedings of the (MOCO) 3rd International Symposium on Movement and Computing* (p. 45), Thessaloniki, Greece, 5 -6 July, 2016.

Paschalidou, S., Clayton, M. & Eerola, T. Mapping physical effort to gesture and sound during interactions with imaginary objects in Hindustani vocal music. Paper to be presented at *the (ESCOM) 25th Anniversary Conference of the European Society for the Cognitive Sciences of Music*, Ghent, Belgium, 31 July - 4 August, 2017, accepted.

Statement of copyright

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors, prof. Martin Clayton and prof. Tuomas Eerola, for their feedback and their ongoing support, trust and encouragement. Having in-depth knowledge from two different disciplines has been valuable. I would also like to thank Mark Leman for advising me on methods of video analysis. A special thanks goes to all musicians who have been so generous in participating to my project by allowing me to record whole ālāp performances, some after travelling several hours on the train from all possible places in India and some even after twisting their ankle.... Without their contribution, this work would have never happened. In specific, I would like to thank: Zia Fariduddin Dagar, Mohi Bahauddin Dagar, Nirmalya Dey, Uday Bhawalkar, Umakant and Ramakant Gundecha, Afzal Hussain, Lakhan Lal Sahu, Aparna Shastri, Marianne Svašek, Pelva Naik, Sagar Morankar, Mamta Tripathi, Manoj Kumar, Manish Kumar Singh, Arjun Kochhar and Vincent Dijkema. Additionally, I would like to thank Nancy Kulkarni for her valuable advice and assistance while in India.

I would like to mention Ustad Zia Fariduddin Dagar in particular, to whom I am especially grateful. He has been such a great source of inspiration and the reason I came up with the idea of this project at the first place: his gestures have been so attractive, informative and clear! Several people and… luck contributed in finding my way to him. It all started with the masterpiece recording of rāga Yaman by Zia Mohiuddin Dagar that Periklis Tsoukalas introduced me to in 2006. Just out of a random chat with Yannis Oikonomou and the organisation of a unique workshop by Ermis Sawantoglou a year after, I found myself sitting in front of his brother (as well as Nirmalya Dey and Marianne Svašek), taking classes in (not even *in*!) a little Greek village on the mountains of Northern Greece learning how to sing Dhrupad! This was one of the highlights of my entire existence and a life changing experience. How lucky have I been! I wonder sometimes… Thank you Ustad!

Two people I owe a big part of this thesis to, without whom this work wouldn't have been possible are: Luiz Naveda for advising me at all stages of the PhD involving movement data (from setting up the MoCap system and collecting movement data to analysing the data in Matlab) and Sam Horlor for appearing as *deus ex machina* in the final stages of the writing and committing himself voluntarily to proof reading the entire thesis; both providing great support too! Thank you both! I would also like to thank at this point choreographers and dancers Dimitra and Dina Kastellou for cross-validating my video annotation data.

I am also grateful to every single person who read parts of my thesis and gave me feedback, advice and/or support; each of these inputs helped me a lot in different ways. Such people include: Simone Tarsitani (for technical assistance with hardware for fieldwork), Sakis Laios (for helping with understanding the ragas), Aleksandra Pawlik (for advising with interview analysis), Chrisoula Alexandraki (for suggestions on my writing), Katerina Tzedaki (for always inspiring discussions about sound), Markus Schmidt and family (for helping with some crucial discussions about Dhrupad and for giving critical advise), Esteban Maestre (for helping with 'THE framework' and giving me a critical piece of advice), Juan-Manuel Loiaza (for putting me off from writing about some tricky issues of embodied cognition), Maria Kouvarou (for some critical support just before field work), the Stavropoulos family (for accommodating me and spending some nice moments with me while in the UK), Eleni Katragkou (for various suggestions about scientific writing), Aris Tsantiropoulos (for an interesting discussion about the term 'agón' ('αγών') as a possible alternative to the term effort), Panos Amelidis (for critical advice), Ilke Kanakci (for making my stays in Durham so pleasant) and Unax Küssner (for putting me off from doing some tricky statistical methods by shouting in the right moment). I would also like to thank Laura Leante for (literally) saving my life once during my field work in India...

I also owe gratitude to members of the Technological Educational Institute of Crete and the Department of Music Technology and Acoustics Eng. in particular, specifically Evangelos Kapetanakis and Nektarios Papadogiannis, who supported me in taking a one-year sabbatical leave in the final period of the PhD.

Certainly, I would like to express my deepest gratitude and love to my parents (Anna and Fotis) and to Roulitsa for being always so supportive. In 1996, when I was in the second year of my physics degree, I was the person to whom my mother would rehearse her PhD viva presentation and I was the one who set up the projection screen for her defence (in physical chemistry). 21 years later it is the other way around. My father was the one who asked me when I first described my PhD topic to him: "So is this effort something related to the production of the voice or to a melodic idea"? At that time I didn't know the answer; now I know more! Not least, I would like to thank my amazing friends Maria Fasoura, Vicky Vamvouri, Vera Paschali, Ilias Koulamas and Yorgos Stylianopoulos for always caring for me and putting up with me at some difficult moments of the PhD. This thesis has been a great experience mostly because of the incredible people I have been surrounded by or have come across on my way to completion.

To
Nana

NOTES ON TRANSCRIPTION AND TRANSLITERATION

In Hindustani music there are seven basic scale degrees (svaras), whose names and abbreviations are listed here:

Note name	abbreviation	symbol	Western equivalent	degree
ṣaḍja	Sa	S	Do (movable in pitch height)	1
ṛṣabha	Re	R	Re	2
gāndhāra	Ga	G	Mi	3
madhyama	Ma	M	Fa	4
pañcama	Pa	P	Sol	5
dhaivata	Dha	D	La	6
niṣāda	Ni	N	Si	7

Table 0.1. Notes in Hindustani music.

For simplicity, numbers are used throughout this text to refer to scale degrees instead of the usual 'Sargam' system of Hindustani music. Low octaves are indicated using the " ' " mark before the note, and high octaves by using the " ' " after the note. The middle octave is not marked. Lowered notes begin with small letters, while the natural notes of the scale start with a capital letter[1]. The symbols "/" and "\" are used before some of the notes to indicate a smooth glide to this note from the previous. The symbol '~' shows an oscillation around a note. When the first note is placed in brackets, it means that this pitch is not part of the melodic phrase itself, but nonetheless is where the phrase starts. Notes in square brackets mean that the note is used as an ornament, while a "…" mark or notes in parentheses mean that the movement has started from a note that is usually lower, not always the same and not always clearly pitched.

The symbol "-/" in section 7.2.1. (4) indicates a stepwise ascent from one note to the other, i.e. a melodic movement where notes are not connected smoothly through a glide. Similarly, a stepwise descent is indicated by a "-\". The '^' symbol in the same section is used to show that emphasis is placed on the pitch immediately prior.

[1] Dhrupad singers do not discuss explicitly about the śrutis (microtonal intervals) that are needed to render various tones in each rāga, but they gain this understanding through repeated listening and practice.

TERMS FOR MELODIC ORNAMENTATIONS AND TECHNIQUES

There are a number of terms used in Dhrupad to describe various basic types of ornamentation or treatment of notes, which are described here and to which the reader can refer. Ornaments in Hindustani music have a grammatical significance and are at least as important as the note itself. These ornaments (some considered vocal techniques) include: *mīṇḍ*, *ghaseet*, *gamak* and *huḍak*.

Mīṇḍ is one of the most significant forms of ornamentation in Hindustani music. In its general understanding it means a smooth, uninterrupted glide from one note to another and could be compared to a glissando in Western music. In its basic form at very slow paces (as in the beginning of the improvisation) it is straightforward and uni-directional and can range from a simple span of two notes to a whole octave or even more. But as the pace gradually increases, the *mīṇḍ* can take more complex structures, such as bi-directional melodic movements, created by an initial ascent and a subsequent descent. The proper rendition of a *mīṇḍ* not only requires a high accuracy of the starting and ending notes, but also an exact knowledge of the implicated tones that need to be traversed according to the specific melodic mode. On a fretted plucked instrument (e.g. sitār, rudrā vīṇā[2], surbahār), a *mīṇḍ* can be produced by first plucking the string to produce the initial note and then bending it with the other hand by pulling it outwards or transversely across a fret. A higher note can be achieved by increasing its effective length and tension. Similarly, a *mīṇḍ* can be performed with the voice.

Ghaseet is the Hindi word that is used to denote an ornamentation produced by 'a sliding along the string of an instrument from one note to another, usually over an interval of several notes' (Miner, 1997: 163). Although the literal meaning of the word in Hindi is "drag" or "pull", according to Miner, in practice the ornamentation is produced by sliding on rather than pulling a string. The musician will first play a note by plucking the string, and while its sound persists he will slide a finger of the other hand to a different fret to alter the pitch (Roychaudhuri, 2000). In fretted string instruments (e.g. sitār) the technique is called *syuňt* and in fretless instruments (e.g. rudrā vīṇā) it is called *āśa* (Miner, 1997). In the voice this could be interpreted as a kind of portamento of Western music.

The difference between a *mīṇḍ* and a *ghaseet* in how they are produced on a fretted instrument, such as the sitār, is that a *mīṇḍ* is produced by pulling the string, while the *ghaseet* is produced by sliding the finger toward the neck (for a pitch rise) or away from the neck (for a pitch descent).

Gamak is a technique, which comprises an oscillation of pitch, either on a single note which is heavily shaken and repeated (Autrimncpa, 2012a) or (in faster movements) on a series of pitches that are rapidly and forcefully intoned. This is achieved by quick and strong sliding

[2] Vīṇā is the general word used to describe a family of stick zithers. Although Dhrupad is primarily a vocal music, the rudrā vīṇā is the only instrument that is traditionally used in this music genre. The right hand is used with a plectrum to pluck the strings in a downward motion, while the left hand is used for stopping and pressing the strings between the frets and then pulling and bending them towards the lower edge of the neck to raise the pitch. The degree to which the strings are pulled alters their tone.

movements through adjacent notes (Ruckert, 2004; Slawek, 1987). Each of the oscillations is articulated in Dhrupad with a different syllable (such as 'ta', 'ra', 'ri', 'na') and stress is always placed on the main melody note. The speed of the oscillation, as well as the amplitude (from a few cents to a whole octave) and the dynamic level (between loud and soft) can vary (Sanyal & Widdess, 2004: 81), but typically a *gamak* exhibits power and depth in the voice production and has a percussive effect. Although a variety of *gamak* techniques exist, one of the main ways to render a *gamak* in Dhrupad involves an oscillation that normally begins above the melody note, falls to the region of the melody note and then rises again back to the starting point (Sanyal & Widdess, 2004), producing a kind of V-shaped oscillation like the one showed in Figure 0.1:

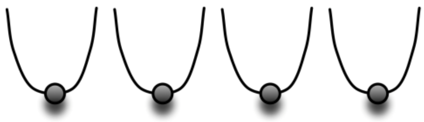


Figure 0.1. V-shaped oscillation (from Sanyal & Widdess, 2004: 81).

Gamaks can be distinguished from āndolans, due to the fact that they are normally faster and with well-defined beginning and ending pitches and their oscillations involve larger pitch intervals rather than microtonal, which is normally the case with āndolans.

Āndolan is a slow and gentle oscillation around a single note.

Huḍak is a vocal technique that is applied on an individual note to add emphasis. In this technique, a note is approached from a lower, but indeterminate pitch of the lower register through a large, but rapid and powerful glissando with a light and breathy voice. For added emphasis, the glissando may be repeated.

To distinguish between a *gamak* and a *huḍak*, Sanyal & Widdess (2004: 166) report that if a 'fully voiced tone' is used and the starting note is definite, the effect is regarded as a *mīṇḍ* or a *gamak* rather than a *huḍak*. Additionally, Sanyal makes the following distinction:

> 'In gamak, the notes spring with a modulating force and are articulated with rounded lips. In *huḍak*, the articulation is heavy and the notes are pulled upwards with extra breath force, creating a humming sound' (Sanyal, 1986: 46-47).

In the annotated material used in the analysis, no examples of *huḍaks* are found, but they are nevertheless mentioned in interviews and are therefore included here.

Mukhṛā is 'a short phrase of conventional but variable melodic and rhythmic structure, which marks the conclusion of each stage of an ālāp' on the tonic (Widdess, 2011: 157).

Chapter 1 – Introduction

In this thesis I will try to ascertain the close relationship 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. My main focus is on how much effort each gesture is perceived to require. Scholars have made strong claims in the past about the relationship between sound and movement, in both the production and perception of music. For instance, Kurth (1922: 1) has argued that melody is motion ('Melodie ist Bewegung') and Jensenius (2007: 1) that 'music is movement'. More recently, embodied approaches acknowledge the fact that the body is not a separate entity to the mind (the Cartesian dichotomy), but that it contributes to the formation of meaning through higher-level mental constructs and cognitive tasks.

The concept of embodiment refers to different ideas. It emerged in the field of phenomenology (Merleau-Ponty et al., 1968, 1945) as an alternative of the Cartesian distinction between mind and body. It later evolved to include various strands that account for the wide distribution of cognition beyond the boundaries of the brain or the skull (Gallagher, 2011; Shapiro, 2011; Hutchins, 2010; Thompson, 2005). Embodiment can now refer to the ecological psychology of affordances (Gibson, 1977, 1979), the formation of bodily schemata and sensorimotor representations (Brower, 2000; Wilkie et al., 2010; Lakoff & Johnson, 1980, 1999; Goldman, 2012), as well as the concept of enactment, which considers knowledge as an emergent phenomenon that is acquired through the action of doing and disputes the notion that cognition involves representations (Noë, 2004; Varela et al., 1991). Leman (2008) was the one who coined the term 'embodied music cognition' for using theories from cognitive science in the field of musical inquiry. He acknowledged the sonic domain and the human body as integral parts of our experiences in music and considered the body as the mediator between mind (the music experience) and matter (sound energy)[3].

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 (Stern, 2010; Eitan & Granot, 2006). From an engineering point of view, Mion et al. (2007) argue that 'when using a physical analogy, force is often subjectively considered as the cause and movement [...] as the effect'[4]. 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. Despite the fact that no real object is involved, we observe repeated patterns of bi-manual effortful gestures, comprising gripping, pulling and releasing phases (Rahaim, 2009). Manual Interactions with Imaginary Objects are the focus of this thesis and are abbreviated as

[3] The approach of the human body as a 'mediator' has been considered by Schiavio (2014: 13) as a fallacy due to the unintentional upholding of the Cartesian body-mind dichotomy.
[4] This might primarily refer to instrumental gestures, but not necessarily as we shall see.

'MIIOs' throughout this text. Additionally, the use of the word 'object' in this thesis refers to imaginary rather than real objects, unless explicitly stated.

Some limited discussion on the subject of MIIOs in Hindustani singing exists, but none of the work has focused solely on these interactions. None has acknowledged the role of effort as central to the subject and none has applied quantitative methods to three-dimensional movement data. Likewise, existing research studies have focused either on the Khyāl genre of Hindustani music (North Indian), or on Karnatak music (South Indian), but none on the Dhrupad genre of Hindustani music. Eitan & Granot (2006) have shown that for slow musical stimuli, a tendency is observed to ascribe motion to the impact of an outside force on an imagined character. Therefore, the precise intonation and the slow rendering of melodic phrases—especially in the ālāp section—of Dhrupad music make this specific genre ideal for studying the role of physical effort in movement–sound relationships.
It is worth asking whether movements during MIIOs relate to the action of forces, and if so, what type of forces and on what types of materials. The central theme of this thesis is the systematic exploration of how the kinaesthetic sensation of effort that is conveyed through movement while interacting with imaginary objects relates to the voice in Dhrupad vocal performance. To meet this end, the analysis takes advantage of both the analysis of hand movements and sound, as well as verbal descriptions of mechanical forces and motor-based metaphors. Despite the specificity of the genre, this thesis aims to address concerns in the study of movement–sound relationships that are of interest to the wider research community and thus outcomes may be extended to other music lineages, such as the Khyāl style, and electronic musical instruments.

Physical effort has been regarded as an important aspect of expression in music performance. Performers 'need to suffer a bit' in order to bring musical tension in the piece and audiences need to visually perceive physical effort in order to recognise particularly intense passages played by the musician (Krefeld & Waisvisz, 1990). In fact, Bennett et al. (2007), consider bodily effort as 'the impetus of musical expression' in the design of electronic musical instruments. However, systematic approaches to the dynamic aspects of movement in music performance—in contrast to its geometry or trajectory in space—have been given little attention by researchers. This is even more the case for non-instrumental music-making—such as singing—but also for non-Western so-called oral[5] music traditions, which emphasise transmission from teacher to student through direct imitation rather than written instructions of music notation[6]. The role of the body in the transmission of music knowledge in other genres of classical Indian oral music traditions (Khyāl/Hindustani and Karnatak) has been examined before by Pearson (2013), Clayton & Leante (2013), Rahaim (2009) and Fatone et al. (2011), but the novelty of this thesis lies in the systematic examination of the concept of effort. The current work offers an original contribution by approaching the Dhrupad genre of Hindustani vocal music as a fully embodied process and by stressing the importance of physical effort. I argue that by looking into other musical traditions, musical styles and music education systems, we might enrich our understanding of the relationship between movement and sound.

[5] As well as aural, visual and corporeal, as will be later explained.
[6] In the rare occasion that notation is used, it usually resembles a series of simple mnemonic symbols above the text.

1.1. Hypothesis, Research Questions and Objectives

MIIOs in Dhrupad singing are voluntary imitations of interactions with real objects and it is for this reason that I consider them of particular interest to embodied music cognition research. Vygotsky and Gal'perin emphasise the fact that 'mental acts origin in material acts' (Parreren & Carpay, 1972: 29), and Bakker et al. (2009) argue that higher functions of logical thinking and memory can only develop by manipulating physical objects. Similarly, by drawing on theories by O'Regan & Noë (2001), Thompson & Varela (2001) and Varela et al. (1991) that consider human cognition as founded on recurrent patterns of interactions with the real world and by extending Gibson's (1977, 1979) ecological theory of affordances to cover the imaginary objects involved in MIIOs, I aim to demonstrate that musical thinking in Dhrupad singing is grounded in the ubiquitous patterns of actions we possess through our ecological knowledge of interacting with objects of the real world. More specifically in the current work, I seek to examine the hypothesis that:

> Hand movements during MIIOs in Dhrupad singing are deeply related to the voice in our cognition through the effort possibilities that these imaginary objects can afford, as experienced by the performer and perceived by an observer. Central to this is an exploration of actions that draw on our prior experience and familiarity in interacting with the real world.

To examine this hypothesis, the following main research questions are posed:

[1] Is effort related to the voice in MIIOs during Dhrupad vocal improvisation in a systematic way or is it just arbitrary?

[2] Does effort simply reflect the mechanical requirements of vocal production or is it also related to cognitive aspects of melodic organisation by the performer?

Furthermore—in case such a relationship indeed exists—several sub-questions arise in the attempt to answer the main research questions:

[3] If such a relationship between effort and the voice exists, is it consistent among different performers or does it apply to only an individual performer?

[4] Which aspects of the voice is effort related to, and how can their relationship be best formally described?

[5] Are MIIO types related to specific aspects of the voice or are they completely arbitrary?

[6] If such a relationship between MIIO types and the voice exists, is it defined by melodic modes (rāgas), is it affected by idiosyncratic factors or does it reflect more generic trends observed among various performers and rāga performances?

[7] Which aspects of the voice are MIIO types related to, and how can their relationship be best formally described?

[8] Is there a systematic relationship between effort and specific MIIO types?

For this, the current work takes an interdisciplinary approach and by drawing on theories and methodologies that exist in music technology, ethnomusicology and embodied music cognition, it offers an original contribution in studies of music and movement. It does this by:

(a) identifying recurrent associations between the perceived (by the observer) level of effort and categorical aspects of hand movements and voice, as well as between MIIO classes and categorical aspects of hand movements and voice; and
(b) devising formalised descriptions for the estimation of perceived bodily effort and the classification of gestures (interactions with elastic vs. rigid objects) based on movement and sonic features.

Singing might make it easier to identify relationships to mental constructions of musical thinking than instrumental playing, as the hands are free to demonstrate the performer's goal (a combination of musical intention and physical target in moving the hands in space) without the mechanical constraints of the instrument. In other words, the absence of a real object that needs to be manipulated may bring to the fore aspects of gesture–sound relationships that are not necessarily defined by the physical laws of mechanical couplings, but are primarily driven by the musicians' conceptual concerns and musical intentions. The spontaneous (without explicit instructions by the researcher) employment of imaginary objects can also be associated with linguistic expressions of motor-based metaphors that musicians frequently use during interviews. Therefore, Dhrupad vocal performances are suited for studying such metaphors in an ecologically valid way. In addition to this, the voluntary imitation of interactions with objects that could, but do not really exist, makes it possible to identify gesture–sound relationships that are grounded in ubiquitous patterns of actions we know from the real world. It also imposes some constraints upon musicians' movements and therefore reduces the otherwise high complexity of full-body motion and the necessity for the use of dimensionality reduction techniques or artificially-induced exogenous constraints (as in Nymoen et al., 2013).

As will be extensively discussed later, the idea of MIIOs is supported by interviewees' frequent recourse to motor-based metaphors and associated imagery. This is also observed in the teaching context, where these images help to clarify how students should use their voice in difficult phrases. Thus, body movements, musical ideas and imagery are three concepts that work in unison (Leante, 2009) and are all employed in the complex process of music-making. Analysing sound, movement and imagery together is important in understanding how artists relate to their music and construct meaning in performance; additional information can be retrieved from movement and imagery that would not be accessed through sound alone (Fatone et al., 2011; Novack & Goldin-Meadow, 2015). Therefore, all three modalities are addressed in the methodology followed: sound through audio recordings, movement through video recordings and motion data, and mental imagery through interviews. To this end, the work combines qualitative and quantitative methods to analyse interviews, audio-visual material and movement data.

By better understanding the role of physical effort in the relationships between movement and the singing voice in MIIOs, I argue that we can significantly enhance our understanding of how performers conceptualise music in Hindustani singing through the specific occasions that they seem to interact with imaginary objects. In addition, this can be of paramount importance for the design and development of new digital interaction paradigms and electronic musical instruments. As will be further discussed in the motivation section 2.4.4. , previously unexplored HCI metaphors inspired by effortful interactions with the real world may prove valuable (Ward, 2013) and contribute to the enhancement of empty-handed artificial interactions with sound, for both novice and expert music performers of new electronic musical instruments alike. In other words, the design of empty-handed digital audio interactions may benefit from findings pertaining to gesture–sound relationships and the role of effort in MIIOs in Hindustani singing. Thus, this thesis can be seen as an attempt to fill a research gap that exists between technology and humanities; on the one hand in the discipline of musicology and music performance studies and on the other hand in the interdisciplinary field of music-related HCI[7].

[7] New Interfaces for Musical Expression (NIME), Digital Musical Instruments (Miranda & Wanderley, 2006), Sonic Interaction Design (Franinović & Serafin, 2013) and Interactive Sonification (Hermann & Hunt, 2005).

1.2. Thesis structure

In Chapter 2 the goals of this study are contextualised by outlining some relevant research work. Chapter 3 presents the methodological framework, justifying the use of mixed methods to systematically integrate qualitative and quantitative data and approaches. It also includes a description of the data collection process in the field in India. The mixed methodological approach leads naturally to the separation of the remainder of the thesis: the first part presents the analysis and findings of qualitative methods, the second describes the analysis and findings of the quantitative methods and the third and final part brings together findings from both methods. Conclusions are discussed in the last chapter.

Part 1: Qualitative analysis

Chapter 4 presents a thematic analysis of the interview material, which involves both the explicit and the tacit knowledge that is encoded in speech through motor-based metaphors.

Chapters 5, 6 and 7 describe the analysis of the audio-visual material in three different case studies: (5) Afzal Hussain of rāga Jaunpurī, (6) Lakhan Lal Sahu of rāga Mālkauns and (7) the Gundecha brothers of rāga Bhūpālī respectively. The main part of each of these chapters is dedicated to the classification of recurrent MIIOs from a third-person viewpoint and the exploration of associations between gesture classes, physical effort levels and characteristic melodic phrases.

Part 2: Quantitative analysis

Chapter 8 reports on the quantitative methods used to explore whether a small set of movement and sonic features would be adequate in describing the complex phenomena of MIIOs for Afzal Hussain and Lakhan Lal Sahu in a compact way. The chapter details the linear models that are fit to the data and the model selection criteria that are followed for (a) estimating effort level and (b) classifying gesture (distinguishing between interactions with elastic vs. rigid objects).

Part 3: Final Results and Conclusions

Chapter 9 offers an overview of conclusions produced through comparison and integration of results from both qualitative and quantitative methods. It then discusses challenges and limitations of the methodology, along with some suggestions for extending this work in the future.

Chapter 2 – BACKGROUND

The scope of this chapter is to situate the thesis against the theoretical frameworks from related research, and at the same time to highlight the innovative approaches by which it offers an original contribution. This includes an overview of the Dhrupad vocal music tradition, the tacit transmission of movement-related knowledge from teacher to students and its relevance to imagery and metaphoricity. The chapter also offers a definition of the term ‘gesture’ as it is used in this thesis, and this carries implications for how expression in music performance is understood in the current context. There follows a detailed elaboration on the meaning of the term ‘effort’ and finally an overview of computational approaches in analysing cross-modal relationships.

2.1. Dhrupad vocal music

In this section I discuss relevant features of classical Indian music and of the Dhrupad genre in particular. First, I present some basic information about this tradition and genre. This is provided as orientation for readers not familiar with the music. Readers familiar with the Dhrupad genre can move directly to section 2.1.3. Next, I discuss research literature on body movement and its tacit transmission in Hindustani singing. This section aims to emphasise the importance of motor imagery in how Dhrupad singers conceptualise music and also to explain why MIIOs are an interesting type of movement to study.

2.1.1. Hindustani music background

Of the two predominant styles of Hindustani (North Indian classical) music[8], Khyāl and Dhrupad, the latter offers unique possibilities for analysing the relationship between movement and sound, since performances start at a strikingly low pace, with slow melodic development. Hindustani music relies heavily on improvisation, which is rule-based and conforms to the 'rāga system'. 'Rāga' refers to a concept of melodic mode lying between scale and tune (Powers & Widdess, 2001)[9]. It follows a movable Do (tonic) system, in which the tonic is defined by the most comfortable pitch area of the singer. The constantly sounding drone instrument (called taňpurā) is therefore commonly tuned on the middle tonic, the middle 5th and the tonic an octave lower. There are seven basic scale degrees (svaras), from which the first (tonic) and 5th are never altered (referred to as śuddha = natural). The 2nd, 3rd, 6th and 7th degrees can be lowered by a semitone (komal = flat) and the 4th degree can be raised by a semitone (tīvra = sharp). In many rāgas, all seven scale-degrees can be used when descending, but some are omitted when ascending. The tonic and the 5th function as strong anchor points in most cases, while two notes (the vādī and samvādī), unique to individual rāgas, gain attention and special treatment and contribute to the special mood of each rāga. Three octaves are used in performance, the low (mandra), the middle (madhya) and the high (tār).

Especially in the Dagar tradition of Dhrupad, singers will frequently refer to the multiple 'shades' of a note, thereby acknowledging that a note is not considered to have fixed pitches (not even the tonic[10]). Instead, a note is conceived of as a flexible and fluent entity that can be rendered over a micro-tonal range of pitches around a central point, and can move and merge into another note. Among the most distinctive features of Dhrupad is the intricate micro-melodic treatment of notes and their long duration. This is not to say that pitch height is arbitrary, but instead that there is some flexibility in the interpretation.

[8] For Hindustani music in general the reader can refer to Wade (1998) and Rowell (1992), for Dhrupad to Sanyal & Widdess (2004) and for Khyāl to Wade (1975).
[9] Loosely, one can say that a rāga refers to a group of important features, including the scale (discrete pitches), melodic movement, specific phrases, mood, ornamentation and even distinct images, as depicted in the so called Rāgamala miniature paintings. For a concrete explanation of the term, following readings are suggested: Widdess, 1995; Bor, 1999; Autrimncpa, 2012e.
[10] Such as the shades of the tonic (Sa) explained by Bahauddin Dagar in this video: https://www.youtube.com/watch?v=iSzNIK4ISzI

2.1.2. Performance setting and ālāp improvisation

Dhrupad is usually performed by a main artist (soloist) with the accompaniment of at least one drone instrument (taňpurā), of the pakhāvaj drum (in metred sections), and sometimes a melodic accompanying instrument. All performers are seated on the ground in particular arrangements, with the soloist at the centre. Typically, a Dhrupad performance consists of two parts: an extended melodic improvisation called ālāp (the improvisatory opening section of the performance, where the rāga is first exposed) and a short composition. The improvisation starts with a slow non-metered[11] section at lower pitches (also called ālāp), in which various melodic features of the rāga are systematically presented for the first time. It gradually builds up in both pace and pitch, leading through a section with a steady and prominent pulse (joṛ) to a quick and highly rhythmical part (jhālā, where the drum may join the voice); in each of these sections the same features are introduced, but with a different rhythmic character. Each part typically starts by first establishing the middle tonic and then it incorporates a gradual expansion of the melody over a total of about 2.5 octaves. The melody starts from the middle tonic, moves down and explores the lower octave range, then ascends and explores the middle octave, it then further ascends into the upper octave (usually it rises up to the 3rd degree) and finally it descends back to the middle tonic, where the melody resolves.

Individual notes are presented one at a time, by first touching them momentarily or just pointing to them through various types of melodic movements, before they gain the full attention of the performer and are established more firmly through different variations of melodic movements. This doesn't happen in a random manner, but according to a set of flexible—but not arbitrary—improvised movements (some of which are characteristic of the specific rāga and are called pakaḍ), that lead to a series of ascending and descending phrases. This series of ascending and descending phrases seems to be used in an effort to build up tension in the listener for the final ascent to the highest pitch.

There is a dramatic character in the gradual ascent towards the pitch climax (i.e. the tonic of the highest octave), with a progressive building-up of melodic tension, which is periodically resolved by longer stops on the tonic and—according to the rāga—shorter stops on the 5th degree of each octave. No lyrics are used in the improvisation, but instead a particular repertoire of non-lexical syllables, called 'nom-tom' (Sanyal & Widdess, 2004: 152), which are also used for the creation of the pulse in the joṛ. The composition involves a few lines of lyrics and encompasses a second, rhythmical improvisation technique, this time on the text. The composition is not part of the analysis and therefore is not discussed in more detail here.

2.1.3. Tacit transmission of music and music-related knowledge

Unlike Western classical music, in Indian classical music students do not rely on written notation and knowledge is transmitted directly from teacher to disciple (the so-called 'guru-śiṣya paramparā'[12]); in this sense Hindustani music can be called an oral tradition. Historically, students would dedicate a significant number of years to living with the teacher, studying with him, practicing (riyāz) under his

---

[11] Some tempo exists, but there is no obvious meter involved.
[12] Paramparā means tradition in Hindi, in the sense of something handed down.

presence and supervision, following him to performances, adopting his style and thus inheriting the style of the particular gharānā[13]. Although this ideal case does not apply any more, guru-śiṣya training remains.

During the teaching process, teachers will often refer to emotions and images of body actions or pictures of nature. Additionally, they will employ a number of manual movements to convey clarifications, especially when they need to explain difficult musical passages or embellishments that a student cannot repeat. Most importantly in the context of the current research, it is not just the sound that is transmitted over music lineages but hand movements too (Neuman, 2004). Although typically gesturing in Hindustani vocal music (which includes Dhrupad) is not taught explicitly, students tend to move recognisably like their teachers (Rahaim, 2009), albeit not as exact replicas; this indicates an additional idiosyncratic element in moving.

In both performance practice and teaching sessions of Hindustani vocal music, observations highlight a number of repeated patterns of gestures (Rahaim, 2009; Clayton, 2007; Leante, 2009 and personal observations) that are related in a consistent way to key phrases (pakaḍ) of rāgas. They seem to be transmitted through practice from teachers to disciples (Pearson, 2013) in the form of evident gestural mirroring (Rahaim, 2009)[14]. For this reason, the term 'oral' has been recently criticised by Rahaim (2009), who has argued that it does not account for the visual (kinematic images of the teacher's 'musicking body') and kinaesthetic (awareness of positions, dispositions and equilibrium of one's own 'musicking body' (ibid.)) sensory channels, which constitute the complete tacit transmission of knowledge.

Rahaim (2009) has coined the term 'paramparic body' (for the Khyāl genre) to describe a set of bodily dispositions that students develop over time following those available in their vocal lineage. The concept draws on Young's 'family body', which refers to the bodily dispositions that family members develop from others in their family (Young, 2002). Thus not only the music style, but also the movement style seems to be inherited, and there is good reason that it might be connected to the style of the particular music lineage (gharānā) and/or the idiosyncratic character of the teacher. This observation 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 of musicians sharing the same teacher. The following section further highlights the importance of 'extra-musical' information, such as imagery and movement in Dhrupad performance and the conceptualisation of music.

2.1.4. Imagery and movement

Ethnographic studies by Clayton (2005), Rahaim (2009), Leante (2009) and Fatone et al. (2011) report that Hindustani music is deeply rooted in cross-sensorial experience and that visual elements are significant for performers (as well as listeners). They associate specific modes of rāga improvisations with dramatic scenes (as illustrated in rāgamālā paintings (Ebeling, 1973)), people,

---

[13] Stylistic school or music lineage.

[14] More on observational learning can be read in (Vogt & Tomaschke, 2007).

deities and emotions and, most importantly for our work, patterns of motor activity. Singers make use of these imagistic elements both while performing and teaching, in order to facilitate a better understanding of the task to be translated into the voice (Fatone et al., 2011). Movement and voice seem to convey the most salient features of these images by projecting diverse qualities associated with the rāga (Leante, 2009). In linguistics, McNeill (1992: 29-30) argues that movement and speech 'arise from a single process of utterance formation' (also termed 'growth point' by McNeill & Duncan (1998) and 'idea unit' by Quek et al. (2002)). This initial image or unit of thinking is created first and it is then transformed into a dynamic process of organisation which is shared over the sensory modalities; both movement and speech are integral parts of this emergent complex structure (McNeill, 2005, 1992: 29-30; Cook, 2001). In other words, body movement does not simply embellish spoken language, but it is part of the language generation process (McNeill & Levy, 1980). Movement and speech employ distinct forms of representation; movement offers a dynamic analogue of the linguistic construction and conveys meaning holistically by relying on visual and motor-mimetic imagery, while speech conveys meaning discretely by relying on codified words and grammatical devices. As a result, although the two modalities can overlap a great deal (match), they do not always contribute identical information, but most often supplement as well as contradict (mis-match) each other (Goldin-Meadow, 2006, 2003; Goldin-Meadow & Wagner, 2005; McNeill, 2005; Kendon, 2004).

In a similar fashion to speech, movement and imagery in singing are not arbitrary and simply decorative. Despite differences between music and speech in creating dynamic sonic analogues (see Zbikowski, 2011), Leante (2009) draws on McNeill's work and refers to a 'single process of expression' in Hindustani vocal performance, which is expressed by both the dynamic resources of movement and the stable constructs of music grammar. Similarly to speech, the two modalities can reflect, contradict or supplement each other and mutually contribute to the creation of meaning in a rāga performance. Imagery, alongside movement, in Hindustani vocal music can reveal information that it would not be possible to gain solely from sound (Fatone et al., 2011); this is where the interview analysis and the observation analysis of movement can prove useful in this thesis. It will be interesting to see whether the results of this thesis will reveal that movement and voice are similar, contradictory or complementary to each other during MIIOs. The (mutual) complementarity of modalities has been also stressed (as more important than simple multi-modal fusion) in HCI and the gestural control of sound by Tanaka & Knapp (2002). Thus, results of this thesis may have implications in strategies by which gesture–sound mappings are applied to interactive sound applications and new electronic musical instruments.

The 'manual cognition' of sound has been widely recognised as an essential tool to musical thinking due to its contribution to conceptual structures of thought (Godøy, 2010: 119). Due to the fact that movement is not usually explicitly acknowledged, novel ideas may be introduced easier through hand gestures than the voice, exactly because they are not the focus of attention (Novack & Goldin-Meadow, 2015; Goldin-Meadow, 2014, 2003). The mental simulation of action (motor imagery), or 'knowing how' (Cox, 2011; Godøy et al., 2006a) is considered as an integral part of music cognition and perception (Berthoz, 2000). Similarly, Godøy (2006) and Jensenius (2007) argue that musical sound has visual (gesture images of kinematics) and motor (gesture sensations of effort, proprioception) components which are based on the biomechanical constraints of what we imagine that our bodies can do. As in speech, where gestures may play a cognitive role in generating the utterances and prosody of the language, in vocal music gestures may also help to configure sound production. In fact, Pfordresher et al. (2015b) have recently argued that mental imagery forms the core of sensorimotor associations that drive vocal pitch imitation in singing as well as other cross-modal associations. I will seek to examine in this thesis whether the voluntary movement constraint imposed by the extension of the body through the imaginary objects in MIIOs is random, or whether it offers an equivalent restriction on what the voice is allowed to do. The next section offers a short discussion on the conception of pitch by Hindustani singers.

2.1.5. Conceptualising pitch

Hindustani music is extremely rich in flowing motivic melodic patterns. There is an explicit discussion among musicians about pitch continua, and especially in Dhrupad smooth and slow melodic glides (mīṇḍ) abound. In fact, according to Van der Meer & Rao (2006), 'the space 'between the notes' is often more important than the discrete notes themselves'. While melodic movement is considered as a sequence of discrete and well-defined notes (svaras) and Dhrupad musicians are especially rigorous about the precision and subtleties of their intonation, at the same time it is also conceived of as an imaginary pitch 'space'[15] (Fatone et al., 2011: 212) in which a vocalist can move smoothly by following paths, trajectories and shapes through a continuum between the scale steps (Battey, 2004). This observation highlights the dual mode by which Hindustani musicians conceive melodic movement and reflects the dichotomy of the human brain with its capacity to process information at multiple levels of temporal resolution, from an analytic to a holistic mode (Hunt & Kirk, 2000). It does not mean that all melodic paths are allowed or that all points can be accessed in the same way; the melodic 'space' is not uniform, but rather it is 'coloured' at different points, reflecting the grammatical rules of the melodic mode (the 'topography' of the rāga, as Rahaim (2009) describes it), as well as the aesthetics of the particular gharānā.

The notion of a rāga as a movement in a melodic pitch 'space' (Fatone et al., 2011: 212) is also accompanied by smooth hand movements in the real space (deliberate or unconscious). However, it is not just about spatial trajectories or geometry (the melographic representation in the 3D Euclidian space as Rahaim (2009) puts it), but also about the sensation of forces. Hindustani vocalists make frequent verbal use of motor-based metaphors and they also employ and manipulate imaginary objects that appear to be somehow related to the melodic motives with which they engage, such as by stretching, pulling, pushing, collecting and throwing. 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. In other words, Hindustani 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 aim of the current research is to examine whether and to what extent the sound may be defined by the know-how of manipulating the object employed and the associated effort required during MIIOs.

---

[15] And sub-spaces, called jāgāh (Rahaim, 2009), which are delimited by specific degrees of the scale (thāṭs).

2.1.6. MIIOs

Hand movements can have different functions in performance or teaching, such as expressing the performers' own engagement with musical sounds (Clayton, 2005; Rahaim, 2009; Leante, 2009; Fatone et al., 2011), communicating with other performers (Moran, 2007), or communicating with the audience (Clayton, 2007) or the students (Pearson, 2013); only the first is of interest in the context of this thesis. According to Rahaim (2009) and personal observations, Hindustani singers seem to engage with the melodic content in two main ways: by keeping the hand(s) open or by closing the hands (forming an initial grip). The open-handed and the closed-handed modes are supposed to reflect two different relationships between the body and the voice. In the open-handed mode, the hands seem to trace curves as trajectories effortlessly[16] in space, providing a melographic representation of the sound; the hand-shape is also assumed to represent the voice's timbre (Rahaim, 2009 and interview testimony in section 4.2.3.). In the closed-handed mode, the hands first form a grip before appearing to manipulate some imaginary object by effortfully acting against the imagined opposing force associated with changing its shape, size, orientation etc. For example, one may notice gestures that look as if the singer is:

- Moving something like a ball around space or throwing it from one hand to the other;
- Bouncing something like a ball on the floor[17];
- Creating the shape of a knot with the two hands;
- Pulling a heavy object or stretching something like a rubber band with the hands

(from Fatone et al., 2011; Rahaim, 2009).

The current work is only concerned with closed-handed interactions with imaginary objects, such as shoving, pulling or pushing a heavy object in space, or stretching and compressing an elastic material. 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. It is not clear whether vocalists have a clear visual imagery of such objects as defined geometrical constructions or if imagery only involves the proprioceptive sensation of their physicality (although this would be interesting to explore in the future). However, in each of these cases, by grasping, holding, extending and then releasing them, vocalists seem to manipulate notes as smooth pitch glides (mīṇḍs). This contact lasts as long as the object serves the purpose of the vocalisation; vocalists seem to unconsciously drop it thereafter.

According to Rahaim (2009: 82), during such interactions the gesture–voice relationship seems to reside in the consistent physicality of the imagined object attached to the hands rather than the spatial and acoustic dimensions. Although it is obviously not expected for a performer to be consciously computing mathematical calculations of physical laws in real-time, it seems that prior recurrent experiences are related to the melodic activity while singing. One of the main assumptions that this work examines is that when Dhrupad singers appear to interact with an imaginary object, the

[16] Or at least with very low effort, if one considers that any gesture essentially carries some level of effort.

[17] Similar descriptions have been reported by subjects in (Eitan & Granot, 2006) as an iconic way to associate an accelerating repeated pitch with images of the descending movement of a ball ('bouncing downhill').

relationship to the sound resides in the interaction possibilities the object affords (according to its physical properties, such as viscosity, elasticity, weight, friction etc.) and the effort it is perceived to require. Affordance in interaction design is a concept that describes the kinds of usage or 'practical opportunities' (Rizzolatti & Sinigaglia, 2008: 34) that a real object invites of the user (Norman, 1988). It was introduced in ecological psychology by Gibson (1977; 1979) as a set of possible motor actions that are evoked by the intrinsic properties of a real object. In other words, although there are multiple ways by which an individual can interact with an object, these are again limited by its physical or geometrical properties and the context. This notion attests to the importance of prior sensorimotor experience and the fact that our knowledge of a real object is not founded on its physical properties and its relation to other (real) objects (as the traditional view of cognitive theory had it), but on the actions we have performed before in relation to this type of (real) objects. This so-called action-based knowledge (or know-how) is fundamental from the point of view of enactive theories.

It is now widely accepted that our ecological knowledge of relationships between our body, (real) objects of the environment and sounds underpins much of our conception of any sound. Music perception (and performance) can be seen as an active process of exploration that draws on a 'reservoir of sensorimotor skills' (Noë, 2004: 27) or movement–sound contingencies that a person develops over time by interacting with the world. O'Regan & Noë (2001), Thompson & Varela (2001) and Varela et al. (1991) have argued that robust movement–sound relationships and cognitive structures develop over time through the cross-modal links that we gain from our everyday life interactions with the environment. Similarly, when playing a musical instrument, strong cross-modal relationships and robust motor programs[18] (Calvert et al., 1988) are developed in the embodied mind of a performer as a result of training and repetition. From a different point of view, based on McAdams (1984) and Bregman (1990), Clarke (2005: 71-72) has also suggested that any movement-sound relationship may be viewed as a perceptual reality that obeys to the same physical laws as actual movement and thus it should not be solely characterised by metaphorical, symbolic or analogical transfer. The only difference between everyday sounds and music, he argues, is that the source is in the first case real and in the second virtual.

According to motor theory, perceiving sound in speech is equivalent to internally simulating the gestures that a listener believes could have potentially produced it (Godøy et al., 2006b). In other words, mental images of sound evoke mental images of sound-producing gestures and vice versa (Mikumo, 1998). Initially originating in linguistics[19], the idea of motor theory has been later extended to include music (Krueger, 2014) and has more recently been supported by different domains such as behavioural studies and neuroscience[20]. The underlying idea behind imitation in music performance is that we constantly and continuously simulate some aspects of sounds we hear, either in the source (the production of sound) or the effect (the actual sound). It is supported by Noë (2004), who claims

[18] For a debate on the matter of motor programs see Schiavio, 2014.
[19] Motor theory in speech production and perception refers to the importance of imagining how speech sounds are produced by the lips, the vocal tract and the air-flow movements (Fowler & Turvey, 2006); Fadiga et al., 2002; Liberman & Mattingly, 1985; Galantucci et al., 2006).
[20] Through the discovery of mirror and canonical neurons. For this, see Rizzolatti & Sinigaglia, 2008; Rizzolatti, 2005; Wilson & Knöblich, 2005; Kohler & Kaysers, 2002; Jeannerod, 2001; Iacoboni et al., 1999; and Gallese et al., 1996.

that the way we understand music in listening situations is based on established ecological knowledge.

Prior interactions with the real world (Noë, 2004) are constrained by the materials involved and determine predictable patterns of co-variation between sonic features and self-movement over time (Freed, 1990; Warren & Verbrugge, 1984). Varela et al. (1991) have emphasised the importance of recurrent patterns of sensorimotor experience (e.g. bouncing a ball or stretching an elastic object) and have claimed that cognitive structures emerge from their slightly varied instances that are attributed to 'image schemata'. According to Johnson (1987), these schemata (which are not exclusively visual) are pre-conceptual, meaning that they are extracted versions of general features of sensorimotor experiences and carry qualities related to the temporal, dynamic and multimodal character of an event. With this in mind, image schemata also allow us to 'ground our description of elusive musical phenomena in concepts derived from everyday experience' (Zbikowski, 2002: 64) and are thus essential to the way musical concepts are connected to other domains. They constrain the possibilities of cross-domain mappings, allowing only a certain range and freedom of possible patterns of reasoning and understanding, which are heavily context-specific (Johnson, 1987). Cox (2011) has argued that implicit knowledge of the physicality of the sound source shapes musical meaning. This ecological knowledge constrains the possible sonic results by allowing us to imagine the gesture that could have produced them and leads to the concept of musical affordance (Menin & Schiavio, 2012; Clarke, 2005; Reybrouck, 2012, 2005; Nussbaum, 2007; Krueger, 2014). Musical affordance is the concept that describes the multiple sonorous possibilities of a gesture, which—although flexible—are still not endless and are constrained by the gestural interaction with the (real) object. In other words, there is a more profound movement–sound relationship that is not just related to strict 'mechanical couplings but to more fundamental, ubiquitous cognitive schemata' (Godøy, 2006: 150). Musical affordance has also been studied in HCI by Tanaka et al. (2012) and has been described as an 'action–sound palette' by Jensenius (2007).

By virtue of our ecological knowledge, we attend to sonic events with certain expectations (Huron, 2006). This means that watching a ball being kicked we anticipate a certain acoustic quality, which depends on the material and the size of the (real) object as well as the way it was kicked; we would be probably caught by surprise if the acoustic quality did not match the visual information. However, it has been also suggested (Küssner, 2014) that we are able to grasp objects figuratively (e.g. the imaginary objects of MIIOs) because we are able to grasp them literally and this shapes our understanding of more abstract concepts too (Gallese & Lakoff, 2005). For instance, if to play with a ball is a learnt skill (from our development since our childhood (Gabbard, 2011)), to imagine the bouncing properties of the ball is also a skill (Haga, 2008). Similarly, if to stretch an elastic band is a learnt skill, to imagine the elastic band being stretched or to manually mimic the stretching during a MIIO is also a skill. Based on this idea, it wouldn't be a crude assumption to also think of an imagined object that is apparently employed in MIIOs as a carrier of certain patterns, general rules and opportunities for behaving (Camurri et al., 2001) that are defined by the imagined resistive force according to size, shape, material etc. and that also afford specific types of sonic results.

In fact, from an engineering point of view, Mion et al. (2007) have argued that 'when using a physical analogy, force is often subjectively considered as the cause and movement [...] as the effect'[21], as displayed in Figure 2.1. Drawing on this idea, Mion et al. (2010) have modelled gesture–sound relationships through the physical measure of admittance Y, which is the inverse of impedance or resistance, such as in elasticity, friction or inertia[22].

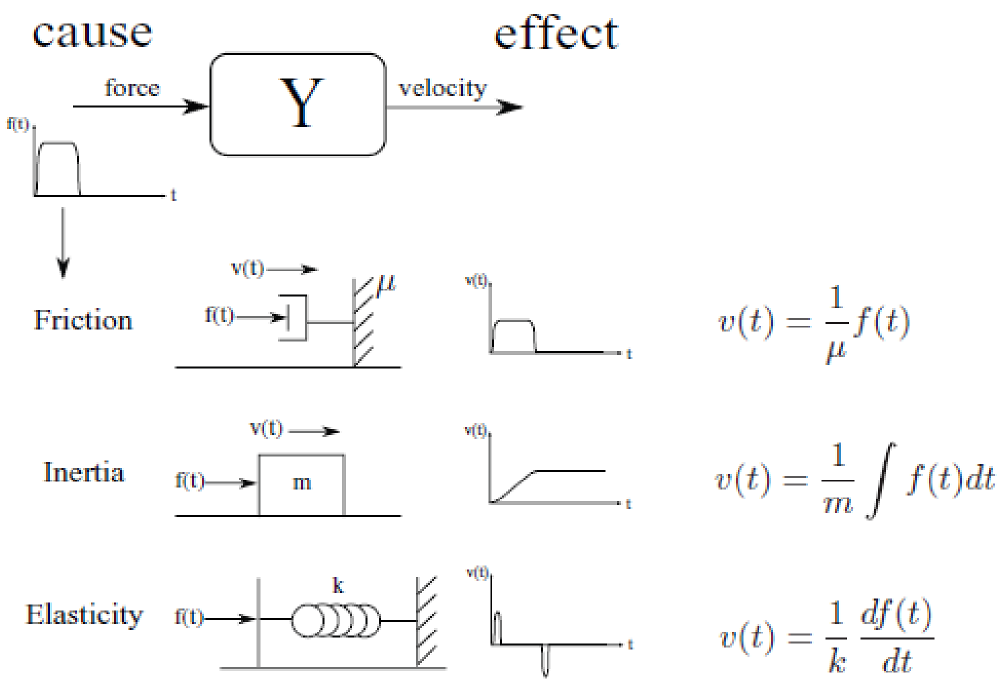


Figure 2.1. Behaviour of the basic linear mechanical systems: friction, inertia, elasticity (Mion et al., 2010: 14: 19).

Although this approach has been developed in the context of instrumental gestures, in which movement is considered as the cause and sound as the effect, gesture–sound couplings of the real world are so deeply rooted in our cognitive system that other types of reciprocal gesture–sound relationships may also be guided by this knowledge (Jensenius, 2007). Therefore, the type and magnitude of resistance/admittance (through elasticity, friction, inertia etc.) may also represent the gesture–sound mappings of MIIOs, where the object is not real and the interaction is only imitated.

The idea of imitation constitutes the concept of 'motor-mimetic music cognition' or 'motor mimicry' by Godøy (2003) and Godøy et al. (2006b) and the 'mimetic hypothesis' by Cox (2001) and provides a robust foundation for the current work on gesture–sound relationships in MIIOs. Mimesis (which overlaps to some extent with imitation or simulation) refers to gestures in both the imagistic (Cox, 2011) and the physical domain (Godøy et al., 2016; Godøy, 2003). Godøy (2006) has suggested a bi-directional model where we actually rely on an incessant process of mentally tracing features of sound when listening to (Lahav et al., 2007) or just imagining (Zatorre & Halpern, 2005) music. During MIIOs performers appear to engage with the melody through an object by imitating the induction of energy that would be required in case the object was real. What is actually imitated is the energetic coupling

[21] This might primarily refer to instrumental gestures, but not necessarily as we shall see.
[22] Inertia and elasticity are considered as those parts of a mechanical system that store energy, while friction is responsible for its dissipation.

between the movements of the hands and the physical state of the imagined object. It is important to note here that imitation does not necessarily correspond to a perfect reproduction, but it can be a simplified abstraction of the original gesture. Therefore, one should not expect to see exact replicas of interactions with real object during MIIOs by Indian singers. Wohlschläger et al. (2003) argue that the imitation becomes more precise around a target point (the goal) and for this reason they refer to the so-called 'Goal-Oriented Imitation' ('GOADI').

As will be further explained in section 2.3.1. , effort is the subjective measure of this (imagined or real) energetic coupling and thus it forms the central point of the current work. Considering that effort is the subjective concept that alludes to the forcefulness of the movement (the resistive forces that cause and constrain it), it is also expected to be somehow reflected on its sonic counterpart. For instance, when we see a musician performing a large, slow, heavy movement as in stretching an elastic object, we expect a similarly heavy, sustained and forceful tone in the voice with a similar shape of evolution, because of the imagined continuous transfer of energy in a hypothetical sound-producing gesture. On the contrary, accompanying a sustained and smooth pitch glide with a short, ballistic-type of gesture, such as a rapid throwing gesture, would create a feeling of mis-match to an observer.

Sensations of elasticity, weight etc. and the acknowledgement of these and other physical metaphors in musical thought are not exclusive to Hindustani music and they are frequently verbally articulated in other musical genres too. A striking example of the notion of elasticity is mentioned by a Canadian bagpipe instructor in (Fatone et al., 2011). However, the deliberate and easily identifiable gestural demonstration of this and other physically-inspired concepts in the ecologically valid conditions of a music performance is not common and Dhrupad singing offers a distinctive case in this sense. Systematic studies on basic concepts of mechanics, such as friction, elasticity, or inertia (the 'FEI metaphor'), exist in Western types of music (and are discussed in section 2.2.2.). However, these studies rely exclusively on designed experiments rather than real performances.

There are several open questions in relation to the role of MIIOs and the amount of effort these are perceived to require in Dhrupad singing. In the first place, it is reasonable to wonder why these specific types of movement are performed in MIIOs and what they have to offer, e.g. whether they function as a refinement tool of the melodic content or as a mnemonic tool for eliciting 'correct' melodic phrases established over year of practice with the teacher; or whether they serve a theatrical role. Although MIIOs can be considered as founded on rudimentary knowledge of interaction with the environment, how these are related to their melodic counterpart and whether there is any level of consistency in the type of interaction (stretch, pull etc.) and the type of sonic event that is simultaneously rendered by the voice are still open questions. An observer would most probably be in the position to perceive some kind of correspondence or similarity between the movement and sonic events. But how is this similarity grounded, if there is no causality between movement and voice, and interactions are just imaginary? Is it useful to talk about similarity at all, or should the apparent likeness between these two modes be thought of as an approximate analogy (similar to Godøy et al., 2016)? Even more fundamentally, is there any relationship at all between voice and hand movement, or are they arbitrarily connected, even assuming that instances do in fact occur simultaneously? How loose or tight is this relationship? In other words, do cross-modal relationships indicate a level of similarity, contradiction or complementarity? Which features of the two modalities (the motor and the auditory) are most strongly related and how could this relationship (or mapping) be analytically described? Which cross-modal features could supposedly reflect the sensation of effort in manipulating objects that are only imagined but absent from the real world? In other words, what are the kinematic/dynamic movement and acoustic features that would enable the prediction or description of how observers perceive the effort exerted by musicians during performance? Finally, is there some level of consistency in the relationships between movement, sound and effort throughout a single performance and between performances of an individual singer and among different performers? From these questions it becomes evident that there is good scope for systematically examining whether and how modalities and physical effort are related to each other, and this forms the central theme of this thesis. For the moment, the term effort remains loosely defined, but a section follows to specify its meaning in the current work.

In section 2.1. I argued that musical ideas, body movements and imagery are three concepts that work in unison and integrate into the holistic experience of musical performance (as in Leante, 2009). In music research, music-related body movement has often been discussed through the term ‘gesture’. However, this is a rather broadly defined term and not easy to pinpoint. Hence, the next section aims to offer an overview of understandings, definitions and taxonomies, as well as to clarify how it is used in this work.

2.2. Gesture

The term gesture has taken various meanings in different disciplines, such as cognitive sciences, experimental psychology, philosophy, linguistics, biomechanics, human-computer-interaction, and musicology. However, these meanings often remain vague. The current section outlines how the term is used in the present research context and how it is implicated in studying gesture–sound relationships, specifically in MIIOs.

2.2.1. Definition

Traditionally, gesture has been broadly regarded as a movement of the body that contains information (Kurtenbach & Halteen, 1990). More recently, the definition of gesture has also been extended to other modalities, to include not just physical movement but also the sonic and imagistic domains. Currently, gesture is regarded as both 'a mental and a corporeal phenomenon' (Leman & Godøy, 2010: 8) that carries expressive information and meaning. Gesture may be now used to refer to empty-handed movements (in which no contact with a real object is required), to movements that involve the manipulation of a real object, to general body movements, to dynamic auditory contours in music perception and performance, to visual imagery, feelings of motion (such as kinaesthetic, tactile, proprioceptive sensations of effort and ideomotor responses), to speech or even to experiences in touch, taste and smell (Luciani, 2004).

Hatten (2004: 95) argues that musical gestures denote inner-musical qualities as 'significant energetic shaping through time with unique expressive force'. Rather than a sequence of particular pitches, music theorist Ernst Kurth (1922) conceives the essence of melody as generated from internal patterns of energetic tension and release. According to Kurth, these patterns make their manifestation in the acoustic world as curves of musical intensification and abatement, resulting in a 'play of [psychological] tensions'. Musical experience is often considered as an ebb and flow of tension that gives rise to emotional responses (Lerdahl & Krumhansl, 2007; Krumhansl & Schenk, 1997; Vines et al., 2006).

These energetic shapes of musical gestures can be portrayed as physical gestures by the hands. According to Godøy (2009: 205), the hands may trace the geometry of the sound (elements such as pitch contours, pitch spread, rhythmical patterns, textures, and even timbral features), but also convey sensations of effort in the music. These so-called 'motion bells' (Camurri et al., 2003) can be conceived as subsequent event units, whose energy is formed around moments of stronger emphasis. According to psychologist Daniel Stern (2010), they form peaks and decays[23] or impulses and rebounds that are supposed to have emerged from our general capacity to move through approach and withdrawal. Stern (ibid.) considers these units as a dynamic form of 'vitality' that emerges as a combination of five aspects, namely force, motion, time, space and intentionality.

Despite acknowledging individual differences between sensory modalities, using a single, common term gives the advantage of surpassing the Cartesian divide between physicality and mind and

[23] The 'peak structure' according to Kendon (1996) in speech gestures.

entering into the realm of embodied cognition. In music, Jensenius et al. (2010) use the term to denote the combined sensations of physical movement and sound, while Delalande (1988) notes that the term 'musical gesture' lies in the intersection between observable actions and mental representations. The observable side of gesture is addressed in this thesis by analysing sound and movement through qualitative (video observations) and quantitative (regression analysis) methods, while its imagistic/mental counterpart is addressed through the analysis of interviews.

Although I do agree with the unifying use of gesture, for the sake of clarity I use the following separate terms for each modality in this thesis:

- 'gesture' or 'physical gesture' for voluntary, goal-directed and coherent movement (event) units or chunks in music performance (even if they don't directly produce sound), which include a prefix, the (main) excitation phase and a suffix (termed as 'action' by Jensenius, 2007);
- 'melodic phrase' or 'melodic movement' for sound gestures (sonic event units)[24].

As I primarily refer to MIIOs (which are well-defined movement chunks) throughout this thesis, I will mostly use the term 'gesture'. The basic premise behind this is that I consider vocalists' hand movements during MIIOs as goal-directed towards salient moments (a combination of physical, acoustic and imagined). I will use the broad term 'movement' only when discussing in general about motion (the displacement of a body part of the human skeleton) and I will keep the term 'action' for goal-oriented coherent body movement units that are not necessarily related to music making.

Event units are considered to be the result of the combined music-related and biomechanical (Hogan and Sternad 2007; Nelson, 1983) constraints and motor-control factors of the body[25], as well as our capacity for attention and short-term memory (Snyder, 2000). It has been suggested that we actually tend to subjectively experience such event units predominantly in terms of dynamics, as caused and guided by some driving force (Runeson & Frykholm, 1983; Stern, 2010). Although most prominent in instrumental excitatory movements, this applies to all kinds of music-related movement and it can be argued that the notion of an imaginary object in MIIOs imposes an additional constraint that is supposed to serve a musical purpose.

2.2.2. Expression through motor-based metaphors

Humanistic theories have stressed the role of movement in communicating expressiveness, as a voluntary and intentional act (Coker, 1972; Gabrielsson, 1995; Kendon, 2004). Expression refers to both the performer's intentionality for communicating musical structure and meaning and the means by which this meaning is conveyed. The means refers to both the different effectors that may be used to produce sound and the temporal dynamics of the multiple modalities. Camurri et al. (2001) argue that physical movement, sound and imagery are supposed to obey the same kind of rules or

[24] Although not needed in this thesis, I would use the term 'imaginary gesture' or 'figurative gesture' (Delalande, 1988) for units of mental expressive intentions.

[25] Biomechanical constraints of the body include: limits and thresholds to speed and endurance, phase transitions with changes in tempi (Todd et al., 1999), the need for alternating between effort and rest, the need to conserve energy (Godøy, 2003) etc. Similarly, schematic constraints on the sonic counterpart are imposed on melodic, rhythmic and textural aspects of musical expression (Godøy, 1997, 1999) and this holds true for the human voice too (such as the need for inhalation, the increase and subsequent release of vocal fold tension in the larynx etc.).

spatial/temporal patterns to convey this expressive information and this is of central importance to this thesis. Along these lines, I seek to understand whether specific patterns of movement involving magnitude and temporal structure—such as the stretching of an elastic band or the shoving of a heavy object in space—might correspond to specific patterns in the voice. Based on similar ideas, the relations between changes in musical tempo and movement dynamics have been systematically explored in the past[26], such as kinematic models for the average timing of final musical ritardandi profiles that are matched to the timing of runners' deceleration towards halt (Friberg & Sundberg, 1999) or relationships between walking speed and preferred musical tempi (Van Noorden & Moelants, 1999).

According to movement theorist and pioneer of modern dance Rudolf Laban (Laban & Lawrence, 1974; Laban & Ullmann, 1971) and other scholars in movement studies (Schiphorst, 2009), expressive power is not only a feature of the kind of movement performed, i.e. the categorical aspect such as punching, but mainly of the way it is performed, i.e. its dynamical aspect such as hard or light. This is the main argument on which this thesis is also drawing and therefore both questions—the 'what' and the 'how'—are addressed by the methodology of the current thesis. In fact, in order to specify those features of nonverbal behaviours that describe how a specific gesture is performed, for example its 'temporal dynamics, fluidity or energy' (Niewiadomski et al., 2013), the terms 'expressive gesture quality' (Niewiadomski et al., 2013), 'expressive movement' (Pierce & Pierce, 2009) or 'corporeal articulation' (Leman, 2008) are also frequently used. In this sense, the expressive content is considered as 'different and in most cases independent and superimposed on any possible denotative meanings' (Camurri, 2004a) or as the 'added value of a performance' (Canazza et al., 2004). Expression in this sense refers to the variation of qualities in natural actions (Camurri et al., 2003; Mion et al., 2010). This means that the same type of movement may be repeated with different expressive intentions and these may have an impact on the way it (and its auditory counterpart) is performed or else on the temporal evolution or its dynamical profile. Many scholars have applied this understanding of expression to its analysis in the realms of both physical movement[27] and sound[28].

From discussions I have had outside the interview context during my fieldwork in India, Dhrupad singers have reported that it is exactly these deviations in rendering the same melodic phrase (which Sundberg et al. (1995) call 'meaningful deviations') that are supposed to make Hindustani music interesting (see the quotation by Mohi Bahauddin Dagar on page 101). Similarly, it is the qualities or patterns in performed actions that make them distinct from others and allow them to be identified as 'structural units that have internal consistency' (Camurri et al., 2004a: 3). According to Camurri et al. (2001: 1), 'it is likely that expressiveness as such subsumes certain universal patterns and general rules'. For example, one would expect that the act of stretching an elastic material carries different movement qualities in comparison to pulling a rigid object in space. Drawing on this idea, the current

---

[26] Honing, 2003; Juslin et al., 2002; Friberg et al., 2000; Friberg & Sundberg, 1999; Van Noorden & Moelants, 1999; Sundberg & Friberg, 1996; Repp, 1995; Todd, 1992; Feldman et al., 1992; Kronman & Sundberg, 1987; Clynes & Nettheim, 1982; Sundberg & Verillo, 1980; Clynes, 1973; Gabrielsson, 1973; Repp, 1993; Gjerdingen, 1994 and Shove & Repp, 1995.

[27] Naveda & Leman, 2010; Godøy et al., 2010; Dahl & Friberg, 2007; Watson, 2006; Wanderley et al., 2005; Suzuki & Hashimoto, 2004; Camurri et al. 2003; Dael et al., 2013; Niewiadomski et al., 2013; Stern, 2010; Chi et al., 2000; Hsieh & Luciani, 2005; Laban, 1974; Camurri et al., 2004b, 2003.

[28] Widmer & Goebl, 2004; de Poli, 2004; Dixon, 2000; Palmer, 1989; Desain & Honing, 1991; Clarke, 1993, 1985; Gabrielsson, 1997; Schaeffer, 1966; Imberty, 1981; Todd, 1985, 1992, 1995; Repp, 1990, 1992.

thesis examines whether the qualities of manual movements accompanying Dhrupad vocal performances may allow us to classify MIIOs and how their movement qualities may be linked to qualities in the voice in terms of distinct patterns.

In the context of non-verbal communication, such as music and dance, there are numerous ways to describe expression, and music can communicate various kinds of messages, each of which contributes only in part to the complex issue of performance. The most common way to study expression in music research is through its affective content, which can be described through the basic emotions (Gabrielsson & Juslin, 2003), the circumplex 'valence-arousal' model of affect (Russel, 1980) or the Kinematics-Energy space (Bigand et al., 2005). Expression can be measured in this respect through the emotional responses of performers in sound and movement (Picard, 1997; Cowie, 2001).

However, an alternative way by which expression can be understood is through a set of linguistic descriptors of motor-based metaphors, such as hard - soft, heavy - light, which are drawn from basic concepts of mechanics (such as elasticity, viscosity, friction etc.). This approach of expression in music performance is of particular interest to the current research, due to the fact that Hindustani singers make a frequent verbal use of physically-inspired sensorimotor labels to describe acoustic properties (Chapter 4 offers a detailed interview analysis). Based on the interviews and the video analysis that follow, it is also reasonable to assume that there is a connection between these labels and MIIOs during performance. Thus, by using observations of MIIOs during Dhrupad vocal performance and by drawing on motor-based metaphors, there is a unique opportunity to approach human experience through the ubiquitous reference of the body. Due to the ubiquity of the body, physically-inspired descriptors displayed through MIIOs may be less culture-specific or stylistic than emotions[29] and therefore they might provide a useful framework for the representation of expressive content in HCI.

The idea of motor-based metaphors has been drawing high attention in music (Eitan & Granot, 2006), such as the verbal description of melody in terms of imagined 'musical forces' (inspired by the physical concepts of gravity, magnetism and inertia) and has been put into use for both the design of interactive systems[30] and the analysis of music[31]; in fact, not only for Western music (Feld, 1981). However, no prior work exists in Hindustani music and none which approaches the subject through real performance recordings.

Previous studies have already used this approach of expression in exploring cross-domain mappings in music research. For instance, Canazza et al. (2003) showed that expressive intentions in playing a music score could be conveyed by a performer tasked with following a list of sensory-type adjectives, and Krumhansl (1997) verified by using perceptual tests that these expressive intentions could be correctly recognised by listeners. Based on these findings, Canazza et al. (2003) have further argued

[29] Gabrielsson & Juslin (2003) concluded that there is no universally accepted set of categorical descriptors for classifying music according to emotion.
[30] Visi et al., 2014; Fels et al., 2002.
[31] Hatten, 2015; Lerdahl & Krumhansl, 2007; Larson, 2006, 1997; Larson & Vanhandel, 2005; Johnson & Larson, 2003; Eitan & Timmers, 2010; Bakker et al., 2009; Zbikowski, 2002, 1997; Saslaw, 1996; Fenza et al., 2005; Wilkie et al., 2009; De Poli et al., 2009 ; Mion et al., 2010.

that sensory-type adjectives can be used to model expressiveness at an abstract level and use this model as a user-interface for producing synthetic sounds from gesture inputs. This idea overlaps with the future aims of the current work in using effort exerted through MIIOs as an intermediate level of gesture–sound mappings; it is discussed in more detail in section 2.2.3.

Additionally, de Poli et al. (2009) conducted extensive studies to verify whether non-linguistic acoustic and haptic descriptors (which they called 'attractors') related to physical concepts and sensory modalities (such as friction, elasticity, inertia, what they called the 'FEI metaphor') could be associated with musical stimuli and the emotional cues that they convey. It has been argued that expression understood through sensorial descriptors of motor-based metaphors consistently clusters with and is hence related to various high-level representations of emotive content on one side (Mion et al., 2010; de Poli et al., 2009 & 2005; Canazza et al., 2003) and to various low-level sonic features of musical stimuli on the other side (Mion & de Poli, 2010). In relation to the current project, this means that it should be possible to deduce physically-inspired descriptors reflecting music expressivity from low-level features extracted from the raw data collected in India. Similar studies exist also in movement research that link motor-based metaphors to high-level emotions and low-level movement features (extracted from video recordings, accelerometers and 3D motion capture systems, e.g. light-point displays and the Kinect sensor). Such examples include the work by Wallbott (1998), Zhao (2001), Lagerlöf and Djerf (2001), Camurri et al. (2003), Volpe (2003), Castellano et al. (2007) and Morita et al. (2013).

Typically, such experiments consider participants' gestural responses to audio stimuli that are supposed to be associated with particular physical metaphors or to listeners' linguistic expressions of motor-based metaphors and sensorimotor descriptors that are induced through sound (Mion et al., 2010; de Poli et al., 2009, 2005; Canazza et al., 2003). Despite the lack of ecological validity in the chosen methodology, what is interesting to draw from the results of the abovementioned studies is that in recognising the expressive content of musical performances, listeners and performers alike may employ features that are not exclusively musical but are grounded in interactions with the real world. These features can be deduced from direct measurements on the raw audio or movement signals. Such findings can have interesting implications for the way gesture–sound relationships in MIIOs are analysed. The most important is that in order to qualify as embodied, the research approach must entail both the sonic domain and the human body as an integral part of a performer's experiences in music (Leman, 2008). In other words, features from both modalities—the sonic and the physical—should be included in the research framework in order to approach expressivity as a motor-based metaphor.

However, the scholars in the studies cited above have made some questionable choices. First of all, none of the studies mentioned has relied on the combination of both sound and physical movement in approaching expression as a physically-inspired metaphor. Additionally, the cited studies lack ecological validity in the conducting of tests. The material used for these studies did not involve the spontaneous hand movements of real performances but mostly designed experiments in which performers were asked to respond when induced by audio stimuli that expressed motor-based metaphors. Moreover, as Scherer (1994) has argued, the use of verbal labels (as the ones used in the cited experiments for priming participants during experiments) is potentially problematic, since it can encourage participants to simplify what they actually experience.

The above mentioned studies are also exclusively concerned with instrumental music playing. Due to the energetic coupling between the gesture (the cause) and the vibrating body of an instrument (the result) that is constrained by physical laws, the relationship that was examined between sound and sensorial descriptors of physical metaphors seems almost self-evident and expected in this context. Temporal patterns of excitation are known to be preserved in the acoustic information and Warren & Verbrugge (1984) have shown that they have perceptual relevance. For example, a study by Mion & de Poli (2008) in playing the violin concluded that the acoustical measure of the residual energy REh[32] is related to the perceived effort and it can be associated with the physical metaphor of strong/weak. However, in the case of the violin, high residual energy in the high range of the frequency spectrum reflects the strong physical pressure of the bow on the strings, and therefore the study simply confirms the physicality of the interaction and the direct causality of the movement on the sonic effect. Finally, the material discussed up to this point has been exclusively drawn from Western tonal music and clearly calls for more research into non-Western music cultures.

To address these issues in the current thesis, I use recorded material of real performances and examine the role of motor-based metaphors when there is no direct energetic coupling between movement and sound, such as in singing. There is definitely some scope for assessing whether similar associations between sound and motor-based metaphors are still valid in the context of vocal music, in which hand movements only accompany the sound and are not directly responsible for its production. MIIOs in Dhrupad singing offer a unique opportunity to study music-related movements that allude to physical metaphors in ecologically valid (non-laboratory) conditions. The performers happen to demonstrate the conceived physical metaphors through identifiable overt hand movements in real performance settings. Dhrupad is a suitable and interesting case of music-making for studying links to the non-discrete nature of movement and phenomena of mechanics, also because smooth pitch transitions in this genre abound and notes are approached through a strong sense of a pitch continuum. Finally, by conducting my study in Hindustani music, I aim to extend movement–sound research into non-Western types of music.

2.2.3. Levels of abstraction in describing movement and sound

Descriptors in both audio and physical movement are defined at different levels of abstraction and temporal contexts. According to Niewiadomski et al. (2013: 3), expression encompasses ‘at least two important aspects of expressive gesture quality analysis, that is, the low-level feature detection and its high-level interpretation in terms of its eventual communicative meaning’. As supported by de Poli et al. (1991), Camurri et al. (2001), Leman (1997), Zannos (1999), and Godøy & Jorgensen (2001), both sound and human movement semantics can be represented at various levels of abstraction and in varying ways of representation, ranging from the purely functional (non-symbolic) to the purely

[32] REh is the part of the residual energy that is computed in the upper part of the spectrum (Serrà, 1997).

symbolic. The symbolic aspect focuses on well-established, fixed symbols and their communicative meanings (such as sign language in physical movements, score in music and text in speech), while the non-symbolic (non-propositional) relies on the fact that distinct emotions or expressive intentions are often associated with distinct qualities of body movements. This thesis does not address symbolic representations of gestures, but 'expressive gesture qualities'.

To address the multiple aspects of expression in the performing arts, Camurri et al. (2001) and de Poli et al. (2005) have developed multi-layered approaches. According to Camurri et al. (2005), information from any modality can be processed at different levels of abstraction: the low-level/sensorial (capturing information from sensors), the medium-level/syntactic (describing gestures in terms of trajectories) and the high-level/semantic (describing music in terms of emotions). Although this multi-layered framework refers to gesture in the sense of physical movement, as shown in Figure 2.2, it can be extended to include any modality.

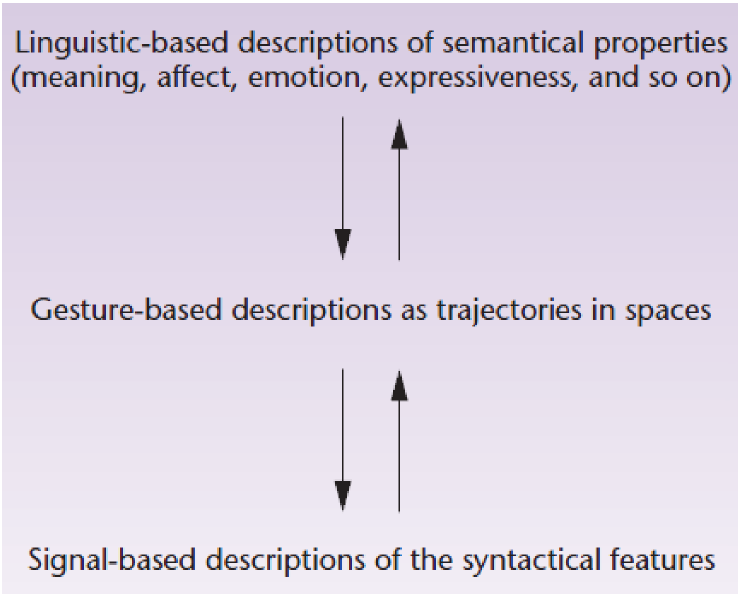


Figure 2.2. A multi-layered conceptual framework for expressive gesture analysis and applications (Camurri et al., 2005).

In the context of the current work, expression is understood at an intermediate level of abstraction (the 'gesture' level) that lies between the high-level affective content described through categorical labels of semantic properties (emotions) and the low-level syntactic, physical layer of the raw signals. At this level, expression in Hindustani vocal music is described through the display of MIIOs and is associated with the perceived sensation of effort (resistance) and physically-inspired verbal descriptors. Therefore, in the current thesis I work with features that are computed on a movement unit of a few seconds and account for expressiveness in terms of events, shapes, patterns or trajectories (Camurri et al., 2001). The challenge in the context of Dhrupad singing is to find whether and how the subjective concept of effort may relate to, and thus can be deduced from, the lower cross-modal syntactic layer, which may include low-level features of movement and sound extracted from the raw captured signals.

Now that I have established the way by which gesture as a carrier of expressivity is understood and approached in the present research context, it is useful to clarify what type of movements I aim to scrutinise. The next section offers a general overview of different types of music-related movements and taxonomies and discusses the specific gestures of interest in this thesis.

2.2.4. Taxonomies and classifications

Several taxonomies for categorising the broad range of movement types—and specifically those related to music—have been suggested over the years. An extensive literature on the topic can be found in music performance, which includes edited books and volumes[33] and individual discussions[34]. However, divergent opinions on the topic still exist and no final consensus on a single taxonomy has been made yet, which reflects the vast range of movement types. The intention of this section is to illustrate the fact that MIIOs form a special case of music-related movement that makes them hard to classify and position in any of the taxonomies that have been proposed up to date. Only two taxonomies—rather than a comprehensive list—are presented here. The first refers exclusively to movements accompanying singing, while the second refers to general music-related movements. Although specific taxonomies for singing movements exist (Clayton, 2007, 2005; Davidson, 2005, 2002, 2001), they classify them in a way that does not necessarily address MIIOs. The second and more general movement taxonomy presented here is discussed in more detail in order to highlight the ambiguity in classifying MIIOs.

1. The taxonomy by Clayton (2007, 2005) focuses specifically on vocalists' movements in Khyāl singing and categorises them according to their function[35] in performance or teaching. Considering the close resemblance of singing movements to speech movements, this taxonomy was developed by following an analogous procedure to McNeill's (2005, 2000, 1992) extension of 'Kendon's continuum' (Kendon, 2004) and the work by Ekman & Friesen (1972) in speech movement studies. It includes 'markers' of musical processes or structure, 'emblems' for communication with other musicians or the audience, and 'illustrators' which are 'tied to the content of singing, appearing analogous to the melodic flow or "motion"' (Clayton, 2007: 75). Only this last type (also called 'gesticulations' according to Kendon and McNeill) is of interest to this work.

2. A widely accepted movement classification has been developed by Godøy et al. (2012), according to which music-related movements can be classified based on their (a) content, (b) timescale and (c) function.

(a) The **content** refers to training level, sociocultural background and performance practice (solo/ensemble) of the performer and is not of special relevance to the current work.

[33] Rojc & Campbell, 2013; Godøy & Leman, 2010; Gritten & King, 2006; Wanderley & Battier, 2000.
[34] Jensenius et al., 2009; Jensenius, 2007; Miranda & Wanderley, 2006; Windsor, 2012; Lakoff & Johnson, 1980; Scruton, 1997; Zbikowski, 1997; Métois, 1997; Luciani, 2007; Godøy et al., 2006a; Davidson, 2001, 2002, 2005; Quek et al., 2002; Cadoz & Wanderley, 2000; Cadoz, 1988; Delalande, 1988; Ekman and Friesen, 1972.
[35] In analogy to speech, 'function' denotes here the way in which singers illustrate or complement musical meaning with their physical movements.

(b) The **timescale** reflects the multiple temporal and structural complexities of movement that are implicated in movement analysis methods (Francoise et al., 2012) and the subdivision of continuous body motion into smaller chunks of varying timescales:

(b1) the *micro*-timescale (0-0.5 seconds duration for single tones and stationary movement features: Melucci et al., 2000; Cambouropoulos, 2001);

(b2) the *meso*-timescale (0.5–5 seconds duration for movement events (gestures) and gesture-sonorous objects: Schaeffer, 1966; Godøy, 2006; Bonini & Rodà, 2001; Mion, 2003; Hsieh & Luciani, 2005);

(b3) the *macro*-timescale (for longer stretches of sections and even of whole works: Camurri et al., 2005; Suzuki & Hashimoto 2004; Canazza et al., 2004; Bresin & Friberg, 2000).

One of the challenges in studying movement-sound relationships is the understanding of the information at different levels of detail (Godøy, 2009). The basic tenet of the current research is that the most prominent movement–sound relationships are expected at a meso-level timescale (Godøy et al., 2016). At this level, coarticulation phenomena are implicated in the fusion of the otherwise separate gestures into larger movement units or trajectories; and vice versa in the chunking of continuous streams of information into holistic movement units or 'atoms', meaningful entities, concrete motor images, trajectories or Gestalts (Godøy et al., 2011; Godøy, 2010; Buxton, 1986; Cox, 2006; Godøy et al., 2006a; Küssner, 2013, 2014; Noyce et al., 2013). This applies to both the sonorous and the movement modalities and is considered as an extension of the sonorous object by Schaeffer (1966) to the concept of the gestural-sonorous object by Godøy (2006). In sound at this level the study focuses on patterns or events, which may include motives, ornaments, melodic shapes and dynamic profiles (Schaeffer, 1952; Bregman, 1990), which 'may be primordial to the singular sound or tone' (Godøy, 2009: 209). Therefore, the analysis concentrates on movements that typically span between 0.5-5 seconds and this stance has implications on the annotation, segmentation and feature extraction processes (Henbing & Leman, 2007).

(c) The **function** addresses the role of movement in music performance and forms the foundation for our discussion on MIIOs. The functional classification can be divided into communicative, sound-producing, ancillary and sound-accompanying.

(c1) *Communicative* movements are intended for communication, which can be further subdivided into endogenous, performer-performer or performer-audience (Jensenius, 2007) and are not of interest to this work.

(c2) *Sound-producing* movements[36] describe human actions that produce (*excitation gestures*) or modify (*modification gestures*) the resonant features of an instrument and therefore bear a clear causal effect to the sound, which is determined by physical laws and a strict energetic coupling with the instrument. Excitation gestures refer to the triggering of the sound and modification gestures to its succeeding adjustment (Cadoz, 1988).

[36] Other terms are also used for sound-producing movements, such as *effective* (Delalande, 1988), *ergotic* (Cadoz, 1994), *technical* (Davidson, 2005 and Delalande, 1988) and instrumental (Cadoz & Wanderley, 2000).

(c3) *Ancillary* (Jensenius, 2007; Wanderley & Depalle, 2004) or *sound-facilitating* (Jensenius et al., 2010)[37] movements co-occur with sound-producing movements, but they are not directly responsible for the generation of sound. Ancillary movements may refer to movements that performers make to add and convey expressiveness to the performance[38], e.g. for the articulation, dynamic shaping and phrasing of sound. This type of movements has been extensively studied by Caramiaux et al. (2012), Nusseck & Wanderley (2009), Wanderley (2002) and Vines et al. (2005). Ancillary movements may also refer to the supporting function of movements, in other words to movements that instrumental music performers make for supporting sound-producing movements in various ways, but without them being directly responsible for sound production. They are primarily intended for physiological comfort, such as avoiding fatigue and strain injury and have been studied for instance by Dahl et al. (2010: 38) in drumming gestures.

(c4) *Sound-accompanying* movements are not involved in the production of sound itself but instead follow music, as in listening situations or dancing and acting to music (Jensenius et al., 2010; Jensenius, 2007). Such movements may refer to the approximate mimicry of the motor activity that underlies a sound-producing gesture (called 'motor mimicry' or 'motormimetic sketching' by Godøy et al., 2006b: 258) or to the tracing of some acoustic aspect following the evolution of a sounding (real) object's resonance (called 'sound tracing' by Godøy et al., 2006a: 27). The first case refers to imitating the excitation gesture (the cause) we believe might have produced the sound, while the second refers to tracing some features of the resulting sound (the effects of what we do, i.e. images of how materials resonate). In fact, Rocchesso et al. (2015) and Caramiaux et al. (2014, 2011, 2010b) have argued that listeners prefer to mimic the imagined sound-producing gesture if they are able to identify the sound source ("causal" character of movement in relation to sound); in the opposite case ("non-causal") they revert to a mode of manually describing the sound. More on this is discussed in the next section.

In practice the abovementioned functional categories are not always so clear-cut and the distinction is not always easy. This means that a movement could possibly fit equally well into several categories or it could be regarded as a combination of different classes. Figure 2.3 by Jensenius (2007: 57) displays the relationship of movement to sound as a continuum that ranges between the extreme cases of quite close to rather loose relationships. The four function classes appear positioned on this continuum, but in practice movements can fall between any of these classes indicating a frequent ambiguity in terms of movement classification.

[37] Also called *accompanying* gestures in Delalande (1998) and *non-obvious performer* gestures in Wanderley (2002).
[38] Also called *expressive* by Davidson (1994) and *expression-supporting* by Jensenius et al. (2009).

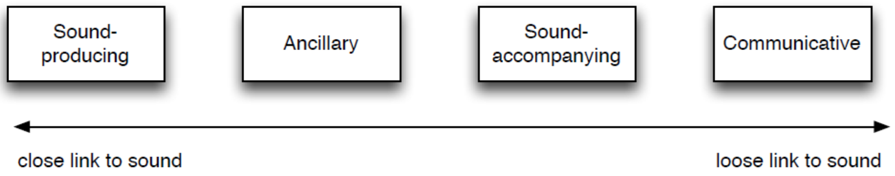


Figure 2.3. The four different music-related movement types (based on function) and their connection to sound on a continuum in respect to their relationship to sound (Jensenius, 2007: 57).

It is also interesting to note, that scholars of the research group that has developed the just cited taxonomy have also recently used an alternative classification (Godøy et al., 2016) that consists of the following three categories: (a) sound-producing, (b) sound-accompanying and (c) amodal, affective or emotive. The last category refers to any type of movement, which includes the physical, sonic and imaginary (motor-imagery) and is associated with global sensations of the sound, such as images of effort, balance and other concepts often encountered in dance. Although this is indeed a category that could fit MIIOs, this alternative taxonomy also shows the complexity of the task in developing a classification system that can be extensive and can have the capacity to capture the broad range of movement nuances.

Considering the general movement classification in terms of function just presented, it seems intuitive and quite straightforward to classify singing movements as sound-accompanying, as they are supposed to add expressivity and support vocalisation (their supporting function has been also mentioned during interviews and is discussed in section 4.2.). However, in the previously cited taxonomy the term has only been used in conjunction with instrumental movements and therefore it cannot be used in the context of singing in the current research. There seems to be a degree of confusion in the definition of ancillary and sound-accompanying movements in terms of their necessity to be related to sound-producing movements or not. This confusion is also apparent by the fact that Delalande (1998) has been using the term 'accompanying gestures' for gestures accompanying sound-producing gestures, while Jensenius et al. (2010) have used the term only as a response to music in a listening context. The deviation of terminology reveals that research in music-related movements has been dominated by instrumental music making. I would argue at this point that the ancillary type of movements in the functional taxonomy that was cited above should not be exclusively associated with sound-producing (instrumental) movements, as they can also be used in vocal performance. For instance, singers can use movements for facilitating voice production (Brunkan, 2015a) or for articulation and phrasing purposes (Pearson, 2016: 200). I would therefore suggest that sound-accompanying movements refer to all types of movements that do not directly produce sound irrespective of whether it is about producing or reacting to sound. In other words, I would consider a movement class called 'sound-accompanying' to include expressive, supporting, motor-mimetic and sound-tracing movements as sub-categories of movements that may accompany movements that produce or respond to sound.

In the next section, I shall endeavour to argue that apart from the fact that the previously cited 'functional' taxonomy falls short of classifying singing movements, MIIOs should be actually considered as a peculiar case that makes their classification even more ambiguous.

2.2.5. Ambiguity in classifying MIIOs

The previously cited functional classification distinguishes between movements that directly produce sound or facilitate and add expressiveness during music performance from those that respond to sound by imitating sound-producing movements or tracing sound features. However, I argue here that this distinction in terms of function is not really possible or easy in the case of singing, and specifically MIIOs. It would be difficult to classify MIIOs as sound-producing movements, but it would also be difficult to classify them as imitating sound-producing or sound-tracing movements. Here I explain why.

The direct production of sounds refers to instrumental (effective) movement, in which a tight energetic movement–sound coupling exists. This coupling is established by the strict physical laws that dictate how the sounding body vibrates according to excitation. In MIIOs no real object is involved and therefore no energetic coupling exists between the hands and the imagined object; or at least this is not an obvious one. Although the performers' movements appear to be directed towards this imagined object, it is not the object that vibrates and produces the sound. In this sense, it would be difficult to speak of a sound-producing gesture. At the same time, a direct or indirect physical effect on the mechanism of voice production might exist. For instance, a quick downward push may produce some desirable wobbling in the voice (as in gamaks) or the intensification of muscle tension in the arms may facilitate the control of the diaphragm. However, there is surely no direct energetic coupling between the hands and what is supposed to exist between them, and therefore it is not possible to talk about sound-producing movements. It could be therefore argued that MIIOs are sound-accompanying movements that rely on imitation.

Godøy (2006b) discriminates between the imitation of sound-producing and sound tracing gestures. It is, however, unclear whether the performer engages with the melody by mimicking gestures that could possibly produce these sounds or by tracing some aspects of the voice. Despite the absence of a real object, the relationship to the sound seems to be mediated by the imagined object. Thus, it is more likely that the performer is not engaging primarily with the effect of his movements (tracing the sound), but mostly focusing on gestures that could be producing the sound. There is an observable match between the voice and the apparent manipulative movements in terms of synchronisation and temporal congruency of various features, such as melody, dynamics and timbre. The vocalists seem to rely on prior sensorimotor experiences from recurrent patterns and familiar interactions with the real world, in order to perform these imitations. Grasping, stretching and releasing an elastic object are examples of such a fundamental embodied experience. However, the imagined object that seems to be employed would not typically produce any sounds in real conditions; an elastic band is not usually expected to sound when stretched. Therefore, it is also not possible to talk about imitating sound-producing movements.

In a similar fashion (referring to imagery), Cox (2011) distinguishes between observed sound-producing gestures, analogous sound-producing gestures and other analogous exertions. The first case refers to imagining playing the violin while listening to a violin piece ('intra-modal MMI'), the second case refers to imagining playing the same thing on another instrument or singing it ('cross-modal MMI'), while the third case refers to imagining the use of any other modality that is not primarily acoustic to match the pattern of the audio stimulus (also cross-modal MMI). Probably the third class by Cox (imagining the use of another modality to match the pattern of the audio stimulus) would be the most adequate way to classify MIIOs. Nonetheless, this case lacks the gist of MIIOs, which involves the mediation of the movement by the imagined object and the imitation of effort as a way to simulate the imagined energy coupling between the hands and the object. Whether this energetic coupling can be extended to include sound and to what extent this sound is actually the one that the imagined object could potentially produce or the real sound of the voice is an open question. This also means that it is not clear whether in MIIOs there is a relationship between manual gestures and the voice and how strong this relationship is. Even more, it is not clear if this relationship (in case it exists) can be potentially detected by an observer. In other worlds, although MIIOs are founded on recurrent patterns of our interaction with the environment, they are not necessarily grounded in learned sensorimotor contingencies. This means that, although the object could be imagined as producing the sound and the gesture could be imagined as potentially sound-producing, how this is finally related to the sound is not clear. Examining this relationship forms the main goal of the current research.

2.3. Effort

The intention of the current section is to clarify the way effort is understood in the context of the current research and in relation to MIIOs. It also intends to provide an overview of previous computational approaches existing in literature in estimating a measure of effort and highlight why these are inadequate for this work.

2.3.1. Definition

Effort is a frequently used word in our daily vocabulary and in its common use it reflects our understanding of how much a person must try, either mentally or physically, in order to achieve an intended goal. In other words, in its reserved, everyday use it refers to the forcefulness or power of movement towards a goal. Still, it seems to be a loosely defined term in different science branches, such as physiology, kinesiology, biology, neuroscience and psychology.

In music the importance of effort has been acknowledged in both music-making (including digital) and listening. According to Krefeld & Waisvisz (1990), designers of electronic musical instruments, effort reflects the musical tension of a piece and is considered as essential to both performers and audiences alike. 'The physical effort you make is what is perceived by listeners as the cause and manifestation of the musical tension of the work' (Krefeld & Waisvisz, 1990: 29). Similarly, Joel Ryan (1991: 6), pioneer of electronic musical instruments, regards effort as 'the element of energy and desire, of attraction and repulsion in the movement of music'. However, we are still missing a definition of the term in music performance.

Dance theorist and choreographer Irmgard Bartenieff (Bartenieff & Lewis, 1980) considers effort as a measure that reflects the active or passive attitude of a person in resisting or yielding to the physical conditions that influence a movement meant to accomplish a task. According to dance scholar Vera Maletic (1987), apart from the physical, effort also reflects the inner attitude of power in terms of motivation and intentionality of the person in respect to an intended goal. Finally, movement theorist Rudolf Laban (Laban & Lawrence, 1974) considers effort as the foundation of any human body movement, and movement scholar Peggy Hackney (2002) argues that it 'exposes the mover's manner, tone and level of energy and is generally associated with emotion'. According to Laban & Ullmann (1971), effort can be described as that subjective quality that is associated with the forces that cause and constrain movement (its dynamics), and at the same time it reflects the inner impulse, intention or source of the movement.

The crucial aspect of effort in all the above mentioned approaches is the intentionality towards achieving a goal, for which the person must fight against and overcome the conditions that may influence and prohibit it. The goal is considered to be both mental and physical. Thus, effort constitutes the interface between the covert (mental) and overt (physical) aspects of the compulsion for movement. It unifies (as Maletic (1987) puts it) the quantitative and measurable side of movement (the physical) with its virtual, qualitative (the mental) side that resides in intention. It is this compound character of effort that makes it difficult to grasp and define.

Although the stance that I am taking in this thesis considers effort as a combination of intentionality and physicality as well, it is the composite nature of the overt aspect of the goal that distinguishes my definition from previous definitions of the term. The concept of effort has always been approached from the realm of movement and kinesiology. However, in this thesis I argue that the overt aspect of the goal should be extended to not only include movement but also the auditory modality of intentionality found in music performance.

In everyday human body movement, action is directed towards designated pragmatic goals (Rosenbaum et al., 2009), such as the highest point in lifting a real rigid object or the largest extension in stretching an elastic object. Our experiences of the world are considered as pragmatic in nature and mediated by goal-directed sensorimotor patterns (Schiavio, 2014). In music performance the goal is supposed to be primarily musical. In other words, action is motivated by an acoustic goal that refers to salient moments (goal-points) of melodic expression, such as accents and peaks (Godøy, 2012) of patterns that regulate or control the sound-producing gestures (Dahl, 2006). In the case of MIIOs these goal-points form a combination between physical, acoustic and imagined; the physical part of the goal (I will call this goal 'pragmatic-like' due to the lack of a real object) refers to a particular target point in space or a particular physical state of the imagined object (either in position, physical condition, size, or orientation), the acoustic to a salient moment in the sound (such as melody, intensity, timbre etc.) and the imagined (visual/pictorial, motor-mimetic etc.) to a mental quasi-perceptual experience or representation bearing intentionality of any sensory modality (including the kinaesthetic, tactile, proprioceptive etc.).

Effort can thus be potentially regarded as a suitable cross-modal (or amodal) descriptor that combines all three modalities (physical movement, mental impulse and musical intentionality) and is close to human perception and musical experience. It can be conveyed and induced by a performer, as well as assessed and experienced by an observer. Therefore, it can be visible in human movement, audible in the singing voice and able to be sensed through proprioception. However, its assessment relies on our subjective experience of how difficult the task seems to be for the person and therefore effort is a perceptual measure that cannot be quantified in a direct way. Thus, in this thesis I introduce an augmented definition of effort for music performance, according to which:

> Effort is a perceptual and subjective measure that reflects the active or passive attitude of a person in fighting against or yielding to the conditions that resist the accomplishment of a task towards a compound goal, which includes mental (imagery), physical (movement) and sonic (musical) aspects.

By extending the ideas by Laban & Ullmann (1971), effort in MIIOs can be described as the subjective quality—as this is experienced by the performer and perceived by an observer—that expresses the balance between forces: on the one hand the resistive force imposed by the object and on the other hand the force exerted by the performer to defy it. I refer to force here (which develops between two objects) as it is defined in physics, i.e. as a quantitative description of an interaction that causes a change in an object's state or motion. Thus, assuming that the performer imitates such an interaction (even if the object is only imaginary), I consider force to be here the quantifiable measure that reflects the proprioceptive sense of strength in case the interaction was with a real object; on one hand the force that is exerted by the performer (the agent), which reflects the inner impulse, intention or source of movement and sound, and on the other hand the force exerted by the object, that is imagined to constrain the performer's movement. Effort cannot be simply equated to force, as I consider effort to be a perceptual and subjective measure of power; I will explain why in the next section. In lines with Laban, I consider effort to describe not only the power (the level of energy expended) of the action but also the way the goal is reached (its dynamic profile) with respect to intention towards the compound goal. Hence, MIIOs can be seen as an active process of tracing energetically shaped event units through forces and 'repeated patterns of embodied experience' (Varela et al., 1991) we have established over years of interacting with the real world, such as grasping, stretching and releasing a real object; albeit the intention is here primarily musical.

By acknowledging the role of MIIOs and the implicated sensation of resistance (force), effort is a key concept that can better reflect expressivity and cross-modal relationships as a dynamic property of music performance and cognition. Which cross-modal features exhibit significant associations between body movement and sound (categorical relationship), and how this temporal congruence is displayed in the magnitude of effort during the imitation of interactions with real objects are questions still open for exploration and are the ones this thesis is concerned with. As will be explained later, this thesis will only be concerned with the first of these aspects, while the second aspect will remain as a future task. However, in the next section on effort in music I will elaborate on both aspects.

2.3.2. Computational approaches

By now it should be clear that effort in this work is considered to be a perceptual and subjective measure that combines the overt (movement and sound) and covert (mental) aspects of power and intentionality in moving towards a goal and against the conditions that might be resisting the movement. Despite the extensive studies on human movement—such as the work by Laban and Bartenieff—there is, however, a shortage of analytic works with a clear focus on the concept of effort with this specific understanding and only few systematic approaches in music exist. Due to its subjective character I argue that it is difficult to quantify effort by using direct measurements, such as by measuring forces. In this section I review some of the attempts to (indirectly) measure effort that I have come across in literature. By highlighting the shortage of successful quantitative methods, I support the qualitative approach taken in this thesis.

According to the seminal treatise 'Theory of Effort' by Laban & Lawrence (1974; see also Bartenieff & Lewis, 1980), effort constitutes one of the four components ('BESS' for Body, Effort, Space and Shape) that form the so-called Laban Movement Analysis (LMA) system that Laban developed (Laban & Lawrence, 1974; Laban & Ullmann, 1971) and Bartenieff (Bartenieff & Lewis, 1980) extended for observing, describing, notating and interpreting human movement. The 'Body' component deals with its moving parts, individually and as a whole; 'Effort' addresses the dynamic qualities of the movement; 'Space' describes how large is the moving person's kinesphere and the spatial trajectories in space and 'Shape' describes the changing forms of the body in space. BESS was later further extended by Hackney (2002) to incorporate the concepts of relationship (to the world, either other people or (real) objects) and phrasing (individual motions combined into a larger, coherent entity) and is now known as the 'IMS' ('Integrated Movement Studies') system.

Effort is regarded by Laban as the qualitative expression used to describe qualities of the movement performed with respect to inner intention. These qualities are expressed through four individual 'motion factors', namely: weight (seen as intention or attitude towards power of the action), flow (seen as progression or readiness to stop at any moment), space (seen as attention or focus in space) and time (seen as commitment in hastening or lingering in time). Each of these 'motion factors' can take one of two extreme values (the 'effort elements') that express the concept of indulging/yielding vs. resisting/fighting against the physical conditions influencing movement (weight: strong/light; flow: bound/free, space: direct/indirect, time: sudden/sustained). By making all possible combinations of the effort elements, Laban developed a 'coding system' of eight 'basic efforts' (categorical movement qualities), namely dabbing, gliding, floating, flicking, wringing, slashing, thrusting, pressing.

The BESS system and the coding of the effort component in terms of the four motion factors is still used today among choreographers, dancers and scientists involved with movement (Broughton & Davidson, 2014; Hackney, 2002; Maletic, 1987). It has also proven valuable in developing computational methods for the analysis[39] and/or synthesis[40] of human movement through the extraction of measurable movement features. However, the actual notation system itself does not provide any explicit quantitative measures, and in the abovementioned studies effort has been only indirectly evaluated by observers through the kinematic properties implied by Laban's motion factors.

Other concepts close to the concept of effort have been previously studied, such as tension, intensity, activation contour or vitality affect. However, none of them expresses the term in the exact way I intend to use it in the current research context. For example, Cox (2016: 97) has argued that musical tension is related to effort. Tension has been studied extensively before (melodic[41], visual[42] or physical[43]), but its validity as an indirect measure of effort is not justified in literature. For instance, Jacobson (1951) has previously argued that muscular tension is not a good indicator of the exerted effort. Intensity[44] is another concept that is supposedly related to the perceptual concept of effort. While physically (at a neurophysiological level) intensity changes are linked to the energetic content of actions, perceptually they can be equated with changes in the magnitude of different sensory modalities, and these may include changes in physical effort, muscular tension and fatigue, sweating, breathing rate, light brightness or auditory loudness[45]. Nevertheless, measuring the intensity change

---

39 Kikhia et al., 2014; Niewiadomski et al., 2013; Morita et al., 2013; Kapadia et al., 2013; Samadani et al., 2013; Mentis, 2013; Glowinski et al., 2011; Hachimura et al., 2005; Zhao & Badler, 2005; Camurri et al., 2003 and Nakata et al., 2002.

40 Hartmann et al., 2005; Allbeck & Badler, 2004; Volpe, 2003; Gibet et al., 2004; Neff & Fiume, 2002; Hsieh & Luciani, 2005; Zhao, 2001; Chi et al., 2000 and Chi, 1999.

41 Melodic tension has been examined by Margulis (2005), Granot & Eitan (2011), Bigand & Parncutt (1999) and Lerdahl & Krumhansl (2007).

42 Visual tension has been studied by Frego (1999), Krumhansl & Schenck (1997) and Sheets-Johnstone (1966).

43 Tension in physical movement has been examined by Neff & Fiume (2002) and (Erickson et al., 1983) in terms of sensed pressure of the vocal folds, that largely determine both loudness and pitch.

44 Intensity in musicology refers to the psychoacoustical understanding that is related to loudness or sound pressure (Rasch & Plomp, 1999).

45 Recent studies in music (Eitan, 2007), supported by developmental studies with infants (Stern et al., 1985; Maurer, 1993; Lewkowicz & Turkewitz, 1980) and studies with adults (Stevens, 1975) have revealed that intensity (or 'activation contour', according to Stern) is an amodal quality and that through its temporal congruence over specific dimensions our senses can be linked inter-modally or intra-modally.

in some other modality would again only be an indirect way to assess effort and not well-justified or useful in the current work.

Some approaches exist that try to relate the sensation of effort to physiologically measurable quantities that are supposedly the prevailing indicators of effort, such as by measuring force, accomplished work, breathing rate, sweating, muscle tension/dynamics and fatigue (Gibet, 2010; Marcora, 2009; Enoka & Stuart, 1992). However, these studies presuppose that the power of an action can be somehow objectively assessed and quantified, which is not the case. Considering the fact that different people have a different capacity (intrinsic factor related e.g. to fitness level) in achieving the same physically demanding goal (extrinsic factor related to task's difficulty level), physical (and similarly mental) effort constitutes a perceptual concept, related to the subjective sense of how intense the specified task is for the particular subject. For instance, the absolute measure of force may be low, but due to low muscle power or the lengthy duration of an action, the sensed level of effort may be high. The calories that are burned may exceed the resting burn rate (measured through metabolic equivalents), but this is an absolute measure that does not account for the level of fitness of the particular person and cannot really reflect the sensation of effort level. Thus, effort can be only subjectively evaluated and not easily quantified in a direct way, such as by measuring real forces.

One of the most tempting computational approaches is offered by quantifying 'weight'. The term, as it is understood here, should not be confused with the force exerted on a body by gravity. It should also not be confused with Laban's 'motion factor' of weight, as the visual assessment of the power of an action by a third person may also include aspects of time, flow and space. I consider weight as a measure of the action's power or the ease and struggle in performing an activity. Moore & Yamamoto (1988: 220) have also used the term 'pressure' instead. Weight (or power) cannot be interpreted in a single, universal way and therefore no generalisable computational method has been established yet for its assessment (Graham & Bridges, 2014). For instance, an evaluation of Hartmann's model of expressive behaviour (Hartmann et al., 2005) showed that power (another term for weight) along with fluidity received the lowest recognition rates when synthesised with the Greta agent (Niewiadomski et al., 2013). This lack of a general agreement in computing the measure of weight is also reflected on the comparison of just a few different approaches for weight analysis and synthesis in Niewiadomski et al. (2013: 17-18). In fact, in some few cases the computation of weight has been even avoided due to its complexity (Santos, 2013; Camurri and Trocca, 2000). Here I will provide a summary of some of the kinematic features that have been reported in the literature as an indirect measure of weight and I will explain why they are not relevant to the current work.

One of the most frequently used features for weight is the physical quantity of acceleration, as it is considered proportional to force in mechanics (Petersen, 2008). A number of alternatives in computing acceleration as an indirect measure of weight have been reported. These include the acceleration of the full body (Castellano & Mancini, 2009) or only the hand (Mancini & Castellano,

For instance, it has been suggested that louder in sound equates to brighter in light (Eitan, 2007; Eitan & Granot, 2006; Neuhoff & McBeath, 1996) and larger in size (Walker & Walker, 2016).

2007), or both the hand and the elbow (Bernhardt & Robinson, 2007), or the vertical component of acceleration (Volpe, 2003; Camurri et al., 2003), acceleration and deceleration[46] (Niewiadomski et al., 2013, 2011; Hartmann et al., 2005), only deceleration (Kapadia et al., 2013)[47], maximum value of the deceleration for the full body (Samadani et al., 2013)[48] or for various upper-body parts (Kapadia et al., 2013), or the first derivative of the motion vectors (Caridakis et al., 2007).

Another measure that has been used for estimating weight based on visual and motion data is kinetic energy: kinetic energy of user's joints (Piana et al., 2013; Hachimura et al., 2005), kinetic energy per unit time (Nakata, 2002) and the peak of kinetic energy calculated from the sum of all moving body parts or only the upper-body parts (Samadani et al., 2013). A few other ways of assessing weight exist also, such as the 'overall body activation' through the 'Quantity of Motion' (Camurri et al., 2004b) or the peak in the time series of 'Quantity of Motion' (Mazzarino et al., 2007), the relative vertical position (Mentis & Johansson, 2013) or the vertical directivity (Zhao & Badler, 2005; Schiel & Heloir, 2008), the optical flow averaged over time and space (Morita et al., 2013), the tension and bias of TCB spline lines for the modelling of squash vs. stretch of the body (Alaoui et al., 2012), even the hand shape, the wrist rotation and the extension of the arms (Hartmann et al., 2005; Kochanek & Bartels, 1984[49]).

None of the abovementioned ways of evaluating weight (as a measure of effort) are suitable for the current work. Specifically, I would like to argue here that weight (and any type of force) cannot be assessed through measures that include kinetic properties (kinetic energy and velocity), in other words measures that reflect the impact on one's movement (Chi et al., 2000). For instance, let's imagine a person trying to move a wall. In the context of the current research, effort (or weight) would be high if assessed as a measure of the forcefulness of the movement, but it would be zero if computed as the kinetic energy, the acceleration or the velocity of the movement. In terms of expended energy, effort could be better reflected by the total biochemical energy expended over time in the form of heat (calories), but this is not always easy to quantify. None of the other measures that were just cited would fit the purposes of measuring effort in the context of the current research. For instance, the vertical component in the abovementioned computations reflects the fact that the only force assumed to be exerted on the body is gravity (which is not the case in MIIOs)[50], while the kinetic energy approaches weight as the observed power or size of the movement. I argue that physical effort cannot be quantified by measuring the work accomplished by the action (such as the displacement of an object), as one could be sensing high levels of resistance (and thus effort) in interacting with an object without ever changing its physical conditions. For instance, one could be essentially pushing a wall really hard in order to throw it down (high sensation of effort) without ever managing to do so (no accomplished work); or one could be holding a weight for a long period of time

---

[46] Higher magnitudes of acceleration and deceleration and less overshoot are expected by more powerful movements (Niewiadomski et al., 2011).
[47] Considered as proportional to the rate of change in kinetic energy.
[48] Indicating the maximum of absorption of kinetic energy, such as at the end of a punching movement. It is considered by Samadani et al. (2013) as a more accurate measure of the weight effort factor.
[49] In synthesising movement.
[50] Studies in measuring 'weight' usually involve free (not interacting with objects external to the body) whole-body movements, in which the only (real) opposing force acting on the limbs is gravity. However, in the context of the current research 'weight' can refer to any resistance imagined or sensed in the hands when the performer seems to manipulate an object by changing its physical condition.

at the same place (high sensation of effort) without moving the hand (no accomplished work). Additionally, only a few of the abovementioned studies have addressed the issue of effort in music performance and even fewer in non-sound-producing movements (Haga, 2008) such as in singing. In fact, computational approaches of effort in music performance in general have been rare (Bennett et al., 2007; Camurri et al., 2007; Volpe, 2003). Finally, I am not aware of any studies that have included both kinematic and sonic features in estimating or predicting effort.

2.4. Analysing movement–sound relationships

In this section I review methods that have been used previously for studying movement–sound relationships in music performance, and explain where this thesis offers a contribution and how its outcomes could become useful in designing new electronic musical instruments.

2.4.1. Definition of the terms *relationship*, *association* and *link*

To avoid confusion, I define the following terms referring to connections between various cross-modal aspects (of movement, sound and imagery) and use them as such throughout this thesis:

- Relationship: It is a general term that I use to describe how movement and sound (and possibly also imagery) may relate to each other in various ways and it may refer to both gesture–sound couplings (those defined by physical laws) and artificially designed mappings (that are developed for instance in new electronic musical instruments); see Jensenius (2007) for a more detailed discussion on this. As I mostly refer to well-defined movement events of MIIOs (thus gestures, as defined in 2.3.1.) throughout this thesis, I will use the terms 'gesture–sound' or 'gesture–voice' (to underline the fact that it is about vocal sounds) when discussing about relationships. I may also use the term 'movement–sound' when referring to body movement in general.

- Association: I use this term to describe relationships in terms of categorical cross-modal correspondences or co-appearances. For instance, I refer to associations in the qualitative part of the analysis when examining whether certain gesture/MIIO types tend to happen along with particular types of melodic movement, or whether particular gesture/MIIO types and/or melodic phrases tend to be performed with a certain level of (perceived) effort. When discussing in a general way about co-appearances between a number of aspects of movement and sound I use the general term 'gesture–sound' associations, and when I need to examine and specify certain aspects of modalities that co-appear I use more specific terms, such as 'gesture class–melodic phrase', 'gesture class–melodic intention', 'effort–gesture class', 'effort–sound', 'effort–melodic phrase', 'effort-mīṇḍ slope type' and 'palm shape–vowel'.

- Link: I use this term to describe the exact way by which cross-modal features are related to each other. A link can be presented as a mathematical expression. When using time-varying features (see Peeters et al. (2011) and Nymoen et al. (2013) for a discussion on this), it describes how they mutually co-evolve in time cross-modally in terms of patterns, shapes and dynamic profiles. When using global features (such as mean, standard deviation, minimum, maximum etc.), it describes the way a feature in one modality will change with the corresponding change of a feature in the other modality. For instance, I use this term when referring to the models I developed to estimate effort levels or classify MIIOs as a linear combination of (global) movement and sound features. When discussing in a general way about mathematical expressions linking a number of aspects of movement and sound I use the general term 'gesture–sound' links, and when I need to examine and specify certain cross-modal aspects that are related to each other I use more specific terms.

As this thesis deals with singing and hand gestures do not come necessarily before sound (as would be the case in instrumental gestures), the order of words in the abovementioned expressions does not imply any particular order in the appearance of modalities.

2.4.2. Existing methodological approaches

Movement–sound relationships in music have been studied for both the strict movement–sound couplings and the looser movement–sound relationships. By couplings I refer here to the causal relationship that is dictated by physical laws between the movement and auditory modalities in instrumental playing or sound-producing gestures in the environment, while relationships can range from couplings to artificial, designed mappings or relationships between sound and sound-accompanying or ancillary movements (Jensenius, 2007). Some aspects of these relationships are literal, while others more metaphorical (Repp, 1993). The first type (couplings) has been studied through sound-producing movements extensively by Rasamimanana (2005), Bianco et al. (2010) and Wanderley (2002). The second type (general relationships) has been studied systematically through gestural responses of listeners to various auditory stimuli by Nymoen et al. (2013), Caramiaux et al. (2011, 2010a), Haga (2008), Godøy et al. (2009), Eitan (2007), Eitan & Granot (2007), Vines et al. (2006), Camurri et al. (2004a) and others. However, only a few studies address movement–sound relationships for physical movements which co-occur during music performance but do not directly produce the actual sound. Such cases include conducting movements (Luck & Toiviainen, 2008), ancillary movements of instrumentalists (Wanderley, 2002; Nusseck & Wanderley, 2009; Vines et al., 2004) or singing movements (Clayton & Leante, 2013; Liao & Davidson, 2007; Davidson, 2001).

It seems quite surprising that singing movements have drawn less attention than instrumental, as there is a serious body of closely related research on speech-accompanying movements in verbal communication[51] that provides a solid ground to start from. The 'oral' tradition of classical Indian vocal music is an exception to this trend, as interest in analytical approaches of body movements in Hindustani and Karnatak vocal performances has been steadily growing in recent years (Clayton & Leante, 2013; Fatone et al., 2011; Rahaim, 2012, 2009). However, most existing studies in movement–sound relationships—including those in classical Indian vocal music—are mainly qualitative and based on McNeill's work. Computational approaches of cross-modal relationships for singing movements have been less in both Western (Françoise, 2015; Brunkan, 2015a,b; Pfordresher et al., 2015a,b; Luck & Toiviainen, 2008) and Indian (Pearson, 2016; Moran, 2007) types of music.

In the current work I intend to devise mathematical descriptions of the way by which cross-modal features are mapped to each other on the premise of 'similarity'. I use the word similarity here to refer to cross-modal links in terms of analogy (Haga, 2008; Hatten, 2004), intensity isomorphism (Eitan & Granot, 2006), or 'integrated intensity flux' (Todd, 1995). According to Karwoski et al. (1942), movement–sound similarities arise as a result of alignment or overlap between different dimensions of the same connotative meaning and are grounded on internal patterns of energetic tension (Kurth, 1922). They are accredited to the way motor imagery (Godøy, 2004; Godøy et al., 2016, 2001; Tubul,

[51] Butterworth & Hadar, 1989; Rizzolatti and Arbib 1998; Kendon, 2004; McNeill, 2005; Goldin-Meadow & Wagner, 2005; Gibbon, 2009; and Novack & Goldin-Meadow, 2015.

2010; Eitan & Granot, 2004, 2006) is implicated in our understanding of music through our implicit, ecological knowledge of sound sources. In this sense, I understand similarity as the temporal congruence of time-dependent cross-modal (and inter-modal) features in terms of magnitude and polarity of change (increase vs. decrease). A summary of such studies can be found in Eitan and Granot (2007). Semiotician Philip Tagg (1999: 24) uses the alternative term 'anaphone' (iconic, kinetic, acoustic) to denote the use of 'an existing model outside music to produce musical sounds resembling that model'. The term 'kinetic anaphone' has been also used by Leante (2009) to denote a sound analogue of movement in Hindustani singing. For instance, in the case of a vocalist producing a sustained melodic ascent that he/she verbally describes as 'stretching the note', the sound would be the kinetic anaphone of a stretching type of MIIO.

A big part of movement–sound studies relies on exploring sound-induced gestural responses by listeners. Godøy et al. (2006b) and Jensenius et al. (2010) have proposed that there are two ways by which a listener may gesturally respond to a sound stimulus in listening tasks and both are of great interest to the current thesis: (a) by 'mimicking' sound producing-gestures and (b) by 'tracing' the spectromorphological features of the sonic stimulus. In this respect, these and other scholars have conducted extensive work on motormimetic sketching through 'air-instrument' studies (Godøy et al., 2006b) and on sound tracing through 'sound-tracing' and 'free dance' experiments (Godøy et al., 2016, 2006a; Caramiaux, 2010b; Glette, 2010; Haga, 2008; Casciato et al., 2005). The 'air-instrument' experiment examined the mimicking of sound-producing movements, the 'sound-tracing' worked on the drawing of spontaneous but restricted gestural responses to audio stimuli on a 2D tablet and the 'free dance' worked on full-body unrestricted, spontaneous gestural renderings of the audio stimuli.

Although these studies have used material from listening tasks (rather than ecologically valid conditions of real performances), their findings are of great importance to the current work, as they address issues of effort in full-body or manual movements that are free rather than limited by the use of some object (as in instrumental playing). Results indicate that both novices and experts alike perceive significant global sonic features and are able to follow coarse gesture renderings of musical sound (Haga, 2008). There is also a considerable amount of flexibility but at the same time a degree of consistency (non-arbitrariness) in the way a listener may perceive music-movement correspondences[52] (ibid.). The observed flexibility relates to both the effectors used (e.g. the hands for imitating vocal sounds or the vocal apparatus for imitating instrumental sounds) and the actual imitation itself (the way movement is performed). More specifically, results on movement–sound relationships indicated that despite the divergence observed between participants in their movement trajectories in space, there is a high degree of convergence between the dynamical profiles of physical movement and the equivalent profiles in their sonic counterpart (ibid.). For instance, speed, force and quantity of motion of hand movements converged with speed in rendering notes, loudness envelope etc. in the sound). In section 2.3. I supported the idea that dynamic profiles of gestures

[52] The term 'correspondence' has been used by Haga (2008) as a general term that refers to a qualitative judgement of the relation between music and movement in an audio-visual example, where the relation is characterised by the way music and gesture are perceived as being similar.

(displayed through various movement features) are associated with profiles of effort (effort envelopes) that we feel through our embodied capacity (Godøy et al., 2016, 2006b; Godøy, 2006; Haga, 2008). Hence, in this thesis I seek to examine whether and to what extent a convergence between acoustic features and movement features that reflect effort envelopes might exist during MIIOs and whether and to what extent this convergence is shared between performers.

2.4.3. Research gap

The mapping between hand gestures and sound that I seek to explore in Hindustani vocal music—and MIIOs in specific—will be the first of its kind and it may allow for a unique methodological approach to be developed in studying issues related to music cognition. Singing forms a special case of music making, as it allows for the analysis to be conducted based on real music performances rather than listening tasks. In fact, it offers a unique opportunity to study voluntary manual gestures of musicians during real performances that are not directly responsible for the production of sound[53]. Studying singing gestures is therefore an effective way to address the trade-off that is usually inherent in experimental approaches between—on the one hand—issues of ecological validity in studying gestures during instrumental performances and—on the other hand—issues related to using listeners' responses rather than sound-producing gestures in order to examine spontaneous free gestures that are not defined by the mechanics of a musical instrument.

In this context, musicians do not respond to an unfamiliar external stimulus, but are allowed to demonstrate the gesture–sound relationships that they have developed as active practitioners over time. MIIOs offer a special case where the hands—although free to move[54]—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.

2.4.4. Motivation: introducing effort in the design of VMIs

The motivation behind the study of movement and sound in MIIOs is to find the function that relates sound and manual gestures. This is ultimately intended to aid the design of artificial gesture–sound interactions by taking advantage of the physical effort induced by familiar gestures associated with handling real objects in our natural environment. The VMI (Virtual Musical Instrument) by Mulder (1998) is an existing example. This digital instrument relies on the concept of 'sound sculpting', in which imaginary objects are sculpted to produce sounds, as shown in the video projection of Figure 2.4. Examples include stretching a rubber sheet, claying, carving and chiselling.

[53] The other cases being linguistics and perhaps—more loosely—conducting gestures.
[54] Apart from biomechanical constraints of the performer's body itself.

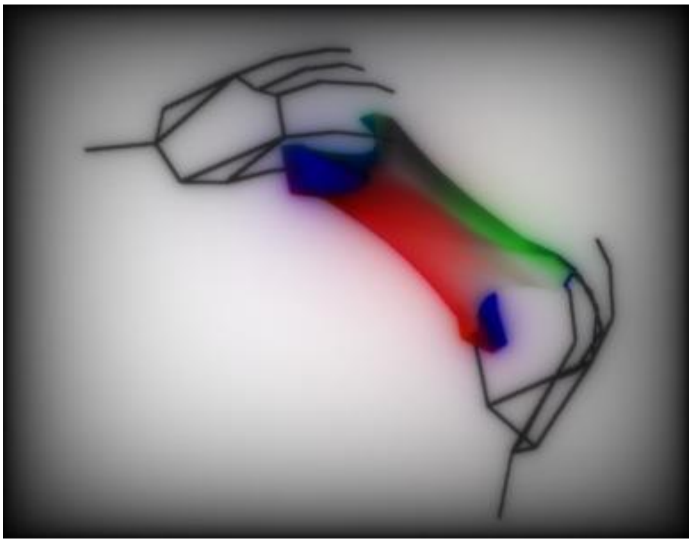


Figure 2.4. Virtual Musical Instrument, adapted from Mulder (1998).

However, this instrument has several drawbacks. The mapping was developed based on intuition and it relied only on the geometrical properties of the imagined object without accounting for the effort that would be required in natural conditions. In designed gesture–sound interactions, d'Escriván (2006: 199) has attributed the lack of expressivity to the rupture of the so-called 'efforted-input paradigm'. Similarly, this problem has been discussed by Godøy (2009), Waisvisz (in Wanderley & Battier, 2000), Vertegaal (1996), Winkler (1995), Ryan (1992) and Krefeld & Waisvisz (1990). It may occur when, for instance, loud sounds can be controlled with very little physical effort, and it is an inherent problem of empty-handed interactions which lack haptic feedback and thus a direct energy transfer link. For instance, Fels et al. (2002) have reported that while claying was not considered a compelling metaphor for shape manipulation due to the absence of tactile feedback, the stretching metaphor proved to compensate for this absence by forcing the player's frame of reference to remain attached to the object.

The current work may enable advances in the effectiveness of sound sculpting; it may be possible to compensate for the absence of tactile feedback by imitating the exertion of effort against forces that are elicited by physically-inspired images of MIIOs. It proposes to achieve this by informing mapping strategies of VMIs through the mathematical formulations that are developed in this thesis to describe gesture–sound links of MIIOs in Dhrupad singing. By using gesture–sound mappings that conform to the amount of effort that such interactions would be perceived to require in MIIOs, the intention is to render the control of artificial sounds more 'natural' (as most of the Hindustani singers put it) or 'physically plausible' (Castagne & Cadoz, 2005: 2). The latter term is used to describe the sense of 'naturalness' in the control of synthesised sounds, where the goal is not to simulate the way real objects would realistically produce sounds (the mechanics of their vibration according to excitation), but to make the performer feel that the sounds could have been potentially produced from the interaction with some real object.

2.5. Summary

The current chapter provided an overview of literature from past and current research in the field. It opened with a brief introduction in the Dhrupad genre of Hindustani music (performance, tacit transmission and conceptualisation of music) and an overview of movements that often accompany vocal performances. There was also a brief description of the types of gesture this work is concerned with. Given the fact that gesture appears in a number of different contexts and is a rather broad and vaguely defined concept (Godøy & Leman, 2010), this chapter also offered a definition of gesture as it is understood in the current thesis. Similarly, as the notion of effort is at the core of this work but is rather ill-defined in scientific terms, a part of the chapter was also dedicated to the way the concept is understood here and how it is implicated in MIIOs. The ambiguity of MIIOs in relation to available taxonomies was also explained, along with the reasons that this tradition and the specific type of movements provide an interesting case for the study of gesture–sound relationships. The chapter closed with a discussion of previous studies, methodological approaches and techniques. This was meant to situate the project's methodology and its intended contribution, and leads smoothly to the main methodology discussion of the next chapter.

Chapter 3 – METHODOLOGY & RESEARCH DESIGN

This chapter discusses the data collection process and justifies the empirical approach chosen for this work, which is a mix of qualitative and quantitative methods (Yin, 2003) and relies on first- and third-person perspectives (Leman, 2010). The chosen methodologies enable the subject to be approached in the most ecologically valid way possible, by avoiding designed experiments and instead recording real performances in the field, despite inherent complexities in the collection process of the data.

As the data in qualitative research is essentially less structured than that generated through quantitative research, the qualitative part of the analysis (interview analysis and video observations) was used first in order to identify and annotate recurrent patterns in the recorded material. These patterns were then organised and classified in order to examine possible cross-modal associations (movement and sound). The quantitative part of the analysis was then used as a way to validate as well as extend the qualitative findings on gesture–sound relationships by making use of exact measurements in terms of quantity, amount or frequency (Labuschagne, 2003), specifically aiming at the ways that dynamic aspects of movement and voice co-evolve in time.

The qualitative analysis started with a thematic analysis of interview material in order to form concrete research questions, help make a selection of performance gestures to look for in the video recordings and inform the coding scheme of the video annotation process. Third-person observations on video material were then used in order to identify, label and later classify the audio-visual material during MIIOs in terms of recurrent types (categorical descriptors) of hand gestures (the 'action-oriented ontology' (Leman 2010: 134) of MIIOs), as well as perceived effort levels (numerical) that appear to be exerted by the performer in a range between 0 and 10. An association analysis was then conducted separately for each of the performers between movement and melodic aspects of MIIOs. The annotations were used in the quantitative part of the analysis as the response values of linear models that were fit to the measured movement and sonic features in order (a) to estimate the annotated effort levels (numerical) and (b) to classify gestures (categorical), specifically discriminating between interactions with elastic vs. rigid objects.

The qualitative approach consisted of:
- Interview analysis of eight performers of the Dagar gharānā (Chapter 4);
- Video analysis for Afzal Hussain's performance of rāga Jaunpurī (Chapter 5);
- Video analysis for Lakhan Lal Sahu's performance of rāga Mālkauńś (Chapter 6);
- Video analysis for Umakant & Ramakant Gundecha's performance of rāga Bhūpālī (Chapter 7).

The quantitative approach included:
- Linear regression for effort estimation and logistic regression for gesture classification (Chapter 8).

The linear models were fit to the data of only two performers, namely Afzal Hussain and Lakhan Lal Sahu, as a proof of concept. However, the methodological approach can be extended in the future to other performers and performances.

3.1. Sequential Use of Mixed Methods

I use the term “mixed methods” to refer to a combinatorial methodological approach that advances the systematic integration of findings emerging from qualitative and quantitative methods on qualitative and quantitative data respectively. In specific, this thesis employs a sequential use of mixed methods, where the two phases occur separately and in a consecutive order; the quantitative approach follows the qualitative in this work. Qualitative methods were favoured at the beginning of the work due to the limited previous knowledge that was available on the subject and a level of abstractness in describing MIIOs at first, which were needed for the application of the quantitative methods.

I argue that by making a sequential use of mixed methods and by combining the rich and complex outcomes of qualitative approaches with the compact results of quantitative methods that are based on exact measurements, this thesis can offer a more rigorous and comprehensive picture of the phenomena under study. The basic premise of a mixed methodology is that such integration has the potential to harness the strengths and counterbalance the weaknesses of each of the approaches, thus offering an especially powerful approach in understanding complex and multifaceted issues in breadth and depth, such as the one investigated in this work. The qualitative methods are more suited to an in-depth investigation of the complex phenomena of human movement and behaviour from a close observation of a rather small number of cases; this approach best fits the perceptual character of the concept of effort as this was defined in section 2.3.1. and allows one to produce results that are grounded in human experience. The quantitative methods are more appropriate for producing dense results that require less interpretative input from an observer, thus complementing the qualitative findings that may be viewed as less rigorous by the hard sciences due to their subjective nature. Mixed methods are also especially useful in highlighting contradictions between quantitative and qualitative findings and in increasing findings’ reliability and credibility through triangulation of different evidence sources.

However, it is not such a straightforward task to bridge the qualitative-quantitative divide and truly integrate and present findings from both types of methods and data in a coherent way. Integration does not equate a simple presentation of a list of outcomes, but it requires innovative thinking to move between different types of data and findings and make meaningful connections between them. It requires both comparing and contradicting findings from both methods, which is not always easy due to the different and potentially incompatible nature of data and method types; one of the main disadvantages stems from the difficulty in combining the multi-dimensional nature of qualitative codes describing human behaviour with the uni-dimensional nature of quantitative data which are produced by reducing the dimensionality and complexity of qualitative data. Indeed, in the last chapter of the thesis the qualitative findings of chapter 4-7 are combined as well as contradicted to the quantitative findings of chapter 8. Finally, following a mixed methodological approach is inherently challenging as it requires rigour and expertise in multiple scientific areas, which is translated in practice into an increase of duration and resources required for carrying out the study.

3.2. Qualitative Analysis (Thematic & Inductive Coding)

At the first stages of the project, I considered a qualitative approach as more appropriate due to the absence of an *a priori* theory from existing literature. Human movement is such a rich source of data, that different observers apply various interpretations and attach different meanings to the gestures being observed (Greenhalgh, 2014; Klein and Myers, 1999). The rationale behind the qualitative approach of the current thesis, therefore, was to develop a reasonable understanding of the phenomena analysed and an identification of the main categories that could be extracted from the data (Seaman, 2008).

This section provides a description of the qualitative methods, namely thematic analysis and inductive coding that were applied to the interview material and the video recordings of vocal performances. As there is no general consensus about how to conduct thematic analysis, a hybrid approach by Boyatzis (1998) and Charmaz (2006) was followed in order to move from a broad reading of the whole corpus of data to discovering particular themes that are related to the research question and finding consistent cross-modal associations. The current research work was therefore heavily data-driven. That is, instead of imposing a pre-defined "code book" to the annotation of MIIOs, coding was at first inductive, allowing research questions, new concepts and themes to emerge from within the actual collected data (Fereday & Muir-Cochrane, 2008). This included the identification of recurrent patterns and ideas, and the progressive focusing on the most significant codes used for synthesising and categorising large amounts of data. As this process relied heavily on the coder's experience and competence in segmenting, labelling and classifying the data into groups of themes, a cross-validation of the annotated material was also conducted.

The qualitative approach aimed:

1. To draw some preliminary results about whether the observed levels of bodily effort may be related to the mechanics of voicing during singing and/or whether they are related to more fundamental concepts of mental and melodic organisation;

2. To identify specific associations between effort levels and classes of gestures and melodic aspects when the singer seems to interact with imaginary objects. The features examined were not the same for all performers and performances, but emerged from conducting multiple views of the video material and thus from becoming progressively more familiar with the content of each specific performance analysed.

The intended outcomes of the qualitative part of the work aim to answer the following questions:
- Are gesture types associated with specific aspects of the melody? If so, how?
- Is bodily effort related to the voice? If so, how?
- Does such a relationship reflect the physical requirements of voice production or the melodic organisation of the rāga?
- Are the abovementioned relationships (in case they exist indeed) limited to a particular performer or rāga or are findings applicable beyond the chosen performance or performer?

As there is potential for the interpretation of qualitative data to be less rigidly defined than for quantitative data, when approaching the former it is essential: (1) to retain rigour throughout these processes and to keep a detailed audit trail, making them as transparent to the reader as possible, and (2) to use an established method of describing how well-defined, consistent and objective these annotations are across the entire corpus. In order to display how the first condition is met, a step-by-step description of the processes involved in the qualitative data analysis is presented. For the second condition, I performed a cross-validation process and inter-coder agreement test that is also described in detail. The three stages of the qualitative part of the analysis are presented here.

3.2.1. Stage I: deciding on sampling and design issues

According to Tucket (2005), it is desirable for the scholar using inductive research methods to avoid exposure to relevant literature at the early stages of the project, as this approach leads in due course to potentially significant findings in the data being more effectively identified. Indeed, I initially raised only an imprecise research question following participation in a musical seminar, during which I first noticed MIIOs in instructions by Zia Fariduddin Dagar (Edessa, Greece, 2007) and the question was only progressively narrowed.

I conducted several readings of the entire corpus (audio-visual material of performances and interviews), and progressively focused on smaller parts of the data set. Following the suggestion by Braun & Clarke (2006), during these readings I kept informal analytical notes of preliminary observations and generic descriptions of movements that I thought would be of interest in the context of the current thesis. This stage also included making full and precise transcriptions of interviews, and these proved beneficial in facilitating the data analysis (Lapadat & Lindsay, 1999) by informing the interpretation of the audio-visual material through motor-based metaphors and concepts of mechanics that are used to express abstract musical ideas.

Based on these preliminary codes and findings from interviews, as well as on the vivid way and frequency by which vocalists executed MIIOs, a subsample (Boyatzis, 1998) of four of the seventeen recorded performers was selected: Afzal Hussain, Umakant & Ramakant Gundecha and Lakhan Lal Sahu. This does not mean that MIIOs were not observed in any other performances/performers, but they were rather rare or they were not as numerous or as easy to identify. This fact needs to be borne in mind when generalising from the analyses to the tradition as a whole.

3.2.2. Stage II: developing themes and a code

The combination of preliminary notes from observations, interview material and a limited body of literature (Fatone et al., 2011; Rahaim, 2009) helped in progressively moving from abstract ideas and broad descriptions of MIIOs to identifying patterns and finally to clustering themes into conceptual categories as a coding scheme (following Charmaz, 2006; Braun & Clarke, 2006; Bryman, 2006). Here I discuss the approaches I used for the analysis of interviews and the recorded audio-visual material.

(a) Interviews

Cross-modal mappings can be expressed in speech and accompanying movements in different ways, although recognising them from the propositions is not a straightforward process. It is the interviewer's task to pinpoint and extract this knowledge from both the explicit and tacit information of verbal reports. The interview analysis aimed at finding ways to best unfold common underlying concepts which are exclusively stated or implicitly embedded in cognitive metaphoric thinking, linking musical features to their visual, motor or imagined counterparts. It also examined the way logical meanings are constructed by merging, contrasting or combining both the 'shariric' (mechanics of voicing and moving) and the 'roopic' (melodic and movement ideas) aspects of sound (Rahaim, 2009), movement and imagery within the framework of a unified concept or idea by the interviewee.

In order to achieve this, both the explicit and tacit linguistic thought of musicians were taken into account. This is a distinction made between knowledge that is acquired and encoded in discourse with or without conscious awareness respectively. Explicit knowledge can be expressed and communicated linguistically. On the other hand, in the case of implicit knowledge, the rules or principles of the ideas that the speaker is trying to convey are so internalised, that he/she is unable to provide a verbal statement without relying on context dependent or demonstrative elements (Chomsky, 1986), in other words imitation and metaphoricity (linguistically or/and gesturally).

Given that interviewees were informed beforehand that the project focuses on music and movement, the first case (explicit knowledge) refers to musicians self-reporting about their own understanding of how movement and sound are related in music performance. The second case, of implicit knowledge, refers to the intrinsically woven visual, imagined and motor-based metaphors of non-auditory domains in speaking and thinking about music and musicking. This extra-modal piece of information embedded in implicit knowledge is often inseparable from descriptions of the auditory domain, and it reveals knowledge about music performance that would otherwise not be available or easily accessible to an observer. As music-making is a poly-parametric process but musicians handle a more manageable subset of these features at any one time, analysing interviews with musicians is a helpful tool for identifying the aspects a singer is focusing on at each time.

A crucial role is attributed to metaphor in our cognitive thinking, seen as a central mental necessity, which helps us perceive and structure otherwise inaccessible abstractions by establishing cross-modal connections. The theory claims that different metaphors may be used to highlight different aspects of the same concept, which could possibly be described as a set of mappings from one to the other domain (Kövecses, 2002). The foundations of such processes have been laid by Lakoff & Johnson (1980) through the conceptual metaphor theory. The authors essentially claim that one understands and conceptualises one domain (the target domain, which is typically unfamiliar or abstract) in terms of another (the source domain, which is most often familiar and concrete). This theory, according to Zbikowski (2002), applies to music too. What, however, the conceptual metaphor theory fails to explain is why some metaphors are preferred over others, and matters become even more complicated due to phenomena of conceptual blending (Fauconnier & Turner, 1998). This means that it is not easy to rely on this theory in order to discover which aspects of movement, voice and imagery are associated with each other in the case of MIIOs, and therefore a more empirical approach must be followed. Findings from the interview analysis were used in order to inform the development of the coding scheme that was later used for the analysis of the audio-visual material, and they are presented in Chapter 4.

(b) Audio-visual material

Through an iterative process of observations, I visually identified, classified and labelled repeated events of manual gestures that allude to MIIOs for each individual performer/performance (which I also discussed with the supervision team). This repertoire of recurrent gesture types (that point to so-called 'catchments' (Rahaim (2009) from McNeill (2005)) constitutes the action-oriented ontology of MIIOs that was used in analysis.

The movement events (gestures) were first manually segmented, resulting in an audio-visual database of MIIOs. Additionally, each gesture was given a manual annotation for the perceived level of effort involved, on a scale of 0-10 (0 being the lowest and 10 the highest for each specific performer). Effort here refers to what is perceived by an observer during each movement event and it is supposed to be indicative of the effort level exerted by the performer. It is important to note here that it isn't easy to use a common scale for annotating effort observed by different vocalists. Therefore, the scale was 'calibrated' instead on each individual performer, which means that effort values represent the range between the lowest and the largest effort level for each specific vocalist. This type of annotation does not allow us to compare mean effort levels between performers, but it permits us to discover whether on average a performer tends to gesture in his/her most effortful or least effortful range. In other words, although the mean effort value is given in the analysis chapters, it only refers to the range of effort levels of a specific performance. Additional aspects of movement were annotated and these were used to facilitate the extraction of features in the quantitative part of the analysis. They include: handedness, direction of movement, distance between the hands, hand shape (open/closed), type of grip and the four basic factors of Laban's effort system (weight, time, space and flow). Although the latter were not used in the analysis, they made me aware of the fact that effort is a complex subjective measure and that I was therefore taking all four Laban factors into account for annotating it.

Unlike when a musical instrument is played, since no energetic coupling exists between the performer and the imaginary object while singing, effort in MIIOs is assessed as a combination of imagined/imitated and real, or overt and covert exertion. Here, real effort refers to that sensed by the hands when moving against the real force of gravity. It is achieved by adjusting the kinematics of the movement and internally activating the muscle tension (through contraction and expansion) that is associated with the hands' fatigue. Imagined/imitated refers to the sensation of effort that is elicited when manually imitating the action that would be required to achieve the intended goal for a real object (e.g. accounting for the imagined stiffness of the elastic material while stretching, as well as for the intended amount of stretching of this material). Despite the combinatory nature of effort in this context, I still expected that a third-person would be able to make judgements on the levels that the performer seemed to be committing to the task through visual observation of the dynamic and kinetic properties of the movement.

Additional movement information was annotated, which included: handedness, direction of movement, distance between the hands, hand shape (open/closed), type of grip and the four basic factors of Laban's effort system (weight, time, space and flow). This information was later used to facilitate automation in extracting features in the quantitative part of the analysis. For instance, distance and velocity were calculated in a different way according to the coding of the 'handedness', and the octave range was used in order to calculate pitch in relation to the tonic of each individual octave.

For each of the gestures, the associated audio content was also coded and classified in terms of the recurrently used melodic phrase type, pitch interval, type of melodic movement (whether ascent, descent or both in succession), 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 features might be associated with effort levels and gestures classes. Therefore, they differed slightly between performers. All movement annotations were made based on the visual information, however the sound was turned on, although not loud, and could have therefore influenced the coding. All audio annotations were conducted without visuals. The video footage was viewed multiple times. Only one of the features was annotated at each time in order to avoid imposing unintentional associations between features by coding them simultaneously for each single event.

Audio (at a 48kHz sampling rate) was coded and segmented in Praat[55] (Boersma & Weenink, 2005). These audio annotations were then imported and integrated into the ANVIL[56] annotation environment (Kipp, 2010). Movement (at a 29.97fps video frame rate) was coded and segmented visually on a frame-by-frame basis (33.3msec resolution). Unique labels (the "code book") were defined in an XML script for ANVIL and in a TextGrid for PRAAT and they were loaded prior to annotation in the corresponding software. The ANVIL file, where all information was integrated, was organised in two groups (as shown in Figure 3.1), one for audio and one for movement, each of which consisted of a number of individual tracks for coding different attributes of each modality separately.

[55] http://www.fon.hum.uva.nl/praat/
[56] http://www.anvil-software.org/

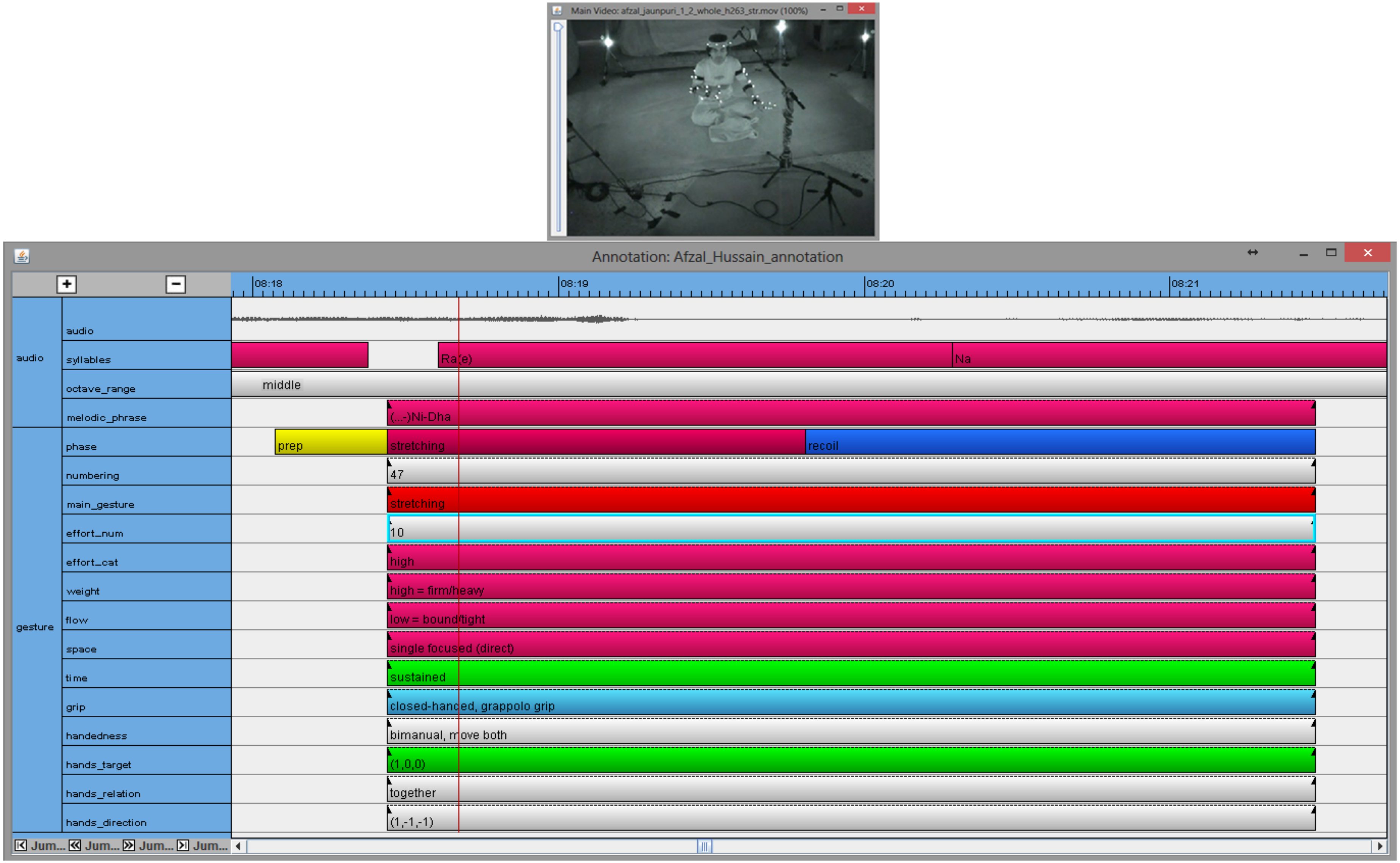


Figure 3.1. Annotation in the ANVIL environment (example taken from annotation of Afzal Hussain).

(c) Gesture identification, classification and coding scheme

Given the focus of this project on gestures that seem to be acted on some imagined object and on the effort that the performer is committing to the task for changing its physical properties, the performer's movements can be separated into two coarse categories:

1. Gestures (of various effort levels) involving the fighting against some resistance which is related to the physical properties of some (imaginary) objects;
2. All other gestures, which do not seem to involve some object and do not suggest acting against some opposing force.

Gestures of the first type are typically closed-handed manual movements and reflect the dynamic aspects of the melody, which is initiated with the closing of the fist and the formation of a grip, while gestures of the second type are typically open-handed and are related to the melographic engagement with the voice or to the keeping of the tempo/beat. In lines with the definition of effort followed in this work, the basic criterion for including gestures in the annotation was the observer's awareness that the performer is acting against some opposing force which is supposedly defined by the physical properties of the imagined object to be manipulated (the first type of gestures that is described in this section). Gestures of the second type are not part of this thesis. Therefore, although an overview of all types of gestures are presented here—given the focus of this work—more detail is placed on the first type.

Gestures that allude to MIIOs

Although the objects are only imagined in MIIOs, it is my assumption that a third-party human observing the process would be able to readily identify the object. The attempt to fully describe the identified gestures which recurrently appear during performances led to a first classification, which is based on the perceived types of interaction defined by the objects' imagined physical properties. These classes of interaction were informed by the interview analysis and led to the development of appropriate coding schemes for the performers' gestures. First the annotation for Afzal Hussain was performed, from which a basic coding scheme for the gesture classes was developed. Although several of Hussain's gesture classes were detected in the other case studies (Umakant Gundecha, Ramakant Gundecha and Lakhan Lal Sahu), some additional gesture classes were also identified and also some of Hussain's classes did not apply. Here I present an exclusive list that covers the gesture classes that were observed for all four performers and I give a detailed explanation of how each of them was defined and was later identified. The exact "code book" that was used in each performance is described for each individual case study in the chapters of video analysis (5, 6 and 7). Although there were several ambiguous cases that did not allow a clear-cut classification, the following list describes on which criteria I based the identification of MIIO classes.

- Stretching an elastic object, such as a rubber band, by deforming it from its original shape and moving against its varying opposing force. It can be performed by one or two hands and in various directions in space. The performer can either grasp the object at both ends and stretch it by moving the hands apart, or grasp it with one or both hands at a single point and stretch it by moving them

together in one direction, in which case the other end can be imagined as being attached to a rigid point in space. The different ways in space by which a stretching can be performed are illustrated in Figure 8.2. A distinct recoil phase is expected immediately after the main part of the movement, during which the hands are apparently pulled back towards the imagined original state of the object.

- Pulling an object of some considerable mass or weight towards oneself in order to overcome the resistance of a constant opposing force. This force can be imagined as either gravity or friction that offers resistance against the performer moving an object in space. It is performed either uni- or bi-manually and there is no expected recoil phase.

- Pushing (or pushing-away) an object of some considerable weight away. It is performed by a single hand while the palm is usually facing the body. This gesture seems to carry the same qualitative properties as the pulling movement, only in the opposite direction (outwards rather than inwards), which means that the hands seem to work on defying a constant opposing force. In the performances by the Gundecha brothers and Lakhan Lal Sahu, this gesture class was named 'pushing-away' to make a clear distinction from pushing-compressing (pushing away a compressible or elastic material).

- Collecting or taking an object towards oneself. This gesture has the same spatial orientation as the pulling (towards the centre of the body) and appears to be performed against some constant force, but in an almost effortless way, as if the object is weightless or extremely light.

- Throwing an object to some point in space. This is an impulsive (rapidly performed) and quite distinctive gesture. It is usually uni-manual and is performed by moving the hand away from the body or to one side.

- Undulation. It refers to a group of quick, repetitive gestures (typically stretching or pulling), which follows a distinct initial gesture. It can be seen as an elongation of this first gesture with some rhythmical character that reflects a repetitive process of intensification and release.

- Side-glide. This gesture comprises two consecutive gestures of the left hand (by Umakant Gundecha), the first moving sideways (to the left) and away from the musician's trunk and the other moving back (inwards) and approaching his body. Although there are several similar gestures of this type, especially in terms of their spatial orientation, a careful observation makes it clear that they vary in terms of which part of the gesture the performer seems to fight against the imagined resistance. In some of these gestures the performer seems to be committing a higher level of effort in the first part (when the hands move away from the trunk), while in others the inward movement (when moving the hands back, closer to the torso) looks more effortful. The varying placement of effort reflects the type of imagined interaction and the "active" or "passive" involvement with the task. The cause of the movement could be either:

(a) Throwing an object of some mass from the right to the left hand; the left hand catches and stops it in the air;

(b) Stretching out an elastic object from its resting point to the left side of the musician.

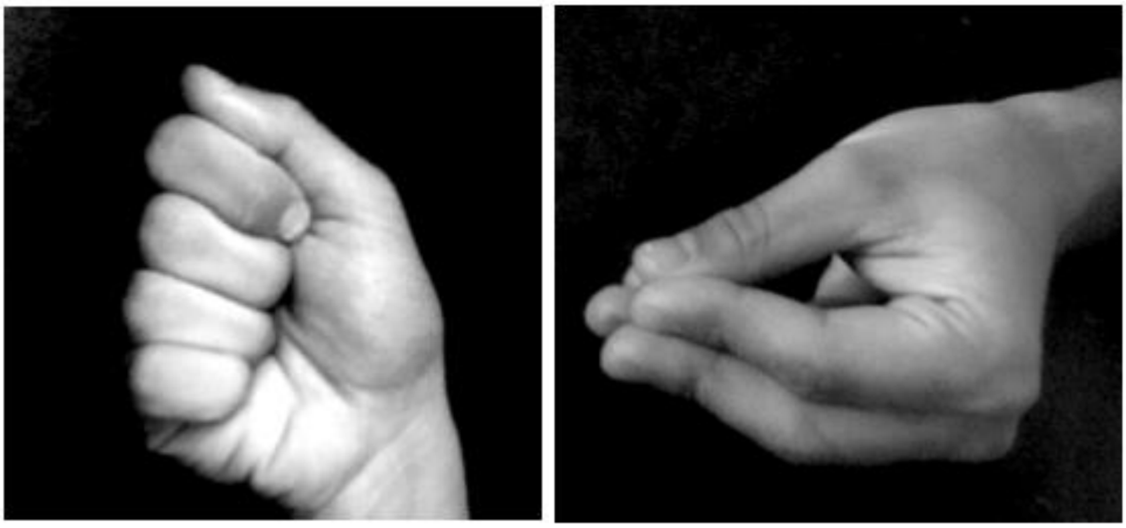
Thus, in the first case (e.g. throwing a ball) the movement of the left hand is considered a "passive" gesture in the sense that the left hand is not the one that sets the imagined object in movement. The right hand of the musician is in contact with the object only at the very beginning. The left hand is the one that traces the position of the object and it naturally retracts towards the body after reaching the limits of its extension. For the second case, the left hand will first perform an “active” movement in the sense that it will extend and cause the imagined (elastic) object to stretch out; in the second part of the gesture the object seems to recoil and pull the musician’s hand back towards the initial position. This gesture could also be imagined as the result of a much more complicated case of interaction, where an object that is attached on one edge of an elastic band is thrown from the right to the left hand.

Thus, the difference lies in the movement of the left hand. The main (first) part of the gesture in the first case is relatively effortless and in the second case quite effortful. The second part of the gesture is in the first case a retraction of the hand towards its rest position (no object involved) and in the second case a recoil of the object towards its original position; it can also be a pull of an imagined object towards the musician’s body. A combination of these alternatives gives the following three sub-classes of the side-glide for the left hand:

- *Passive-passive*, i.e. throwing and retraction (the left hand only traces the position of the object in space and then returns effortlessly towards the body);
- *Passive-active*, i.e. throwing and pulling (the return is done through a dynamic pull);
- *Active-passive*, i.e. stretching out and recoil (a stretching movement, but to the left side of the musician’s body).

To distinguish between a passive and an active first part of a sile-glide, in case this was hard to do, I had to rely on either an initial short flick of the left hand (as if throwing) or the dynamics of the first part of the movement.

- Pushing-compressing. This type of gesture refers to a pushing movement that is performed in order to compress an elastic object, which consequently will expand backwards during the recoil phase. Usually it is performed by pushing down, towards the ground. The recoil phase is a criterion that allows it to be distinguished from pushing-away (moving a rigid object in space).
- Lifting (when the hands only ascend) and ***lifting & letting fall*** (when the hands first ascend and then descend). Lifting has similar dynamic properties to pushing away an object against some constant opposing force, only in this case the gesture is performed upwards and against gravity. In some cases, the hand remains in the highest position after lifting the object, while in other cases it is allowed to fall back to its original position.
- Grappolo grip (see Rahaim, 2009 for details; similar to Kendon, 2004). It is a light grip formed by bringing together only the tips of the fingers of each hand and it is performed while moving the hands apart in opposing directions, fingers facing each other. It is used mainly to smoothly twist, tilt and modify notes (such as glides) as if an object were attached to the hands and manipulated.
- Kite-flying grip (see Rahaim, 2009 for details). This refers to a bi-manual gesture of firmly closed fists, showing the tight holding of an imaginary object. It is used in order to hold a note in place and sustain it without fluctuations (Rahaim, 2009). The two types of grip can be seen in Figure 3.2.



(a) (b)

Figure 3.2. (a) The 'kite-flying grip', usually for melodic glides, and (b) the 'grappolo grip', usually for the precise sustain of notes (Rahaim, 2009: 184).

As suggested by Rahaim (2009), the gripping hand shapes used by Hindustani singers seem to reinforce Kendon's view that human movement is 'rooted in the construction and manipulation of objects' (Kendon, 2004: 360). Apart from being separate gestures, a grip similar to kite-flying was sometimes observed in the beginning of MIIOs.

Although gestural interactions with an object can be visually identified[57] with great ease when the item really exists, it is not always a trivial task to do this in its absence. For instance, it was often difficult to distinguish between stretching and pulling gestures or between pulling and collecting/taking gestures. The following criteria were taken into account in the attempt to make a consistent distinction in case gestures were not clear-cut:

- When an elastic object is being stretched, an increase of effort (due to Hooke's law[58]) is expected and thus a non-uniform decrease of velocity with distance from the object's supposed equilibrium. Additionally, it is expected that a recoil phase will follow, during which the hands tend to be pulled back forcefully towards their initial position (in an initially accelerated movement), due to the imagined retraction force applied by the deformed elastic body.

- For pulling it is expected that the effort remains relatively constant with distance and thus the movement is uniform with a constant velocity vs. displacement during its largest part (excluding the initial and final parts of the movement, which are defined by the biomechanical constraints of the performer's body). Additionally, no recoil phase ('pulling' back) of the hands towards the equilibrium position is expected, although—as with any gesture—the hands normally return effortlessly back to the rest position before executing the next movement.

[57] Certain types of biological movement patterns can be unambiguously recognised not as static points in space, but due to their own organization of motion (Johansson, 1973).

[58] Hooke's law: $F = k \cdot x$ (holds true only for small displacements). F is the force applied on the cross-sectional area of the material at the contact point(s) to stretch it (='stress'), k is the stiffness of the material and x is the displacement (its distended length minus its relaxed length; called 'strain'). According to this law, force and displacement are proportional, with stiffness being the constant of the analogy.

- Collecting/taking is the act of bringing something weightless or extremely light towards oneself through a uniform movement. It could be possibly confused with pulling a light object, albeit when collecting/taking, gestures are almost effortless.

Despite these criteria, there were cases which were not clear, especially in distinguishing between stretching and pulling gestures. Therefore, mixed labels were also allowed in case of ambiguity. For instance, from the point of view of physics, stretching gestures should be performed with a varying velocity, while pulling gestures should be performed with a constant velocity. However, due to the necessary phases of initial acceleration and final deceleration imposed by biomechanical constraints of the body on any gesture, discriminating between stretching and pulling has not always been possible, especially when the gestures were far too short (under 0.5sec). Additionally, stretching was sometimes confused with pushing in the case of movements that were performed in an outward direction from the performer's body, as it was hard to discern whether the performer was interacting with a heavy load that needed to be moved away or with a stiff elastic object that needed to be extended and deformed in the same direction. The formation of a grip could not be used as a safe indicator for identifying stretching gestures, as apart from stretching an external object to the body, the performer could be also causing an internal stretching of the muscles.

Gestures unrelated to MIIOs

- Worshipping a deity (pantomimic gestures according to McNeil, 2005). It looks as if offering something (to gods) by bringing the two hands together and it is often used with either a single, discrete note or with an ascending or descending pitch glide to the tonic or the 5th degree. This is one of the most prominent symbolic movements, it has been brought up several times during interviews, and it is related to the fact that Dhrupad singing is aimed at praising God.

- Melographic gestures[59] (illustrators according to Clayton, 2007). They are open-handed uni-manual gestures, performed effortlessly by the right hand either (a) inwards and towards the centre of the body or (b) outwards and upwards. This type of gesture suggests a melographic engagement of the hand with the sound and is associated with a smooth pitch ascending or descending glide to the tonic or 5th degree. In the case of the Gundecha brothers, who have displayed this type of gesture frequently, there is a spatial asymmetry in the association between the position of the hand and the pitch height; pitch ascent is associated with the vertical position of the hand, while pitch descent is associated with the sagittal position of the hand (in relation to the singer's body).

- Keeping the pulse or marking focal points (markers according to Clayton, 2007). They are open-handed uni-manual gestures, performed as straight downbeat movements (for the repetition of the tonic or the 5th through a number of syllables) or as cyclical light movements performed inwards (towards the body) and with a downbeat at the end of each cycle. They are usually associated with a characteristic closing phrase, called mukhṛā.

- Resizing gestures (illustrators according to Clayton, 2007). They are bi-manual gestures, performed to show the distance between the hands in an effortless way. They seem to be associated with a change of brightness (an 'opening') in the voice and in accordance to the used syllable. For

[59] Term suggested by Rahaim (2009), meaning the tracing of pitch over time illustrated by the position in physical space.

very low pitches the size may also be depicted as the distance between the one hand and the floor. This type of gesture could possibly be associated with the mechanics of voice production rather than the melodic content of the music, and more specifically with the sense of expansion or 'fullness' in the epigastric area and the lower ribs, which is noticed by singers when producing notes at their lowest pitch range.

(d) Chunking and Segmentation

A cross-modal analysis necessitates the segmentation of musical sound and corresponding movement (Godøy, 2008a,b) into meaningful chunks. This was achieved by discriminating movements from ‘key-postures’ (Godøy, 2009) or so-called ‘phase-transitions’ (Godøy et al., 2012). These phase-transitions can be thought of as the boundaries for each of the MIIO events. They are defined by the biomechanical constraints of the imagined object and shape the exertion of effort (imitated and/or real); a bell-shaped effort pattern can be seen as an inherent feature of MIIOs and can potentially guide the segmentation process of MIIO events. Segmentation was thus carried out manually[60].

Following the analysis of gesture phrases employed by Kendon (2004: 112) and Jensenius (2007: 46), here the start and end points of a gesture were taken to correspond to the boundaries of the nucleus, which consists of a stroke (the main body of the gesture) and—in most cases—a post-stroke (a prolongation) phase. As shown in Figure 3.3, the preparation and the recovery phases were excluded from the segmentation[61]. This ensures that analysis is directed only towards the part of the movement in which the hands appear to be in contact with the object.

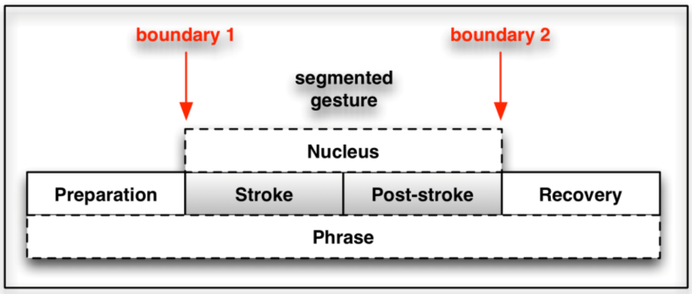


Figure 3.3. Gesture segmentation based on the construction of an action phrase (gesture according to my definition in 2.2.1.)
by Kendon (2004: 111) and Jensenius (2007: 46).

The stroke constitutes the main goal-directed movement associated with the engagement with the sound and it starts at that point in space from which the MIIO appears to originate, i.e. the end of the preparation phase. During the performance, a stroke is most often followed by a post-stroke, i.e. the halt and/or oscillation around a point in space, or a retraction in the opposite direction to the main stroke. In the case of a stretching gesture, the post-stroke will coincide with the recoil phase back to

---

[60] Automatic segmentation of ancillary gestures is also possible, e.g. Caramiaux et al., 2012.
[61] Assuming that any movement starts from the rest position (hands lying on legs in front of the centre of the body), the preparation appears as a displacement of the hands to the point in space from which the main stroke will start. The recovery phase refers to the return of the hands to the rest position.

the original position. Some exceptions do exist in which there is no post-stroke phase, but these are relatively few. Often, but not always, during preparation the hands will form a grip before they start manipulating the object. Although this will often demarcate the beginning of a MIIO, it is not always a safe indicator. Additionally, although vocal interruptions for inhaling constitute an important phrase marker in solo Hindustani singing performances (Rahaim, 2009: 155), they are not good segmentation markers, as a gesture can often continue across several breaths (Clayton, 2007).

Segmenting a MIIO movement event in a real performance situation is not a trivial task, largely due to coarticulation issues between adjacent movement events, discussed thoroughly in Godøy et al. (2010, 2008c). Coarticulation refers to the smooth transition and blending from one gesture to the other, sometimes even without an intervening retraction towards the rest position. Indeed, the gestures observed only rarely appeared as separate entities, but instead the ending part of one gesture usually ran into the beginning of the next, lending some of its energy (or momentum) to it. Therefore, in practice the observer may visually identify the corresponding boundaries of MIIO events by considering changes in geometry (the spatial pattern of the interaction), effort and velocity (a pattern by which there is a peak between two minima); velocity and perceived effort are minimised when the hands return to their rest position.

In an effort to ensure some level of consistency when identifying the boundaries of the gesture nucleus, the task was approached with a checklist of four criteria. Due to the complexity of the movements and the difficulties of analysing the video footage (in identifying crucial points in space from the specific angle by which the video was shot), these points were followed in order of priority, with points 1 and 2 being usually the clearest and points 3 and 4 functioning as supplementary information in case the first points were insufficient for segmentation purposes. Using these criteria, the identified gestures were manually segmented in ANVIL and all annotations were made upon these segments.

1. Abrupt direction change of the hands, marking quite clearly the beginning of the gesture nucleus.
2. Momentary zeroing or minimisation of velocity at points of sudden direction change (at least as seen in the x-y plane of the two-dimensional video image) at the beginning and end of the movement segment.
3. Maximum or minimum extension from a reference point (the performer’s body or the perceived object’s initial rest position).
4. Formation of an initial grip (although this can be misleading in the absence of a real object). Had the object been real rather than imagined, this would have been a far stronger indicator. With imagined objects, the grip was not always formed exactly at the beginning of the gesture.

The segmentation of the audio was carried out separately. Although the stroke is supposed to be timed to co-occur with its vocal counterpart, the two streams are not exactly aligned in singing. This means that the sound would be cut before completing a meaningful phrase, if it had been segmented based on the boundaries of the movement event. Therefore, the voice was segmented separately (in Praat), based on the combined criteria of syllable boundaries, silence and pitch extremities (the maximum and the minima) of glides.

3.2.3. Stage III: validating and using the code

(a) Code validation (inter-coder agreement)

The aim at this stage was to assess through comparison whether findings between various data sources and different coders converge (Guest et al., 2012). An inter-coder validation of the visual material in terms of gesture class (MIIO type) and perceived effort level was carried out by two professional dancers/choreographers for the performance of Afzal Hussain. No validation was carried out for the other performances, since the high cross-coding correlations of the first case study were a good indicator that gesture type and effort levels can be reliably annotated by a single annotator. The annotators were chosen due to their expertise and professional backgrounds, which allowed them to observe each gesture with an analytical and detailed approach. Only the visual material was cross-validated, as this is the only domain in which observers were expected to disagree significantly. The subjects were informed prior to the annotation sessions that the project involved an investigation of the relationship between gesture and sound in cases where the singer seems to be interacting with an imagined object. The coders (both of Greek nationality) were asked to identify instances of this kind of interaction and to assess the level of imitated effort that they perceived the performer had committed to the task. Clearly stated guidelines were provided orally, in order to ensure a common understanding of the coding scheme and to minimise variation across the raters' coding procedures. There was first a clarification about how these gesture types are defined. The purpose of this was to make sure that the coders annotate according to the same criteria.

Each coder annotated the material during a separate session, while I was present to clarify possible misunderstandings. The process was carried out in Greek and the coding scheme was translated from English into Greek with a detailed explanation of the terms (although both coders are fluent English speakers). The sound was on but at a very low level during the sessions, in order to keep the annotation conditions consistent with the way I had been coding. Thus, the coders' annotations may have been influenced by the sound. However due to their professional occupation as dancers/choreographers and the fact that the sound was hardly audible, I have been observing during the process how their annotation decisions were based exclusively on the visual information. Due to the large size of the corpus and the highly demanding task, only one third of the samples were presented to the annotators, i.e. 30 samples. They were selected as to represent the entire duration of the performance and include equal numbers of all types of MIIOs. The samples were shuffled and played in a random order. The coders were first given the same coding scheme for the identification of the interaction types that I had used, except that it excluded the two types of grips, which are specific to Hindustani singing. The coding scheme included the five unambiguous gesture classes and three ambiguous classes that I detected in the case of Hussain, that are all shown in Table 5.4. Both coders were also asked to assess the effort levels on a scale of integers between 0 and 10 (10 being the highest value). They were allowed to play back each excerpt as many times as necessary. The annotations were made on a prepared form, which required the annotators to fill in the gesture type and effort level for each numbered sample. This selection of annotations was then compared against the original codes made by me. The degree of correspondence between annotators was measured by

Cohen's kappa value and it was tested by examining whether the kappa value exceeds chance levels (Kipp, 2010; Kohen, 1960). Thus, although the analysis described in this section was qualitative, a quantitative approach relying upon statistical methods was followed in order to check the robustness of the annotations.

(b) Code application

What followed was an investigation of consistency in the co-occurrence of particular melodic phrases with certain types of gestures. Pair-wise associations between various annotated features were examined, most importantly the association between melodic phrases and effort level. Findings are presented in chapters 5, 6, and 7. Timing issues (duration and alignment) between hand movements and their melodic counterparts were also examined.

3.2.4. Credibility assessment of qualitative methods

'Validity' is often discussed in the assessment of quantitative work, a concept that hinges on the repeatability of results. For qualitative research, which cannot reflect the same standards of objectivity, the term credibility is often used as a criterion to assess the confidence that should be placed in conclusions. Credibility may be tested by means of participants' agreement, analysis of multiple sources of data and inter-coder cross validation (Suter, 2011).

It is argued that strategies to ensure the trustworthiness of the whole research process are more important measures of credibility than *post hoc* evaluations (Rolfe, 2006). To this end, the eight strategies proposed by Easterbrook et al. (2008: 306-307) for improving credibility were followed to a large extent in this thesis. This meant: collecting data from multiple sources, performers and rāgas during a prolonged stay in India as well as on other occasions; documenting processes through a detailed audit trail along with collecting a significant amount of actual data; addressing biases in the interpretation of data by ensuring clarity of definition during the inter-coder validation process; and reporting discrepancies from clear-cut cases as ambiguous or unclassified gestures and melodic phrases.

In order to allow for credibility to be evaluated in qualitative analysis, as suggested by Miles and Huberman (1994: 11–12) the rich and complex audio-visual material was reduced into meaningful classes and codes of recurrent movement events and melodic movements. Following the same guidelines, findings were displayed in organised and compact ways through tables, graphs and charts, and conclusions were drawn from the findings and presented in the end of each of chapters 4, 5, 6, 7 and 8.

3.2.5. Generalisability of findings in qualitative research

Since qualitative studies are typically conducted based on a relatively small number of subjects, considerations on the generalisability of findings from one subject to a larger population are often raised. Therefore, the purpose of qualitative research is not to provide statistically significant findings for generalising from a sample to the whole population. Rather, what is sought is an in-depth insight into whether particular concepts and ideas can be applied outside of the study across different contexts, for instance with different participants, groups, situations etc. (Suter, 2011). Thus, the term 'transferability' is often used as an alternative to 'generalisability'. It hinges on judgement of whether the detailed descriptions of one case could find a "fit" in other contexts, such as through cross-case comparisons, i.e. cases that yield similar findings (ibid.). In this thesis, qualitative case studies II (Chapter 6: Lakhan Lal Sahu, rāga Mālkauńś) and III (Chapter 7: Umakant Gundecha, rāga Bhūpālī) were analysed based on common principles and codes that were developed in the first case study (Chapter 5: Afzal Hussain, rāga Jaunpurī), and findings were compared.

3.3. Quantitative Analysis

A number of computational measures and tools have been proposed for the analysis of cross-modal similarity in the way the term was defined in section 2.4.2. These methods require the use of time-varying features that can be represented as time-series or curves and can be extracted from continuous multidimensional data streams that are captured by sensing technologies (the raw audio and movement data in this case). Approaches in the analysis of movement–sound similarity include both analysis-by-synthesis and analysis-by-measurement methods. Examples of analysis-by-synthesis methods include the works by Fenza et al. (2005), Jensenius (2007), Françoise et al. (2012), Caramiaux et al. (2010b) and Schwarz & Caramiaux (2013). Examples of analysis-by-measurement methods include t-tests and Spearman's correlation (Nymoen et al., 2013), HMMs (Caramiaux et al., 2010c), Kullback-Leibler probability as an appropriate divergence measure (Caramiaux, 2010a), Functional Data Analysis methods (such as Canonical Correlation Analysis[62]: Nymoen et al., 2011 and Caramiaux et al., 2009, 2010a, and B-spline functions: Bianco et al., 2010) and pattern recognition classifiers (Nymoen et al., 2010, 2013). An overview of machine learning techniques can be found in Caramiaux & Tanaka (2013), which also includes a description of regression models that were used in the context of this thesis.

This section details the quantitative approach of the current work, which aimed at validating and formalising the relationships that were found during the qualitative analysis. At the centre of the quantitative method is regression analysis. Although regression has been previously used by Caramiaux et al. (2009) in estimating movement feature values from acoustic features (and vice versa), this work offers a novelty by including a combination of both movement and acoustic features in each single model, as a way to incorporate an embodied approach in the analysis.

Regression refers to the task of estimating the unknown relationship of a set of output values (the response) to an input or a set of input values (the features) through a mathematical function (James et al., 2013). In its simplest form, this function is assumed to be linear, as in this work. The method is referred to as 'simple regression' when we use a single feature to infer the response and as 'multiple regression' when we use several features simultaneously (a linear combination, in this case). It is a statistical method that falls under the category of supervised machine learning techniques, which means that for each observation there is a known associated response (here the annotations). The response (target) variables can be either quantitative (here the annotated effort level numbers) or qualitative/categorical (here the annotated gesture classes). Quantitative responses take on numerical values, while categorical responses take on codified numbers that do not exhibit any natural numerical ordering (such as low-medium-high) but correspond to the pre-determined categories. Assuming that the function is linear, we tend to refer to regression problems as 'linear' or 'least squares regression' for quantitative responses and as 'logistic regression' or 'classification' for qualitative/categorical responses. The difference between them is that in logistic regression the linear function estimates the probability of the response falling into one of the available categories, which is

[62] Canonical Correlation Analysis (CCA) has been used both as a dimensionality reduction method and as a way to find intrinsic relationships between two sets (movement and sound) of time-varying features.

then used to determine the output class according to a given threshold. This work makes a distinction between 'Linear Models' (LM) for the estimation of effort levels and 'General' or 'Generalised Logistic Models' (GLM) for the classification of gestures.

There are two ways that the outcomes of a regression method (linear or logistic) can be used in a study: (a) for prediction, i.e. to accurately predict the response for future observations and (b) for inference, i.e. to better understand the relationship between the response and a set of features in the available data (James et al., 2013). The aim here was to achieve a simple interpretable model that could explain a good part of the observed responses (1. effort level, 2. gesture classes) with a reasonable accuracy through a small number of features. To take an embodied approach, a combination of movement and acoustic features was considered.

Typically, the formalised description of a relationship is inferred during a training phase by using part of the observed data (the training dataset) and this is then tested on the remaining data (the testing dataset). However, in the present study, the size of the dataset was quite limited for training purposes. Hence the goal was not so much to produce statistically significant results, but rather to explore the role of effort in gesture–sound relationships and to discuss possible convergences as well as discrepancies in the results between qualitative and quantitative approaches, both among different performers and during different performances. More specifically, the following questions were addressed through the quantitative methods:

1. Is there any simple (linear) function between the features (audio and movement) and the response (effort level/gesture class) that can adequately summarise their relationship?
2. How well does the model fit the data?
3. How does each individual feature relate to the response (the direction and strength of the relationship)?
4. Which (movement and audio) features have the strongest correlation to the response? In other words, do we need all features that can be extracted from raw data to explain the response or is it possible to achieve good results with a limited set of features? This process can also be seen as a dimensionality reduction method.
5. Which is the linear model that can be developed to best describe the response of an individual performer and which is the (most shared) linear model that can best describe inter-performer responses?

The following sections explain how each of these questions (numbered) was approached, separately for each of the LM and GLM methods.

3.3.1. Linear Model (LM)

1. *Linear Function:* Whether a feature-response relationship exists can be examined by estimating the coefficients of the linear model. In the case of a multiple linear regression, the model takes the form of:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \cdots + \beta_n \cdot X_n + \varepsilon$$

, where β_0 is the intercept (in other words the null model, i.e. the one that does not include any features) and ε is a random error term, which is independent of the features and its mean is zero. The existence of ε indicates that even if we knew the real regression line (i.e. even if the 'real' coefficients and intercept were known), we would not be able to perfectly estimate the response from the features X_i, due to the existence of noise in the measurements. The unknown variables (the coefficients and the intercept) are calculated by fitting the model to the training data based on the principle of least squares or maximum likelihood, i.e. methods that minimise the error between the estimated and the real values. In practice, the regression method learns the coefficients and intercept that govern the exact shape of such a linear expression.

In order to determine whether there is a relationship between the response and the features of multiple regression, the null hypothesis (all coefficients are zero) is tested against the alternative hypothesis (at least one coefficient is non-zero). The hypothesis test is performed by computing the so-called F-statistics, which provides a way to assess the significance of all independent variables (James et al., 2013). However, there is no well-defined threshold for assessing the strength of the relationship and the criterion is based on the number of features and the data size. It is important to note here, that the approach of using the F-statistics to test the strength of the feature-response association can only work in cases where the number of features is small in relation to the number of observations.

2. *Assessing the accuracy of the model fit:* There are different methods for measuring the goodness of fit, in other words the extent to which the model can adequately capture the response variance. The R^2 (R squared) statistical measure was used here, which is an estimate of the standard deviation of ε. Roughly speaking, it reflects the average amount by which the response deviates from the true regression line. R^2 is calculated as a proportion of the variability of the response that can be explained through the available features, with values close to 1 indicating a good fit and values close to 0 indicating a low model fit (James et al., 2013). Nevertheless, there is no well-defined threshold for assessing the model accuracy. In simple linear regression (with a single feature) the R^2 is identical to the calculation of the correlation between individual features and response (which was also examined).

3. *Correlation between response and individual features:* The strength and the direction of the correlation of each individual feature to the response were rated based on the absolute value and the sign of each individual coefficient. One important thing to note about multiple linear regression is the fact that the coefficients and the goodness of fit produced by fitting a separate linear regression model for each individual feature cannot be simply extended to a model that uses the combination of these features. In practice the results would differ tremendously due to the combinatory effect of the features on the response. This means that adding a feature that would produce a good fit if used on its own, might actually lead to a drop in the model fit if used in combination with other features.

4. *Feature selection:* Depending on the aim, more flexible (smaller number of features) or more restrictive (larger number of features) linear models can be fit to real data. By using less features, models have a more generic power on a wider range of data (therefore flexible) but they are less

precise in estimating the response. By adding more features, the goodness of fit for the specific data set used may be raised, but the generalisability in applying the model to different data may drop (hence restrictive). As a general rule, when we use a more flexible method, the variance is expected to increase and the bias to decrease and this is often referred to as 'bias-variance trade-off' (James et al., 2013: 33). Variance refers to the amount by which the response would change when estimated using a different training data set, while bias refers to the error that is introduced due to the approximation of a complicated real-life problem by employing a simple model (this being the best possible attempt to describe it). More specifically, by using a larger number of variables for estimating the response we could potentially raise the accuracy of the model fit, but there are two disadvantages:

(a) Complicated interpretation of the model: The more features we fit the model to, the more difficult it becomes to interpret the results, with some of the features having complicated combinatory effects on the response. A more compact model can be more effective in explaining a phenomenon.

(b) Over-fitting: This means that the learning procedure run on the training data picks up patterns specific to the training data samples rather than as a reflection of more general properties of the phenomenon under examination. This in turn means that if the model were run on new (test) data (e.g. additional data from the same performer/performance or other performers/performances), there would be a high error level, as the model would be explaining patterns that do not necessarily exist in the test data.

The good performance of a statistical method on the test data requires both low variance and low bias. Due to the trade-off between them, however, it is a challenge to find the ultimate number of features that will produce the best compromise between a flexible and a restrictive method. In the case of this thesis, linear models were fit for inference purposes; therefore I was more inclined to develop a rather restrictive method (with fewer features). This would be useful in explaining a good part of the response's variance, so that it could be interpreted more easily and described in a more compact way.

The refinement of the model towards a more compact alternative with a lower number of features was achieved in this work by assessing the contribution of each individual feature to the model's response. Phrased differently, it was important to find out whether the coefficient (of each individual feature) is non-zero. For this, I relied on the probability value (p-value) of each feature, which—roughly speaking—can be interpreted as an indication of likelihood in observing an association between feature and response due to pure chance in the absence of any real association (James et al., 2013: 67). The smaller the p-value, the more significant the parameter and the less likely it is that the actual parameter value is zero. Finally, a decision can be made about whether the null hypothesis can be rejected in favour of a non-null hypothesis by comparing the p-value to a threshold, called the significance level α; if the p-value is smaller than the significance level α, one can safely reject the null hypothesis and conclude that the estimated coefficient is non-zero and thus that it significantly contributes to the estimation of the response. Typical levels for α are between 0.001 - 0.05 (Moore and McCabe, 2006). In the current thesis I used an α value of 0.05 as a threshold. As there were a number of model candidates of similar sizes that could be deduced based on this criterion, the final selection was based on their accuracy[63]. Finding a good feature subset is an exploratory process that can be approached in two ways:

(a) Through the progressive subtraction of features from a feature set, i.e. by starting with a full model that includes a high number of features and progressively reducing it to only include the most significant features while raising the fit;
(b) Through the progressive addition of features to a basic feature set, i.e. by starting with a compact model that includes a small number of features and progressively extending it to include a wider range of alternative features while raising the fit.

Due to the limited size of observations in this study, the first approach was not considered. Instead, the features that are reported in Nymoen et al. (2013) were used as the core set from which the trials started. In regression analysis there is a 20:1 (Jackson, 2007) or a more relaxed 10:1 (Kline, 2015) rule of thumb in defining the ratio of sample size to number of features. In the opposite case, i.e. if the number of features is too large in comparison to the sample size, there is a danger of over-fitting a model that accurately describes explicit characteristics of the specific dataset but cannot be easily extended to other datasets. The core feature set by Nymoen et al. is described in detail in section 8.1.1.

5. *Intra- and Inter- performer modelling:* To address this issue, it is necessary to find the best compromise between a high accuracy in the model fit and a small number of easily interpretable features. This involves a process of exploration in testing various combinations of possible features. Section 8.1.2. describes step-by-step how the best trade-off between model accuracy, compactness (small number of features) and simplicity in feature extraction was achieved.

3.3.2. General Logistic Model (GLM)

1. *Logistic function:* Similarly to linear regression, whether a feature-response relationship exists can be examined by estimating the coefficients of the model. This involves applying a logistic function to the relationship between response probabilities (log-odds or logit) and features. A full discussion can be found in James et al. (2013: 132). The coefficients are estimated by fitting the model to the training data based on the principle of least squares, in order to reduce the errors to a minimum.

2. *Assessing the accuracy of the model fit:* Several methods for capturing the goodness of fit exist, from which I chose to use AUC (Area under Curve from ROC curve) and the rate of correct classifications. The use of k-fold cross-validation was also considered but was not used, as it requires a larger data size that was not available. The two chosen methods are outlined here.

- **AUC:** This measure reflects the area under the so-called Receiver Operating Characteristics curve (Hanley & McNeil, 1982). ROC is a plot that illustrates the performance of a binary classifier system by plotting the true positive rate (fraction of correctly identified classes) against the false positive rate (fraction of falsely identified classes) at various threshold settings or criterion changes. An ideal ROC ascends steeply in the left side of the plot, creating a greater area between the curve and the

[63] R^2 for LM and AUC for GLM.

horizontal axis and representing a better classifier. A straight diagonal line (with AUC=0.5) represents a classifier that performs no better than chance (thus poor). AUC can be useful for comparing different classifiers/logistic models.

- **Rate of correct classifications:** This measure is calculated as the fraction of correct classifications among the total number of classifications. It can be easily extracted from a 'confusion matrix', a table displaying the number of real response instances against the number of correct and incorrect classifications of the model. It is an easy way to recognise if the system is getting confused in classifying gestures correctly.

All of the abovementioned measures take values between 0 and 1, with numbers close to 1 indicating a good fit. Model selection in the quantitative analysis of the current thesis relied on AUC due to the invariance of ROC against class skew of the applied dataset (i.e. ROC will yield the same AUC for different percentages of positive labels).

Questions 3, 4 and 5 were assessed in the same way as in linear regression.

3.3.3. Potential pitfalls

There are several potential pitfalls to account for when fitting a regression model. These are explained here.

(a) Non-linearity of feature-response relationships

Although starting from the assumption that a simple, linear expression can describe the relationship adequately, it is also possible that the true relationship lies far from this model. Thus, in addition to looking at the R^2 statistics just discussed, it may be useful to produce some diagnostic plots, including:

- "Fitted samples vs. Residuals" to search for pronounced or identifiable patterns in the residuals;
- "Quantile to Quantile" (QQ) to examine whether data displays a standard normal distribution (which is one of the assumptions of least squares linear regression).

In the current analysis, graphs displaying the residuals and the quantiles were always plotted in order to develop a sense of whether the assumption for linearity could be considered as valid. In most of the cases, some patterns in the residuals were identified indicating an element of non-linearity, however these patterns were limited enough to consider the linear models as suitable for estimating effort and classifying gestures for inference purposes.

(b) Co-linearity of features

This problem refers to cases in which two or more of the explanatory variables are strongly correlated to each other. Correlation is a measure of the direction (sign) and strength (absolute values between 0 and 1, a value close to 1 showing a strong correlation) of the relationship between two quantitative variables (Moore and McCabe, 2006). The presence of such features can pose problems, as it masks the true relationships among variables by making it difficult to discern the individual contribution of each feature to the response. This causes problems in developing a reliable model and can lead to a misleading rise of the overall accuracy in the fit and complications in model inference. The detection of co-linearity problems can be achieved by producing a correlation matrix of the features. In the current analysis, in case heavily co-linear features were detected in the models with the help of the correlation matrix, one feature from each pair was always removed (pairs of features with absolute correlation values>0.9).

(c) Non-normal distribution of sample data

The statistical tests used here assume that the data samples (the extracted features) display a normal distribution, something that is however not always the case in real life. Although this condition was not always met by our (limited) dataset, due to the explorative nature of the analysis it was not considered prohibitive.

(d) Small sample size

In case of small datasets, it is not always possible to separate the data into a training and a testing group of samples. Therefore, the model fit can only be tested on the actual data that was used for its development rather than any other (testing) data. In this case, the goodness of fit of the model can be overly high and possibly misleading, as in practice the same observations are used for both training and testing the model. This can be counted as a limitation of this project and stems from the peculiar difficulties of the data collection process (the decision to use ecologically valid real performances in India rather than responses to designed stimuli). On the other hand, as the aim of the project is focused on inference, it is here believed that the method still reveals the most important trends in the relationships between features and responses.

3.4. Data collection

Qualitative information was gathered from a variety of data sources, including interviews and non-participant observation of audio-visual materials. The quantitative part relies on continuous data streams of captured motion and recorded audio material for the numerical analysis of extracted time-varying features. The material of the current thesis is based primarily on my own recordings. I made these recordings exclusively for this work during my field-work in India in 2010-2011 (audio, video, movement). I also conducted a few interviews in the UK and the Netherlands during the same period. The only exception is the video material of a performance by the Gundecha brothers that was recorded by Martin Clayton, Laura Leante and Andrew McGuiness in India in 2007 and is used in Chapter 7. No mocap[64] data is available for this recording. Although I have my own recording by the Gundecha brothers of the same rāga (Bhūpālī), it is a short performance and therefore the number of MIIOs detected is too limited (about 20) to draw statistically significant conclusions. The next section describes in detail the whole data collection process of my own recordings and includes some information on this separate recording of the Gundecha brothers.

3.4.1. Recruitment of musicians

A purposive (or theoretical) sampling method was used for these studies. This is a deliberately non-random method of selecting participants relevant to the research question, with the aim of allowing the researcher to focus on the particular area under study (Bowling, 2014). This section includes a general description of the data collection processes and methods, while exact details for the data collection of each study are presented in the respective chapters of the individual studies.

Seventeen (17) professional and non-professional vocalists of various backgrounds and musical experiences were recruited in India between December 2010 and January 2011. Four (4) professional musicians were contacted prior to the beginning of the field work in India, all of whom I had personally known beforehand (through various musical activities). Another four (4) professional musicians were approached while in India as a result of personal contacts in the field, while seven (7) non-professional musicians (students of the above) were also recorded during fieldwork; only one of them was not Indian. An additional recording of Marianne Svašek that was made in the Netherlands during an earlier period of fieldwork was also used. Only one (Mohi Bahauddin Dagar) is not a vocalist but an instrumentalist. However, he was recorded during classes where he instructed a student by singing. Some musicians were recorded only once, while for others I had the chance to run more than one session. The total duration of recordings is about eighteen (18) hours, however some mocap recordings are not acceptable due to excessive optical noise.

All selected participants belong to the 'Dagar gharānā' (a stylistic school of Dhrupad). This means that they are first- or second-generation disciples of the renowned vocalist Zia Fariduddin Dagar; the maestro himself was also recorded. In principle this means that all of the musicians have been trained according to the traditional guru-śiṣya paramparā principle of living for a few years with the teacher in his house. This deliberate choice in the recruitment of musicians aimed to retain a level of cross-

[64] Motion capture.

performer consistency in gesturing (see section 2.1.3. for a discussion on the 'paramparic body'). Participants belong to both genders, are of different ages, levels of musical experience and nationality (although almost all are Indian)[65]. The diagram in Figure 3.4 shows guru-śiṣya relationships between the participants.

Recordings were exclusively made during specified sessions for which the subjects were informed beforehand and never for the normal domestic music activities. For each of these sessions, written informed consent forms and recording agreement release forms were obtained from the participants before each performance or interview (Appendix C), that specified both the collection and the use of the data (including publication). The performers were always informed by me about the fact that the collected material will be used solely for research and educational purposes and not for profit. I considered important to avoid giving a specific institution or person the exclusive rights to the use of the recorded material, but rather to be able to make it public and accessible for the international research community. This involves also my future plans to upload the whole material to a public space, after cleaning the mocap data and aligning all data sources (audio, video and mocap). I also considered important to give performers copyrights of their recordings for non-profit purposes.

[65] All musicians were paid a small amount per hour for the recording of performances but not for interviews.

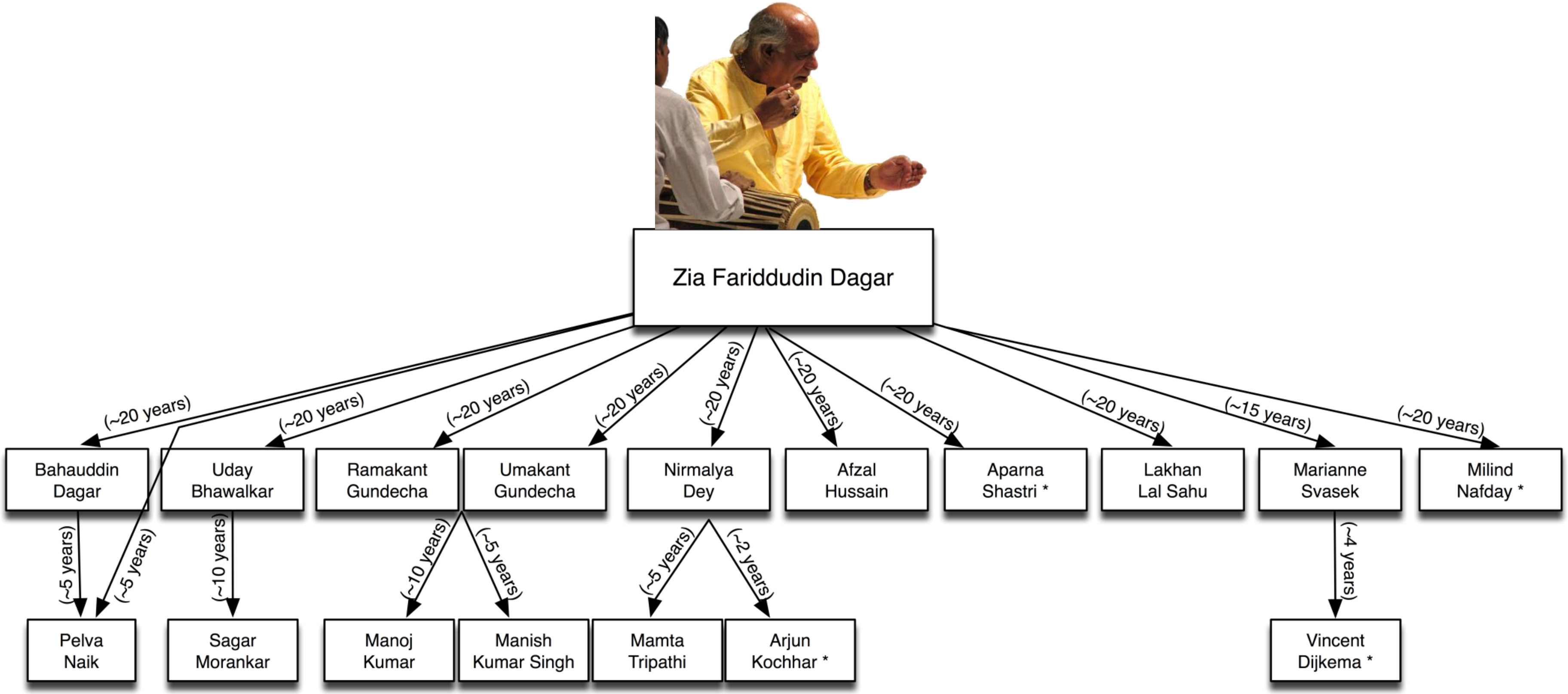


Figure 3.4. Recruited musicians (= not a full-time musician), showing the relationships in the transmission of music knowledge and the years of study.*

3.4.2. Interviews

Conducting interviews is considered a valuable method for obtaining rich and descriptive data (Robson, 2002) when no prior hypothesis exists. Interviews in this project were meant to provide an initial understanding of the subject and to inform the annotation process of the audio-visual material that was recorded during performances. Of the seventeen musicians that were recorded singing, eight were interviewed, all senior performers. Subjects were only informed about the fact that the research project was related to music and movement, but no more specific details were given. Questions were asked in English and interviewees responded either in English or a mixture of English and Hindi if they did not feel confident enough to rely entirely on English. A translator was used in these cases. All interviews were recorded in audio and video, with equipment placed at fixed positions. Some interviews were conducted immediately after a performance and while the subject was still wearing the mocap markers, therefore some mocap recordings of interviews are also available.

All interviews followed a semi-structured method of open-ended questions, starting with general topics and moving to more specific ones. Open-ended questions are used when the area of discussion is pre-defined, but the interviewees are allowed to digress and raise issues that they need to elaborate in more detail (Britten, 1995). This means that although a set of questions was prepared beforehand (as a guideline to insure that all the important topics would be covered), unexpected discussion subjects could also be explored and the interview could be guided by the musician's answers rather than being constrained to the initially designated topics (Denscombe, 2003; Robson, 2002).

Interviews with the musicians began with questions about their musical background and experience. They would then typically move to more music-related issues, including general topics of rāga improvisation, as well as more specific questions about the rāga that was just performed (in cases where the interview was conducted immediately after a performance). As soon as the musician started making use of motor-based metaphors in an attempt to describe abstract melodic ideas, the questions gradually moved to more focused issues related to the research topic of the project. Some of the musicians (usually junior students) hesitated at points in their responses, suggesting that the questions should be better directed to their guru for a more precise answer. When this happened, I persisted in trying to elicit the musician's personal view on the issue raised. The interviews lasted between 10' - 90', depending on the availability of the interviewee and his willingness to provide information which could extend from purely musical aspects to issues related more specifically to body movements.

The following questions are an example of how an interview would proceed from general to more focused questions:

> 'Could you please tell me a few things about this rāga'?
>
> 'As I am interested in hand movements during Dhrupad singing, I would be interested to know if Ustad (Zia Fariduddin Dagar) would verbally explain you how to move when he was teaching you; and if yes, could you give me some examples of what he used to tell you'?

Usually the answers to this question would involve pictures of nature, colours, images of discussions between people and moods. My next questions would then concentrate more on the melody, such as characteristic phrases of the rāga or important melodic movements (such as mīṇḍs or gamaks) that would eventually elicit the verbal use of motor-based metaphors and sensorimotor descriptors. Then my questions would focus on the actual hand movements that I would see the vocalist performing while speaking and I would refer back to the performance that I had just been observing while recording. For example, I would ask:

> 'While you were doing a mīṇḍ you moved your hands in a way that gave me the impression that you were holding something. Do you really feel that you are holding something between your hands when you do this mīṇḍ'?

From that point the discussion would remain mainly focused on MIIOs.

3.4.3. Recordings of ālāp performances

The recording sessions were conducted in the field in different places in India during a period of five weeks between December 2010 and January 2011. The recording sessions included solo singing as well as vocal teaching and jugalbandī (duets between two solo musicians). Parts in which compositions were sung were excluded from the study in favour of improvised sections. This is to focus only on the expressive qualities of the sound (Cowie et al., 2001), avoiding the possibility of semantic content from the text shaping the hand movements. Similarly, the metered part of the improvisation was not used in the analysis, in order to avoid hand claps and strokes that sometimes occur to mark moments in the tāla (the metric cycle of Indian music). In fact only the very first and slowest part of the ālāp (non-metered part of the improvisation) was used, as it better serves the research purposes of the current work. Therefore, the musicians were asked to perform only an ālāp improvisation of a rāga of their own choice, without a composition that usually follows.

Recordings varied in length according to the willingness and mood of the musicians and the peculiar recording conditions (power cuts, ambient noise, other scheduled activities in the domestic spaces where the performances were conducted), ranging between 10 and 70 minutes. The audience also varied in size from just me (when an electronic taňpurā was used and thus no other musician was required to be present) to a number of people attending a session (with the peak being the jugalbandī of Zia Fariduddin Dagar and his nephew Mohi Bahauddin Dagar, where an additional seven people were seated in the 7x4m room, behind the calibrated motion capture and audio recording system).

The choice of collecting data from real performances in domestic spaces in India—in contrast to conducting designed experiments such as by asking the vocalists to respond to given stimuli of individual melodic phrases–was made in order to ensure ecological validity in the data collection process. I support this approach, as in the opposite case the performance of individual melodic phrases would have been most probably disconnected from their foundation (the rāga system), which is essential in Hindustani singing. Additionally, it would have been harder to perform meaningful and consistent gestures within the short duration of 0.5-5sec as to correspond to typical movement events, similar to the ones finally recorded, identified and analysed here.

There were no remarks made by any of the musicians about the capturing of their movements that might imply that this might have affected their performance in practice. In fact, when asked, some musicians reported that wearing the reflective markers might have felt a bit awkward or uncomfortable at the beginning, but that they became familiarised and completely forgot about them quite quickly. Only maestro Zia Fariduddin Dagar felt limited by the elastic bands on which the markers were attached during the first recording and therefore I just had to loosen them for the following ones. There were also no comments by musicians about being affected by the setup, time of the day (in relation to the chosen rāga), place or dark conditions of the recordings. Indian musicians are used to being exposed to bigger or smaller audiences and/or being recorded in sound and video, and they are also used to all types of distraction.

Overall, I consider the effect of the peculiar recording setup on their performance to be only limited, perhaps mostly on the intensity rather than the way of moving or singing. Thus, despite the unusual (mocap) setup, I consider this choice to be the closest possible to a true ecological approach in making exact 3D-measurements of hand movements in Hindustani singing. Taking into account the busy schedule of the most senior musicians, the limited travelling opportunities of the most junior musicians outside India and the age and medical condition of others, it would have been difficult to collect this data outside India or through designed experiments. For instance, Maestro Zia Fariduddin Dagar reported once during my stay in Palaspe in 2010 that he only felt comfortable to perform in his own house and that he would have denied the idea of being recorded in a different space. However, the use of such data can be challenging (such as the lack of a common ground for comparing results between performers, such as rāga and length of the ālāp) and is discussed in more detail in 9.6.1.

3.4.4. Recording spaces

Recordings were carried out in four different places in India. It was not practical to install the motion capture equipment in a single fixed place under controlled laboratory conditions, but instead it had to be set up on each occasion (different musician and city) in spaces where conditions were partially uncontrollable (subject to power cuts, displacement of cameras according to domestic activities etc.). Calibration processes had to precede all recording sessions, even those scheduled in the last minute. Therefore, in order to ensure access to the equipment at all times, all recordings were conducted in domestic spaces, usually musicians' so-called music rooms, i.e. the living rooms in which they held musical activities. In fact, this was an efficient way to ensure ecological validity in the data collection process. Maestro Zia Fariduddin Dagar reported once during my stay in Palaspe in 2010 that he only felt comfortable to perform in his own house and he would have denied the idea of being recorded in a different space. In order for the equipment to remain in place for some days, the setup had to be adapted to the specificities of the room in such a way that it would interfere in the minimum possible way with the normal activities of the space.

Specifically, recordings were conducted in the following four places/spaces (as shown in Figure 3.5):

(1) Palaspe/Panvel, Zia Fariduddin Dagar's house/gurukul[66] (music room);

(2) Bhopal, the Gundecha brothers' gurukul (one of the school's student accommodation rooms);

(3) Delhi, Nirmalya Dey's gurukul (music room);

(4) Pune, Uday Bhawalkar's house/gurukul (music room).

(1) *(2)*

(3) *(4)*

Figure 3.5. Recording spaces in the cities of (1) Palaspe, (2) Bhopal, (3) Delhi and (4) Pune.

The recording of the Gundecha brothers by Clayton, Leante and McGuiness took place in Aikatan Auditorium, Salt Lake, in Kolkata, India in February 2007.

3.4.5. Data, equipment and setup

My own recordings include the following streams of data:

1. Audio stereo recording of the voice and the taňpurā (close-miking);
2. Audio stereo recording of the ambience;
3. Motion data capturing of the performer's movements;
4. Video recording focusing on the vocalist.

[66] Music school.

I used the following equipment for capturing the data:

- For 1: A TASCAM HD-P2 two-track recorder and a paired AKG C 480 B set of two microphones, recording at 196kHz, 24bit;
- For 2: A Sony PCM-D50 recorder with on-board microphones;
- For 3: A 10-camera (3.5'' optical angle, i.e. wide) high-speed infrared point-light system, namely Optitrack by Naturalpoint[67], recording at 100 fps and controlled through the Arena software environment;
- For 4: A Sony DCR-SR65 handy-cam recording in night shot (infrared) from a fixed position.

Figure 3.6 shows equipment setup in the music school of Zia Fariduddin Dagar in Palaspe, which represents a typical recording arrangement.

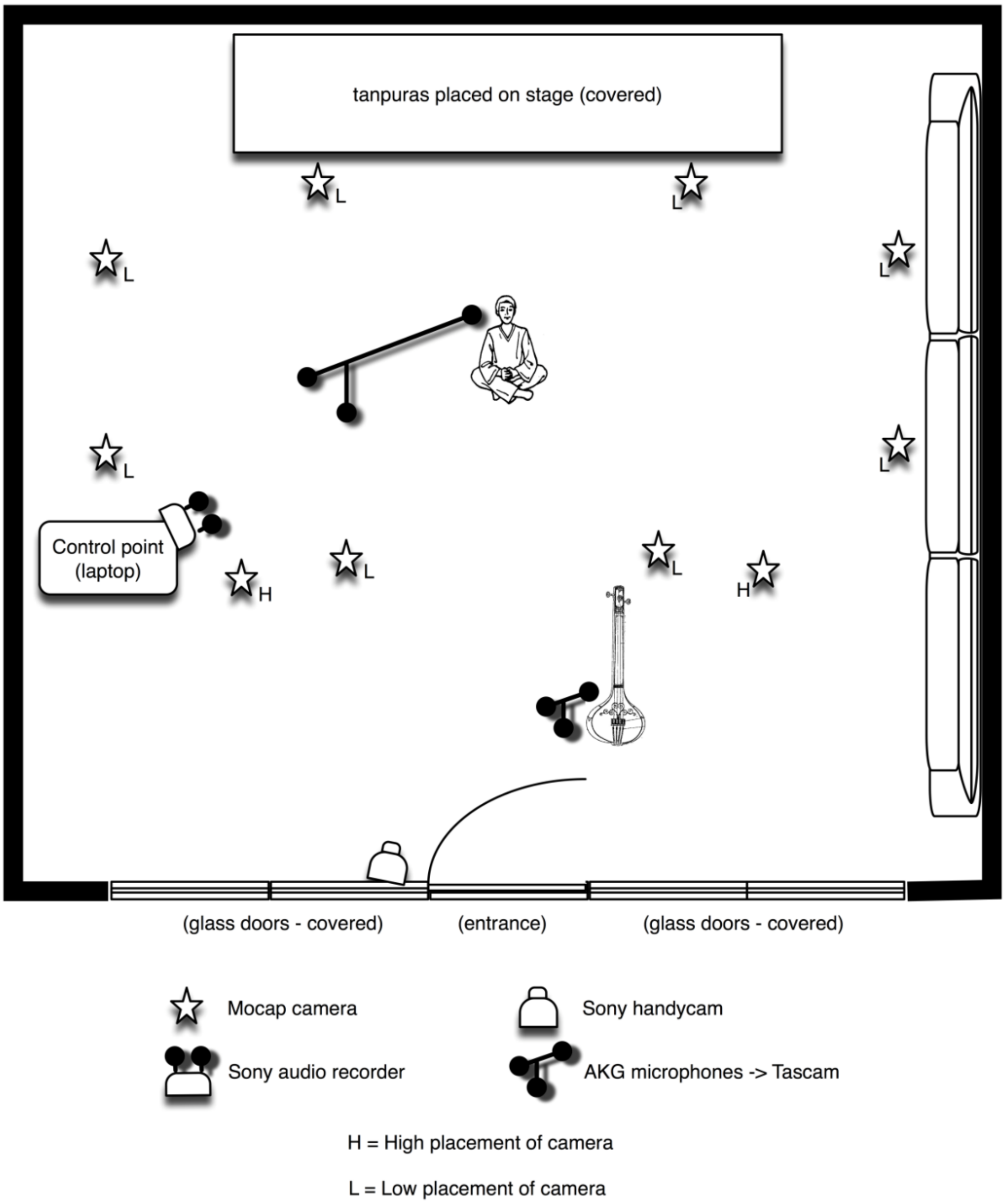


Figure 3.6. Recording setup in Palaspe, displaying a typical setup for the equipment.

[67] http://www.optitrack.com/

The setting includes the solo musician and the person playing the drone instrument (usually another student). Due to the requirements for avoiding occlusions in capturing the performer’s movement and audio leakage from the drone instrument into the voice, the taňpurā(s) had to be placed across the musician[68]. Despite the fact that this is not the typical performance setup, most of the time musicians reported that this didn't disturb them[69]. Due to the low lighting conditions (preferred in order to limit occlusions in the mocap system), the video was recorded in night-shot (infrared light). Both the voice and the taňpurā were recorded with a separate microphone, in order to remove the taňpurā leakage from the voice channel, while the ambience of the space was also recorded in audio for the same reason. A real taňpurā was replaced in a few cases by an electronic taňpurā box or an iphone. To avoid external noise (e.g. cooking, conversations, dog barks and traffic (especially in the Palaspe music school, which is placed on the Bombay-Goa highway)), several recordings took place in the very early morning (before sunrise) and in the late evening.

Despite the 3.5" wide camera lenses and a multitude of arrangements of the cameras (some at high and some at low positions), the limited size of the spaces meant that the effective capture volume (the common space seen by all 10 cameras) was quite limited and occasionally some additional reflections from other objects in the space would appear, that were not always possible to avoid. To limit these arbitrary reflections, recordings were carried out in darkness (blacking out sunlight and switching off lights), while a dim light enabled visual contact between musicians and the researcher. Setting up equipment in such spaces meant that large furniture had to be first removed, cameras and microphones had to be placed at points which wouldn't disturb the normal everyday use of the space, sunlight had to be blocked out and reflective surfaces had to be covered (any locally available material, such as heavy curtains, bed linen, newspapers was used).

The motion capture system was restarted every 10 minutes, in order to avoid possible instabilities of the software or unexpected power cuts, which would result in the loss of lengthy recordings. In order to be able to later (during analysis) align the motion, audio and video data streams (manually), I used a hand clap in the beginning of each of these 10 min sessions while wearing one reflective marker on each hand. These markers were otherwise hidden and not visible by the system. The gap between stopping and restarting the mocap recording was as long as required for saving the file and restarting the system, usually around 15 seconds.

Multiple reflective markers were placed at 13 points of the performer’s upper-part of the body, at points that were thought to offer a good representation of the bone and skeleton structure that was later recreated.

For each of the points of interest I created a unique custom-made geometrical shape by three reflective markers (for the head I used four), that was visually identified by the system due to its distinctive form. For each of these triads the mean 3D-position (x,y,z) was calculated through triangulation and was used as the position of the specific body part. A stick-figure of the skeleton was later attached to these points, as shown in Figure 3.7.

[68] Taňpurā/s is/are normally placed directly behind the singer.
[69] A single exception to this was the jugalbandī between Ustad Zia Fariduddin Dagar (singing) and his nephew Mohi Bahauddin Dagar (playing the rudrā vīṇā), when the first reported that he could not hear the taňpurās clearly.

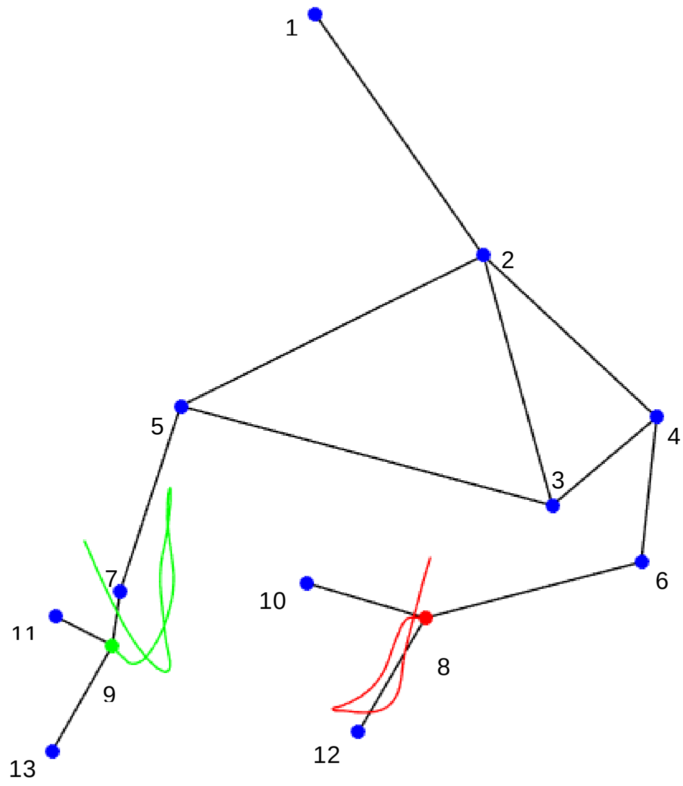


Figure 3.7. Stick figure from mocap data of seated Indian singer.

The following are the points of the body were rigid bodies (triads) of reflective markers were attached and tracked:

- 2 rigid points on the performer's back (low and high);
- 1 on the head;
- 2, one on each shoulder;
- 2, one on each elbow;
- 2, one on each wrist;
- 4, on two fingers of each hand.

The recording of the Gundecha brothers by Clayton, Leante and McGuiness included:

- Three Cameras: 2x PD170 and 1x HVR-A1E for video recording in HDV 1080i from different angles;
- An audio close-miking recording of 4 channels via live mixer to Korg D888 in 16bit and 44kHz (one on each of the brothers, one on Amita Sinha and two channels on each of the sides of the double-sided pakhāvaj drum).

3.4.6. Data extraction and pre-processing

(a) Interview material

All interview material was first fully transcribed and then multiple 'readings' of both the written and the visual material of the interviews followed, during which notes for describing the content were taken. Due to the exploratory nature of the work, this process allowed for some first ideas and patterns in the understanding of the phenomena to evolve, and gave rise to ways of organising the material into concepts and themes.

(b) Audio-visual material

For each piece of equipment (audio and video), the recording automatically moved on to further tracks after a set duration of time, which was different for each. Therefore, all data (two separate stereo audio, video and mocap) had to be manually aligned. All data streams were first merged, then they were manually aligned by aligning the (common) hand claps of every 10 minutes and finally they were manually segmented into chunks of 10 minutes. The need for chunking the data into 10 min segments is imposed by Matlab's limitations in processing large data files. Figure 3.8 shows the automatic track change for each piece of equipment (white) and hand claps of every 10 minutes (red).

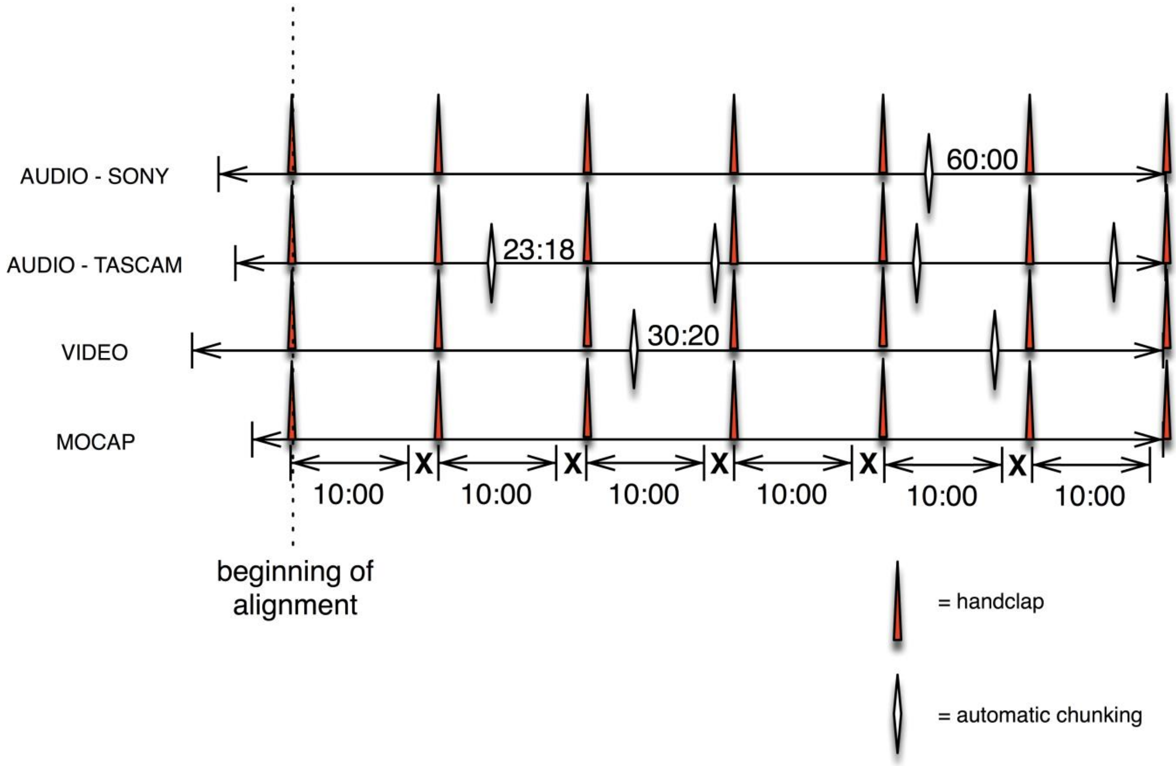


Figure 3.8. Manual alignment and segmentation of data streams.

Basic acoustic features were extracted from the raw audio file of the voice (TASCAM, 96kHz), primarily by using the MIR Toolbox in Matlab (Lartillot et al. 2008), but also Praat (for pitch).

(c) Movement data (motion capture)

Raw motion data are multi-dimensional streams of data in time that correspond to a unique position in the chosen 3D coordination system (typically Cartesian, (x,y,z)). I carried out the recording of the mocap data in the proprietary motion capture environment of the NaturalPoint Optical Tracking system that was used at the time, namely Arena. However, I encountered a problem while capturing mocap data in Arena. Despite the rigorous setup and calibration of the system and the careful selection of distinct geometrical shapes for creating 'rigid bodies', Arena failed at identifying each of these shapes as unique and therefore it constantly confused them to an extent that made the data unusable. More specifically, neighbouring markers from different rigid bodies occasionally formed a new geometrical shape that was confused with the ones that were initially defined in the system. These errors appeared as short and abrupt jumps of identified markers that changed numbering, especially when the hands of the performer moved closer to each other. Using the latest software environment by NaturalPoint, namely Motive, the issue of identifying the rigid bodies as distinct from each other was resolved. However, while migrating from the older Arena to the newer Motive software, there was an automatic change of frame rate from 100 to 120 fps, which was enforced on the data without changing the total file size (the number of samples). This resulted in a loss of about 20% of the mocap data from the last part of each of the files. Therefore, MIIO movement events that were annotated in this part of the data were not included in the quantitative analysis. Raw mocap data was exported from Motive into c3d format. Multiple salient movement features were then extracted from these low-level time-series by using the MoCap Toolbox (Burger & Toiviainen, 2013) of the Matlab environment. Recorded mocap data was first pre-processed (see Nymoen et al., 2013 for a good overview of methods) by:

1. Filling gaps: Due to occasional optical occlusion of the reflective markers in infrared marker-based mocap technology, mocap data streams often display gaps in the recordings. Short gaps were filled through linear interpolation between adjacent points in time.

2. Smoothing: Even in ideal capturing conditions, mocap systems that are based on identifying unique geometrical constructions of reflective markers (rigid bodies) can still be occasionally confused, especially with a large number of rigid bodies that move close to each other. Through low-pass filtering these were smoothed out. A Savitzky-Golay of a 90msec long window was used on the raw mocap data. Additionally, extracted kinematic features were also smoothed with a Savitzky-Golay FIR filter, by using a window of 90msec on features of position and its first derivative (velocity) and a window of 130msec on the second derivative of position (acceleration). The choice of the specific window sizes was made on the basis of minimal noise in the data streams after simple visual inspection, with the second derivative of position introducing a higher amount of noise and thus requiring a longer window than the first derivative and the position data.

3.5. Summary

This chapter has provided justification for various choices of approach in the research, along with a detailed description of the research design. It has reported on data collection methods, as well as on the way themes, concepts and ideas emerged from the data through thematic analysis.

Additionally, the chapter has outlined the basic principles of interpretive inquiry and the preconditions for credibility and generalisation of the findings, along with details of the ways in which these principles and conditions were met in this project. The qualitative approaches that I followed in the first part of the analysis for data collected from interviews and video observations were first presented. The quantitative approaches that I later followed for both (a) the qualitative (codes) and (b) the mocap and audio (quantitative) data were also described.

Chapter 4 – Interview Analysis

The current chapter presents an analysis of interviews with performing musicians, and explores the relationship between their singing, upper-body movements and imagery. The central questions that the interview analysis aimed to explore are:

- *In which cases are the body and its extension through imaginary objects used as a mediator between mind (melodic intention) and voice?*
- *What is being melodically grasped and manipulated when imaginary objects are being stretched, pulled, pushed and released?*
- *Does the facilitating role of movements—and MIIOs in particular— reported by vocalists refer to the cognitive process of melodic activity or to the mechanical process of voice production?*

To address these questions, the analysis was constructed in two parts. The first part focused on information cues which are related in some way to the interaction of performers with imagined objects. As this table lacked transparency in how various codes and themes relate to each other, the second part focused on pertinent issues of how movement relates to the voice. Although the two parts of the interview analysis were conducted in a sequence, in practice they constitute two independent processes that contributed to informing the video analysis that followed.

More specifically, in the first part thematic analysis (Boyatzis, 1998) was applied to sensorial descriptors (adjectives, verbs and nouns) of motor-based metaphors or pictorial terms (Sanyal & Widdess, 2004: 126) of performer-object interactions that are used recurrently when talking about sound and music. These were extracted, coded and organised 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.

The second part 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. It refers to both explicit articulations of conscious choices made by the musicians, as well as implicit information embedded in descriptions of melodic processes. This part includes discussions on the implication of multi-modal phenomena in both the cognitive construction of meaning as well as the mechanical production of sound is also discussed. Issues of movement–sound relationships raised by participants in terms of timing, trajectories and other morphological features are also included, as well as statements regarding the awareness of the impact that gestural inheritance from teachers and imitation from instrumental gestures has on their own gesturing habits. Most importantly, this part concentrates 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 may be facilitating engagement with melodic ideas.

All interview videos are available at the following links:

Interview media file playlist		
https://n2t.durham.ac.uk/ark:/32150/r1tx31qh68k		
4.1.	**Afzal Hussain, 10.01.11**	https://n2t.durham.ac.uk/ark:/32150/r1fx719m45d
4.2.	**Uday Bhawalkar, 04.08.10**	https://n2t.durham.ac.uk/ark:/32150/r1ft848q62z
4.3.	**Lakhan Lal Sahu, 16.01.11**	https://n2t.durham.ac.uk/ark:/32150/r1v118rd54x
4.4.	**Bahauddin Dagar, 28.09.10**	https://n2t.durham.ac.uk/ark:/32150/r1j098zb10h
4.5.	**Umakant Gundecha, 16.01.11**	https://n2t.durham.ac.uk/ark:/32150/r1m613mx572
4.6.	**Marianne Svašek, 27.09.10**	https://n2t.durham.ac.uk/ark:/32150/r13r074t931
4.7.	**Pelva Naik & Bahauddin Dagar, 21.12.16**	https://n2t.durham.ac.uk/ark:/32150/r22r36tx52c

Table 4.1. Interview media files

4.1. Themes, cues and codes for imaginary objects and interactions

In this part of the chapter, I first include a detailed audit trail of the way thematic analysis was applied to both the explicit and implicit knowledge embedded in musical discourse. All interview material was first transcribed and then printed out. The transcript was read a number of times and the left-hand margin was used to annotate what seemed to be relevant and significant in respect to the questions that were posed initially. The transcript was then read again from the beginning, and the other margin was used to document emerging theme titles.

Then, samples of words related in any possible way to music embodiment, such as motor-based metaphors and descriptors of imagined objects and interactions, were collected and organised in various ways, in an effort to classify them effectively. Using an iterative process, the transcript was read a number of times and with each iteration the initial notes were gradually transformed into themes that were descriptive of the data at a higher level of abstraction. Theme titles emerged from the repetition of converging themes in the annotation material. Initially, these were listed chronologically, but then in an attempt to make better sense, they were re-arranged and clustered in a more analytic and higher level ordering. There was constant close interaction between the transcript and these codes, in order to ensure that the connection to the primary source material was retained. Finally—and for reasons of transparency—two tables of organised themes were produced in order to display the complex material in a most compact way.

The first table (Table 4.2) includes a listing of all keywords that fit to these themes and it reflects the initial organisation of the annotated material into nouns, verbs and adjectives of music embodiment. Due to the tight entanglement that was revealed in the first part of the analysis between the cognitive, visual, corporeal and auditory aspects of music making, there was no sense in organising the material into sensory related domains, so these codes are simply listed in the table without mentioning their source modality. These codes are arranged and presented here in classes of (a) objects/materials/body parts (anything that is been referred to as being involved in the actual interaction or its corresponding image), (b) types of interactions expressed through verbs, (c) movement descriptors (physical measures or other expressions referring to physical phenomena) and (d) movement qualities (as adjectives, expressing the way a movement may be performed).

Objects/materials/body parts	**Interaction types**	**Movement descriptors**	**Movement qualities**
Spring or rubber	Move (body/hands)	Elasticity	Dynamic
Ball	Dance	Pressure	Intensification
Water	Stretch (elastic band)	Weight (contra-)	Heavy/Light
Ocean waves	Apply pressure	Balance	Tight/Free
Liquid	Intensify	Friction	Fast/Slow
Worm	Scratch	Time, timing	Slow /Quickly
Vīṇā	Lift (sound)	Speed	Restless/Fixed
Strings	Press (sound)	Load (heavy)	Delicate

Body Hands Stomach Breath Suchi mudra Tai-chi	Pull back Pull (sound) Push / push away Throw Catch Take (one note to another) Take out (sound) Keep (inside) Stay Hold (something, sound) Open (hands) Turn Go (with the sound) Go, come Get back Go away Leave Flow through Slide Loose balance Point Precede Stop Support (breath, voice, control of weight and time) Control (breath, volume of sound, weight, timing) Restrict (timing) Facilitate voice Measure (weight of phrase) Enter (my body) (never) get stuck Expand (my stomach) (stomach) contract Show (distance of note)	Distance (between hands/notes) Vibration Tension – restriction Initiation – growth – release Prefix-suffix Pattern Shape (of phrase) Path/trace Curve Space / big-bang	Before/After Open/Closed (voice/hands) Up/Down Low/High Away/Backward Inside Jerky

	Describe (by hands) Slide (up) Visualise See Watch Focus Control volume (of sound) Change (sound) Oscillate (sound) Cut (phrase)		

Table 4.2. Discourse classification-1.

The first column intends to reveal the kinds of objects and materials that might be imagined by musicians in association with the melody and are good indicators for the types of interaction one might observe during performance. The second column offers a list of verbs or actions related to types of interaction that these imagined objects may afford. Some words are labels of manual types of gestures (press, pull, throw, lift, catch, take), others refer to the use of space (the 'where'), while some refer to the functionality of the interaction (control, support, restriction, measurement). Going through this list, it is almost inevitable to further identify sub-themes; these are addressed more precisely by the second table (Table 4.3). The functionality part of this list is an interesting aspect that may offer insight into why musicians may make such spontaneous associations to physical interactions. The third column reveals some of the physical forces that might be conceptualised in association with manual interactions. Finally, the fourth column refers to the fundamental aspect of movement quality. The words in this list refer to the way gestures are performed, which makes them distinct from others. The most important to this thesis are the light/heavy descriptors that refer to the intensification of the action and therefore also the exerted effort. Based on these reports, I expected such qualities to be imprinted on the dynamic and kinematic profiles of MIIOs that I later analysed through video recordings.

In the second table (Table 4.3), the codes are re-organised in order to further reduce the complexity of the themes by classifying the material into types of interaction (both metaphors and iconic representations of movement–sound relationships), as well as questions about the 'what', 'where' and 'how'. Although this classification emerged from the actual data, it somehow converged on the way with the IMS system by Hackney (2002), which is an extension of Laban's BESS system (see section 2.3.2. for details). The new table captures the themes, sub-themes and classified codes, but it should not be regarded as a complete reference. It is presented here in support of the coding scheme that was used in the annotation and analysis of the performances which followed.

BODY (what?)	SPACE (where?)	EFFORT (how?)				SHAPE	RELATIONSHIP	PHRASING	Ambiguous & other
		Weight	Time	Flow	Space				
Body: Hands Stomach Breath **Objects (extensions):** Spring or rubber Elastic band Ball Water Ocean waves Liquid Worm Vīṇā Strings	**Descriptors:** Away/Backward Inside (Push) away (Take) out (Pull) back Forward Up/Down Low/High **Verbs:** Go away/come Leave Enter (my body) Loose balance Stay Point	**Descriptors:** Heavy/Light Delicate Tension – restriction – release Dynamic **Verbs:** Intensify	**Descriptors:** Slow/Quickly/ Fast Speed Before/After	**Descriptors:** Restless/Fixed Tight/Free **Verbs:** Get stuck Flow through	Jerky	Pattern Path/trace Curve Open/Closed	**Verbs:** Stretch Scratch Press Expand/contract (stomach) Apply pressure Lift Pull Push Throw Hold Slide Catch Take Keep Turn Get back **Physical Forces and Measures:**	Initiation – growth – release Prefix-suffix Cut (phrase)	Support (breath, voice) Control (breath, weight, time, volume of sound) Restrict (timing) Measure (weight of phrase) Balance Vibration Show (distance of note) Describe (by hands) Dhrupad = Visual art Visualise See Watch Control volume (of sound) Change (sound) Oscillate (sound)

							Elasticity Pressure Weight (or load) Friction Time/timing Speed Distance (between hands/notes)		Facilitate the voice

Table 4.3. Discourse classification-2.

The first column lists types of objects and materials that might be imagined by musicians in association with the melody. The list indicates that some of these objects are thought of as an extension to the performer's body (rubber, ball, water, vīṇā strings), while others are internalised conceptions of materials (hands, stomach). In the external objects we mainly find cases which are geometrically defined and therefore graspable. From these 'external objects', I consider water as a special case, as one can interact with a liquid in an effortful way but without being able to grasp it. This idea has implications on the annotation of the video material, meaning that the formation of a grip might not always be an exclusive indicator of a MIIO. Internal materials on the other hand—although geometrically less defined—can be thought of as affording similar modes of interaction to externally employed objects; one can internally compress or stretch a muscle without grasping it. However, it might be more difficult to visually detect a covert interaction than an overt manual interaction, and it is for this reason that the current project relies on identifiable MIIOs.

The second column includes descriptors related to the use of space (that answer to the question of 'where'). It is striking that space is described predominantly in reference to the performer's body (moving away vs. approaching the body) rather than in absolute three-dimensional coordinates. Based on these labels, I later expected to observe a large variety of trajectories in space for the same type of gesture during video analysis, most of which performed away from or towards the vocalists' trunk.

The third column refers to the movement qualities (in other words the 'how' of a movement) and these are further divided into the four motion factors of effort by Laban, namely: weight, time, flow and space. It is remarkable that despite the fact that effort is understood in this thesis as the power of the action (which is closer to Laban's weight factor), during the video annotation process I was made aware that in fact I was indeed taking into account all of these movement aspects in assessing the level of effort exerted by the performer.

The fourth column refers to the use of space in terms of 'how' a point in space might be accessed. In other words, the labels of this column describe gestures as trajectories or shapes, which are essential in music performance (Küssner, 2014) and are especially relevant in Hindustani music due to the strong conceptualisation of pitch as a kind of space by Hindustani musicians (see section 2.1.5.).

The fifth column refers to the relationship between the gestures and the manipulated objects, in terms of types of action (such as press, pull, throw, lift, catch, take) or physical forces (elasticity, friction etc.) and measures that might be conceptualised in association with the melody. The nature of the forces alludes to the types of manual interactions and offered a good source of information for guiding the identification of MIIOs that was conducted in the later stages of video analysis.

The sixth column is added to the table to include a list of ambiguous keywords, which represent essential aspects of the interview, but which cannot be classified clearly under any of the categories.

4.2. Discussion on interview material

The discussion of the material is organised in five different topics that include general observations on motor imagery, self-reports on its role in music, explicit and metaphorical gesture–voice relationships, gesture inheritance from teachers and explicit and tacit information about imaginary objects and their interactions.

4.2.1. General observations on visual imagery and motor-activity in musical discourse

The visual element and bodily engagement with melodic activity prevails in the discussions with all participants, and movement is described as a concept of primary importance.

(a) Movement & music

Explicit statements by participants emphasise a strong conception by musicians of music as movement. This is what the Dutch Dhrupad singer Marianne Svašek expressed too during an interview in Rotterdam:

> 'I do a lot of thinking about movements.
> [...] For me music is not notes, it is movements.
> [...] So, you never get stuck at one point, but you go somewhere to go away from it again'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

The awareness of the importance of movement in the conceptualisation of music and sound is paramount in this expression. Restlessness in sound is expressed here through restlessness in moving in the physical space. The vocalist is not dealing just with notes, but also 'places travelled by the hands' (Neuman, 2004: 221).

It is interesting to note that musicians do not compare music to movement by saying 'music is *like* movement', but they provide a statement of equality with no prevalent hierarchy between the auditory modality and the concept of gesturing. To stress this even more, the opposite to the previous quotation was also stated by Dhrupad vocalist Uday Bhawalkar during an interview in London:

> 'How a beautiful lady will put kajal, the timing of that, the dot on the forehead, taking the veil, *moving* is music'.
> (Uday Bhawalkar, interview, London, UK, July 27, 2010)

Here the restlessness of music is expressed through visual imagery of non-auditory content, in which the aspect of time—of something always changing—rather than static pictures, is prevalent. The same with another statement by Marianne Svašek in conceiving notes as something that never stops moving:

> 'Also a note moves, everything moves, there is nothing static in sound. It shouldn't'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

Even in expressions that do not justify the contribution of the body in moving, there exists a primordial conception of music as something restless, such as in the following example by Dhrupad instrumentalist Mohi Bahauddin Dagar, son of the renowned rudrā vīṇā player Zia Mohiuddin Dagar and nephew of vocalist Zia Fariduddin Dagar:

> 'There must be constantly a dynamic aspect to the sound,
> something restless that will move all the time'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

(b) Visual imagery

Participants explicitly testify the importance of visualisation in their conception of music through concrete statements, such as in the following excerpt from an interview with rudrā vīṇā player Mohi Bahauddin Dagar in Rotterdam:

> 'I believe that Dhrupad is primarily a visual art - its' not just about the sound - you see it'!
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

Images related to nature, colours, people, deities and various emotions are reported by several participants, such as the following excerpts by Dhrupad vocalists Lakhan Lal Sahu and Uday Bhawalkar:

> 'I visualise light and colours of rāga'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

> 'I visualise things, like nature, people, life, human feelings, love, anger, passions, work,
> moving body. [...] Perhaps I am moving my hand, but I am thinking of a person, rain etc.
> Maybe I can see the waves of the sea, that time I am not thinking of a note,
> I see something and it helps me, it has to be like that.
> [...] Because it has to be inside, I have to see that, I have to visualise it,
> it has to enter into my body'.
> (Uday Bhawalkar, interview, London, UK, July 27, 2010)

In the last report it is striking to notice, how tightly intertwined the domains of visualisation, imagery, sound and body movement appear in a single statement. The visual element is not reported just as a process that is limited to the eyes, but as a concept that is embodied ('it has to enter my body').

Additionally, instructions by teachers are not limited to guiding students in ways by which they should use their voice (and move their hands or upper-body, as we shall see in section 4.2.4.), but they also extend to metaphorical uses of visual representations (such as 'birds flying').

> 'Ustad[71] sometimes used [...] other examples too, like birds flying in the sky.
> Sunset, sunrise, how the sun moves. Clouds passing in front of the moon.
> Also the way you may expose your love to your lover, with love, with fighting, with anger.
> He gives always new examples'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

[71] The master/teacher/guru, Zia Fariduddin Dagar in this case. The term 'ustad' is used for Muslims, while the term 'pandit' is used for Hindus.

These images are used by teachers either as metaphors for clarifying what students need to do with their voice, such as how to precisely stress and punctuate melodic phrases, or as a support to the techniques of sound production (Fatone et al., 2011). This is exactly the process in which movement and metaphorical or pictorial terms (Sanyal & Widdess, 2004: 126) can work as a mediator for enhancing the understanding of the complexities involved in a melodic phrase.

Apart from pictorial images such as the previous, there are also plenty of reports about melody being conceived and visualised as curves and dynamic patterns:

> 'I can see the notes [he draws a trajectory with one hand in the air].
> When it [the sound] leaves and arrives to another thing, [...] there is a certain curve to it or a weight to it. I can see *that*' [he moves his hand drawing a parabolic shape in space].
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

It is interesting to point out here, that although there is an explicit reference to 'notes', attention is placed on the design of a curve, which passes smoothly through all graduations of the participating notes, starting at the first pitch and finishing on the last one. Thus, the hands do not engage with the melody by provisionally pointing to discrete notes one-by-one, but instead by tracing the melodic movement in terms of a curve. As discussed in section 2.1.5. , improvised melody is conceived in two complementary ways by Hindustani musicians: as a succession of discrete notes and, at the same time, as coherent shapes. Even more, within the same statement the image of a curve is combined with the physicality of a gesture that causes this (sonic) curve to be created, namely weight. Again, the visual and the embodied elements are here combined into a single entity that attempts to describe a melodic feature. The physical metaphor of 'weight' for expressing some quality in the auditory domain[72] was encountered in many instances throughout the interview with Mohi Bahauddin Dagar, as well as Marianne Svašek. This term, as well as other sensorimotor descriptors which are associated with the idea of resistance or some force opposing an imagined (as well as performed) action are further discussed later in this section, as an attempt to interpret their linguistic and cognitive functionality in musical discourse.

Another statement that underlines the importance of visualisation and movement in music cognition is the following:

> 'Whatever you are doing, you are trying to purely watch one image. I have to project this image of this particular rāga. It is like whatever lines you are trying to draw on a canvas. You are trying to make one particular picture. For the ālāp it won't be a particular, but an abstract picture. [...] So for me... Until there is some visualisation... you cannot improvise... What are you going to improvise about? Even abstractness at any point is a visualisation... a whole pattern [he moves his hand like painting in the air]. You keep on doing the strokes again and again and when you do that, at the end of it the painting is ready. But you need to keep going with them... When you see a painter moving...

[72] It is not exactly obvious to which sonic aspect(s) the term 'weight' might be referring.

> There is a gesture that is following'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

The musician refers initially to the process of unfolding the rāga development during improvisation through successive phrases, which he conceives as strokes of a painter. In this interview extra emphasis is placed on the fact that these strokes are not imagined as static, fixed shapes of a completed painting, but as end-products of a dynamic gesture that unfolds over time, similarly to the dynamic designs of melodic phrases. Mohi Bahauddin Dagar continues elaborating on the same thought by stating:

> 'What is the weight of the gesture and how much you want [he makes a large gesture]... it is a very complicated matter, because... it is very difficult in a sense, because your eyes decide in a certain moment that you want a particular effect, but your mind lets go so much of it.
> So, it is all controlled, maybe from here [he points at the diaphragm],
> from here [he points at the heart] and from here [he points at the head]'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

This is an explicit report on how tightly music cognition (the mind) is entangled with the visual element (associated with the function of eye-sight), sensorimotor experience (the sense of weight) and the real, tangible cause of voice production. The activity of 'control' here does not refer explicitly to the aural domain of vocalisation, but rather it describes a cross-modal process. Despite the fact that vocalisation is obviously fundamental in the sense that a vocalist is supposed to sing, this statement does not explicitly qualify any of the implicated modalities as the target domain. The imagined, performed or conceived action of fighting against or giving in to some opposing force (here 'weight') has a central role in the conception of melodic movement.

Dagar will later continue elaborating on this same thought by adding:

> Because that weight won't come if you don't do that action. The weight gets displaced in the sound. I have to *support*! [he moves his hand twice: once upwards through a quick gesture with an open palm while the voice sounds "lighter" and once through a slow, horizontally movement away from the body and with the palm kept closed while the voice sounds "heavier" or more supported].
> (Mohi Bahauddin Dagar, ibid.)

During the same interview, Mohi Bahauddin Dagar also acknowledges quite clearly the concept of the 'growth point' (see section 2.1.4. for a discussion on this) in Hindustani musical thinking as an important initial process or musical idea (a 'big-bang') which is progressively enriched and expanded by shaping it and adding details. This initial process is given so much emphasis, that a wrong start is expected to lead to a wrong progression which cannot be later corrected. Here is how Mohi Bahauddin Dagar articulates these ideas:

> 'When you start singing, you create a space - it's like big-bang. You initiate the phrase and then take it and shape it and add little details… embellishments, svaras etc.

> If you start wrong, the whole piece will go wrong and you won't be able to correct it later on'.
> (Mohi Bahauddin Dagar, ibid.)

This section attested that movement is not an exclusivity of the physical world, but instead that it is an inherent and fundamental concept of human cognition. According to the interview material that was discussed here, sound, movement and imagery exhibit such close bonds that any method of analysis would need to approach musical thinking in a holistic way rather than isolate a specific modality. The next section offers a more concrete discussion that focuses on how musicians acknowledge the role of visual and motor imagery in the understanding and conception of music.

4.2.2. Self-reflections by musicians on the role of motor imagery in the conception of music

The importance of gesturing in singing is often attested by Dhrupad musicians, such as in the following excerpt from an interview with vocalist Lakhan Lal Sahu:

> 'And these hand movements are very important in Dhrupad gayan[73].
> According to hands, the sound is coming.
> And if I will sit like this without movement [he shows it by sitting still],
> the sound will not come'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

The quotation stresses the awareness of a strong relationship between the hands and the voice as a kind of match between modalities. It also suggests a clear hierarchy and directionality in the cross-modal relationship, with the sound being the target domain that is controlled or directed by the hands. In other words, it means that the voice is determined by the hands, which logically also means that no sound is produced in the absence of hand movements. Such a report attests to the primordial role of hand movements in 'manual cognition', a recent theoretical development inspired by work in linguistics and verbal communication (Goldin-Meadow, 2003, 2006; Rizzolatti & Arbib, 1998). It gives an almost instrumental functionality to singing movements, as if the singing voice is directly driven by gestures.

At a second stage, the same musician attempts to become more specific about this cross-domain relationship, replacing the word 'control' by the word 'support'.

> 'With hand movements I feel much support; without, the sound won't come.
> When people [move] this left hand down and right hand up,
> then we are expanding the sound. It gives support'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

Initially, the musician uses the term 'support' without clarifying whether it refers to a cognitive and/or mechanical functionality. Later, he moves his attention to the effect of the gesture, i.e. the sound. However, 'expansion' in relation to sound cannot be easily explained as a term that refers to the

[73] Singing.

immaterial entity of a wave that is created by vibrations in the body and transmitted in an elastic medium. The hands can by no means expand the sound waves, but what is described as expansion here may refer to the iconic representation of the expansion of the rib cage, which results in the diaphragm to be lowered, tensed and pulled flat and the lungs to be pulled and stretched downward (in which case a kind of vacuum is created and air pushes into the lungs). Thus, 'expansion' could be referring instead to the iconicity of a mechanical functionality. 'Support' may be hinting to the physical support provided by the body trunk, around which this expansion of the ribs is mainly placed. Or it could also be associated with the notion of louder sounds reaching larger distances away from a sound source and the wavefront expanding with the increase of the distance, so it could be also possibly associated with the parameter of loudness in the auditory domain.

The term 'support' could also take on a cognitive rather than real, mechanical functionality, according to the following statement by Dhrupad instrumentalist Mohi Bahauddin Dagar:

> 'The movements do not control the voice, but they support it.
> They may intensify a feeling or point to something, to him, to yourself.
> There is also a facilitating reason and function of gestures [he moves while singing a higher vs. lower pitch]; you cannot do it in the opposite way'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

In the same lines, the next excerpt by Marianne Svašek seems to refer again to a cognitive facilitating process. The clarification emanates from Svašek posing a question that she immediately attempts to answer.

> 'But what if you are born without arms? [she moves a bit her body as if without arms]
> The hand can help you focus. So, it helps.
> Because you visualise it or feel it or something'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

Following her path of thinking, the support functionality seems to be a matter of focusing on sound or imagery, which is again strongly interwoven with the domain of vision ('because you visualise it') and/or kinesthesis and embodiment ('or feel it or something').

This section was intended to expose the extent of musicians' awareness of their own movements during performance. The interview material reveals that musicians conceive the relationship between their hands and their voice in two different ways at the same time:
(a) Through a causal relationship, where the hands control or affect the voice;
(b) As two parallel processes, in which the hands play a facilitating role of supporting the voice, either mechanically or cognitively.

This finding justifies why it might be difficult to classify MIIOs as movements mimicking sound-producing gestures (causal relationship to sound) or tracing acoustic features (the cognitive facilitating aspect of movement in music).

Despite the different views, what is clearly acknowledged by Dhrupad musicians is that hand movements are implicated in vocal performance and this finding encourages a further examination of gesture–voice relationships.

4.2.3. Relationship between vocalisation and hand movements

In terms of synchronicity between the voice and the hand movements (which has implications in decisions for segmenting movement and audio event units), the following statement by Mohi Bahauddin Dagar during an interview in the Netherlands affirms that in (vocal) performance gesture will usually come before sound (starts and finishes before sound), which also reveals that musical thought exists and is manifested before its actual auditory counterpart:

> 'Always the gesture will come... even for a fraction of a second, it will come first'.

Or at another moment during the interview:

> 'Usually, the gestures precede the sound and they may stop before the sound stops.
> As an example, the small finger designates the start of an embellishment;
> like what happens on the string of the vīṇā'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

However, Dagar adds a striking comment on the fact that 'non-symbolic' or 'fixed' gestures[74] do not have any strict boundaries, but represent only one part of musical thinking:

> 'Because a [melodic] phrase is never ending. I feel, the gestures are not complete.
> They create a path/trace and then they leave the imagination grow;
> unlike symbolic gestures, which are fixed and cannot change'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

During a long interview with Marianne Svašek in Rotterdam in 2010, interesting issues about the complexity of multiple-mapping strategies between voice and movement were gradually revealed. A part of this interview is presented here in the exact order of the original material, as a way to trace the unfolding of her thought while trying to elaborate on an explicit description of musical–kinetic associations. Small chunks of this excerpt are commented individually.

> 'What I always feel is that you have these two poles [she moves her hand up and down in front of her body to trace its central vertical line]. So, it is like this, it is physically also and you say you have to earth yourself... It is true in the sense, that when you sing low, you think high. And when you sing high, I think low. And it is also with movements'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

Western musicians will often report of the need to think of low pitches when they need to produce a high pitch. This is mostly described as a cognitive and psychologically facilitating mechanism of avoiding the stress and fear of producing a pitch in the highest pitches of one's vocal range. Marianne Svašek is a Dutch disciple of Zia Fariduddin Dagar, so whether this is a Western influence on her

---

[74] The way the term 'symbolic' is used by the performer needs to be considered with caution.

musical thinking or something that she is encouraged to do by her Indian teacher is not clear. Interestingly, directionality in thought is also brought in the physical space by the hands.

Nevertheless, right after this statement, Marianne Svašek tries to further elaborate on the association of voice and movement and adds the following:

> 'So, I always tell students your sound has to be as if your hand is under water, there is a heavy load on both sides [she moves one hand horizontally, left and right].
> So the movement can also help in the sense that if e.g. you make a slide up,
> you give the contra-weight. Otherwise, it sounds lighter'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

The role of water in motor imagery is striking in this quotation and has been previously used by Zia Fariduddin Dagar during group teaching sessions I had in Greece in 2007. Movement is not performed loosely according to this quotation; it should be bound due to the awareness of some resistance that hinders the hand from moving freely and is described as a 'contra-weight'. This means that weight is acting against the movement. The concept of weight is mentioned by other musicians too in reference to sound and will be further discussed in section 4.2.5. In physics the above expression would refer to a vertical movement in ascending direction, where the hand needs to defy the opposing gravitational force. But the prevalent resistance to the hand under water is not the gravitational but the viscous resistance of the liquid, which does not have any directionality preference. Suddenly, physical terms are merged under a common significant concept; resistance, which can be quite indifferently related to either gravity or viscosity. And what this should bring in the voice is a sound which should not be rendered in a 'light' fashion. What 'light' means in musical terms is not clear here and it could be referring to a number of different musical attributes. However, it is definitely an interesting metaphor for explaining a musical feature through a descriptor coming from the physical domain. I would even suggest here that the idea of 'weight', or something constantly resisting and not allowing the intended movement to be performed easily by the performer, is an inherent aspect of the Dhrupad music genre and could be approached as a metaphor.

According to this quotation, directionality of the hands in space is somehow blurred in the attempt to self-reflect on these phenomena. But what is expressed with confidence is the important role of movement in facilitating vocalisation (whether mechanically or cognitively is again not clear), which students are in fact encouraged to follow. The implication of gesture inheritance and education of students is an aspect that is discussed in section 4.2.4. Further developing her thoughts on the relationship between voice and hand movements, Marianne Svašek gives a few more examples and explains:

> 'Sometimes it helps that you sing high but move your hands lower, other times that you go before or after it moves [she experiments by moving her hands in two ways while singing an ascending pitch glide: once as if pulling something towards her body which resists to her movement while producing a "tighter" voice, and once moving her hand upwards and without any resistance while producing a "lighter" voice]. I loose… you hear? The sound becomes less. It's a contra. And then the hand can also do… [She experiments on the same pitch glide

> by moving her hands on the horizontal axis, giving the impression that she has a heavy load (as she described it before) on both sides of the hand while making her voice sound "tight"]. It is a slow motion thing, but it is just before or after... I don't know... [she starts exploring a series of similar ups and downs]. But what I really tell students is also if you go… what helps is if you go up [the hands] while you go down [the voice], it helps for the voice production'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

Here Svašek insists on this awareness of a resistance on the hands and tries to describe how it may have an effect on sound (described as something which should not sound 'less'). It is important that the vertical axis is asymmetrical, comprising two poles, with downward movements producing 'heavier' sounds and upward movements producing 'lighter' sounds. Later she introduces a similar asymmetry in the horizontal axis (for left vs. right).

Finally, it seems that it is not so much about the vertical or horizontal axis, but any direction in space that is asymmetrical and comprises two poles for displaying resistance felt on the hands. Movements of away vs. closer or left vs. right that are coupled with a sensation of increasing vs. decreasing resistance can illustrate the difference between pitch ascent vs. descent and at the same time the voice becoming 'heavier' vs. 'lighter'. However, Svašek finally goes back to the idea of verticality, according to which moving the hand downwards while making an ascending pitch glide should produce a more 'robust' (or perhaps supported) voice than moving the hand upwards, which makes the voice sound 'lighter' and less supported. It is not clear at this point whether this imagined opposing force affects only the timbre of the voice or if it also changes the timing in approaching the highest pitch of the ascending glide. However, it is clear that pitch height and physical vertical position are best matched in opposite directions. After exploring a few more gesture–sound relationships by moving the hands first horizontally (to the left) and then vertically (upwards) while singing, she adds:

> 'Actually this [she shows how the hand is moving up while the pitch is descending] feels better. Finally, there is no up, down, or sides or this any more... There is *movement*! You should be like in a big sea, you don't know any more about up/down, here/there, in scale nor in moving, but it moves'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

In this excerpt Marianne Svašek indicates that she does not cope with space through its objective Cartesian coordinates, nor does she have any particular sense of an objective orientation. Most fundamental is the notion of reaching at or leaving from a point in space regardless of direction, as long as there is a movement relative to this reference point and the movement is not happening freely, but in the presence of some resistance (here visualised as an elastic medium with a higher density than what we are normally used at in the air).

In a sense, this last description contradicts the first excerpt from the same interview in terms of directionality in physical space and reveals phenomena of multiple mapping strategies (Eitan & Granot, 2006). It suggests that there is no unique and solidified way of associating modalities; while a pitch ascent may be associated with the hand moving downwards, it could also be performed with a side or lateral gesture. Additionally, no bi-lateral symmetry exists (definitely not for vertical movements, but presumably not even for the horizontal ones) and the importance is placed here on something restless (a constant change), something that is forced to change despite the applied resistance.

This engagement with sound by acting against some opposing force is also central in the case of Mohi Bahauddin Dagar:

> 'The gesture supports this control of weight and time.
> [...] The important thing is the *goal*. And what you do with the movements is to constantly measure the weight of the phrase leading to the end in the suitable [presumably for the rāga] timing. It controls the timing after all. If it takes more or less time, the mood of the rāga will be ruined. The gesture points to the end of the phrase.
> [...] How much weight has to be put to displace that thing is a matter of gesture;
> a bit like doing tai-chi'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

In contrast to the previous excerpt from the interview with the same musician, this time the term 'weight' refers to the melodic counterpart of the movement. Once again here the physical metaphor of 'weight' is employed for some acoustic quality. A combination of weight and time in moving the hands is reported as the relationship to sound. Thus, the gesture–sound relationship is not simply manifested by the hand's position in physical space producing trajectories and shapes to represent melodic designs, but it is also imprinted in the dynamics of the gesture as a temporally evolving entity from an initial state to a final. There is a raised awareness of the final 'state' (the 'goal') and the combination of 'weight' and time reflects the melodic intention of the performer.

It is interesting that Mohi Bahauddin Dagar is conscious of the fact that a goal point also designates the boundary to a consecutive phrase, thereby acknowledging the complexity imposed by phenomena of coarticulation:

> 'The goal isn't the end. It designates the start of the next phrase.
> There is a kind of prefix-suffix'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

In the next excerpt, Dagar details this through some concrete examples. He gestures while singing variations of a double pitch glide, which always starts from the 2th degree, moves up to the 5th and then returns to the 4th:

> 'Here I am showing the timing of the Ma [4th degree. He gestures with the second finger]. This is called suchi mudra[75]. So I am taking the Ma to the Pa [5th degree] and then getting it back. I make a variation [he gives a few different examples, where the ascending movement, the highest note and the descending movement differ in time, although the entire duration remains the same].
> All these phrases are timed in the same meter, but they are cut at different phases and times. It is the same thing in ten different ways, looking into it like this [he tilts his hand several times

[75] Mudra is a symbolic gesture employed in the iconography of spiritual practice in Indian rituals, religions and dance.

in different angles]. You can look at it directly or from the side.
But here there is an advantage, because all my phrases have a Pañcam [5th degree] in them.
I am moving between Re [2nd degree] and Pa; those are my parameters.
And I am trying to show the various colours of Ma and Pa with various gestures'.
(Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

In all cases demonstrated, the goal is to rise to the 4th degree by first passing through the 5th and there are various alternative paths or pitch curves in doing so. Two things are fixed in all variations: the pitch anchors of the 4th and 5th degrees (or their pitch distance) and the entire duration of the melodic movement, which can be broken in sections of different individual durations (for ascent, holding and descent). What varies is the magnitude of pitch change within a given duration, in other words the rate of pitch change, which allows different notes and pitches to receive different levels of emphasis. To connect this to the previous statement, the rate of pitch change could possibly be related to the concept of 'weight', which is shifted on different pitches according to the expressive intention of the musician. 'Weight' or 'vazan' in Urdu (Sanyal & Widdess, 2004: 125)—a term that is also used with a musical focus—means in this context melodic emphasis, although in a wider context it could also imply other methods of placing emphasis, such as rhythmically.

The next report by Umakant Gundecha extends this idea about the aspect of time:

'What is the feeling in me, I show with the hands. E.g. the speed of the note [he gives two examples of different hand movements following the speed of pitch change].
If the speed is increasing, then the speed of the hand movements is also increasing'.
(Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

The rate of change in the position of the hands is associated here with the rate of change in the pitch of the sound. But as already mentioned, there is no univocal gesture–sound relationship; shortly after, the same musician reports of the fact that (in a particular situation of showing the tonic during the ālāp) the distance between the hands is imprinted on, translated to or associated with the magnitude of the melodic movement in pitch space (the size of the melodic interval). The most typical example refers to achieving precision in the tonic.

'At the beginning of the ālāp, when the note Sa (tonic) needs to come, as it comes from lower, it is like holding the sound [he appears to grasp something, hold it and alter the distance between his hands] and when the Sa comes the hand stays, which means that the note has come to the perfection. And when I stay on that note, the hand is holding'.
(Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

Along the explanations by Umakant Gundecha on the link between the pitch height of the voice and the distance of the hands, the observation of the interview video supports the fact that the pitch of the tonic is brought to a 'perfection' by holding 'something' (the note) with the hands and carefully adjusting its size. Then the tonic can be further prolonged by holding the hands at that same distance, as if by holding (the note) through a grip he can keep it under control and prevent it from fluctuating beyond that precise pitch height that was just achieved. The video analysis of rāga Bhūpālī performance confirms this statement and is discussed in detail in Chapter 7. To illustrate this concept, Umakant Gundecha gives a vocal example of a rising pitch and repeats it in three different ways (in the first it seems like taking something with the left hand from the right hand and moving it out and upwards, in the second time he moves his two hands horizontally and away from each other, and in the third time he takes something from the left hand with the right hand and moves it vertically (upwards) and explains:

> 'I might do one movement and my teacher a different one, but basically what is the thing there... to show the distance of the note with the hands going away from each other.
> It is a note distance showing by hand'.
> (Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

So, here the musician tries to justify something that is shared between him and his teacher in moving their hands regardless of the fact that the gesture is of a different type. This is a concept that is further discussed in section 4.2.4. (the effect of gestural transmission in teaching Dhrupad singing).

The abovementioned statements refer to a somewhat clear relationship between melodic movement and (hand) movement in the approach of the tonic. But an additional comment by Umakant Gundecha clarifies that this is not the only melodic feature that can be handled by the hands:

> 'Especially this movement [he moves his two hands apart and upwards, with open palms]… I feel that the voice is free, and especially when you go to the Sa; upper Sa [the tonic]… The voice is free. And showing with hands… this open feeling… feeling of openness.
>
> [...] When I want to say that the voice is open, very clear, you know... staying there... so the hands are open. And also if I am putting open voice… [he sings a rising melodic movement for approaching the tonic by pronouncing the 'a' vowel and keeps both his palms open]. Whereas if I… [he sings the same rising pitch by pronouncing the syllable 'num' and moves as if stretching an elastic band, by closing the palms and moving his hands apart]. If I am doing 'num', then I will not do this [he moves his two hands apart and upwards, with open palms]. Because the voice is closed [he illustrates the dependence of the palm shape to the used syllable by repeating the same melodic phrase with either the 'num' syllable and palms closed or the 'a' vowel and palms open]. So, it depends on the variety of notes… if it is open, if it is closed, or if it is half-closed [he sings 'ri' and adds]. Like 'ri', what is 'i'? It is half-closed. 'Num' is totally closed. 'A' is totally open. 'Ri' is half-open or half-closed'.
> (Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

Here a quality of timbre in the voice is represented by the palm shape and is associated with the used syllable. A closed voice produced by pronouncing the 'num' syllable is demonstrated by closing the palms and apparently grasping and keeping the hands attached to an imagined object that is manipulated. On the contrary, an open voice produced by pronouncing the 'a' vowel is demonstrated by opening the palms and moving the hands in a way that does not allude to handling any object. Here, the articulated syllable seems to dictate the timbre of the voice and guide the shape of the palm. The importance of the used syllables is also supported by Uday Bhawalkar in the following interview excerpt:

> 'The sound changes in our system with the syllables. And then it is used in a particular way and it changes the meaning and the feeling of a note'.
> (Uday Bhawalkar, interview, London, UK, July 27, 2010)

This section aimed at presenting how musicians self-report on the relationship between their own body movements and their voice. Aspects of time and synchronicity were discussed, which indicate that gestures are typically expected to precede the voice (similarly to instrumental playing). This conclusion is also meant to inform the video annotation and segmentation processes. Sound seems to be related to movement based on a concept of bipolar dimensions. Opposite directions in space (up vs. down, away vs. close or right vs. left) or the change in perceived resistance and effort (intensification to abatement) or speed ('increase' vs. 'decrease') can be associated to pitch change (high vs. low) and vocal timbre ('heavy' vs. 'light' or 'open' vs. 'closed').

Reports support the fact that the spatial relationship between (single-handed) gestures and pitch in the voice is not conceived primarily through the vertical axis (as in Western music), but is often displayed in terms of the distance between the hands (in bi-manual gestures) or the position of the hand in reference to the vocalist's body (in uni-manual gestures) in any direction. This is an important report that was meant to guide the video annotation process that followed and to assist the choice of movement and acoustic features to be extracted and used in the models developed by the quantitative methods in the later stages of the analysis. Additionally, there is a raised awareness of resistance (described by two different musicians through the terms 'contra' or 'weight') which represents a dynamic aspect of the hands' movements and is reflected on the 'heaviness' of the voice (so that 'it does not sound light'). It is interesting to examine whether the quantitative methods of the thesis can capture this relationship.

Finally, it was reported that movement qualities are more important than simply the type of gesture. To what extent verbalised gesture–voice relationships of this kind are a matter of idiosyncrasy or bodily disposition inherited from the musicians' teachers remains unanswered and forms the subject of the next section.

4.2.4. Gestures in music pedagogy

The following excerpt from an interview with Mohi Bahauddin Dagar, nephew of Zia Fariduddin Dagar, explains how important the visual engagement with music making is; at a small age it might be even more important than the music itself.

> 'Right since childhood I've been seeing these things [talking about movements]. When you are small, you *don't listen* to that music. You *see* all these *actions* and that they are happening. You are looking at it. You are seeing these things that are growing into your eyes. And it's so much easier if you follow those things! It just conveys effortlessly what you want to

say'.

(Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

According to reports by Dhrupad vocalists, explicit instructions on how to move during vocal teaching sessions are typically limited to helping novice students to clarify how to produce a difficult melodic movement. Gestures represent a way to help teachers embody and express qualities of a difficult musical passage or embellishment that would be otherwise difficult to articulate verbally to the student (Fatone et al., 2011). In the next example, Marianne Svašek, student of Zia Fariduddin Dagar who has been also receiving training by his nephew Mohi Bahauddin Dagar, reports of how she managed to produce a melodic technique she was finding hard to do by following specific instructions about her hand movements.

> 'Once Bablu [Mohi Bahauddin Dagar] gave me a lesson and he said "do like this" [she sings a kind of portamento from one note to another while suddenly twisting her left wrist with two fingers extended]...
> And I could not do a certain gap and just by moving the hand, twisting...
> So, it helps. Cause you visualise it or feel it or something'.
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

Drawing on the ideas of Goldin-Meadow & Wagner (2005) about how our hands help us learn in linguistics, it is not a serious leap to also argue that gesturing during singing encourages students to produce gestures of their own, which, in turn, leads them to learn how to improvise. The progressive learning and refining that follows repeated gesture–sound simulations is one of the basic assumptions about the way the hands might help a performer to explore sound and it is supposed to be achieved through 'a kind of "hermeneutic circle" of progressively becoming closer to the sounds by attaching them to our gestures' (Godøy, 2010)[76]. In Hindustani music, students are supposed to assimilate progressively longer sample material through a trial-and-error process of imitation[77], most often without much verbal explanation from the teacher on technicalities of sound production or theory. Physical action has the capacity to elicit and energise the memory system (Rieser et al., 1994; Butterworth & Hadar, 1989) and recall of music (Godøy et al., 2005) and facilitate learning (Martin & Schwartz, 2005). Godøy (2013: 16) has acknowledged its importance as both a 'mnemonic tool for performance' as well as a 'pedagogical visualisation of the music'.

Gestures are acknowledged as an important part of the learning process by both teachers and students, as the following quotation from an interview with Dhrupad vocalist Lakhan Lal Sahu in India reveals:

> 'Ustad [Zia Fariduddin Dagar] also does this. Ustad taught me that hand movements are very important in Dhrupad gayan and that they should not be wrong'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

[76] Recent work on (Western) singing supports the idea that using movement in conjunction with singing tasks can affect intonation, timbre and loudness, as well as perception tasks (Brunkan, 2015a,b). For instance, solo singers are more in tune when singing is accompanied by hand movements vs. when it isn't, and specific types of hand gestures (circular, arched etc.) are associated with lowered formant frequencies and increased vs. decreased amplitude.

[77] Schmidt (2012) argues that this trial-and-error process of imitating progressively longer melodic phrases is in fact at the core of learning to improvise Hindustani music.

What is also important to keep from this interview excerpt is that the student will be corrected for moving in the 'wrong' way. Indeed, vocalists often admit that teachers may discourage students from moving in the 'wrong' way, meaning movements that do not "match" what the voice should do. This observation has also been discussed in (Pearson, 2016) on Karnatak music. For instance, Fariduddin Dagar once reported during an interview:

> 'Students, when they do wrong things, I tell them how to move correctly'.
> (Zia Fariduddin Dagar, interview, Palaspe, India, January 11, 2011)

Similarly, Mohi Bahauddin Dagar, Dhrupad musician and son of the renowned rudrā vīṇā player Zia Mohiuddin Dagar, once said:

> 'You cannot sing with the wrong movements. I myself correct gestures as well.
> And without the movements you wouldn't be able to sing'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

What 'wrong' means in this context is not exactly clear. However, what is emphasised here is the fact that hands and voice are not two arbitrary parallel processes, but there are certain ways of matching them, that should be followed. A similar idea is reported by Dhrupad vocalist Lakhan Lal Sahu, who also adds that this "match" should feel 'natural':

> 'If you do a wrong movement with your hands, then the sound will be also wrong.
> It should be very natural'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

Terms like 'natural', 'automatic' etc. are common with Hindustani singers in describing hand movements which are not symbolic, codified or deliberate (Rahaim, 2009; Clayton, 2007). The next two quotations by maestro Zia Fariduddin Dagar and vocalist Marianne Svašek are also in support of this non-deliberate use of hand movements:

> 'The movements are not fixed, but they come automatically'.
> (Zia Fariduddin Dagar, interview, Palaspe, India, January 11, 2011)

> 'But it should be natural in a way... I don't know...'
> (Marianne Svašek, interview, Rotterdam, the Netherlands, September 27, 2010)

On the other hand, in the next excerpt from the interview with Umakant Gundecha the word 'automatic' is associated with a process of gestural transmission from teacher to student that is achieved through visual engagement rather than volition by the individual.

> 'I never explain [to] the student how to move. Also our teacher didn't tell us. It is not part of our teaching tradition. What we do is just watching the teacher and then it comes automatically, maybe a bit different, but it comes from our teacher'.
> (Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

The shift from the term ‘automatic’ to the term ‘traditional’ in the following remark emphasises and better reveals how musicking bodies, despite their idiosyncratic element, are also passed through teaching lineages:

> ‘But if he is our teacher and we are learning from him, seeing him, watching him, so these movements you can say they come traditionally.
> Although they are not fixed, but at the same time in the whole prospect we will see that the movements are of same quality’.
> (Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

These words seem to be an attempt to express both idiosyncrasy and personal identity (something that is not taught or inherited), as well as something unintentional (which is not consciously planned or inherited), i.e. something both spontaneous and unique. In a similar fashion, immediately after referring to weight and time as the important concepts that relate movement to sound, Mohi Bahauddin Dagar declares:

> ‘The gestures are different according to the individual,
> but the *qualities* of the movement are important’!
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

The term ‘quality’ is a key concept for describing multi-modal features at a higher-level of abstraction, where the link to the lower-level constituent features is not easy to describe. This statement on ‘quality’ implies that regardless of idiosyncrasy and personal style, what seems to matter even more than the exact patterns of movement (and the interlinked imagination) is the way these are performed (rather than the ‘what’), meaning their temporal and dynamic aspects. This observation can be seen in lines with Hatten’s argument (2004: 95) about sound gestures conceived as ‘significant energetic shaping through time’. Linking this to previous statements, it could be hypothesised here that it is exactly the physical metaphor of (contra-) ‘weight’ or ‘resistance’ against some imagined opposing force that functions as the source or cause of these amodal emerging dynamic shapes (that are shared between movement and sound or among different musicians), in the same way that the balance between the weight of an object and the force applied by a person in the task of elevating it determines the way the movement is performed.

Sahu explains how the transmission of movement is achieved primarily through visual engagement during year-long apprenticeship:

> ‘From the start,17 years ago, in 1993, I have been observing how Ustad [Zia Fariduddin Dagar) is doing these movements with the singing and I always try to learn’.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

Nevertheless, performance practice can also move away from these strict prescriptions; personality is allowed and even encouraged to unfold. The following three reports by different Dhrupad vocalists, including maestro Zia Fariduddin Dagar, can partly explain why:

> ‘Every artist has his own movement, so I don't tell them how to move.
> While learning, I teach them [how to move], but with time and while they move away from me,

they change.
(Zia Fariduddin Dagar, interview, Palaspe, India, January 11, 2011)

'While you teach, it is more technical and once they [the students] have learned it, then he [the teacher] tells them about getting that feeling into their singing'.
(Afzal Hussain, interview, Palaspe, India, January 11, 2011)

'We see our guru and then we follow.
[...] But there are a few things I may not follow, because I cannot or don't want to be exact copy of my guru.
You have to create your own thing, because you are different person
and he is different person'.
(Uday Bhawalkar, interview, London, UK, July 27, 2010)

The purpose of teaching is to aid the understanding of students in conceiving and conveying technically demanding parts in the musical structure. Volition, which leads to a divergence from the teacher's habits in both music-making and gesturing, becomes more evident during performance practice of mature musicians. This can also lead to students who only bear a fair resemblance to their teachers. Thus, apart from the conscious or unconscious inheritance of singing-accompanying gestures from teachers, there are also several examples for the deliberate choice of deviation in their gesturing habits.

The common qualities of movement that are shared among members of the same music family may also take the form of physical interactions with something that extends beyond the musician's body. In an interview with Uday Bhawalkar about music and movements, he commented that:

'There is volume control, throwing, keeping it inside,
this is all part of your learning from your guru'.
(Uday Bhawalkar, interview, London, UK, July 27, 2010)

According to this quotation, although movements (and their verbal description) may have a metaphorical function in respect to some abstract melodic idea, they can also be of iconic (or pantomimic) nature, representing something concrete, as in controlling some auditory feature. Furthermore, for musicians who are primarily instrumentalists, a number of instrument-related sensorimotor skills may be drawn from the performers' reservoir and brought into their singing through a process of motor translation. Rudrā vīṇā player Mohi Bahauddin Dagar reported the following during an interview in the Netherlands:

'It is a very typical movement on the vīṇā, because the hands are moving [he moves again his left hand vertically in both directions]. Because you practice for hours with the vīṇā, if you close your eyes you have the picture of the scale. Or sometimes [he moves his hands inwards and outwards] you have the pulling of the string. Because I associate it with... when I pull from here to here... then the Ma is going to hit [he gives several examples of pulling the

string]. Then I know, I am going to the right place'.
(Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

In a similar fashion, a motor translation from vīṇā instrumental gestures to the voice may apply to students learning to sing from teachers who are primarily instrumentalists but use their voice in instructing them. The following reports from interviews in India are enlightening:

> 'I have only studied voice, but our guru was a vīṇā player, so sometimes these vīṇā playing movements come. Because when he was teaching us, he did both with singing and through playing the vīṇā, and he used to do this kind of movements.
> And especially my teacher Zia Fariduddin Dagar also used'.
> (Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

> 'With these words (ti-ta-ra-na) [we have a] very close relation with these movements [he gives examples]. With fingers also like vīṇā [he moves his small finger in the air and sings a few syllables[78]], with hand also'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

But are these gestures always of pantomimic nature? Is it simply the grasping, pulling and plucking of strings that a musician imitates? Movements of the hands and fingers in Dhrupad can go beyond straightforward pantomimic gestures of direct sound production, as in real instrumental playing, and can also take the status of MIIOs, which would normally not produce any sounds under real conditions.

> 'He [his teacher, Zia Fariduddin Dagar] says "you pull your note" [he gives a vocal example]. For explaining actually. It is not rubber really… [he smiles].
> But for new students, he gives some examples for understanding. But it needs practice'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

Indeed, explicit gestural instructions in training sessions of Dhrupad seem to be more frequent with beginners, specifically when teachers need to clarify matters of difficult music passages or embellishments, which students might find hard to follow otherwise. I should also note at this point that movements in a performance situation might differ from those of a teaching context, as the following quotation by maestro Zia Fariduddin Dagar clearly attests:

> 'Concert and teaching is different'.
> (Zia Fariduddin Dagar, interview, Palaspe, India, January 11, 2011)

Indeed, according to my experience from participation in music classes and seminars as a novice student of Fariduddin Dagar (and others, such as Nirmalya Dey, Marianne Svašek and Uday Bhawalkar) and my observations on the recorded material of performance sessions, teaching and performance practice can deviate in respect to the degree of bodily resemblance to the teacher. The difference may refer to the functionality of movements (whether they reflect the musician's own engagement with sound or an effort to communicate with other performers or the audience), as well

[78] The small finger of the right hand is also used for plucking the strings when playing the rudrā vīṇā.

as their frequency and/or power. A one-to-one musical instruction is a practical scenario, where a teacher amplifies his movements and verbally articulates the way a melodic phrase should be performed, often by employing visual images. However, mainly due to practical reasons[79], the present work is essentially based on material from performances[80].

This section presented issues of gestural transmission from teachers to students. Contradictory statements were reported on bodily dispositions that are inherited from the teacher and support the fact that although movements are not explicitly taught, they are nevertheless considered as important and they are corrected from being wrong. Although gestural inheritance is acknowledged, at the same time idiosyncrasy and conscious decisions of deviating in the gestural habits are also reported and are expected to be observed in the video recordings.

Additionally, musicians emphasised the similarity of movements more than their exactness. Thus, it was expected that in the video analysis that followed different performers would be observed performing a MIIO not in the exact same way but in similar ways that share some common movement qualities. This further means that in quantitative analysis, movement features extracted from the raw data must be capable of capturing these qualities rather than simply describing the exact way a gesture is performed.

Finally, pulling and releasing was acknowledged as a fundamental concept of pitch control in both voice production and rudrā vīṇā playing. The extension of the musicking body and the engagement with melodic ideas through motor-based metaphors in interacting with imaginary objects, such as gripping, stretching, pulling, pushing and other types of manipulating and releasing such materials, is discussed in the next section.

4.2.5. Imaginary objects, materials and physicality of interactions

The current section provides an insight into the role of imaginary objects and their physical properties (such as elasticity, viscosity, gravity etc.), as well as the association between sensory-motor activities and patterns of melodic activity in these instances, as articulated by the musicians.

Pulling and stretching are verbs of imagined (as well as performed) actions that were fairly common during interviews, both as a self-reflection on the musician's own practices, as well as something they are instructed to do during their musical training. For example, Afzal Hussain during an interview in India explained how a student is encouraged to imagine an action, in which 'something' is cognitively equated to the voice:

> 'While teaching the mīṇḍ they [the teachers] tell that you have to pull your voice; imagine that you are pulling something and pull your voice'.
> (Afzal Hussain, interview, Palaspe, India, January 11, 2011)

---

[79] Such as the need for the availability of two musicians recruited concurrently, namely the teacher and a beginner student (usually such gestures are stronger and descriptions of imagery more frequent with beginner rather than senior students).
[80] A limited number of teaching sessions has also been recorded.

Something (the voice) that is immaterial (as form) or not easily accessible through our senses (the body parts which participate in sound production) is replaced by something that is tangible by a human being (an object). In other words, what is grasped and pulled after all by the hands while singing is the voice itself. It is not clear whether Hussain uses the term 'pulling' to refer to an elastic or rigid object. Sahu is even more explicit in the following statement:

> 'Ustad sometimes explained: "You do like rubber this sound" [he gives one short example and explains by slowly moving the hands], sometimes, not always, but sometimes he gave examples for understanding: "you pull the note like this, like rubber".
>
> […] One time mātājī[81] also said: "These notes are like rubber".
> When we do practice, these notes go like rubber… We can do [he gives more examples]'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

Here the 'sound' or the 'notes' (the immaterial aspect of voicing) take the form of an object with elastic properties. The type of object and its physical properties dictates the type of interaction and indirectly the use of the voice. But 'sound' cannot be simply equated to 'notes'. 'Sound' is the oscillation of some physical quantity, such as pressure, displacement and velocity, in an elastic medium and its auditory sensation. Here it takes the form of an object with elastic properties that needs to be manipulated and sculpted by the musician into a pitched sound, i.e. a 'note'. Thus, 'sound' might be expressing here the aspect of 'flesh' (Rahaim, 2009), while 'note' may be denoting the aspect of 'form' in the voice. Both are metaphorically related to the sensory-motor activities that an object with elastic properties may afford.

In response to my question of whether the musician indeed imagines having an object in the hands that he needs to manipulate, Uday Bhawalkar replied:

> 'Yes, perhaps ball... No... [laughing], but there *is* always something, when this happens… it *IS*… [he tries to describe with the hands, which look like indeed holding something]. You feel like elasticity, like taking... like a small worm [he gestures with one hand like a moving worm]. Elasticity is like one note is coming, taking it and then it is happening, you feel that you are taking it and it goes like this… [he gestures like taking something from the left hand with the right hand and extending it]'.
> (Uday Bhawalkar, interview, London, UK, July 27, 2010)

At a conscious level the object might not have a particular shape or form, but what is central here is the interaction it can afford as a metaphor of how the note (the 'form') needs to be treated, again in this case through the concept of elasticity. A similar explanation was given by Umakant Gundecha:

> 'Mīṇḍ is the travelling through śrutis [microtones]… And when you travel through śrutis, the movement is slow… you cannot go fast… you cannot show śrutis in fast movement. So when you go very slow, then the two hands help you to describe the distance between notes.

---

[81] 'Respected mother' in Hindi. Here it refers to Zia Mohiuddin Dagar's (brother of Zia Fariduddin) wife, Pramila Dagar, a musician (sitarist) herself, who has been always involved in music making in their music school.

> […] In the mīṇḍ it is like stretching the elastic band… [he moves his hands apart] and it goes up to a point. I feel that I have a rubber band in my hands and I am stretching and I stretch till there, where note comes to the perfection and when sometimes I feel it is not good, then it is a problem. Basically, the hand movements come with the feeling of notes'.
> (Umakant Gundecha, interview, Bhopal, India, January 16, 2011)

'Travelling through śrutis' is a description for melodic movement. According to this and a previous quotation by Umakant Gundecha (reported in the previous section) as well as observations of the interview video, the distance between the hands in the imagined presence of an elastic band reflects pitch height in the voice and therefore describes the temporal evolution of a pitch glide. However, the hands in this case are not free to describe distance, but are confined by the idea of manipulating an elastic band. The visible—as well as kinaesthetically perceptible—distance of the hands works as an analogy to pitch height, thus allowing the intonation of the final note to be adjusted accurately.

Elasticity is a common key concept in musicians' reports, but not the only one. In fact, some musicians are able to report on clear associations between distinct vocal techniques and modes of physical engagement with specific imagined objects. During an interview in India, Sahu used the verbs pulling (a rubber band, i.e. stretching), pushing and throwing in order to describe a mīṇḍ, a gamak and a huḍak respectively:

> 'With a mīṇḍ I feel like pulling a rubber band.
> With gamak it feels like applying pressure.
> When we throw our hands, it is huḍak'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

It is interesting to examine whether these pairs of gesture classes and vocal techniques will be revealed in video analysis of Hussain's performance. Similarly, Hussain explains how gestures are associated with a mīṇḍ, a ghaseet and a gamak:

> 'When you are doing mīṇḍ or ghaseet, it is like pulling it [the voice]. I feel like I am holding something. It is like softly pushing it away and again pulling it back. It is a combination of both, when you have the gamak as well as the mīṇḍ. Gamak is when you are pulling actually and there is lot emotion and anger, but it's also got mīṇḍ in it. When it is softer it is mīṇḍ. Dada[82] [here referring to Mohi Bahauddin Dagar] will be able to explain to you'.
> (Afzal Hussain, interview, Palaspe, India, January 11, 2011)

The voice takes the form of an object that is held between the hands and is either pulled or pushed away, depending on whether a mīṇḍ, a ghaseet or a gamak should be produced. According to Hussain, the difference between a mīṇḍ and a gamak lies mainly on the power, a gamak being more powerful than a mīṇḍ. It will be interesting to examine whether a difference in bodily effort will be confirmed by the findings of video analysis and regression. Likewise, Sahu explains that it is the sound that is handled by the hands in various ways:

[82] The literal meaning of the word in Bengali is elder brother.

> 'It is like lifting and pressing my sound.
> It is like pulling, pushing sound, because I have to take out some sound, to pull sound.
> And this is why this sound [the gamak] is coming'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

Sound takes the form of a tangible object that resists (through gravity, friction or elasticity) to the attempt to manipulate it with the hands by lifting, (com-)pressing, pulling or pushing. Although it still remains somewhat ambiguous whether sound refers here unequivocally to the 'flesh' or 'form', it is here topologically more defined. Sound is something that lives in the musician's body and needs to be brought out through some body act. Nevertheless, the verb that expresses this action is not always consistent. The following are two different statements, again by Sahu:

> 'This whole sound gives inside some pressure to our body.
> […] I see also something, I am expanding the sound, like a spring or rubber that I am pulling'.
> (Lakhan Lal Sahu, interview, Palaspe, India, January 6, 2011)

The immaterial sound (or voice) becomes a tangible object of interaction, which—based on its spatial, visual and kinaesthetic information—is manipulated with the hands with bigger ease. The idea of expansion by Sahu was already discussed in the previous section, only now it takes a more tangible form, something elastic that is enacted by the musician in order to be expanded. The visual, imagined and embodied aspects of sound are all again tightly entangled in this expression. Pulling is specified here too as something that refers to an elastic object that needs to be expanded rather than a heavy, rigid body that needs to be displaced. So both compression (pressure) and expansion are verbs that reflect the physical property of elasticity. What is then the discriminating factor? Are these interaction types supposed to refer to different cross-domain mappings or to a single metaphorical concept (elasticity) that differs only in terms of direction in space? Again here, do these concepts reflect purely the morphology of melodic ideas (the 'form') or the source of vocal production (the 'flesh')? It still remains unclear if movement serves an imagined or real, mechanical purpose in voice production. Mohi Bahauddin Dagar gives a quite illuminating explanation:

> 'There is a cycle of movements: tension, restriction, release.
> […] It is often said, that the human body resembles the body of the vīṇā; the strings are running in front of the body and you take them and pull them. It affects the chakras. When you pull, you have a slight intensification in the hands and the stomach at the same time; it's all about supporting the breath after all. In the opposite movement [the release] it's not the same; stomach and hands are not tight any more. I can feel my stomach contracting [he makes another movement]. So I am controlling the breath of it [he does another gesture, where he starts in a heavy way and then moves in a lighter way and adds] and then I let it go over'.
> (Mohi Bahauddin Dagar, interview, Rotterdam, the Netherlands, September 28, 2010)

At first, the statement seems to refer to an imitation of gestures performed while playing the vīṇā, but then there is a shift to something physiological, related to the mechanics of voice production. Instead of interacting with something external to the body (through a pantomimic or a metaphorical gesture), focus is placed on an internal sensation, in which the hands reflect a process of intensification, restriction and release, which is achieved through an equivalent contraction, restriction and relaxation of the stomach. Pulling and releasing may have a simple metaphorical functionality or they can be regarded as a process of motor equivalence[83] (Kelso et al., 1998) between muscle activity and pitch. In singing, the control of pitch height is achieved through the adjustment of the diaphragm, cricothyroid and strap muscles (Sundberg et al., 1989, Erickson et al., 1983; Roubeau et al., 1997) and the regulation of laryngeal/vocal fold tension (Pfordresher et al., 2015b) or larynx height (Ohala et al., 1972), while on the rudrā vīṇā it is achieved by pulling and releasing the strings (adjusting their length and tension).

But there seems to be a striking contradiction here between Sahu and Dagar: the first refers to a sense of expansion, while the second to contraction. What also remains unclarified is why—according to a previous statement by Sahu—mīṇḍs are associated with stretching gestures (extending an elastic object), while gamaks are performed with pushing gestures (compressing an elastic object). At this point, it is essential to take into account a more detailed description of the voice production mechanism.

During inhalation two groups of muscles pull against each other; these are the abdominals and the external intercostals, which are located between the ribs. The external intercostals pull upward and outward, expanding the ribs. In contrast, the abdominals pull downward and inward, against the open rib cage. Thus, a line of opposing pressures is created. In normal exhalation the diaphragm is left to simply relax. But in singing during exhalation (voicing) a fine control of outflow air is required, which is achieved by coordinating the balance between these opposing forces, so that the diaphragm is not allowed to simply relax and return to its rest position. For this reason the singer keeps the abdominal muscles contracted. As the abdominals are fixed at the ribs and the pelvis, the contraction causes the viscera, the diaphragm and finally the bottom of the lungs to be pushed upwards. Therefore, the diaphragm yields only gradually and creates a smooth and even pressure against the bottom of the lungs, which allows for a controlled and even outflow of air. This means that two antagonistic muscle groups are employed simultaneously, one which is related to contraction and pushing inwards (abdominals, with a tendency to bring the diaphragm back to its rest position) and the other to expansion and stretching out (intercostals, with a tendency to keep the diaphragm tensed). Whether a powerful or delicate sound needs to be produced, singers are trained to well coordinate these antagonistic muscles for a forced (stronger use of abdominals) or delicate (lighter use of abdominals) airflow respectively, which in turn drive the lungs (which cannot move by themselves) and provide singers with a good command over a wide range of musical expression.

Going back to the initial question related to contraction vs. expansion and pulling vs. pushing in relation to patterns of melodic movement and embellishment it could be suggested, that a double cross-domain parametric mapping blends into a single conceptual metaphor, which can be

[83] Motor equivalence refers to one's ability to achieve the same goal (both physical and cognitive) by switching from one set of effectors to another.

shifted from one to the other according to the needs of the performer. The expansion and compression of something elastic could be reflecting a shift of focus and awareness by the vocalist between the two esoteric opposing forces of the antagonistic muscles which are employed in singing; while the abdominal muscles are contracting, the rib cage is expanding and both are employed simultaneously for the production of the voice. But if a fundamental necessity of a gamak is the heavy and powerful oscillation between notes (adjacent or distant), then a strong contraction of the abdominal muscles is required for forcing a rapid and strong airflow. On the other hand, for slow and gentle mīṇḍs a more delicate use of the abdominals is required, with a stronger sensation of keeping the rib cage expanded for a longer time. Thus, contraction and expansion can be seen as two complementary iconic cases of corporeal processes related to the mechanics of voice production, and the singer can shift from one to the other according to the purpose they better serve. For forced airflow, as in gamak, the hands refer to the increased contraction of the abdominals, while in mīṇḍ the hands reflect an increased awareness in the tension created through the expansion of the ribs. Thus, movement cannot be taken as a unidirectional iconic representation of the sound. Nevertheless, what is shared is the notion of increased tension in the internal organs and a raised awareness of the voice production mechanism.

In summary, linguistic descriptors used during interviews attest to the idea of MIIOs (such as stretching, pulling, pushing, throwing), which I first encountered through self-participation at a Dhrupad music seminar by maestro Zia Fariduddin Dagar in 2007. There is therefore good scope in examining how these appear in practice during performance and how they might be related to the voice. The sounds or notes take the form of specific types of imaginary objects, such as rubber, ball or water that musicians need to manipulate. Some musicians even reported of distinct associations between types of melodic ornamentation in the voice and types of interactions with imaginary objects. However, the distinction between some of these interactions, especially the difference between stretching a note and pulling a note, was not clarified during interviews. All types of interaction imply a rise in tension and this can refer to either an external object to the body or to the literal control of tension (intensification vs. abatement) of corporeal processes that are required during voice production in order to regulate airflow and adjust pitch height. Tension in physical movements is associated with perceived effort, as already discussed in section 2.3. Therefore, I expect that closely observing how effort is exerted by a performer is a way to identify gesture types and boundaries of MIIOs.

4.3. Discussion

The following gives an overview of some of the most important findings:

- Performers displayed a high visual element in the conceptualisation of music and a high level of awareness in the importance of upper-body movements in vocal music-making.

- The analysis of the interview material showed a strong entanglement between musical parameters, imagery and movement, that makes them almost inseparable in linguistic thought, but is nevertheless not hard-wired and fixed. As music-making is a multi-parametric process, musicians may shift between a number of possible cross-domain mappings according to their specific needs. Associations are either explicitly described and illustrated through concrete examples or interpreted through metaphorical uses of visual imagery and sensorimotor descriptors.

- Visual engagement and discouragement from wrong movements by teachers lead to gesturing habits that are unconsciously inherited. The conceptualisation of a demanding musical movement can be refined through the additional insight provided by extra-aural information. Not the type of movement, but the way it is performed in relation to the melodic activity is passed down through teaching lineages. Therefore, despite the idiosyncratic element, melodic qualities that are attributed to the sound are embodied and articulated through visual images that are shared among disciples of the same teacher. Imitation of instrumental gestures can also impose specific gestural manners.

- The absolute direction of the hands' movement is not important on its own. Considering that change can be represented in a 'bipolar quantitative continuum' (Eitan, 2007) in two directions, what is more important is the fact that space is rendered asymmetrical. This means that while the movement of the hands in one direction reflect a process of intensification, moving in the other direction is translated as abatement (proximal vs. distal, left vs. right etc.). Metaphors of this asymmetry can be found in both the mechanics of voice production, the production of a rising note on an Indian stringed instrument such as the sitār, as well as the interaction with any other object of our surroundings, such as an elastic band. While one direction reflects the intensification of the interaction, the other direction refers to its release.

- Musicians reported about the significant facilitating role of movements in 'supporting' singing, which may refer to both the body (the tangible side: the mechanics of voice production and gesturing) and the mind (the immaterial side: the cognitive aspect of melodic and gestural activity as forms and shapes). This distinction was not always made explicit and clarification would shift from one to the other, meaning that both cases can be valid for a single performer.

- The object employed imposed the impression of some resistance that opposes the performed action (explicitly described as '(contra-)weight' and implicitly expressed through motor-based metaphors). Resistance was a concept that could either indicate the internal sensation of tension and opposing forces created by the antagonistic muscles—of which a high degree of control is required by the singer—or a means of conceiving and conveying melodic movements of specific qualities in relation to melodic movement and time.

- The relationship between sound and movement does not reside on the exact points in space that the hand traverses, but on the way these are accessed.

- Similarly, not only pitch but also its time derivatives (speed and acceleration) that are expressed in terms of time spent and emphasis placed while moving between specific pitches are conveyed through similar aspects in the movement. These aspects can be regulated through the employment of a suitable imagined object, that resists to the change and restricts the flow of the movement according to the interaction that it can afford in an appropriate for the rāga way. What is then grasped and moved in space is the sound or the note itself.

- For this, vocalists made frequent use of concepts taken from physics or mechanics, such as elasticity, viscosity or weight and friction, or imagined objects, such as elastic rubber band or heavy object, as well as sensory related linguistic descriptors, such as heavy and light, in an attempt to describe qualities that should be brought in the voice.

- Explicit relationships between materials and melodic techniques or ornamentations were occasionally provided by musicians, such as rendering a mīṇḍ while stretching an elastic band, a gamak while compressing an elastic object and a huḍak when throwing. Vivid manual co-verbal gestures accompanied these descriptions and in some cases explanations of concrete examples of melodic phrases were given with the aid of the hands. However, it should be noted that these associations should be regarded with some caution, as it might be the case that musicians use multiple and flexible cross-domain parameter mappings and that they shift from one to the other according to their needs.

4.4. Summary

The current chapter presented findings about gesture–sound relationships as well as visual images, which were revealed by analysing both the explicit and tacit linguistic thought of performers, with a focus on interacting with imaginary objects. The scope of the interview analysis was to provide a better understanding of the implicated phenomena and to allow the development of codes and themes that form the basis for the annotation and analysis of performance videos.

The chapter opened by illustrating in an organised and compact way codes of frequently used sensory-related linguistic descriptors and motor-based metaphors which are related to some imagined objects by Dhrupad singers. There was a frequent reference to images of objects, like a rubber/elastic band, water, ball, and a notable awareness of the resistance one senses while moving against or with the opposing forces related to their properties, be it elasticity, viscosity, friction or gravity. These were expressed through various verbs which betray the types of interactions these objects and their physical properties may afford. Such descriptions seemed to fulfil an ambiguous purpose; either an abstraction of the mechanics involved in the vocalisation or a metaphorical description of melodic qualities. In the second part of the analysis, selected excerpts from the interview material were discussed and interpreted in terms of both the explicit and implicit knowledge that is often embedded in musical discourse. The analysis was dedicated to a detailed exploration of how participants are making sense of these imagined interactions through a process that involved an interpretation element.

Interviews were a useful tool in investigating both explicit and unconscious knowledge in music-making and offered many significant details which allow us to have a better insight into the role of upper-body movements in Dhrupad singing and the relationship between melodic and bodily activity on the occasions that a singer seems to interact with some imaginary object. One of the most important outcomes of the analysis was the table of themes and codes, which were extracted and classified from the transcribed interviews and which were intended to inform the annotation process of the audio-visual material of the actual performances that followed. This means that the codes presented were not strictly applied to the annotation of these performances; the coding scheme of performances was developed based on the visual observations, but it was driven by the findings of this chapter.

Chapter 5 – VIDEO ANALYSIS I – AFZAL HUSSAIN (RĀGA JAUNPURĪ)

In this chapter I present video analysis of an ālāp vocal improvisation of rāga Jaunpurī by Dhrupad singer Afzal Hussain, exploring associations between bodily effort, MIIOs and melodic content. He is a senior disciple of maestro Zia Fariduddin Dagar, who has been trained under guru-śiṣya paramparā (see section 2.1.3.), and a professional Dhrupad performer and teacher.

First, I identified and segmented MIIO movement events visually, on the timeline of the ANVIL annotation environment. I then annotated the identified events using a coding scheme that was informed by the analysis of the interview material but that emerged progressively during multiple viewings; the exact coding scheme is described in section 5.1.3. and can also be seen in the snapshot of the ANVIL annotation environment in Figure 3.1. The movement coding included gesture classes, perceived effort levels and other attributes useful for the quantitative analysis. The audio coding included classes of recurrent melodic movements, mostly characteristic pitch glides of the specific rāga, and the melodic intention following each movement event. Finally, I examined cross-domain associations. The association analysis focused upon gesture classes, melodic phrases, musical context (melodic intention in rising towards the tonic), as well as effort levels. A detailed audit trail of the annotation process and the results of the analysis is presented, along with a discussion of the findings.

5.1. Analysis

5.1.1. Data collection: Ālāp recording and Interview

The recording session with Afzal Hussain (Figure 5.1) took place in the field, in a room of the music school run by Zia Fariduddin Dagar in Palaspe/Panvel India, during the day time of January 11, 2011. The musician was informed beforehand that the recording was to be made for a research project regarding music and movement and he was asked to perform the first part of an ālāp improvisation without the accompaniment of a percussion instrument; the length (45 min) and the rāga (Jaunpurī) were of his own choice. The recording includes the solo performance, as well as an interview session that followed immediately after, all captured in audio (separate close-miking of voice and drone instrument, as well as ambience), video and motion. The musician was asked to sing only an ālāp improvisation (no composition), and only the first, slowest, un-metered part of the improvisation (18 min) was selected for analysis. Despite the fact that the school is situated on the Mumbai-Goa highway, the ambient sound conditions were good for a field recording. The taňpurā was played by a junior student of the school, who sat across from the musician in order to avoid leakage of the drone instrument's sound into the singer's microphone. As Hindustani music follows a movable tonic system, the taňpurā was tuned by the musician at his most comfortable pitch area, with the tonic at 140.1 Hz, approximately a C#2 in Western terms. The other strings were tuned to the tonic and 5th degree of the scale in the low octave. According to Serrà et al. (2011) and van der Meer (2001/1999), the frequencies of the other notes to which the taňpurā is tuned can be reliably deduced from the tonic by using the equal temperament scale. This gives 104.2 Hz for the lower 5th and 70.1 Hz for the lower tonic. Based on the tuning of the tonic and the equal temperament scale, all other degrees can also be deduced. The teacher was not present during recording[84]. The interview was conducted immediately after the performance and while the musician was still wearing the reflective markers.

Figure 5.1. Afzal Hussain recording session.

[84] Sometimes students hesitate or find it hard to perform in front of their teachers.

5.1.2. Analysis of rāga Jaunpurī by Afzal Hussain

Although a rāga can be described in terms of theoretical rules, it needs to be clarified here that variations of the mood and specificities in its rendition may be observed among different music styles (gharānā). Therefore, I first present the theoretical background of rāga Jaunpurī and then highlight deviations or phrases of special interest in the particular performance by Afzal Hussain. The following points summarise some of the most significant features of rāga Jaunpurī according to Bor (1999) and Autrimncpa (2012b):

1. Tone material: 1, 2, *b*3, 4, 5, *b*6, *b*7.

2. The tonic and the 5th degree are prominent notes. The 5th degree is a resting note, which means that the melody often resolves upon it.

3. The 6th degree is the dominant and most important note (called vādī in Hindustani music) and the 3rd degree is the second-most prominent and frequently used note (called samvādī); soloists oscillate heavily upon both.

4. The 3rd degree is omitted in ascent, but used in descent, especially in closing phrases.

5. Both *b*7 and *b*6 are sustained in descending movements.

6. Occasionally, the 6th and/or 7th degree may be avoided during a mīṇḍ, such as in 2-4-5/*b*7, 4-5~*b*6/1' or 4-5/1'\~*b*6-5.

7. Most melodic activity in rāga Jaunpurī happens in the upper half of the middle octave.

8. Jaunpurī has a descending character, which means that even in ascending movements, target notes are usually approached through smooth pitch glides (mīṇḍ) that first go above them before descending on them. For example, in an ascending movement from a lower note to the 6th degree of the scale, the goal pitch (the 6th) is typically approached by first moving up to a higher pitch (the 7th degree) and then descending to the 6th. Usually the movement does not conclude there, but it descends further and resolves on the 5th, which is the resting point in rāga Jaunpurī: 2/*b*7*b*6-5. In other words, the melody is first attracted to a higher note (the 7th degree) before being allowed to reach the actual target (the 6th degree)—however through a descending movement—and finally drop upon the 5th degree, where the movement will typically resolve. This idea aids the interpretation of movement analysis that follows. Not all notes of such a pitch glide receive the same amount of emphasis (or 'weight'), but the scope of the improvisation is to explore a variety of ways and "paths" in the pitch space for approaching the same goal pitch.

A general outline of the ālāp melodic progression of rāga Jaunpurī as interpreted by Afzal Hussain is presented in Figure 5.2. The dark blue line is the pitch graph of the performance (extracted from Praat), the black line reflects a (subjective) assessment of the pitch focus (the progressive presentation of individual degrees of the rāga) and light blue areas show the pitch range used in blocks of time (which indicate the melodic expansion, called vistār[85]). The horizontal red lines show the steadily sounding pitches to which the drone instrument is tuned. The time scale is in seconds.

[85] The gradual widening of range to include successively higher and/or lower pitches. This process can be seen both in the development of an individual phrase, and in the structuring of each large section of a performance.

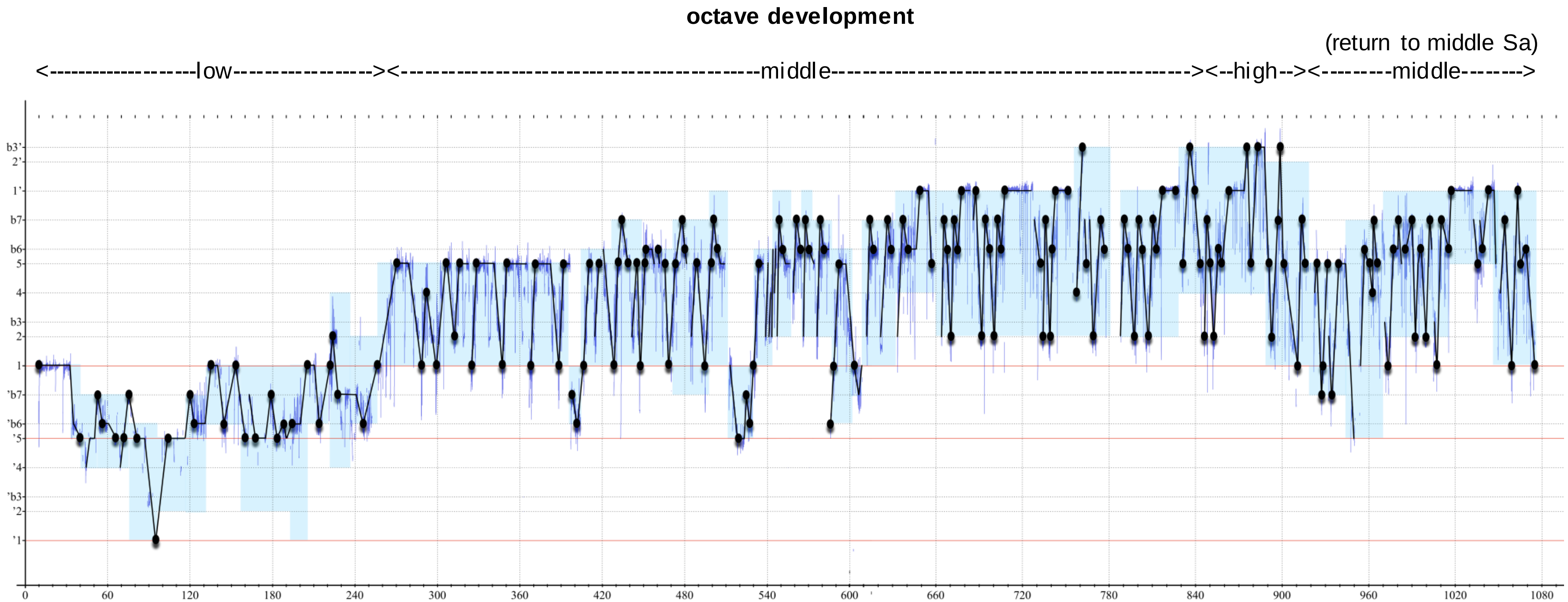


Figure 5.2. Graph giving an overview of the ālāp melodic progression in rāga Jaunpurī by Afzal Hussain; time is in seconds (inspired by Widdess, 2011).

As is typical of a Dhrupad performance, the singer makes use of an almost two-and-a-half octave range. He starts at the middle tonic, explores the lowest octave, then the middle octave and finally rises to a climax (the upper tonic), extending this further by briefly reaching to the 3rd degree of the highest octave, before finally descending to the middle tonic again. Hussain provides an unexpected but distinctive second (although shorter) exploration of the middle octave before finally concluding the ālāp, although this is not typical of the rāga.

In Figure 5.3 the pitch distribution of the ālāp is plotted, reflecting the tonal hierarchy of rāga Jaunpurī as rendered by Afzal Hussain:

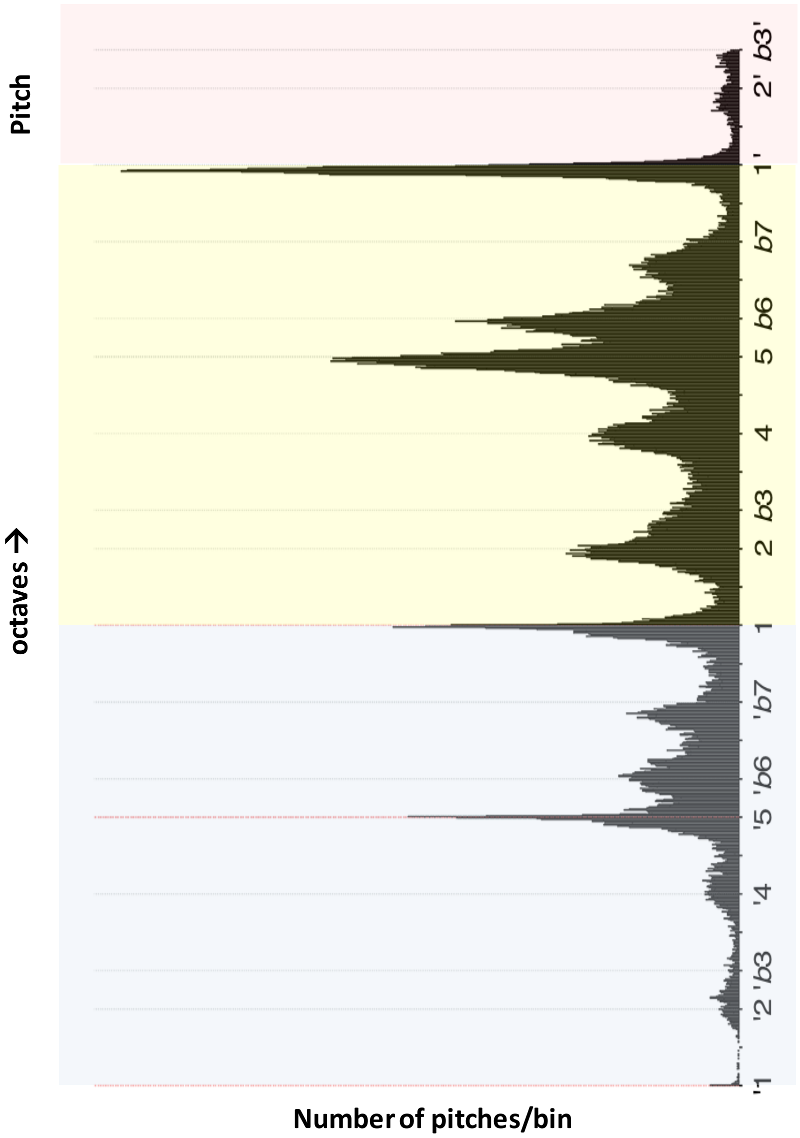


Figure 5.3. Afzal Hussain: Pitch distribution (in cents) of Jaunpurī ālāp performance (Praat). The red lines indicate the frequencies of the taňpurā tuning.

Some of the most important findings, supported by the pitch distribution graph, are summarised here:

1. Afzal's rendition of rāga Jaunpurī is similar to the pitch-class distribution of rāga Jaunpurī as computed by Chordia & Rae (2007) and the tonal hierarchy profile of rāga Āsāvarī (a rāga often confused with Jaunpurī) produced from subjects' ratings of tonal hierarchies by Castellano et al. (1984). The treatment of the 7^{th} degree is also closer to rāga Āsāvarī (according to Bor, 1999). As expected by van der Meer (2001/1999), peaks in the distributions fall close to the equally-tempered intervals, apart from the 7^{th} degree which is lowered compared to its position on the equal temperament scale; thus the *b*6-*b*7 interval of the mīṇḍ is less than a tone.

2. As expected, the tonic and the 5^{th} are the prominent degrees, on which the soloist focuses and spends more time.

3. In line with typical music theory, the 3^{rd} degree is only used in closing phrases of *b*3-2-1, but despite being the samvādī note, it is rarely used by the performer.

4. In contrast to what would be expected by music theory, high melodic activity is observed in the lower part of the middle octave.

5. As expected, peaks on the tonic and the 5^{th} are quite sharp, as they are the most sonant notes and therefore well-defined pitches. In contrast, the distribution around most of the 2^{nd}, 4^{th}, 6^{th} and 7^{th} degree peaks is relatively flattened and spread out. This is because these pitches are approached through smooth glides that start at adjacent notes and move to these degrees, with the eventual pitches often varying by a few cents. The 6^{th} degree of the middle octave forms the only exception to this, which may suggest that in the middle octave glides of .../*b*7*b*6 place emphasis and spend more time on the 6^{th} degree during descent.

6. The most striking examples of mīṇḍ, which also constitute the most recurrent pitch phrases (in association with MIIOs) over the entire rendering of the ālāp, are:

- In the upper part of the octave: .../*b*7\~*b*6 for the approach of the 6^{th} through the 7^{th} degree, which finally resolves by either descending to the 5^{th} or moving up to the tonic. Frequently, pitch glides in the upper part of the octave (the *b*7-*b*6 interval) are repeated several times before the final ascent to the tonic, which seems to be a technique used by Hussain for building up anticipation and melodic tension.
- In the lower part of the octave: 1/4\2 for the approach of the 2^{nd} through the 4^{th} degree and 2/5\4 for the approach of the 4^{th} through the 5^{th} degree. These two glides usually sound in succession.

For each repetition of a mīṇḍ in the lower part of the octave, the weight of the glide is placed on its highest pitch (the 4^{th} and 5^{th} degree respectively). Repetitions of the mīṇḍ in the upper part of the octave are rendered with the emphasis placed on a pitch that may vary by some cents around the peak distribution of the note and by approaching the 7^{th} degree in a multitude of ways (which is an important aspect of Dhrupad performance).

5.1.3. Coding scheme

The following basic features were included in the coding scheme for Afzal Hussain: {MIIO gesture class; effort level; melody phrase class; melodic intention}.

The gesture classes used for Afzal Hussain are listed in Table 5.1, along with links to video files of examples and a video of the entire ālāp performance:

Afzal Hussain: media file playlist (Jaunpurī ālāp performance) https://n2t.durham.ac.uk/ark:/32150/r1r494vk17g			
5.1.	**ālāp performance**	**times**	https://n2t.durham.ac.uk/ark:/32150/r10p096691p
5.2.	**stretching**	07:13:13 – 07:15:87	https://n2t.durham.ac.uk/ark:/32150/r19p290932t
5.3.	**pulling**	12:17:67 – 12:19:67	https://n2t.durham.ac.uk/ark:/32150/r12j62s4877
5.4.	**pushing (away)**	04:28:20 – 04:29:77	https://n2t.durham.ac.uk/ark:/32150/r1p5547r361
5.5.	**throwing**	04:02:54 – 04:03:34	https://n2t.durham.ac.uk/ark:/32150/r1db78tc02t
5.6.	**collecting/taking**	15:05:81 – 15:06:91	https://n2t.durham.ac.uk/ark:/32150/r108612n53g
5.7.	**kite-flying grip**	11:52:68 – 11:54:08	https://n2t.durham.ac.uk/ark:/32150/r105741r71c
5.8.	**grappolo grip**	17:15:97 – 17:19:11	https://n2t.durham.ac.uk/ark:/32150/r1t148fh122

Table 5.1. Afzal Hussain: Media file playlist

Some gestures were hard to classify as a single type of gesture and therefore they received a mixed code. These three ambiguous cases include non-clear-cut instances lying between [stretching ad pulling], [stretching and pushing], or [pulling and collecting/taking]. However, towards the end of the annotation it became apparent that the ambiguous annotation of stretching/pushing is actually a distinct form of gesturing that represents (in contrast to pushing away) the compression of a compressible and elastic object away from the body. Therefore, in the quantitative methods that followed the video analysis, the stretching/pushing gestures were classified as interactions with an elastic object.

Melodic phrase classes include six melodic glides (most of which are characteristic of rāga Jaunpurī), a typical cadence for closing long melodic phrases, as well as a small number of melodic movements that remained unclassified:

{…/*b*7*b*6; 1/4\2; 2/5\4; 5/2'*b*7*b*6; tonic-approach and hold; 5th-approach and hold; cadence (*b*3-2-1); unclassified melodic movements}

Finally, the melodic intention of ascending to the tonic (reaching the tonic or just moving towards the tonic) immediately after each coded gesture/melodic event was also annotated as a Boolean parameter (yes or no).

Additionally, the annotation included the numbering of the MIIO event, the effort level as a categorical variable (as low-medium-high), the four Laban motion factors (weight-flow-space-time), the type of grip (closed-, open-, semi open-handed), the handedness (unimanual-bimanual/using both or using only one), the target of the movement in space (the point in space where the hands reach their maximum extension from the original position; see 'hands_target' track in Figure 3.1), the relation between the hands in case of bimanual gestures (together-apart-approaching), the direction of the hand or hands (a 3d coding system was used here, see 'hands_direction' track in Figure 3.1), the syllable sung and the octave range.

5.1.4. Inter-coder agreement test

The process of conducting the inter-coder agreement test is described in detail in the methodology section of this thesis (3.2.3.). It was carried out by two dancers/choreographers and refers only to movement annotations (gesture classes and effort levels), both clear-cut and ambiguous. Table 5.2 presents the annotations of the three coders (my own and of the two additional raters):

PERFORMER: Afzal Hussain, Jaunpurī (18min)						
ANNOTATION DATE: 5.1.2015						
Sample number	**GESTURE TYPE**			**EFFORT**		
	S.P.	**Dim.K.**	**Din.K.**	**S.P.**	**Dim.K.**	**Din.K.**
2	collecting	collecting	collecting	1	2	1
4	stretching	stretching	stretching	7	5	5
5	stretching/pulling	stretching/pulling	stretching/pulling	3	3	3
6	stretching	stretching	stretching	9	7	5
7	stretching/pulling	stretching	stretching	7	4	4
8	stretching	stretching	stretching	9	9	10
9	pulling/collecting	pulling/collecting	pulling	0	3	4
10	pushing	pushing	pushing	7	8	8
11	stretching/pushing	stretching/pushing	stretching/pushing	9	10	7
12	stretching/pushing	stretching/pushing	pushing	8	8	9
14	stretching/pulling	stretching/pulling	stretching/pulling	5	6	2
15	pushing	pushing	stretching/pushing	8	7	7
17	stretching/pulling	stretching/pulling	pulling	6	5	4

18	stretching	stretching	stretching	7	9	10
19	pushing	stretching/pushing	stretching/pushing	6	6	5
20	throwing	throwing	throwing	1	3	1
21	pulling/collecting	pulling/collecting	pulling/collecting	1	3	2
22	stretching/pulling	stretching/pulling	stretching/pulling	3	5	2
26	pushing	stretching/pushing	stretching/pushing	8	6	5
34	stretching	stretching	stretching	10	8	7
40	stretching/pulling	pulling/collecting	stretching/pulling	7	5	4
44	pulling	pulling	stretching/pulling	4	4	3
58	pulling/collecting	pulling/collecting	stretching/pulling	3	3	2
63	pulling	stretching	stretching	9	9	10
64	collecting	pulling	pulling	2	2	1
65	stretching/pushing	pushing	pushing	7	7	8
73	throwing	throwing	throwing	1	3	4
74	throwing	throwing	throwing	1	2	1
76	collecting	collecting	collecting	2	3	2
80	pulling	stretching/pulling	stretching	5	2	1

Table 5.2. Afzal Hussain: Annotation summary of the three coders.

This means that the calculation of inter-coder agreement coefficients is based on three coders, 30 cases, and thus 90 decisions.

(a) Effort levels

For the ordinal values of the effort level (on a scale of 0-10) the computation of Krippendorff's method gave: Krippendorff's alpha = 0.81.

(b) Gesture classes

The task of calculating reliability in the coding scheme of the above mentioned gesture classes is a challenging one. It contains ambiguous annotations which fall between two discrete nominal categories and which are not taken into account in typical calculations of agreement coefficients. To tackle this obstacle, I used a mix of two approaches, each representing an extreme in possible ways of calculating reliability, one being the most restrictive and the other the most flexible. The first approach requires that the exact same annotation (be it clear-cut or ambiguous) be given by all three coders in order to consider the coding as validated, while the second only necessitates that all three annotators share at least one code (even if it is part of an ambiguous, mixed case). The following describes these two approaches in more detail:

- In the first case it is assumed that all categories, including the ambiguous cases, are considered as distinct classes (neglecting the fact that three of the cases in practice are a mix between some of the other labels). For instance, sample number 15 in Table 5.2 is considered as a disagreement, despite the fact that all three coders have annotated as a pushing gesture.

- In the second case, the events for which at least one coder used an ambiguous annotation require an explanation. In the second case, if one class is shared by all three annotators (e.g. ‘pulling’ for sample 58), the coding is considered in agreement between the three coders. This is of course a simplistic approach, but it is effective for the purpose of the project. A simple inspection of Table 5.2 reveals that there is only one event (the last sample) in which there is no common label shared between all coders where ambiguous labels were used by at least one coder.

Finally, it is expected that the agreement coefficient values fall between these values. Although this is a rather imprecise method, it allows us to gain some understanding on the level of agreement between the three coders and to draw some tentative conclusions.

Table 5.3 shows the results for both calculations of the nominal values of gesture classes:

Inter-coder Agreement coefficient	ambiguous classes considered as discrete	ambiguous classes corrected according to shared labels
average pairwise Cohen's kappa	0.62	0.91

Table 5.3. Afzal Hussain: Cohen’s kappa inter-coder agreement coefficients for two ways of calculating coding agreement.

(c) Evaluating inter-coder agreement coefficients

In the absence of knowledge for establishing a robust threshold for inter-coder agreement, I decided to rely on typical values used in the social sciences; a≥0.8 to consider the data as reliable, a≥0.667 to draw only tentative conclusions and lower values to completely discard the data as unreliable. The alpha value for the annotation of effort levels lies clearly over the limit and shows a high degree of agreement. Cohen’s Kappa values for the annotation of gesture classes lie between 0.62 and 0.91 (0.62 without taking into account the ambiguity of some common labels and 0.91 considering shared part of a label as coding agreement). Taking into account the fact that the identification of interaction types is carried out in the absence of a real object, these values are high enough to allow us to consider the annotation of gesture classes reasonably reliable. Thus, for the purposes of this work, the manual annotations I conducted are considered valid and can be used reliably in the qualitative and quantitative parts of the analysis.

5.2. Results

The presentation of the results follows the order in which the analysis was performed. First, I present and discuss the distribution of effort level values, as well as the types and frequency of appearance of gestures and melodic movements that were identified during the annotation process of the entire performance by Hussain. Then I report on the results of the gesture–sound association analysis, which is organised in two parts. In the first part (presented here), I focus on the clear-cut cases of gesture classes. In the second part (presented in Appendix A), I focus separately on ambiguous gestures, unclassified melodic phrases and special cases in the rendering of melodic phrases. The second part aims at retaining rigour in the analysis by examining the validity of the results drawn exclusively from the clear-cut cases.

5.2.1. Effort levels

From a simple visual observation of this and other performances of Hussain that I have attended in the past[86], I would describe his movement as the most powerful of all performers I have been working with. I consider this an idiosyncratic element of Hussain's performance style. Figure 5.4 displays the histogram of Hussain's effort values.

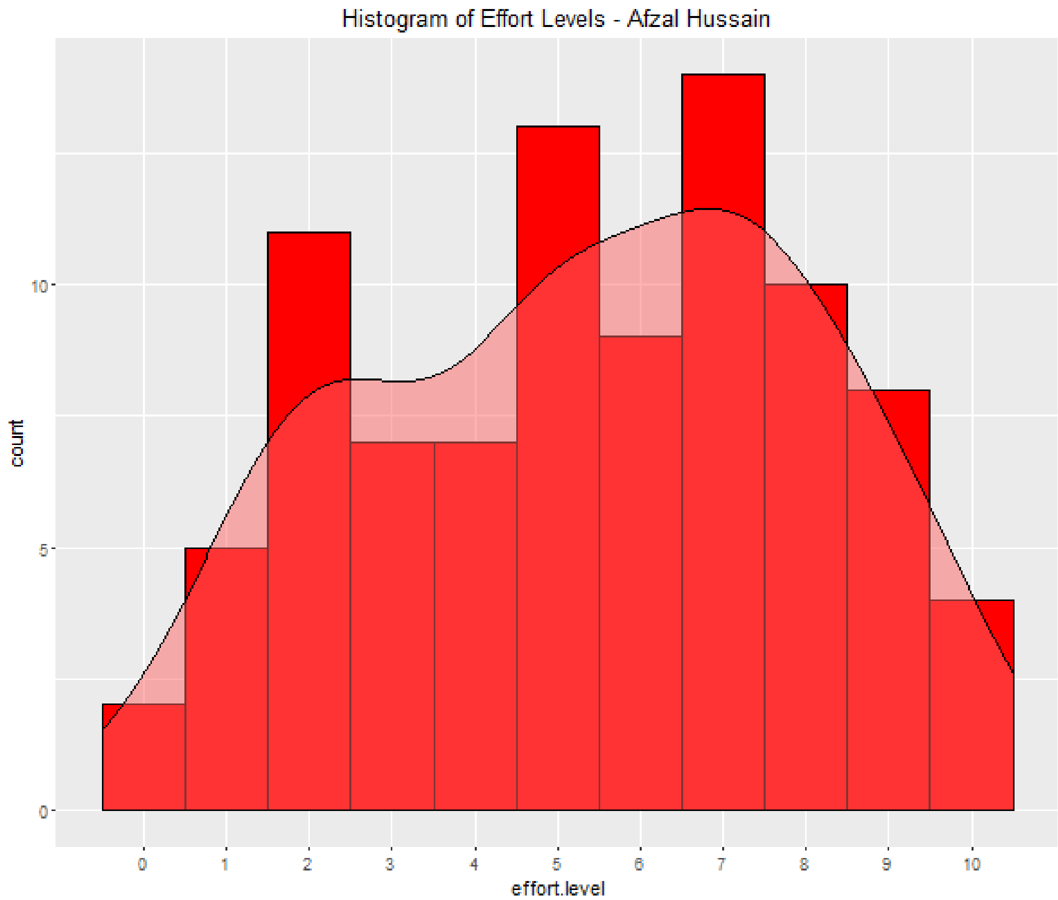


Figure 5.4. Afzal Hussain: Effort distribution.

[86] One during a Dhrupad festival in Udaypur, India in December 2010 and one during a Dhrupad festival with participants exclusively from this Dagar gharānā (disciples of maestro Zia Fariduddin Dagar) in Pune, India in August 2011.

By plotting the effort histogram for Hussain it is possible to see that although the probability curve deviates to some extent from a standard normal distribution, it can still be assumed that the requirement for normality in the distribution of response values is not violated in the linear regression methods used in quantitative analysis. There is a general trend for the effort values to be distributed around medium levels (mean=5.4). It should be stressed again here that this mean value is computed according to the range of effort levels of the particular performer and cannot be compared to other vocalists.

5.2.2. Gesture classes

The movement annotation of the whole ālāp includes 145 gesture events, which belong to seven classes of MIIOs and three categories of ambiguous classes, mostly lasting between 0.4 and 10sec.

Gesture classes Hussain		number	percentage (%)
Unambiguous classes	stretching	10	11.11
	Pushing	8	8.89
	Pulling	18	20
	Throwing	4	4.44
	Collecting/taking	5	5.56
	Grappolo grip and holding	3	3.33
	Kite-flying grip and holding	1	1.11
Ambiguous classes	Stretching/pulling	23	25.56
	Stretching/pushing	8	8.89
	Pulling/(collecting/taking)	10	11.11

Table 5.4. Afzal Hussain: Gesture classes.

There is a high degree of ambiguity when discriminating between pulling and stretching. Of these 145 gestures, 55 belong to 26 groups of undulation gestures (see section 3.2.2.) that accompany āndolans, as shown in Table 5.5:

Undulation Gestures Hussain	number	percentage (%)
stretching/pulling	40	72.7
pulling/(collecting/taking)	4	7.3
pushing	2	3.6
collecting/taking	9	16.4

Table 5.5. Afzal Hussain: Undulation gestures (groups).

Their short duration (60% have durations of 0.2-0.5sec) and their high level of coarticulation, renders their identification and segmentation prone to uncertainty and therefore they are excluded from the analysis. Thus, the final corpus of data for Hussain consists of 90 movement events.

5.2.3. Melodic phrase classes

The melodic annotation of the whole ālāp includes 137 melodic phrases, which belong to seven identified classes of pitch intervals, while a few phrases are unclassified and are therefore annotated as 'other'. Their durations usually vary between 0.4 and 7.2sec; the three exceptions of over 10sec durations correspond to a holding of the tonic. Of the 137 melodic phrases, 55 belong to 26 groups of undulations/melodic oscillations and are therefore excluded from the analysis, resulting in a corpus of 82 melodic phrases. This figure is eight fewer than the total number of gesture phrases (90), and this can be explained by the fact that although there is mostly a one-to-one correspondence between melodic phrase and gesture, in these eight cases a single melodic phrase was accompanied by the use of two consecutive gestures. For the purposes of pair-wise association analysis these peculiar cases are broken down into two parts giving a total of 90 phrases. Table 5.6 presents annotated melodic phrases.

Melodic Phrases Hussain		number total	percentage (%) total	number (analysis of ascents/descents)		
				double slope	ascent only	descent only
Melodic glides (mīṇḍ)	.../*b*7*b*6	40	44.45	25	7	8
	1/4\\2	10	11.11	9		1
	2/5\\4	9	10	9		
	5/2'*b*7*b*6	5	5.56	3		2
	tonic - approach and hold	10	11.11			
	5th - approach and hold	3	3.33			
	cadence (*b*3-2-1)	3	3.33			
	other (unclassified)	10	11.11			

Table 5.6. Afzal Hussain: Annotated and classified melodic phrases, including 10 non-classified phrases.

Some of the most characteristic melodic phrases of rāga Jaunpurī appear in the classification of Table 5.6, including double-sloped pitch glides (.../*b*7*b*6, 1/4\\2, 2/5\\4, 5/2'*b*7*b*6), which consist of an ascent and a subsequent descent, the holding of the tonic or the 5th (usually approached through a portamento over an interval of a 5th), and the cadence *b*3-2-1. Pitch glides can be divided into those that are performed in the lower part or in the upper part of an octave. The 5/2'*b*7*b*6 pitch glides specifically are performed in the upper part of the octave, but they move beyond the tonic and into the next octave. It is important to note here, that these glides are not always repeated in the exact same fashion. Therefore, although the typical rendering of a glide would involve both an ascent and a subsequent descent, in some cases only one part (only the ascent or only the descent) of the glide is performed; this is discussed in Appendix A in more detail. The melodic phrases which coincide with two consecutive gestures include four occurrences of double-sloped .../*b*7*b*6 pitch glides, with one gesture each for the ascending and the descending parts, and four occurrences of a pitch glide (predominantly double slope), the starting note of which (annotated as 'other' - 'single note') is performed through a separate gesture. Thus, in practice the performance includes 29 double-sloped .../*b*7*b*6 pitch glides (only three cases are single ascents and four are single descents), while four 'single notes' belong to the 'other' unclassified category (no specified pitch). The ten unclassified ('other') melodic phrases consist of individual cases that do not appear with any significant frequency throughout the performance and involve either non-classified melodic phrases (of non-typical pitch intervals) or individual notes that are part of a melodic phrase but are performed with a separate gesture.

'Other' melodic phrases Hussain	number
single note (of various pitches) [2,1,1,7,1,2]	6
mel. movements of non-classified intervals [5/*b*6\4, 2/*b*6\5, 5\2, 1/1' glide]	4

Table 5.7. Afzal Hussain: Unclassified ('other') melodic phrases.

The remaining 26 āndolan groups (excluded from the analysis) refer to the oscillation on a single note of one of the following phrases:

Āndolans (oscillations) Hussain	number	percentage (%)
.../*b*7*b*6	14	53.8
1/4\2	6	23.1
2/5\4	4	15.4
Other	2	7.7

Table 5.8. Afzal Hussain: Undulation (oscillation) melodic phrases.

The findings suggest that groups of undulation gestures are always associated with āndolans on the final note of the main gesture/melodic phrase. Due to the extremely short duration of each individual phase of an undulation group and the higher risk of uncertainty, this category is excluded from the analysis.

5.2.4. Gesture–sound timing comparisons

Although melodic phrases are annotated within the time boundaries of the gestures, they are not necessarily aligned, which means that they may start before and end after the corresponding gesture. Table 5.9 and Table 5.10 capture the temporal relationships between melodic phrases and corresponding gestures.

Duration of voice vs. gesture Hussain	
shorter	25.50%
equal	23.49%
longer	51.01%

Table 5.9. Afzal Hussain: Voice vs. gesture durations.

Beginning of voice vs. gesture Hussain	
after	75.00%
in synchrony	4.55%
before	20.45%

Table 5.10. Afzal Hussain: Beginning of voice vs. gesture.

Overall, gestures tend to be shorter than corresponding melodic phrases and to precede the voice.

5.2.5. Effort–melodic phrase & octave range associations

To address the question of whether physical effort might be related to pitch height and the higher mechanical strain required for producing high pitched sounds (or extremely low pitches), Table 5.11 captures (un-weighted) mean values of physical effort levels for each of the three octaves used in the performance.

Effort vs. melodic phrase & octave Hussain	octave		
	low	medium	high
other	8	2.6	1
tonic - steady & hold	8	3	6.3
5th - steady & hold		7	
1/4\2 (no descending glides)		5.9	
2/5\4 (no descending glides)	4.5	6.3	
.../*b*7*b*6 (no descending glides)	6.7	6.9	
5/2'*b*7*b*6 (no descending glides)		9.3	
mean of each octave	**6.8**	**5.9**	**4.2**

Table 5.11. Afzal Hussain: Effort vs. melodic phrase by octave.

By looking at the mean values it seems that there is no clear significance in the association of effort with the octave range, although some tentative conclusions could be drawn; it seems that the production of low pitches requires higher levels of effort and strain in the voice than pitches of the middle (the comfort pitch area of the singer) and high octave. However, the performer spent only a limited amount of time in the high octave (only five movement events exist), which does not allow any safe conclusions for the high pitch region. The trend observed by the mean effort values in requiring higher effort levels for producing low pitches is also supported by the effort values for the establishment of the tonic; indeed, higher values are observed for the low and high octaves rather than the middle. However, not all melodic phrases seem to follow this pattern. Thus, whether physical effort is somehow related to the mechanical strain of voice production for high and low pitches is an open question that needs to be investigated further by looking into other performances and rāgas by Afzal Hussain.

A stronger association is observed between effort and the interval of ascending melodic movements (or the ascending part of double-sloped pitch glides), as can be seen in Figure 5.5.

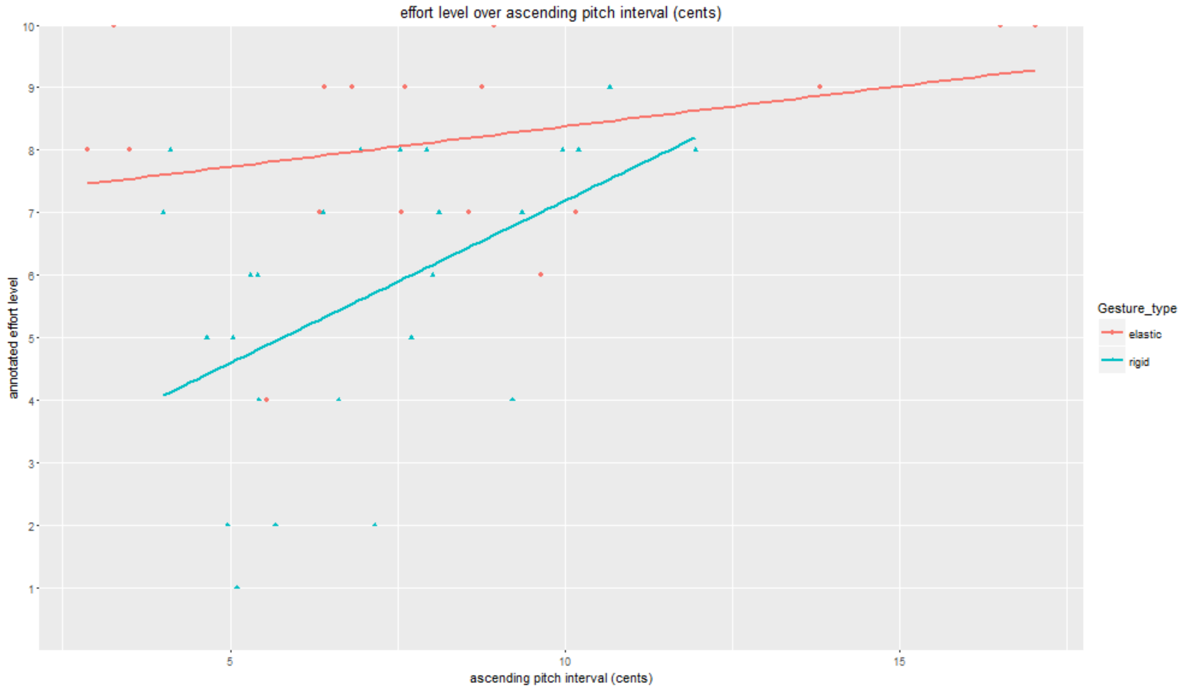


Figure 5.5 Afzal Hussain: Effort level vs. pitch interval (cents)

Next, I analyse the associations between gesture types and various aspects of the melody. As there were a number of ambiguous gesture classes that made the movement—sound association analysis difficult to present, I have decided to divide the analysis in two parts: the main one that refers to the clear-cut gesture classes and the more detailed one that deals with ambiguous gesture classes. All the main results (of the clear-cut gesture classes) are presented in this chapter, while the more detailed side of the analysis (for the ambiguous gestures) is presented in Appendix A1, the results of which support and confirm the trends observed by the clear-cut gesture types presented here.

5.2.6. Gesture class (unambiguous)–melodic phrase associations

Table 5.12 displays results of the association analysis between the 49 pairs of clear-cut gestures and melodic phrases.

Gesture vs. melodic phrase Hussain	Melodic phrase							
	octave range							
	lower part			5^{th} degree	upper part		tonic	unclass.
	3-2-1 cadence	1/4\2 glide	2/5\4 glide	5^{th} approach & hold	.../*b*7*b*6 glide	5/2*b*7*b*6 glide	tonic approach & hold	other
stretching					8	2		
pushing				3	4		1	
pulling		5	5		5		1	2
throwing								4
collecting/taking	1						2	2
grasping/grappolo and holding							3	
grasping/kite-flying and holding							1	

Table 5.12. Afzal Hussain: Unambiguous gesture class vs. melodic phrase.

The information on the clear-cut cases included in Table 5.12 suggests the following:

1. There is a tendency for stretching and pushing movements to appear only in the upper part of each octave and in particular with .../*b*7*b*6 or .../2*b*7*b*6 pitch glides. Pushing gestures are also associated with holding the 5^{th} degree of the scale.

2. Pulling gestures seem to be equally distributed over the entire range of an octave and to be associated with the three main pitch glide classes.

3. Both grips—the kite-flying and the grappolo—as well as the collecting gestures are associated with approaching the tonic and keeping it tightly in place. In fact, kite-flying grips were often followed by a gentle and very slow movement in one direction while the hands were kept together, a movement which was associated with a gentle and almost imperceptible unidirectional movement of the tonic by a few cents in pitch.

4. The 3-2-1 cadence does not tend to be associated with any particular gesture class.

Although this analysis reveals some fundamental trends in the association between gesture classes and melodic phrases, it is not yet clear under what circumstances each of the gestures of stretching, pushing and pulling appear in the upper part of the octave.

5.2.7. Gesture class (unambiguous)–melodic phrase & melodic intention associations

In order to further disambiguate these classes, the musical context is also taken into account. This means taking account of the musician's intention to rise towards the tonic following these melodic phrases and gestures.

Table 5.13 illustrates the intention of the melody in rising towards the tonic (the tonic above the glide itself) after a .../*b*7*b*6 pitch glide.

Gesture class vs. melodic intention after .../b7\b6 Hussain	total number	Melodic intention of reaching the tonic	
stretching	8	6	no / fall
		2	yes / rise, but after great hesitation
pushing	4	3	no / fall
		1	yes / rise, but after great hesitation
pulling	5	4	yes / rise, but after great hesitation
		1	no / fall

Table 5.13. Afzal Hussain: .../b7\b6: unambiguous gesture class vs. melodic intention (to tonic).

Table 5.14 displays the trends of associations between unambiguous gesture classes and the melodic intention in moving to the tonic after a .../*b*7*b*6 pitch glide, as deduced from Table 5.13.

Trends: gesture types vs. mel. Intention in .../b7\b6 Hussain	melodic intention of reaching the tonic
stretching	no / fall
pushing	no / fall
pulling	yes / rise, but after great hesitation

Table 5.14. Afzal Hussain: association trends of .../b7\b6 unambiguous gesture class with melodic intention (to tonic).

Similarly, the associations for the 5/2'*b*7*b*6 pitch glides are:

Gesture class vs. melodic intention after 5/2'\b7\b6 Hussain	total number	melodic intention of reaching the tonic	
stretching	2	2	no / fall

Table 5.15. Afzal Hussain: 5/2'\b7\b6: unambiguous gesture class vs. melodic intention (to tonic).

Table 5.15 clearly reveals that 5/2'*b*7*b*6 glides, after briefly moving beyond the tonic, descend and finally resolve on a lower 5th. This observation highlights the descending character of rāga Jaunpurī.

A similar analysis for 1/4\2 glides is captured in Table 5.16. None of the melodic phrases reach the tonic, so instead I examine here whether the melody moves towards the tonic, from the lower to the upper part of the octave.

Gesture class vs. melodic intention after 1/4\2 Hussain	total number	melodic intention of reaching towards the tonic (moving from the lower into the upper part of the octave)	
pulling	5	3	yes / rise to upper part of octave (*b7*)
		2	no / rise to upper part of octave

Table 5.16. Afzal Hussain: 1/4\2: unambiguous gesture class vs. melodic intention (towards the tonic).

Although at this stage there is no significant bias in the melodic intention toward the tonic that follows 1/4\2 glides performed with pulling gestures, I will return to this issue when I discuss variations in the rendering of the melodic phrases in Appendix A.

A similar analysis for the 2/5\4 glides reveals the following associations:

Gesture class vs. melodic intention after 2/5\4 Hussain	total number	melodic intention of reaching towards the tonic (moving from the lower into the upper part of the octave)	
pulling	5	4	yes / rise to upper part of octave (*b7*)
		1	no / rise to upper part of octave

Table 5.17. Afzal Hussain: 2/5\4: unambiguous gesture class vs. melodic intention (towards the tonic).

According to Table 5.17, pulling gestures with 2/5\4 glides display a strong ascending character in rising into the upper part of the octave.

Table 5.18 captures the overall trends of gesture classes in respect to melodic phrases and melodic intention in moving to or towards the tonic.

Gesture class vs. melodic phrase & melodic intention Hussain	Melodic phrase & intention					
	octave range					
	lower part of octave		5th	upper part of octave		tonic
	1/4\2	2/5\4	5	.../*b*7*b*6	5/2'*b*7*b*6	1'
stretching				fall	fall	
pushing			approach & hold	fall		
pulling	rise to upper part of octave	rise to upper part of octave		rise to tonic		
throwing						
collecting/taking						approach & hold
grip (kite-flying or grappolo)						approach & hold

Table 5.18. Afzal Hussain: Unambiguous gestures class vs. melodic phrase and melodic intention.

Table 5.18 indicates that pulling gestures tend to be followed by rising melodic phrases into higher regions (either the upper part of the octave or the tonic), whereas stretching and pushing gestures are associated with descending onto a lower stable note (the 5th).

Overall, the association analysis between gestures and melody reveals a tendency for melodic activity in the upper part of the octave to be performed with either stretching or pushing gestures without an intention to reach the tonic (usually these glides descend and resolve upon the 5th) or by pulling gestures which follow an ascent to the tonic. Melodic activity in the lower part of the octave is strongly associated with pulling gestures that tend to be followed by an ascent into the upper of the octave (toward the tonic that is higher than the actual phrase). Pulling gestures performed with .../*b*7*b*6 glides tend to be repeated several times before finally ascending to the tonic; it looks like this hesitation is used as a way to build up tension and anticipation.

5.2.8. Effort–gesture class (unambiguous), melodic phrase & melodic intention associations

Table 5.19 shows (un-weighted) mean effort values for each of the gesture-melodic phrase occurrences.

Effort levels vs. gesture class & melodic phrase Hussain	**Melodic phrase & effort levels (mean values)**					
	octave range					
	lower part of octave		**5th (single)**	**upper part of octave**		**tonic (single)**
	1/4\2	**2/5\4**	**5**	**.../*b*7*b*6**	**5/2'*b*7*b*6**	**1'**
stretching				8	9	
pushing			7	7		
pulling	4.25	6.2		6		7
throwing						
collecting/taking						1.5
grasping/grappolo and holding						4
grasping/kite-flying and holding						6

Table 5.19. Afzal Hussain: Mean effort levels vs. unambiguous gesture class & melodic phrase.

By bringing together the information from Table 5.18 & Table 5.19, the following table is obtained, which combines gesture classes, melodic phrase types, melodic context and mean effort levels:

Gesture class vs. melodic phrase Hussain	Effort (mean values), melodic phrase & melodic intention					
	octave range					
	lower part of octave		5^{th}	upper part of octave		tonic
	1/4\2	2/5\4	5	...\♭7\♭6	5/2'\♭7\♭6	1'
stretching				8 fall to 5^{th}	9 fall to 5^{th}	
pushing			7 approach & hold	7 fall to 5^{th}		
pulling	4.25 rise to upper part of octave	6.2 rise to upper part of octave		6 rise to tonic		7 approach & hold
throwing						
collecting/taking						1.5 approach & hold
grasping/grappolo and holding						4 approach & hold
grasping/kite-flying and holding						6 (medium) approach & hold

Table 5.20. Afzal Hussain: Effort vs. unambiguous gesture class, melodic phrase & melodic intention.

Table 5.21 summarises all findings of unambiguous gestures, melodic phrases and melodic context according to the range of required effort levels.

GESTURE	MELODIC PHRASE & INTENTION	
stretching	…/b7\b6 tonic	
pushing	…/b7\b6 tonic	
pulling	1/4\2, 2/5\4 upper part of octave	…/b7\b6 tonic
throwing	other	
grip (kite-flying or grappolo)	tonic - steady and hold	
collecting/taking	other	tonic - steady and hold

Table 5.21. Afzal Hussain: Unambiguous gesture class vs. melodic phrase class according to effort levels.

As we shall see in the next section, these trends that were deduced from analysing clear-cut gesture cases are confirmed by the analysis of ambiguous gestures and special cases of pitch slope types. Final results drawn from combining all annotated information, from both clear-cut and ambiguous cases, are discussed in detail in the summary section of this chapter.

5.3. Discussion

The first qualitative study of the thesis focuses on a performance of rāga Jaunpurī by Dhrupad vocalist Afzal Hussain. It is worth first discussing and then summarising the findings of this analysis in order to use them as a point of comparison with other performances in later sections.

- The kite-flying grip is used for grasping a note and holding it firmly in place with no fluctuations. This is also the gesture used at the opening of the performance and when establishing the tonic.

- The grappolo grip is used as if to hold the note with the fingers, while it is sustained in place through a straight and gentle airflow, while the hands are moved apart. The voice does not sound as 'tight' as with the kite-flying gesture.

- Stretching is the most effortful and most frequently used gesture type. It is solely used in the upper part of each octave and specifically with .../♭7\♭6 pitch glides. It is never associated with an intention to rise to the tonic. Perhaps it is useful to imagine the parallels between the manipulation of a deformed object with elastic properties and how the notes are approached with the descending character typical of rāga Jaunpurī. The opposing force that can be imagined acting against the expansion of an object with elastic properties is perhaps what drives the performer to exert such a high level of effort that the goal pitch is initially transcended and 'pulled' up to the 7th degree, forming the ascending part of the glide. This pitch is then left to be pulled back during the accelerating recoil phase caused by the recoil force in the opposite direction (towards the original position) to the 6th degree. Finally, it resolves due to inertia upon the 5th degree, which is the resting point in rāga Jaunpurī. According to how forcefully the performer decides to act on the imagined object, the highest pitch of the glide (the 7th degree) can vary in microtonal intervals by several cents, just as the displacement of the deformed elastic object is variable. The various tones or micro-tones of the pitch glide interval on which emphasis is placed might possibly be reflected by a force agent (the performer) acting in a variety of ways and intensities upon the elastic object, by fighting against or giving in to the opposing force that is defined by the object's physical properties (how stiff the object is supposed to be). Although the stiffness of the imagined elastic object is not transparent, the kinetic properties of the movements allow the observer to deduce this information indirectly. The concept of an increasing recoil force (and thus required effort) with displacement may also explain the failure of .../♭7\♭6 pitch glides to reach the higher tonic. Thus, it can be argued that stretching is an active process of controlling time and accentuation of pitch (time and 'weight' according to Mohi Bahauddin Dagar, as discussed in section 4.2.3.), a kind of acceleration or gradient of the pitch slope. Analysis of motion capture data may be useful in investigating whether, how and to what extent the emphasis in the voice resembles the dynamics of the body movements.

- Pushing is performed as a variation of stretching, with the performer effortfully (medium-high levels of effort) "pushing away" notes in the upper part of the octave towards the tonic, but against an opposing force that prohibits the final ascent and pushes it back down to the 5th degree. It may also be the case that this opposing force is felt internally within the hand muscles.

- Pulling is less effortful than stretching. It is performed with medium levels of effort. It reflects the displacement of an object in space when moving against a constant imagined force (friction/gravitation) and it is largely related to the intention of reaching the tonic when in the upper part of the octave, and the absence of this intention when in the lower part.

- Throwing is associated with a single note that stands out from the rest of the melodic phrases by being higher in pitch.

- Collecting/taking is an action performed with a uniform movement of approaching the tonic without any particular emphasis in its rendering.

In summary, findings indicate a tendency for the most effortful gestures to be performed in interactions with elastic objects (stretching or pushing/compressing) which are performed in association with a double pitch glide that ascends in the upper part of the octave (typically to the 7th degree, which is the most unstable in rāga Jaunpurī, over an interval of three to eight semitones), but does not finally reach the higher tonic; instead, it is followed by a descent upon a lower, stable note (usually the 6th or 5th degree). The next most effortful gestures (medium-high effort levels) are interactions with rigid objects (pulling or collecting/taking) which are again associated with double pitch glides ascending to the upper part of the octave (maximum pitch) and are followed by a rise to the tonic (where melodic tension is released). Even less effortful (medium-low effort levels) are interactions with rigid objects (pulling or collecting/taking) which are associated with melodic movements in the lower part of the octave (ascending over a five semitone interval to maximum the 2nd or 4th degree) and are followed by a rising melodic movement to or towards the tonic (either to the upper part of the octave or to the 5th degree, where the melodic tension is momentarily released). Glides in the lower part of the octave (for pitches that are more stable in the specific rāga) are typically rendered (through pulling gestures) in a lighter manner than those of the upper part of the octave (through stretching/compressing gestures), and are most often used with the intention to rise into the upper part of the octave.

According to the findings, effort is associated with both the highest degree of the scale that is reached by the ascending part of a pitch glide and by the melodic intention to rise to or towards the tonic following the melodic movement under study. Higher effort levels are required for ascending (or the ascending part of double-sloped pitch glides) rather than descending melodic glides. In specific, melodic movements that ascend to the higher degrees of the scale (the upper part of the octave) are associated with higher levels of bodily effort than those reaching only the lower degrees (the lower part of the octave). Higher effort levels are also required when the melodic movement is not followed by the intention to rise to or towards the tonic, while lower effort levels are observed when the melodic movement is followed by an ascent (to the tonic if the movement happened in the upper part of the octave or to the upper part of the octave if the movement happened in the lower part of the octave). Overall, interactions of deforming elastic objects tend to be more effortful than those of moving rigid objects in space.

Thus, not all gestures and melodic phrases are performed with the same level of physical effort, but the observed tendencies indicate that bodily effort levels are more closely related to the rules and the organisation of the rāga as well as the melodic context, rather than to absolute pitch height (requirements of vocalisation). The analysis has revealed three aspects that define the amount of physical effort, which include: (a) the melodic tension of notes according to the specific rāga (higher for unstable rather than more stable notes), (b) the melodic intention of moving towards more stable notes (such as the tonic or the 5th) and (c) the pitch interval of the ascending part of the melodic movement. Additionally, just before the performer finally reaches the middle tonic, on most occasions there is a hesitation, manifest as a repeated use of the *b*7-*b*6 interval. This seems to be used to build up melodic tension and anticipation.

For instance, the powerful stretching/compressing gestures of elastic objects are associated with the characteristic .../*b*7*b*6 double-sloped pitch glide of rāga Jaunpurī, in other words with the increased melodic tension of melodic movements ascending to the unstable 7th degree of rāga Jaunpurī (the highest degree of the scale); tension in this case does not get released by rising to the tonic, as it falls back to the lower 5th degree instead. This observation highlights the descending character of rāga Jaunpurī and may suggest that the melodic tension created by the unstable character[87] of the 7th degree in rāga Jaunpurī requires higher levels of effort from the musician than the approach of the 2nd and the 4th degree, which are more stable pitches in this particular rāga. Medium levels of effort in the approach of the 7th degree are only then used when the performer intends to ascend to the tonic, at which point tension associated with the unstable 7th degree is released. In other words, .../*b*7\~*b*6 glides are performed in a relatively light manner (through pulling) when the intention is to rise to the tonic, or in a much heavier and slow way (through stretching) in cases where the melody is supposed to descend upon the lower 5th; the feeling here is as if not enough effort is exerted for a rise to the tonic. Pulling gestures are related to melodic glides that are performed with the intention of finally releasing melodic tension through a subsequent rise to the tonic or the 5th and are less effortful when performed in the lower part of the octave (more stable pitches) than when performed in the upper part (with the highest pitch being the unstable 7th degree).

Based on these findings and the conceptualisation of melody as an activity performed in the imagined pitch space of the rāga, melodic movements could be ascribed to some force agent (the performer) who acts against or gives in to some resistive force in order to move to higher parts of the octave, and finally to the (stable) tonic. The following is a summary of these imagined resistive forces:

(a) The attempt to rise in pitch in the upper part of an octave is performed against two kinds of opposing force:

(a1) A recoil force applied by a deformed elastic object (stretching or pushing/compressing), when the performer has no intention to ascend towards the tonic (high effort); the higher effort levels occurring in this case can be justified by the strong recoil phase that imposes the return of both the hands and the melody in the opposite direction to the initial phase of the gesture/melodic phrase respectively.

(a2) An opposing friction or gravitational force (pulling) following the decision to arrive at the tonic after some prior unsuccessful attempts (lower effort, as if the opposing force relaxes at this decision to perform the final movement).

[87] Meaning the demand for the melodic tension of the 7th degree to be resolved through a descending motion to a more stable note, such as the 5th.

(b) The attempt to rise in pitch in the lower part of an octave is performed against an opposing friction or gravitation force and is associated with a failure in moving in the higher part of the octave. A recoil force (and therefore also the required levels of effort) is expected to increase with pitch height (distance from the starting point), a gravitational force is expected to decrease and friction to remain constant (apart from a short moment at the beginning of each movement).

Similarly to the fact that a larger amount of effort is required for extending an elastic object over a larger distance, higher effort levels tend to be required for larger ascending pitch intervals. On the one hand this reveals a common underlying concept between the gesture (hand distance) and the melodic phrase (interval size). On the other hand it reveals that effort may also directly reflect the physical effort required by the mechanics of voicing, with larger ascending melodic movements requiring more effort than smaller ones. Thus, these imagined opposing forces reflect the qualities of the melodic movement and they are neither arbitrary nor of the same nature or magnitude over the whole scale, but they reflect the rāga structure and serve the potential needs for melodic expression. This makes the performer switch between object types according to the interactions that these afford, choosing those that are most suitable in expressing the character of the particular region of the scale in the specific rāga (stretching: varying recoil force, pulling: constant gravitation or friction). The following summarises these consistent associations.

5.3.1. Gesture class–sound associations

Results indicate that body gestures are related to their melodic counterparts based on:
(a) the mental organisation of the rāga pitch space and
(b) the cross-modal analogy in the asymmetry between increase vs. decrease (pitch and e.g. distance between the hands).

Interactions with elastic objects are on average more likely to be associated with pitch glides that ascend to the upper part of the octave, while interactions with rigid objects can be performed in both parts of the octave, but they are more effortful in the upper part than the lower. Double-sloped pitch glides are associated with gestures performed in two opposing directions in the physical space, while monotonic pitch glides are associated with gestures performed in a single direction. More specifically:

- Stretching (out) and pushing-compressing gestures of an elastic object are primarily used with double pitch glides in the upper part of the octave and with no intention to ascend to the tonic.

- Pulling gestures of a rigid object are mainly associated with double-sloped pitch glides that are immediately followed by higher pitches; either reaching the tonic when the double pitch glide happens in the upper part of the octave, or rising into the upper part of the octave when the double pitch glide happens in the lower part of the octave.

- Throwing gestures are associated with single, short notes.

- Collecting/taking gestures are performed when approaching the tonic through a monotonic ascending melodic movement without any particular emphasis in its rendering.

- Both types of grip (grappolo and kite-flying) are used for holding and prolonging a note.

5.3.2. Effort–sound associations

The effort level exerted by the vocalist's body is associated with:

(a) The pitch range of the octave, in other words the highest degree of the scale that is reached through the ascending part of the melodic glide (upper or lower part of an octave);

(b) The melodic intention of ascending to or towards the tonic in what follows the melodic movement;

(c) The size (interval) of the (ascending part of the) melodic movement;

(d) The asymmetry between ascending and descending melodic glides (ascent vs. descent in pitch associated with intensification vs. abatement in effort respectively).

5.3.3. Effort–gesture class associations

Interactions with elastic objects (stretching or pushing-compressing) tend to be more effortful than interactions with rigid objects (pulling gestures: medium levels, throwing: low, and the two types of grips: effortless). Hussain has exhibited the most effortful gestures compared to the other three performers analysed in this thesis.

Figure 5.6 captures the overall trends of effort in relation to gesture–sound associations observed in Afzal Hussain's performance:

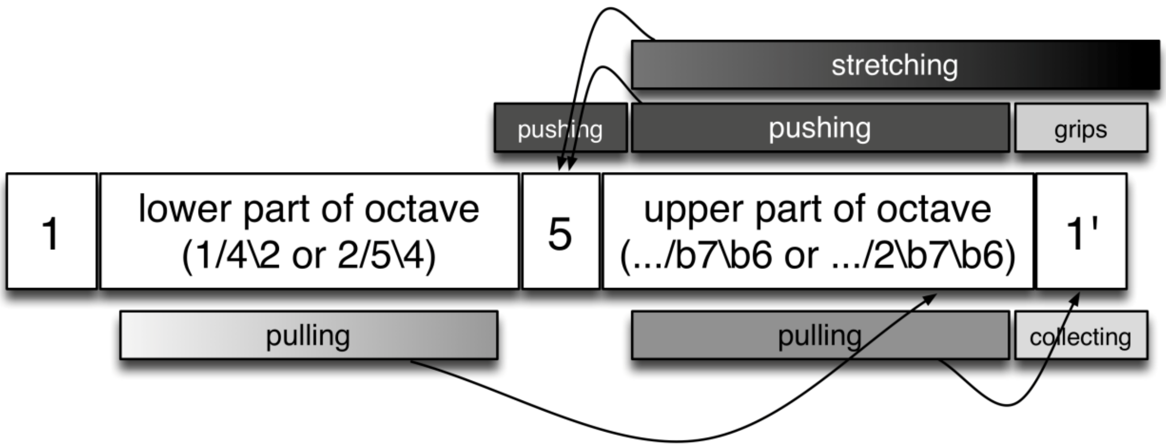


Figure 5.6. Afzal Hussain: Associations between octave pitch range, gesture classes and effort levels. The shades of grey indicate the effort levels and the arrows show the melodic intention. Numbers stand for the degrees of the scale.

Figure 5.7 and Figure 5.8 present the transcription of two short excerpts from the Jaunpurī ālāp performance by Afzal Hussain. The annotated gesture types and effort levels are added on these transcriptions to illustrate these points.

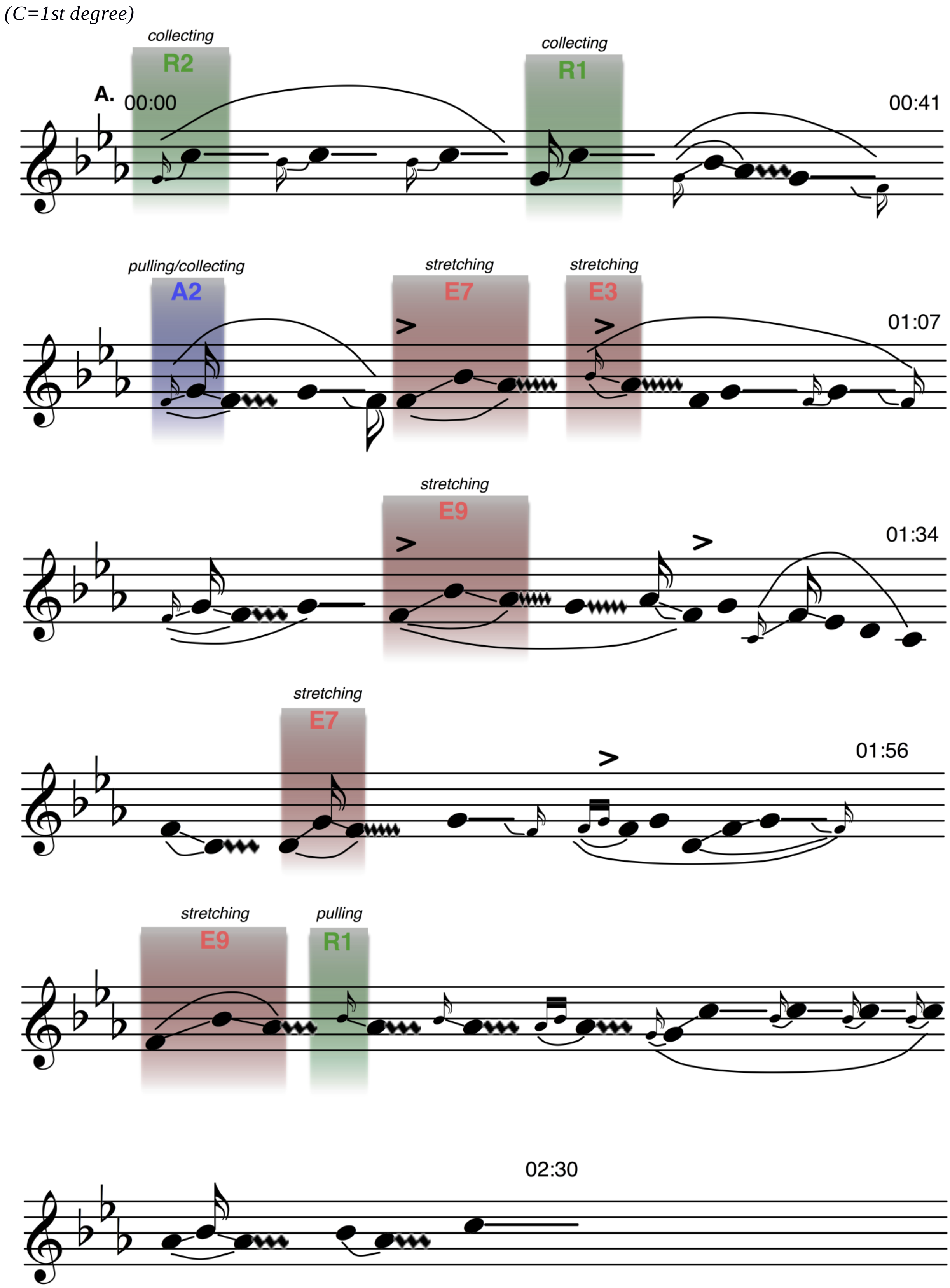


Figure 5.7. Transcription of a short excerpt [A] of rāga Jaunpurī by Afzal Hussain showing gesture types and effort levels. Red shading corresponds to interactions with elastic objects (E), green corresponds to interactions with rigid objects (R) and blue to ambiguous cases (A). The number corresponds to the effort level. The width of each shaded area reflects the duration of the MIIO.

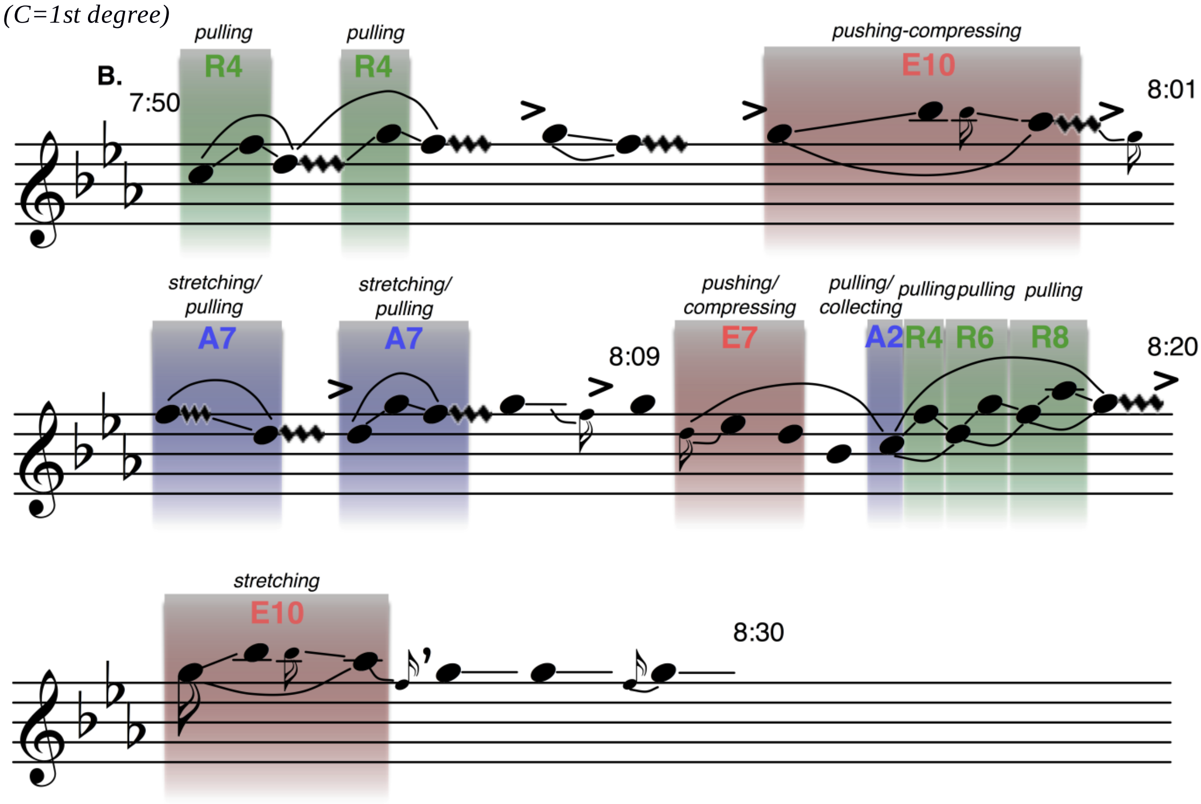


Figure 5.8. Transcription of a short excerpt [B] of rāga Jaunpurī by Afzal Hussain showing gesture types and effort levels. Red shading corresponds to interactions with elastic objects (E), green corresponds to interactions with rigid objects (R) and blue to ambiguous cases (A). The number corresponds to the effort level. The width of each shaded area reflects the duration of the MIIO.

5.4. Summary

Findings for Hussain in rāga Jaunpurī indicate that MIIOs are not combined in an arbitrary way with the melody, and that bodily effort is not uniform over any pitch range. Instead, they imply a systematic association to the pitch space organisation of the rāga. According to these results, physical effort in MIIOs is not only a feature related to the mechanics of voicing and the requirement for higher strain in far high or low pitches, but it is also a more fundamental concept of mental organisation. More specifically, body gestures and the perceived physical effort are associated with particular areas of melodic activity and characteristic melodic movements (glides) of the rāga. Hence, it could be argued that a rāga cannot be conceived of as a uniform space of body activation, but that it comprises pitch-effort regions of potential interactions.

The next chapters aim to examine whether the findings of the current analysis are limited to the idiosyncrasy of the performer and/or the rāga, or alternatively if they may refer to basic concepts shared among different performers and rāgas. When similar gestures are observed in other performances, are they related to similar musical features? This investigation calls necessarily for the analysis of additional performances.

Chapter 6 – Video Analysis II – Lakhan Lal Sahu (rāga Mālkauńś)

Video analysis is presented here for Lakhan Lal Sahu's performance of rāga Mālkauńś. Sahu is a senior student of maestro Zia Fariduddin Dagar who has been trained under guru-śiṣya paramparā. Here again, the analysis consists of first annotating the audio-visual material of the performance and then analysing associations between various aspects of movement and sound. The annotation started by identifying and segmenting MIIO events and then developing a coding scheme from multiple readings of the video. The coding scheme is similar to that of Afzal Hussain. However, some new gesture classes emerged and the melodic content was coded in respect to the characteristics of the specific rāga. Then an association analysis was conducted that examined how systematically various gesture and melodic features appeared together. Specifically, the analysis explored how gesture classes, effort levels, characteristic melodic phrases, the size of pitch intervals, mean and maximum frequencies of melodic movements and elapsed time of the performance are associated with each other. The chapter provides a detailed audit trail of the annotation process and then presents results and findings.

6.1. Analysis

6.1.1. Data collection: Ālāp recording and Interview

The recording session took place at the music school of Zia Fariduddin Dagar in Palaspe/Panvel, India on January 6, 2011 in the very early morning (around 04.00am in order to avoid excessive environmental noise), a snapshot of which can be seen in Figure 6.1. The musician was asked to perform mainly an ālāp improvisation, but also—if willing—a joṛ, for a rāga of his own choice. He was informed that the project is related to Dhrupad and movement. The vocalist performed rāga Mālkauńś for 33 min, 23 of which involved the ālāp, while the remaining 10 were dedicated to the joṛ. The recording includes audio (separate inputs for voice, drone instrument and ambience), video and motion. The taňpurā was played by the musician's wife, who was seated facing the artist. It was tuned by the musician at his most comfortable pitch area, with the tonic at 137.2 Hz, approximately a C#2 in the Western system. The teacher was not present in the room. An interview was conducted immediately after the performance and while the singer was still wearing the reflective markers of the motion capture system. Only the first 23 min of ālāp were used in the analysis.

Figure 6.1. Lakhan Lal Sahu recording session.

6.1.2. Analysis of rāga Mālkauńś by Lakhan Lal Sahu

First, I present the theoretical background of rāga Mālkauńś and then I provide some information on the particular rendering by Sahu. Mālkauńś is based on a pentatonic scale, it has a serious and solemn expressive mood, and is played the late at night. According to Bor (1999), Autrimncpa (2012c) and the Indian classical music portal by Sakis Laios (Laios, 2008) the following are its most important aspects:

1. Tone material: 1, *b*3, 4, *b*6, *b*7 (a pentatonic scale, both in ascent and descent).

2. The rāga is developed mostly in the lower octave.

3. The 4th degree (vādī) and the tonic (samvādī) are the prominent notes of the rāga. The 4th degree is the pivotal tone of this rāga and it is so important and so frequently used that it almost gives the impression of being the tonic. Therefore, the interval 4-1 (a fifth) is used so more frequently than the 1-4 (a fourth), that it becomes more prominent. In fact, the taňpurā is tuned to the 4th rather than the more normal 5th degree.

4. Both the ascent and descent can be direct.

5. Movements are slow and often connected through glides, such as in *b*7*b*6, *b*6\4, 4*b*3, *b*3\1.

6. In slow movements, the 3rd degree is held with an oscillation (āndolan) and the 6th degree can be sustained as well.

7. A notable feature is that in ascending movements the 3rd, 4th and 6th degrees are usually approached from above; the melody first ascends and momentarily moves beyond these pitches before finally descending to reach them.

8. There is a symmetry in the pitch intervals of the two tetrachords, the lower (1-*b*3-4) and the upper (4-*b*6-*b*7). Nevertheless, the melodic movements in the lower tetrachord lead to the highest note (4th), while those in the upper tetrachord typically either lead to its second note (the 6th degree) or ascend to the higher tonic (Laios, 2008).

9. Characteristic phrases: *b*6-*b*7/*b*3'\1, *b*3*b*7-1'*b*6, *b*3-4-*b*6*b*3-4-*b*3\1, 1-*b*3-4-*b*3\1, 4-*b*6-*b*7-1'-*b*6-*b*7-*b*6\4.

Lakhan Lal Sahu follows the structure and organisation that is typical in the development of rāga Mālkauńś. The singer uses a range of about two-and-a-half octaves; he starts from the middle tonic, he moves immediately to the lowest octave and mainly explores its upper part. Then, he establishes the middle tonic and moves into the middle octave by first briefly touching upon and then revealing one-by-one all of the scale degrees of the scale in ascending order. He finally reaches the upper tonic, touches briefly the upper 3rd degree and then moves back to the middle tonic, in order to begin the joṛ (which is not analysed here). Sahu's rendering of rāga Mālkauńś in the performance I recorded is outlined in the following points:

- beginning: short establishment of middle tonic (only once) and immediate fall to the *b*6 of the lower octave through a typical *b*7*b*6 slide.

- beginning – 5.25: exploration of the upper part of the lower octave, usually resolving on the 4th degree of the lower octave or the middle tonic.

[4.53: first touch upon the lowest tonic].

- 5.50 – 10.15: momentary touching and then showing of the 3rd degree of the middle octave, often resolving on the middle tonic. Short descent into the lower octave again.

[7.00: first touch upon the 4th degree in the middle octave].

- 10.15 – 13.15: showing of the 4th degree in the middle octave.

[12.40: first touch upon the 6th degree in the middle octave].

- 13.15 – 16.03: showing of the 6th degree, with occasional resolutions on the middle tonic or showing of the 3rd degree.

[15.50: first touch upon the 7th degree in the middle octave].

- 16.03 – 18.50: showing of the 7th degree in the middle octave.

[17.50: first touch upon the upper tonic].

- 18.50 – 23.00: establishment and showing of the upper tonic.

- 20.38 – 20.45: a few gamaks

[20.45: first (brief) touch upon the upper 3rd degree].

- 21.20 – 22.20: short descent to the middle tonic and ascent again into the upper octave.

- 22.20 – 23.00: brief touches upon the 3rd degree in the upper octave.

[23.00: return/descent to the middle tonic].

- 23.40: joṛ starts.

In the next graph of Figure 6.2 the pitch distribution of the ālāp is plotted, which reflects the tonal hierarchy of rāga Mālkauṅś as rendered by Sahu:

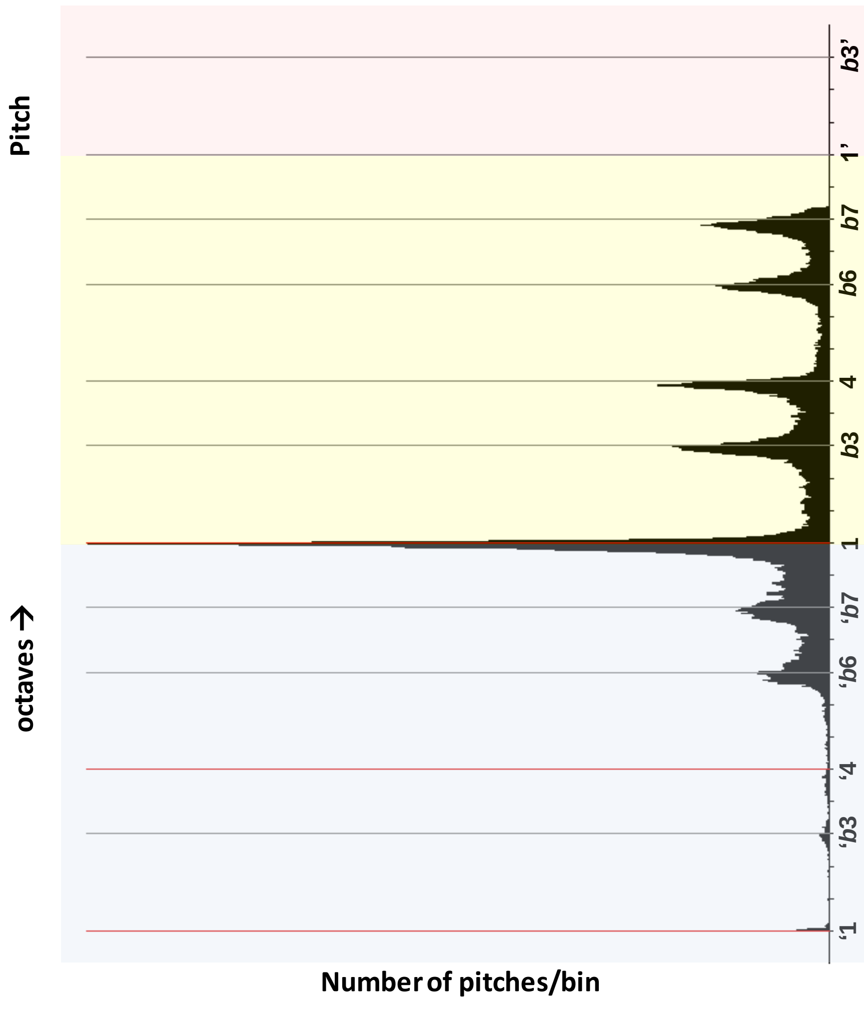


Figure 6.2. Lakhan Lal Sahu: Pitch distribution (in cents) of Mālkauṅś ālāp performance (Praat). The red lines indicate the frequencies of the taňpurā tuning.

Some important aspects of the Mālkauns̈ rendition by Sahu can be drawn from the pitch distribution graph:

- As is to be expected from reference to music theory, the 4th degree is a prominent note, although it is used far less than the middle tonic.

- The lower part of the lower octave and the upper octave are not explored very much.

- In contrast to the theory, ascents and descents do not tend to be very direct. Instead, each time the performer highlights one particular degree of the scale, the approach to this note is organised through a group of chained (consecutive) melodic double-sloped pitch glides, which form an overall stepwise ascent and/or descent toward the pitch in question. The following is an example of how the tonic might be approached: b6/b7\b6, b6/1\b7, b7/1\b6, b6/b7\4, 4/b6\b3, b3/b6\4, 4/b7\b6, b6/1\ b7, b7/1\ b6, b6/1\b7, b7/1. 'Dhuran and muran'[88] is a phrase used in Dhrupad when describing a group of phrases in which a melodic ascent from a lower to a higher emphasised pitch is followed—either immediately or after some intervening material—by a 'balancing' descent (not necessarily concluding on the initial pitch) (Sanyal & Widdess, 2004).

- As expected, there is a plethora of b7\b6, 4\b3 and b3\1 glides. Especially, in line with the theory, the 6th degree is often approached in ascent through the 7th degree, i.e. through a …/b6\b7 double pitch glide.

- Both the b3 and the b6 are sustained. In fact, in the lower octave the b6 is sustained and highlighted to the extent that it receives even more attention than the 4th degree (which is supposed to be the vādī in Mālkauns̈). This is not predicted by the theory. Similarly, in the middle octave, the 3rd degree is highlighted and often sustained.

6.1.3. Coding scheme

The coding scheme was adapted from that used for Hussain. However, here pushing gestures are divided into those referring to the transposition of a rigid mass in space ('pushing-away') and those reflecting the compression of an elastic object ('pushing-compressing'). The following are the features of the coding scheme used for Lakhan Lal Sahu:
{MIIO gesture class; effort level; melodic phrase class; pitch interval}.

The gesture classes used for Lakhan Lal Sahu are listed in Table 6.1, along with links to video files of examples and a video of the entire ālāp performance:

Lakhan Lal Sahu: media file playlist (Mālkauns̈ ālāp performance) **https://n2t.durham.ac.uk/ark:/32150/r1wm117n965**			
6.1.	**ālāp performance**	**times**	https://n2t.durham.ac.uk/ark:/32150/r12b88qc178
6.2.	**stretching**	6:59:390 – 7:01:020	https://n2t.durham.ac.uk/ark:/32150/r108612n54t

[88] From *dhurhna*, which means to explore, and from *murna*, which means to turn back (Sanyal & Widdess, 2004: 163).

6.3.	**pulling**	20:48:350 – 20:49:366	https://n2t.durham.ac.uk/ark:/32150/r1nk322d345
6.4.	**collecting/taking**	21:18:879 – 21:20:553	https://n2t.durham.ac.uk/ark:/32150/r1xw42n791z
6.5.	**pushing-away**	7:42:930 – 7:44:060	https://n2t.durham.ac.uk/ark:/32150/r141687h44h
6.6.	***pushing-compressing***	20:33:826 – 20:34:595	https://n2t.durham.ac.uk/ark:/32150/r13j333224r

Table 6.1. Lakhan Lal Sahu: media file playlist

Pushing-compressing is displayed in italics because it is a distinct gesture type that was not coded as such in the case of Hussain. The reader can refer to section 3.2.2. for a detailed discussion of why this is so.

Melodic phrase classes were the following:
{double pitch glide, only ascent, only descent, gamak, steady}.

Pitch intervals were organised in number of semitones.

6.2. Results

In this section, I first present a discussion on effort levels distribution, as well as the type and frequency of appearance for the annotated classes of gestures and melodic phrases. Then I outline results of the association analysis between various aspects of the performance and the corresponding effort levels.

6.2.1. Effort levels

From a simple visual observation of Sahu's performance, his gestures are notable as the least powerful of all performers I have worked with. Although a few movement events were annotated as effortful, most MIIOs were coded with low effort values. This may reflect the idiosyncrasy of the performer or the solemn mood of rāga Mālkauṅś. Although no systematic annotation has been conducted for the second ālāp performance by Sahu, I tend to believe that the difference lies primarily on the idiosyncrasy of the performer rather than the rāga; however a larger database of performances by the same vocalist in different rāgas would be required to be able to draw safer conclusions. Figure 6.3 displays the histogram of Sahu's effort values.

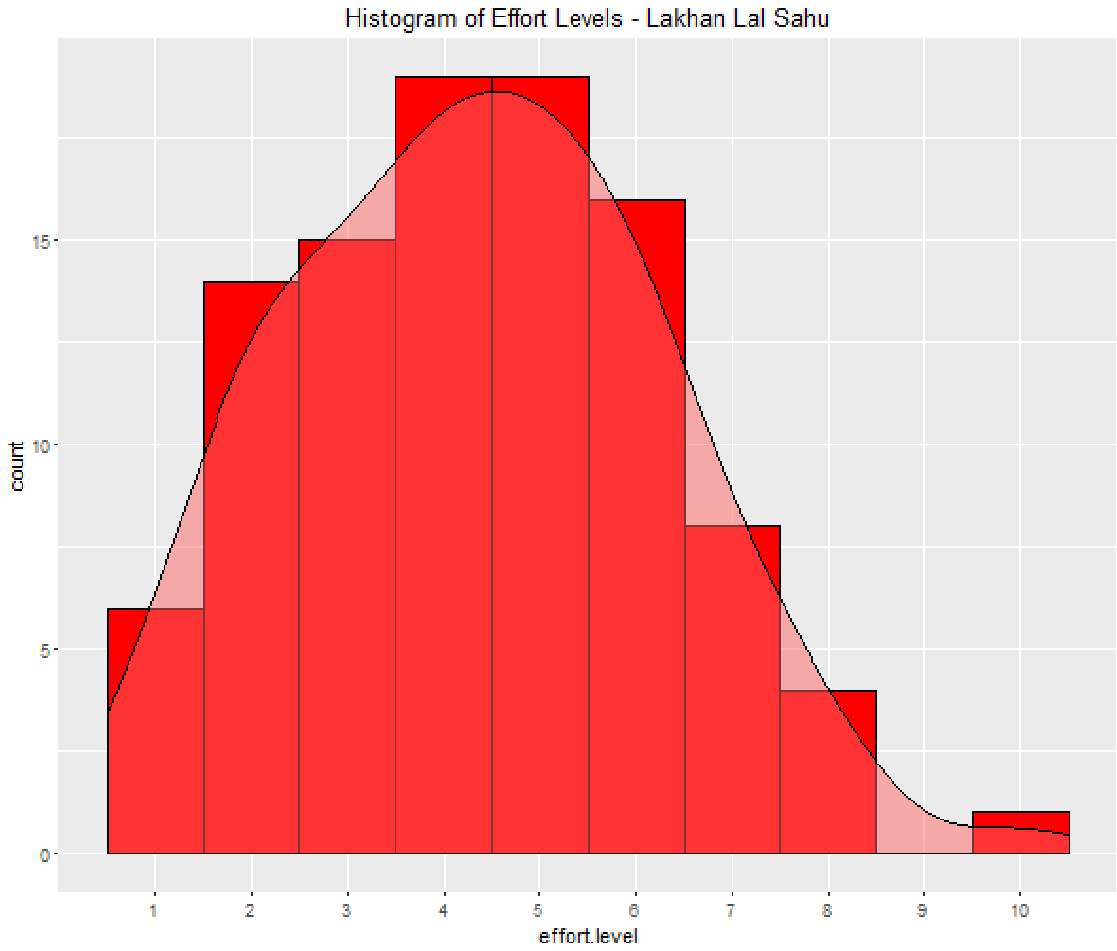


Figure 6.3. Lakhan Lal Sahu: Effort distribution.

The probability curve of Sahu's effort histogram does not deviate extensively from a standard normal distribution. Despite a light skew towards the lower range of values, it can be assumed that the requirement for normality in the distribution of response values is not violated in the linear regression methods used in quantitative analysis. Effort is distributed around values of medium levels (mean=4.3) of the range used by the specific performer.

6.2.2. Gesture classes

The movements repeated most often are groups of chained gestures. In each of these groups there is a pair of consecutive gestures consisting of an outward movement followed by an inward movement: (a) outward: pushing-away/compressing and (b) inward: pulling/stretching. In total, I identified 31 such groups throughout the entire performance. As will be discussed in more detail in section 6.2.4. , these are associated with consecutive double-sloped pitch glides. In each one, a different target note is emphasised each time. The relatively short duration and the high degree of coarticulation of these consecutive gestures often make it difficult to classify them with confidence. It is especially difficult to distinguish between interactions with elastic and rigid objects, due to the difficulty in separating a recoil (following the stretch or compression of an elastic object) and a retraction phase of the hands to the rest position (which naturally follows any gesture). Therefore, a mixed label of 'stretching/pulling' and 'pushing-away/compressing' is used in the majority of these cases. The few clear examples that are finally used in the annotation and analysis refer primarily to isolated gestures which are not part of such gesture groups.

The gesture annotation of the whole ālāp includes 233 movement events, which belong to seven classes of MIIOs and last between 0.33 and 7.11 sec. Table 6.2 captures the frequency of occurrence for each gesture type:

Gesture classes Sahu	number	percentage (%)
stretching/pulling	105	45.06
pushing-away/compressing	82	35.19
pushing-away	2	0.87
pulling	14	6.01
stretching	10	4.29
collecting (or taking)	8	3.43
pushing-compressing	7	3
pulling/collecting (or taking)	5	2.15

Table 6.2. Lakhan Lal Sahu: Gesture classes.

It should be noted here that the pushing-compressing gestures involve the two hands, as if the performer were holding an imagined elastic object in the bottom hand and compressing it with the other hand with a small downward movement from the top. Pushing-away/compressing gestures (ambiguous) are always performed with both hands moving downwards and away from the performer's body, as if the object was placed on the floor.

6.2.3. Melodic phrase classes

The audio annotation includes 233 melodic segments of 39 different melodic phrases, which are further classified into 28 classes of pitch intervals and five coarse classes of melodic movement (double pitch glide, only ascent, only descent, gamak and steady).

Melodic phrases Sahu	1st part of movement	melodic interval	melodic phrase	number of occurrences	total number
double pitch glides	2/	2/ 2\	b3/4\b3	6	185
			b6/b7\b6	20	
		2/ 4\	b7/1\b6	13	
		2/ 5\	b3/4\1	7	
			b6/b7\4	6	
	3/	3/ 3\	1/b3\1	5	
			4/b6\4	1	
		3/ 5\	1/b3\b7	4	
			4/b6\b3	6	
	4/	4/ 2\	b6/1\b7	37	
	5/	5/ 2\	1/4\b3	21	
			4/b7\b6	19	
		5/ 3\	b3/b6\4	12	
			b7/b3\1	12	
		5/ 3\ \2	b7/b3\1\b7	10	
	7/	7/ 2\	4/1\b7	2	
		7/ 3\	b6/b3\1	1	
	12/	12/ 1\	1/1--\b7	2	
	15/	15/ 3\ \2	1/b3'\1'\b7 (1.5 octave)	1	
only ascent	2/	2/	b7/1	2	6
	5/	5/	b3/b6	2	
			4/b7	1	
	12/	12/	1/1'	1	
only descent	2\	1\7	1\b7	2	18

		4\3	4\b3	1	
		7\6	b7\b6	15	
	3\	3\1	b3\1	5	8
		6\4	b6\4	3	
	4\	1\6	1\b6	1	1
gamak	9\	9\ 5/	1\'4/'b6	1	5
		9\ 21/	1\'b3/1'	1	
	7\	7\ 19/	1\'4/1'	1	
	12\	12\ 8/	1\'1/'b6	1	
	17\	17\ 5/	b6\'b3/'b6	1	
steady	-	0	1	2	10
			b3	3	
			4	1	
			b6	3	
			b7	1	

Table 6.3. Lakhan Lal Sahu: Annotated and classified melodic phrases.

As can be seen in Table 6.3, double pitch glides abound in the performance; these are most often interlinked within groups of phrases, like the ones already discussed, rather than appearing as isolated melodic phrases. Also, five gamaks are performed near the end of the ālāp (in two groups, consisting of two and three gamaks respectively), a pitch graph of which can be seen in Figure 6.4.

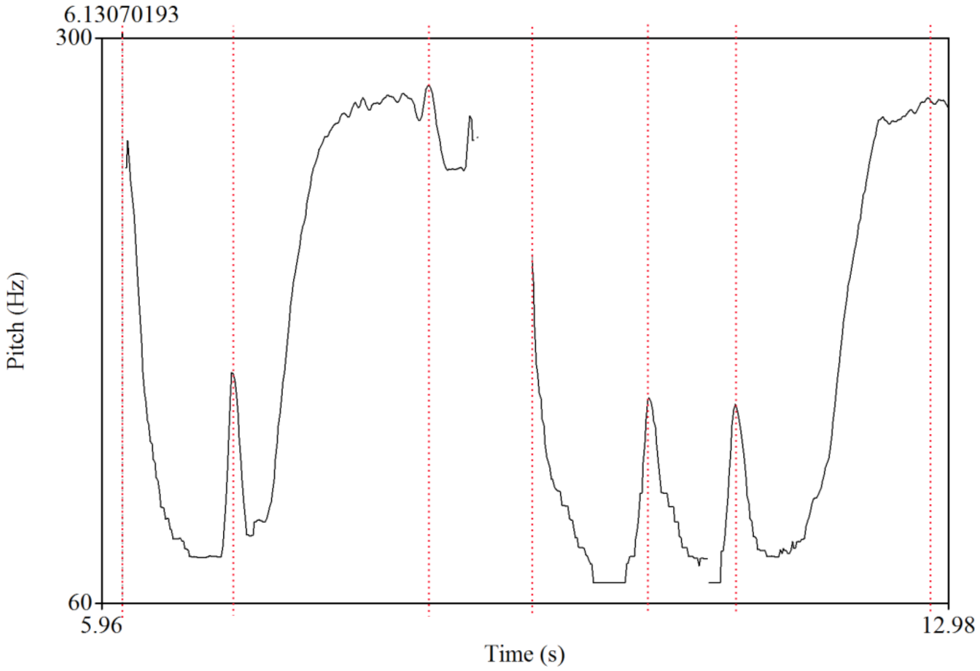


Figure 6.4. Lakhan Lal Sahu: Pitch graph of the five gamaks towards the end of the ālāp performance.

6.2.4. Gesture class–melodic phrase associations

In order to find possible associations between gesture types and melodic phrase classes (in terms of melodic movement and melodic context), the number of occurrences is counted for each gesture class–melodic phrase co-appearance and is presented in Table 6.4:

Melodic phrases vs. gesture class Sahu	gesture class	number	total number
double pitch glide	stretching/pulling	98	185
	pushing-away/compressing	74	
	stretching	9	
	pulling	3	
	pushing-away	1	
straight ascent	pushing-away/compressing	4	6
	pulling	2	
straight descent	collecting/taking	6	27
	pulling	6	
	pulling/(collecting/taking)	4	
	pushing-away/compressing	6	
	stretching/pulling	5	
gamak	pushing-compressing	5	5
steady note	pulling	3	10
	collecting/taking	2	
	pulling/(collecting/taking)	1	
	stretching	1	
	stretching/pulling	2	
	pushing-away	1	

Table 6.4. Lakhan Lal Sahu: Gesture vs. melodic phrase classes.

According to Table 6.4, it is possible to draw the following conclusions for gesture class–melodic phrase associations:

- **Double pitch glides:** From the 98 stretching/pulling gestures appearing in this table as associated with double pitch glides, 46 cases are associated with chained gestures performed with groups of melodic phrases, while the remaining 52 are performed with mukhṛā phrases.

(a) More specifically, the chained consecutive gestures of stretching/pulling and pushing-away/ compressing are heavily associated with double-sloped pitch glides of the type a/b\c (where is b is higher than a and c). Typically these chained gestures start with pushing-away/compressing (moving the hands away from the performer's body) and there is a transition to stretching/pulling gestures (bringing the hands back towards the performer's body) as the melody rises towards the target note of the melodic group. Usually, in the first pitch glides of each group (mostly associated with pushing-away gestures) the last note (c) is lower than the first note (a), giving the impression of an overall descending character to the glides. The remaining glides of the group exhibit an ascending character with the last note of each glide being higher than the first, which in the end leads to the target note of the group (tonic or 3rd). Additionally, the double-sloped pitch glides are longer at the beginning and they tend to become shorter as the melodic line approaches the target note of the group, therefore pushing-away/compressing gestures tend to be longer than the stretching/pulling gestures of the later stages of each group. A typical example can be seen in the pitch graph of Figure 6.5.

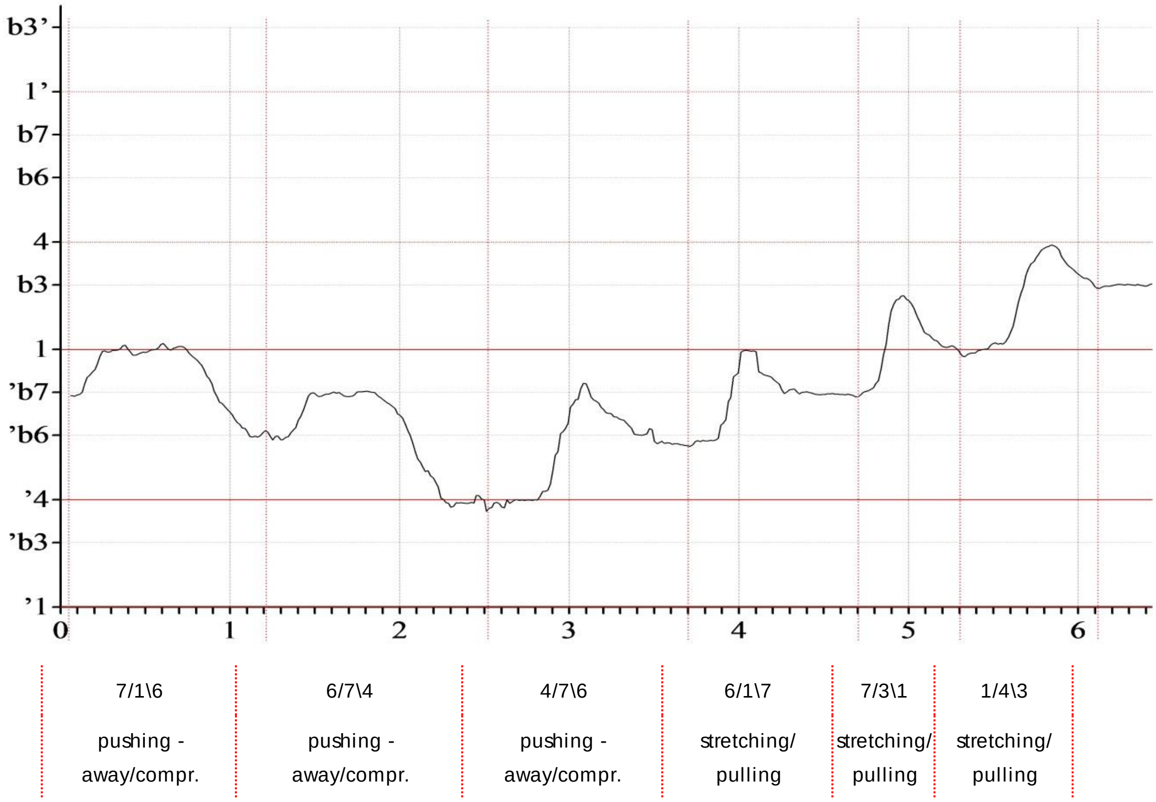


Figure 6.5. Lakhan Lal Sahu: Example of chained pitch glide group and associated gestures.

The impression given by this graph is that the choice between pushing and pulling is associated with the aspects of time and weight (see the interview with Mohi Bahauddin Dagar in section 4.2.3.). In the case of pushing-away, the pitch ascent is longer than the descent and the middle note (the highest pitch) is the one stressed the most. In the case of pulling/stretching, the descent from the middle to the final note is longer than the ascent and the final note is stressed more. There are also some cases that do not include a retract phase and instead a pushing gesture is immediately followed by a pulling gesture. Here the effect seems akin to equally distributing the 'weight' and duration of the melodic glide on the ascent and the descent.

Gestures belonging to groups such as the one given in Figure 6.5 are usually ambiguous due to strong coarticulation phenomena. Despite this, considering the annotation of the entire performance, it is possible to suggest that interactions with elastic objects (such as in stretching or pushing-compressing) are associated with double-pitch glides in which the first part (pitch ascent) coincides with the hands moving away (from each other or the performer's body) and the second part (pitch descent) with the hands moving back (closer to the each other or the performer's body), as in double-sloped pitch glides (of the form a/b\c) and gamaks (of the form a\b/c); 'b' being higher than 'a' and 'c' in the first case, and lower in the second case.

(b) The remaining 52 cases of stretching/pulling gestures belong to a number of groups consisting of three consecutive gestures, which are associated with the last three phrases of a mukhṛā, such as in: *b*6/*b*7*b*6, *b*6/1*b*7 and *b*7/*b*3\1*b*7.

(c) Additionally, there is a small number of isolated (not belonging to a group) stretching gestures associated with double pitch glides, while pulling and pushing-away gestures are used only scarcely.

- **Straight ascents** (without a succeeding pitch descent)**:** Only a few such examples occur and they are performed with either pushing-away/compressing or pulling gestures.
- **Gamaks:** They are clearly associated with compressing an elastic object between the hands. The first part (the compression of the imagined object) is performed with the pitch descent and the second part (the release of the compressed of object) co-occurs with the pitch ascent.
- **Straight descents:** They are performed with a variety of gestures without any particular bias.
- **Steady notes:** They do not show any particular bias toward any type of gesture.

6.2.5. Effort–gesture class & melodic phrase associations

In Table 6.5 (un-weighted) mean effort levels are displayed for the annotated gesture types.

Effort level vs. gesture class Sahu	Mean effort levels
stretching	5.9
pushing-compressing	5.4
pushing-away/compressing	4.5
pushing-away	4
pulling	3.7
stretching/pulling	3.2
pulling/(collecting/taking)	2.2
collecting/taking	1.5

Table 6.5. Lakhan Lal Sahu: Gesture class vs. mean effort levels.

According to this table, gestures that are related to the concept of elasticity as in stretching or compressing a malleable object (stretching: approaching the performer, and pushing-compressing: moving away from the performer) are the most effortful, while gestures that are related to interactions with rigid objects (pulling: approaching the performer, or pushing-away: moving away from the performer) are less effortful. The ambiguous cases follow the values that would be expected if the values of the clear-cut cases were mixed (un-weighted). Gestures related to collecting/taking are annotated as the most effortless gestures, with the ambiguous case of pulling/collecting receiving a slightly higher value, as expected.

The following table displays mean effort levels according to the melody class.

Melodic phrase vs. effort levels Sahu	Mean effort levels
gamak	6.2
straight ascent	4.8
double pitch glide	3.9
straight descent	3.3
steady note	3

Table 6.6. Lakhan Lal Sahu: Melody class vs. mean effort levels.

The mean effort values of each melody class in Table 6.6 indicate that the forceful sound production of gamaks is reflected in the performer's body movements; these account for the most effortful gestures of the entire ālāp. Straight ascents and double pitch glides follow in terms of required effort levels, while straight descents and steady notes are the least effortful melodic phrases in terms of body movements. Nevertheless, the small differences in the mean effort levels do not allow us to confidently justify a strong bias of effort in relation to melodic movement classes.

Table 6.7 displays (un-weighted) mean effort levels vs. gesture and melodic phrase classes.

Melodic phrase vs. gesture class & effort level Sahu	Mean effort & gesture class		
	total mean effort	**gesture class**	**mean effort**
gamak	6.2	pushing-compressing	6.2
straight ascent	4.8	pushing-away/compressing	5.3
		pulling	4
double pitch glide	3.9	stretching	6.2
		pushing-away	5
		pushing-away/compressing	4.4

		stretching/pulling	3.2
		pulling	3
straight descent	3.3	pushing-away/compressing	5.2
		pulling	4.3
		stretching/pulling	3.2
		pulling/(collecting/taking)	2.3
		collecting/taking	1.2
steady note	3	stretching/pulling	4
		pulling	3
		stretching	3
		pushing-away	3
		collecting/taking	2.5
		pulling/(collecting/taking)	2

Table 6.7. Lakhan Lal Sahu: Melody class vs. mean effort levels and gesture classes.

Additionally, (un-weighted) mean effort levels are calculated in relation to the pitch interval of the (first) ascending part of a pitch glide (for those including one) and are displayed in the following table:

Pitch interval vs. effort levels Sahu	**mean effort levels**
2/	3.8
3/	4.6
4/	3.6
5/	3.8
7/	5
12/	6
15/	10

Table 6.8. Lakhan Lal Sahu: Melodic interval of ascending part vs. mean effort levels.

According to Table 6.8, for large ascending melodic movements (of four-five semitones and above) there is a (positive) dependence between the size of the pitch interval (for ascending movements) and the effort levels; higher levels of effort are required for larger melodic movements. Smaller intervals do not show any clear bias.

In order to visually examine possible associations, scatter-plots of effort levels against various acoustic features are produced separately for each gesture class (interactions with elastic vs. rigid objects). The acoustic features were extracted from the audio signal in Praat. The scatter-plots reveal a (positive) dependence of effort levels on:

(a) The elapsed time of the ālāp performance (Figure 6.6);

(b) The mean frequency of each melodic movement (Figure 6.7);

(c) The maximum frequency of each melodic movement (Figure 6.8).

All three acoustic features are related to each other, as in a typical Dhrupad ālāp performance the melody rises towards the climax with time. According to the scatter-plots, effort required by the performer rises in a similar fashion. These illustrations reveal an additional association between effort levels and gesture classes, according to which interactions with elastic objects (red) are more effortful than those with rigid objects (blue).

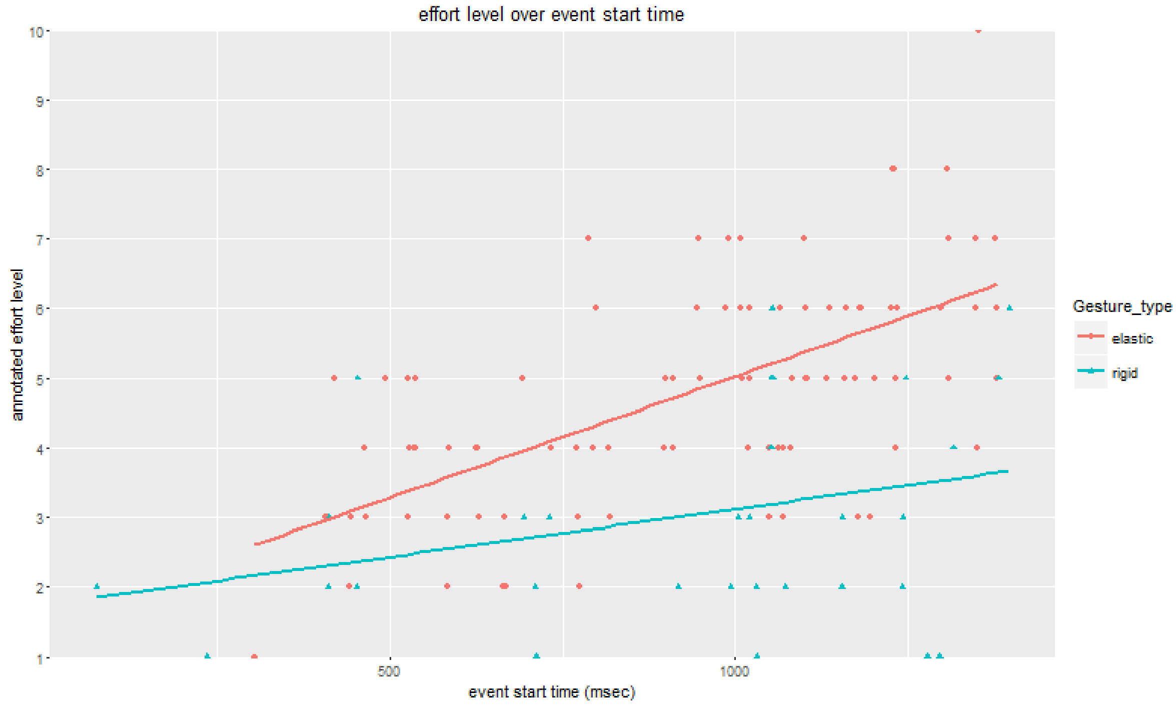


Figure 6.6. (a) Lakhan Lal Sahu: Effort level vs. elapsed time (event start time) according to gesture class.

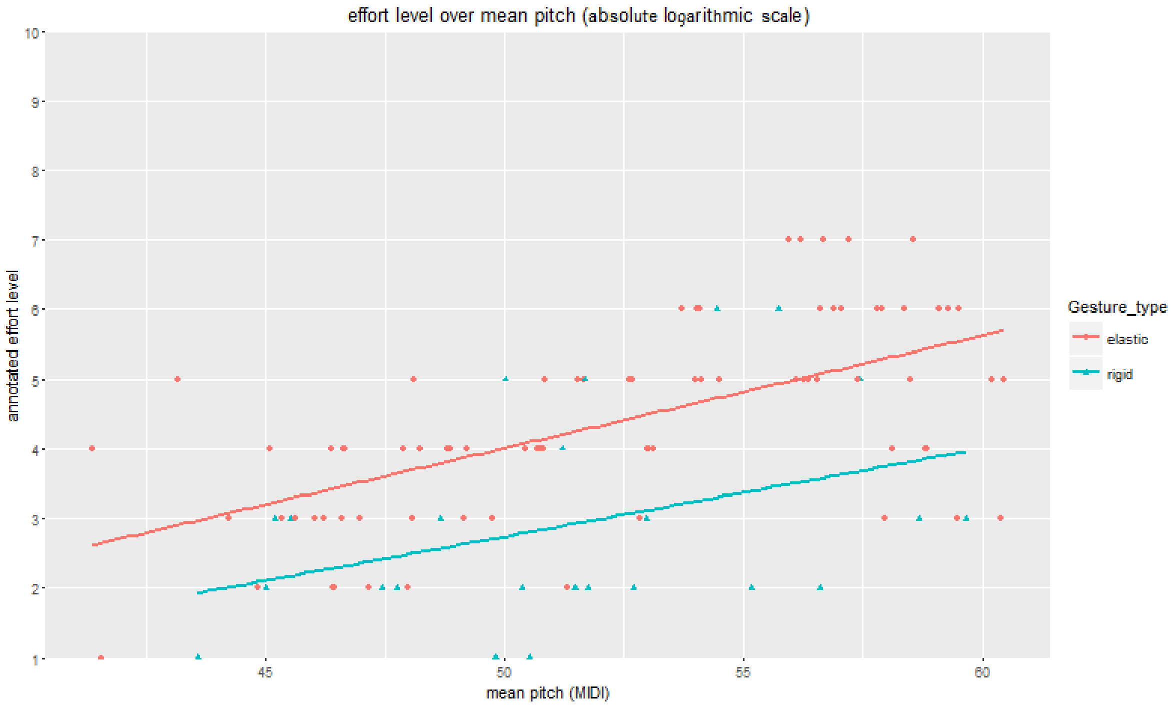


Figure 6.7. (b) Lakhan Lal Sahu: Effort level vs. mean pitch (absolute logarithmic scale) according to gesture class.

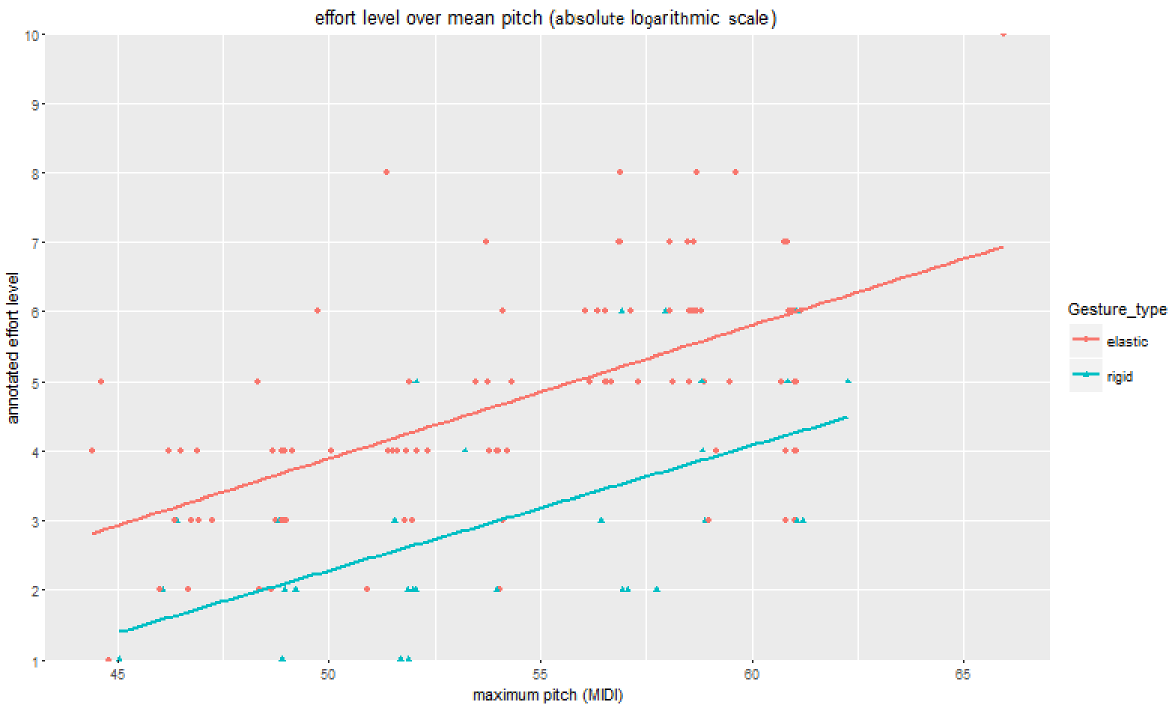


Figure 6.8. (c) Lakhan Lal Sahu: Effort level vs. maximum pitch (absolute logarithmic scale) according to gesture class.

6.3. Discussion

The second study in the qualitative part of this thesis focused on an ālāp performance of rāga Mālkauṅś by Dhrupad vocalist Lakhan Lal Sahu. The findings of the analysis point to some level of consistency in the ways gesture and sound are associated with each other. It is worth discussing the findings of Sahu's performance. The following offers an overview of findings for Lakhan Lal Sahu:

- Gamaks and double-sloped pitch glides are the most effortful melodic movements, they tend to be associated with the concept of elasticity and they can be performed in any of two directions: by moving closer (stretching an elastic object) or away (compressing an elastic object) from the performer's body. In specific, gamaks are associated with the compression and double-sloped pitch glides with the stretching of an elastic object. This confirms interview testimony by Lakhan Lal Sahu in section 4.2.5. , according to whom a mīnd feels like 'pulling a rubber band' and a gamak feels like 'applying pressure'. As already discussed, this conclusion is reached despite the presence of ambiguous kinds of gestures arising in the annotation process (due to the high degree of coarticulation), described as a combination of [pushing-away and compressing] gestures, and [stretching and pulling] gestures. In most cases these are organised in groups of chained melodic phrases and chained accompanying gestures.

- Simple ascents (without a descent) are few and although no safe conclusion can be drawn about their association with gestures, those in which the first note is emphasised and those covering more than an octave are often performed by interacting with an elastic object. Overall, they tend to be performed with medium levels of effort and they are the second most effortful gestures.

- Simple descents do not exhibit any notable bias toward any particular type of gesture and they are less effortful than pitch descents or any other melodic movement.

- Steady notes are performed with a variety of gesture types and effort levels. This is taken to imply that the high amount of resistance felt in these cases is more important than the actual nature and type of the interaction. This is reflected in the note remaining unaltered.

The following summarises association results for the performance by Lakhan Lal Sahu in rāga Mālkauṅś.

6.3.1. Gesture–sound associations

Based on the results, there is a considerable degree of consistency in how melodic phrases and coded gesture classes are associated with each other throughout the entire ālāp performance. This consistency seems to be based on common underlying principles and more specifically on analogous cross-domain morphologies that are shared between movement and acoustic aspects. In particular, melodic movements comprising two slopes (ascent and descent or vice versa) are associated with gestures comprising two opposing phases (intensification vs. abatement, or a change of direction in the physical space, e.g. away vs. closer). For instance, interactions with elastic objects (such as stretching out or pushing-compressing) tend to be mostly associated with gamaks or double-sloped mīṇḍs or with an ascending glide in which the first note is emphasised. In a few cases, they accompany large ascending glides that cover more than an octave, as if the size of the melodic movement is so large that the hands—due to their limited length—fail to demonstrate it through a monotonic movement in a single direction (pulling or pushing), and therefore they need to use gestures that include a second phase; the return to the body. Interactions with rigid objects (comprising only one phase in a single direction) are mostly associated with monotonic descending or small ascending melodic movements.

6.3.2. Effort–sound associations

On average, interactions with objects of elastic properties (stretching/pushing-compressing) tend to be more effortful than gestures of transposing a rigid object in space (pulling, pushing-away or collecting/taking). Ambiguous cases show effort values that might be expected if the results (the effort level values) for each of the unambiguous cases were combined (un-weighted) with each other.

6.3.3. Effort–gesture class associations

Findings indicate a robust relationship of effort levels with:
(a) The elapsed time of the improvisation, hence also the mean pitch height as it progressively ascends towards the final climax (the 3rd degree of the upper tonic);
(b) The mean pitch height and the maximum pitch reached in ascent during a melodic movement;
(c) The pitch interval of an ascent or the ascending part, in case of large melodic movements (over four-five semitones); for small melodic movements this association is blurred.

Thus, in the case of Sahu effort is independent of the specific rāga. The most robust and straightforward associations are reflected on the elapsed time and the gradual pitch rise towards the climax in the development of the ālāp improvisation; with time and as pitch rises, more effort is required by the performer. This does not come as a surprise, as the ālāp in Dhrupad follows a progressive intensification of various musical aspects, such as tempo and pitch height; although not in a linear fashion (Sanyal & Widdess, 2004: 144). Additionally, effort tends to increase with the size of the (ascending part of a) melodic movement. Whether bodily effort is somehow related to the mechanics of voicing and the progressively higher effort demands for the production of higher pitches or whether it is simply a time indicator for tracing the unfolding of the rāga over the course of the ālāp improvisation is not completely clear from the findings. There is additionally a tendency for gamaks to be performed along with the most effortful gestures; ascending glides with medium levels of effort; and double pitch glides, descents and steady notes with less effortful gestures.

6.4. Summary

In conclusion, both effort levels and MIIO classes are not combined in an arbitrary way with their melodic counterpart, but instead findings suggest an association of the body with either the mechanics of voice production or the macro-level structure of the ālāp improvisation. In any case, bodily activity for Sahu does not seem to be associated with the conceptualisation of the rāga as a pitch space with regions of particular interest and potential activity, as was the case with Afzal Hussain.

Chapter 7 – VIDEO ANALYSIS III – GUNDECHA BROTHERS (RĀGA BHŪPĀLĪ)

In this chapter I present video analysis of an ālāp vocal improvisation of rāga Bhūpālī by the Gundecha brothers (Umakant and Ramakant). The Gundecha brothers are two leading exponents of Dhrupad vocal music. They have both received training from vocalist Zia Fariduddin Dagar as well as his brother Zia Mohiuddin Dagar (the renowned rudrā vīṇā player) under guru-śiṣya paramparā in the Dhrupad Kendra[89] in Bhopal and they have always performed together on stage as a duo. Their common background (same music lineage) and musical experiences (always performing together) offers a unique opportunity to examine the impact of idiosyncratic factors in performances of the same rāga. Through the systematic treatment of coded features, it is possible to draw conclusions on whether there are any underpinning qualities that are shared between the two performers, or if, alternatively, movement is primarily idiosyncratic.

The analysis consists again of first annotating the audio-visual material and then performing an association analysis between various aspects of movement and sound. Specifically, the analysis explores how gesture classes, melodic phrases, musical context (melodic intention in rising towards the tonic), and effort levels are associated. Although the coding scheme that emerged from the multiple readings of the video is close to that for Afzal Hussain's performance, some new gesture classes emerged that are not necessarily shared between the two performers of this recording. The melodic content was also annotated in terms of melodic movements (classes) that are characteristic of the specific rāga. The current chapter first includes a detailed audit trail of the annotation process and then it presents the results for each of the performers separately. Findings for the two brothers are compared and conclusions are drawn in the last section of the chapter.

[89] Centre.

7.1. Analysis

7.1.1. Data collection: Ālāp recording and Interview

The recording took place in Aikatan Auditorium, Salt Lake, in Kolkata, India on February 10, 2007, a snapshot of which can be seen in Figure 7.1. There were five people on stage during performance: the two brothers (Umakant and Ramakant Gundecha) as leading vocalists positioned at the front, two taňpurā players behind them (Amita Sinha (also providing vocal support) and Renu Gundecha) and the pakhāvaj drum player (Akhilesh Gundecha) on the left hand side. The taňpurā was tuned by the musician at his most comfortable pitch area, with the tonic at 138.5 Hz, which is approximately a C#2 in Western terms. Only the first, slow part of the ālāp, lasting 25 min, is used in this work.

Figure 7.1. Gundecha brothers recording session.

7.1.2. Analysis of rāga Bhūpāli by Umakant and Ramakant Gundecha

First, I present the theoretical background of rāga Bhūpālī and then I discuss some issues in the way the Gundecha brothers rendered the rāga. The most significant features of rāga Bhūpālī, according to Bor (1999) and Autrimncpa (2012d) are the following:

1. Tone material: 1, 2, 3, 5, 6 (a pentatonic scale, both in ascent and descent).

2. Both the ascent and descent are direct and most of the elaboration in Bhūpālī takes place in the upper part of the lower octave and in the middle octave.

3. The 3rd and the 6th are prominent notes, and phrases often conclude on the 1st, the 3rd and the 5th degrees.

4. A notable feature is that in ascending movements, the 2nd, 3rd and 6th degrees are usually

approached from above, i.e. through a higher pitch. For example, to approach the 2nd degree in ascent, the singer briefly and softly touches the 3rd degree before falling back onto the 2nd degree.

5. In descending movements the 3rd and 6th are frequently connected through smooth glides, as in 5\3 and 2\6/1. In performance a sharpened 4th or a natural 7th can sometimes be used within these glides. These two notes are not typically heard in rāga Bhūpālī, and their appearance hints at the related rāga, Kalyāṇ.

6. Characteristic phrases: 3-2\6/1 and 3-2-5\3.

The following points in the way rāga Bhūpālī was rendered by the Gundecha brothers are worth mentioning:

- In contrast to music theory, double pitch glides are limited and notes are mainly approached through direct ascending glides.

- More specifically, a major part of the ālāp rendering is dedicated to approaching and establishing the most stable pitches, i.e. the tonic, the 3rd or the 5th degree, through repeated ascending monotonic melodic glides of 5/1, 6/1, 3/1, 1/3, 5/3, 1/5 and 3/5. In such glides there is an impression of the first pitch being attracted to the last (the target note); as soon as that is reached, it is held steadily in place.

- As might be expected with reference to music theory, melodic activity is mainly concentrated in the upper half of the lower octave and in the middle octave.

- Not only the 3rd and 6th degrees, but also the 5th and sometimes the 2nd are often approached through a monotonic descending glide.

The singers make use of an almost two-octave range. They start by establishing the middle tonic, then they explore the upper part of the lowest octave followed by the lower and the upper part of the middle octave, before finally rising to the upper tonic. They briefly touch the 3rd degree of the highest octave and descend again to the middle tonic. The general outline of the melodic progression of the ālāp can be summarised as follows:

- beginning – 2:30: approach and establishment of the (middle) tonic starting from the lower 5th or 6th degree (5/1, 6/1) with brief touches upon the 4th degree.

- 2:30 – 4:00: exploration of the upper part of the lower octave, resolving on the middle tonic.

[4:00 first touch upon the 2nd degree in the middle octave].

- 4:00 – 7:30: showing of the 2nd degree (exploration mainly of the upper part of the lower octave).

[6:30: first touch upon the 3rd degree in the middle octave].

- 7:30 – 12:30: showing of the 3rd degree (exploration of both the upper part of the lower octave and the lower part of the middle octave).

[10:30: first touch upon the 5th degree].

- 12:30 – 17:30: showing of the 5th degree (exploration of the lower part of the middle octave).

[16:30: first touch upon the 6th degree].

- 17:30 – 19:00: showing of the 6th degree (exploration of the lower part of the middle octave).

[19:00: immediate move to the upper tonic].

- 19.00 – 22:00: exploration of the middle octave (resolving on the upper tonic).

[22:00: first touch upon the 2nd degree of the upper octave].

- 22:00 – 24:30: exploration of the upper octave, showing the 2nd degree.

[24:00: first touch upon the 3nd degree of the upper octave].

- 24:30 – 25:00: return of the melody to the middle tonic and beginning of the joṛ (not included in this work).

In Figure 7.2 the pitch distribution of the ālāp is plotted, which reflects the tonal hierarchy of rāga Bhūpālī as rendered by Umakant (and Ramakant) Gundecha:

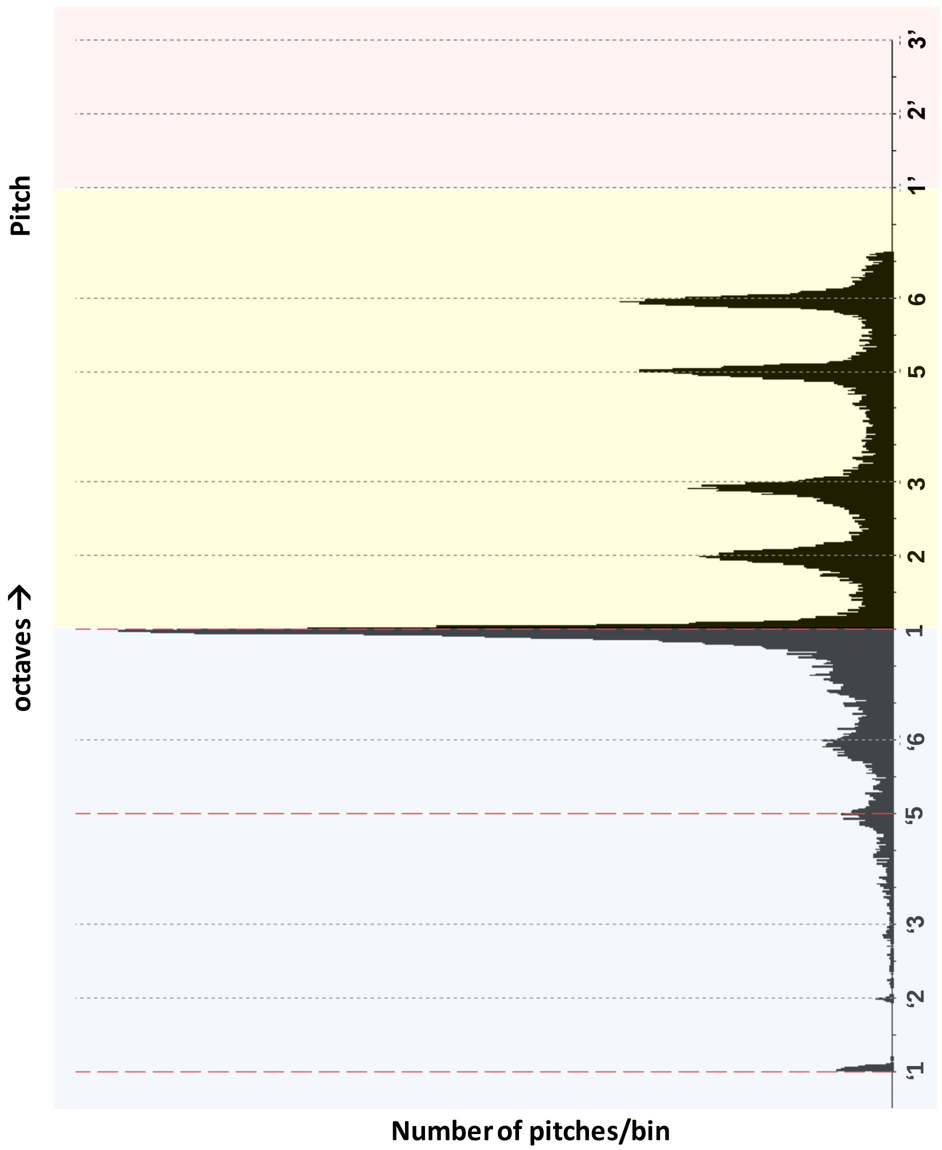


Figure 7.2. Umakant Gundecha: Pitch distribution (in cents) of Bhūpālī ālāp performance (Praat). The red lines indicate the frequencies of the taňpurā tuning.

From the graph, it can be deduced that the ālāp is indeed primarily developed in the upper part of the lower octave and in the middle octave, while the highest octave is explored very little. The importance of the 6th and 3rd degrees can also be inferred, along with the 5th where phrases usually conclude. The graph indicates that the 2nd degree of the middle octave is also prominent in this performance, although this is not to be expected with reference to music theory. The strong link between the tonic and the 6th degree of the scale through descending glides is quite clear, although it seems that the other degrees of the scale are also interlinked through glides.

7.1.3. Coding scheme

The coding and segmentation process was carried out separately for each of the performers. The annotation scheme was adapted from the one used for Hussain, the only additions being two distinct gesture types not observed by Hussain: a *side-glide* gesture by Umakant Gundecha and a *'lifting'* gesture by Ramakant Gundecha. Additionally, in contrast to Hussain (see section 5.1.3.), for the Gundecha brothers I discriminated between 'pushing-away' (a rigid object) and 'pushing-compressing' (an elastic object). The following provides a list of annotated features for each performer.

(a) Umakant Gundecha

The following are the features of the coding scheme for Umakant Gundecha:
{MIIO gesture class; effort level; melodic phrase class; pitch interval; syllable (IPA system)}

The gesture classes used for Umakant Gundecha are listed in Table 7.1, along with links to video files of examples and a video of the entire ālāp performance:

Umakant Gundecha: media file playlist (Bhūpālī ālāp performance) **https://n2t.durham.ac.uk/ark:/32150/r16h440s45b**			
7.1.	**ālāp performance**	**times**	https://n2t.durham.ac.uk/ark:/32150/r15m60qr90s
7.2.	**stretching**	23:09:44 – 23:12:84	https://n2t.durham.ac.uk/ark:/32150/r1f1881k902
7.3.	**pulling**	00:51:28 – 00:53:16	https://n2t.durham.ac.uk/ark:/32150/r1mp48sc76m
7.4.	**collecting/taking**	02:29:52 – 02:30:88	https://n2t.durham.ac.uk/ark:/32150/r1t148fh13d
7.5.	**throwing**	24:08:40 – 24:08:72	https://n2t.durham.ac.uk/ark:/32150/r17h149p845
7.6.	**holding steady**	00:04:56 – 00:15:44	https://n2t.durham.ac.uk/ark:/32150/r1b8515n36g
7.7.	**pushing-away**	03:48:00 – 03:49:40	https://n2t.durham.ac.uk/ark:/32150/r1pz50gw08x
7.8.	***pushing-compressing***	23:59:60 – 24:02:84	https://n2t.durham.ac.uk/ark:/32150/r1qz20ss49f

7.9.	*side-glide*	*passive-passive*	09:17:68 – 09:20:20	https://n2t.durham.ac.uk/ark:/32150/r1pz50gw098
7.10.	*side-glide*	*active-passive*	02:17:28 – 02:19:64	https://n2t.durham.ac.uk/ark:/32150/r1cf95jb468
7.11.	*side-glide*	*passive-active*	00:02:16 – 00:04:56	https://n2t.durham.ac.uk/ark:/32150/r14t64gn17r

Table 7.1. Umakant Gundecha: media file playlist. Gesture types in italics denote that they were not observed by Hussain.

Side-glides were further classified into:
{passive-passive; active-passive; passive-active}.
For a detailed description and discussion on gesture types the reader can refer to section 3.2.2.

Pitch intervals (classified in numbers of scale degrees) and melody phrase classes were also annotated. The following are the coded melody phrase classes:
{double pitch glide; straight ascent; straight descent; one-octave ascending glide; weight & release; single note (no movement)}.

Syllables in the IPA system and palm shape were also included in the coding scheme of Umakant Gundecha. This was as a result of interview reports and observations implying the possibility of a systematic association being revealed between the openness (form) of the left palm and the "openness" of the voice. The left palm can take one of the following shapes: {open; half-open; half-closed; closed}.

The flick of the left hand and the right hand involvement in the gesture (both Boolean values indicating whether they were observed in the beginning of a MIIO or not) are additional movement aspects that were annotated as indicators to facilitate the distinction between throwing and stretching gestures (see section 3.2.2.), that was required for automating computations in the later stages of quantitative analysis.

(b) Ramakant Gundecha

The following are the features of the coding scheme for Ramakant Gundecha:
{MIIO gesture class, effort level, melodic phrase class, pitch interval}

The gesture classes used for Ramakant Gundecha are listed in Table 7.2, along with links to video files of examples and a video of the entire ālāp performance:

Ramakant Gundecha: media file playlist (Bhūpālī ālāp performance) **https://n2t.durham.ac.uk/ark:/32150/r16h440s45b**			
7.12.	**stretching**	11:20:04 – 11:21:84	https://n2t.durham.ac.uk/ark:/32150/r19s1616189
7.13.	**pulling**	19:00:24 – 19:01:96	https://n2t.durham.ac.uk/ark:/32150/r1g732d8964

7.14.	**collecting/taking**	10:58:48 – 10:59:48	https://n2t.durham.ac.uk/ark:/32150/r1fn106x92b
7.15.	**holding steady**	19:43:60 – 19:45:28	https://n2t.durham.ac.uk/ark:/32150/r17d278t00j
7.16.	***lifting***	11:30:60 – 11:31:36	https://n2t.durham.ac.uk/ark:/32150/r15x21tf41k
7.17.	***lifting & letting fall***	19:52:36 – 19:54:32	https://n2t.durham.ac.uk/ark:/32150/r1474299139

Table 7.2. Ramakant Gundecha: media file playlist. Gesture types in italics denote that they were not observed by Hussain.

Coded melody phrase classes are the following:
{double pitch glide; straight ascent; straight descent; 1-octave+ ascending glide; weight & release; single note (no movement)}.

Pitch intervals were organised in numbers of scale degrees.

7.2. Results

First, I discuss findings on the effort level distribution, as well as the type and frequency of appearance of annotated gesture and melodic phrase classes. Then, I report on the results of gesture–sound associations. The two brothers were analysed separately and therefore the results of each are presented individually. Findings are compared only at the end of the chapter.

7.2.1. Umakant Gundecha

(a) Effort levels

Figure 7.3 displays the histogram for Umakant Gundecha's effort values.

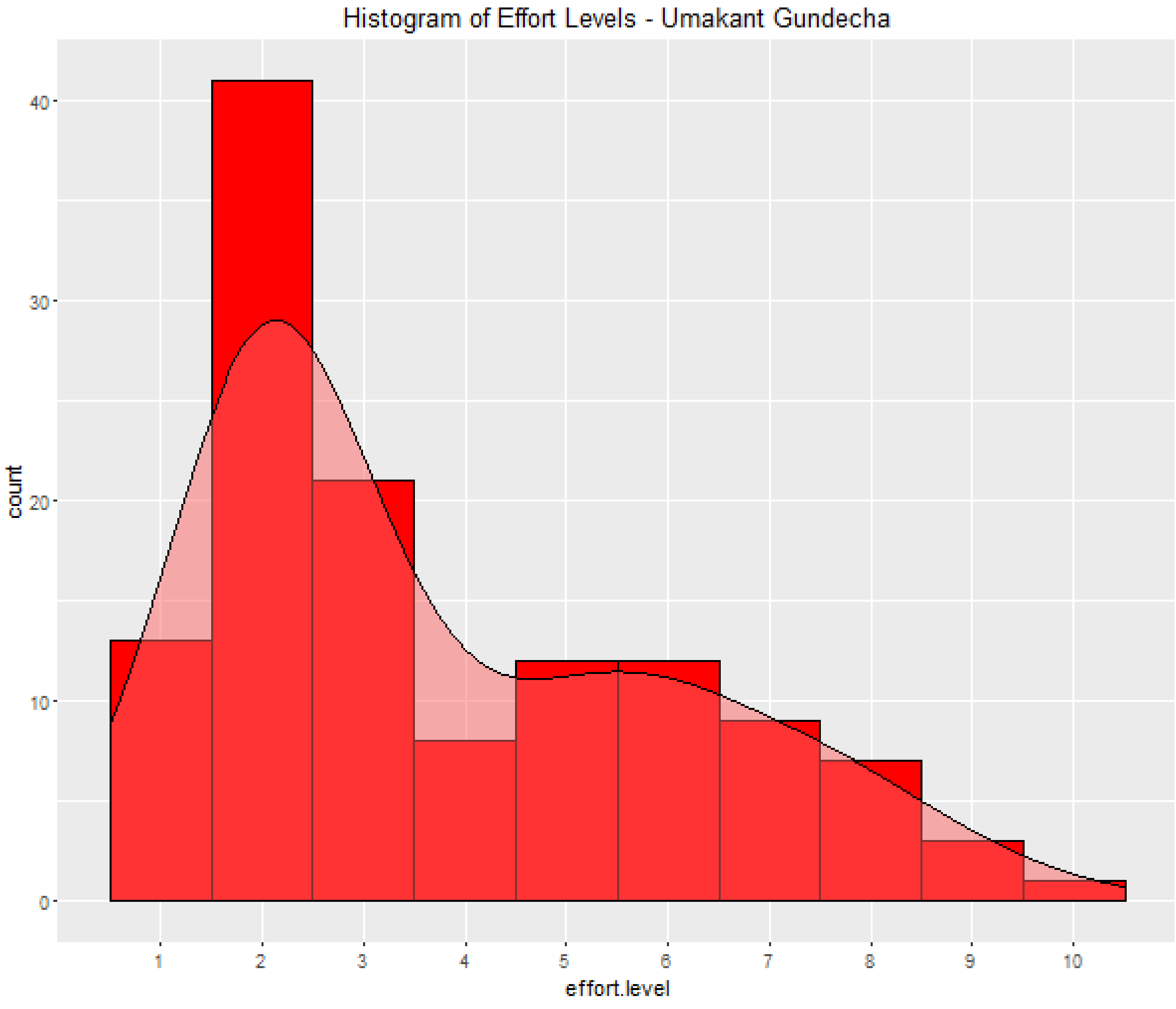


Figure 7.3. Umakant Gundecha: Effort distribution.

The probability curve seems to be skewed towards lower effort values and deviate to a certain extent from a standard normal distribution. Despite the skew, it can still be assumed that for inference purposes the requirement for normality in the distribution of response values is not seriously violated in the linear regression methods used in quantitative analysis. The mean effort value lies at around 3.8, which means that the vocalist mostly uses the middle and lower part of his effort level range.

(b) Gesture classes

The movement annotation of the whole ālāp includes 125 gestures, which are coded in eight classes of MIIOs and last between 0.28 and 14.76 sec. Table 7.3 captures the frequency of occurrence for each gesture type:

Gesture classes Umakant Gundecha		number	percentage (%)
side-glide	passive-passive	20	16
	active-passive	2	1.6
	passive-active	3	2.4
stretching		11	8.8
pushing-away		7	5.6
pushing-compressing		11	8.8
pulling		19	15.2
throwing		3	2.4
collecting (or taking)		22	17.6
holding steady		27	21.6

Table 7.3. Umakant Gundecha: Gesture classes.

Side-glides, the holding of the hands steady and the pulling of an object are the most frequently used gestures throughout the performance by Umakant Gundecha.

(c) Melodic phrase classes

The melodic annotation of the whole ālāp includes 125 melodic segments of 30 different melodic movement types, which are classified into six pitch classes.

Melodic phrases Umakant Gundecha			number	percentage (%)	total
1. straight ascent	.../1	5/1	6	4.8	41
		6/1	15	12	
	.../5	3/5	12	9.6	
	.../3	1/3	3	2.4	
		2/3	1	0.8	
		5/3	2	1.6	
	.../2	5/2	1	0.8	
		1/2	1	0.8	
2. straight descent	...\1	3\1	1	0.8	8
	...\5	1\5	2	1.6	
		2\5	1	0.8	
	...\3	1\3	2	1.6	

	...\6	1\6	1	0.8	
	...\2	1\2	1	0.8	
3. 1 octave+ ascending glide		1/1’	4	3.2	6
		5/5’	2	1.6	
4. double pitch glide		6/1\5	1	0.8	2
		5/1\6	1	0.8	
5. weight & release	(...-/) ^...\...	(3-/) ^5\2	2	1.6	9
		(6-/) ^1\5	1	0.8	
	(...-\) ^.../...	(2-\) ^5/1	1	0.8	
	(...-/) ^.../...	(5-/) ^6/1	2	1.6	
	(...-\) ^...\...	(3-\) ^1\6	1	0.8	
		(3-\) ^2\6	1	0.8	
		(5-\) ^3\2	1	0.8	
6. single note (either short or prolonged)		1	32	25.6	59
		5	15	12	
		3	9	7.2	
		6	1	0.8	
		2	2	1.6	

Table 7.4. Umakant Gundecha: Annotated and classified melodic phrases.

Table 7.4 reveals that the main melodic activity in the ālāp is concentrated around the approach and establishment of the most important and stable notes of rāga Bhūpālī, as can be seen from various types of pitch glides. It is worth considering where melodic emphasis (‘weight’) is placed and how this may be reflected on its movement counterpart. The class called ‘weight & release’ is a melodic movement, the first pitch of which is emphasised as a separate note, followed by a (light) monotonic ascending or descending slide (the release) to another note. This type of glide is different to the usual pitch glides already discussed extensively, because of the distinct emphasis placed on the first note. In the usual pitch glides, emphasis is placed on the target note, which gives the impression of the first note being attracted by the target note during the ascent.

(d) Gesture class–melodic phrase associations

The results displayed in Table 7.5 are produced by performing an association analysis between gesture and melodic phrase classes for the 125 annotated movement events.

Gesture vs. melodic phrase Umakant Gundecha	Melodic Phrase Class					
	straight ascent	straight descent	1 octave+ pitch glide	double pitch glide	weight & release	single note
active-passive (like stretching out)	2					
passive-passive	19		1			
passive-active (like pulling in)	3					
stretching	3		2	2	1	3
pulling	9	2*[90]				6
collecting/taking		4				18
pushing-away	3*					2
pushing-compressing			3		8	
throwing						3
hold steady						27

Table 7.5. Umakant Gundecha: Gesture class–melodic phrase associations.

From this association table, the following conclusions can be drawn:

- The approach to a note through a straight monotonic ascent is most typically accompanied by a hand gesture performed against a constant imagined force, such as a passive-passive side-glide (e.g. throwing and retracting to rest position) or a pulling gesture.

- A melodic descent is usually accompanied by either collecting or pulling gestures, i.e. through gestures directed towards the body.

- Large pitch glides with a range of an octave or more are most often associated with the manual imitation of stretching or compressing an imagined elastic object (therefore with the concept of elasticity). The single exception is a large pitch glide associated with a side-glide gesture (throwing an object from one hand to the other). In fact, this particular side-glide defers from the rest, as it deviates in direction (it was performed upwards and to the front rather than to the side) and instead of retracting immediately after the first part, the hands remained at the same position for some time, as if holding something in place.

[90] The '*' symbol denotes that the number was adapted due to an exception to the usual annotation process. For most of the annotated material, there exists a one-to-one gesture to melody correspondence, but in two cases a single melodic phrase was found to expand over two consecutive gestures. This incident was observed in two pairs of pushing-pulling gestures (aligned with an ascent-descent melodic movement, a 2/3\1 and a 1/2\5), which were removed from the table.

- The two double-sloped pitch glides are associated with stretching gestures, in the same fashion that this is observed for Afzal Hussain.

- The 'weight & release' melodic phrases are associated in most of the cases with abruptly pushing down and compressing an elastic object that naturally expands towards its original rest position. Therefore, they are associated with the concept of elasticity.

- To keep a note steady without any fluctuations, the hands can either hold an imagined object in the same point in space for a longer span of time or bring the object in as if collecting something. The latter is usually the case when this is the concluding note of a larger phrase. Less frequent is the use of pulling, pushing and stretching gestures, while throwing gestures are usually associated with individual rather than prolonged notes that are part of a glide.

To summarise, the following gives an overview of gesture class–melodic phrase associations by Umakant Gundecha:

- The opposition to a varying force as in stretching or compressing (and thus the concept of elasticity) is associated with 'weight & release' phrases, double pitch glides, large pitch glides (one octave or more), as well as straight ascents when the stretching is performed to the side. A few cases also exist of prolonging a note and keeping it steady without fluctuations.

- The opposition to a constant force (as in pulling in (toward one's body) or pushing-away) is associated with either straight ascents or prolonging a note and holding it steady, as well as a few straight descents.

- Collecting gestures as well as gestures of holding something without moving are related to ascending melodic movements and prolonged single notes.

- A few throwing gestures denote individual short notes (not connected through glides).

(e) Palm shape–vowel association in the use of side-glides

One of the most striking features of the side-glide gestures is that they are performed with a variety of palm shapes for the left hand—from completely open to half-open, half-closed and completely closed—implying a parallel between the openness of the hand and the openness of the voice. This observation is supported by the performer's own statements (see section 4.2.3.), as well as by reports of other Hindustani singers (for example Sudokshina Chatterjee reported in Fatone et al., 2011) and is examined here in a more systematic way.

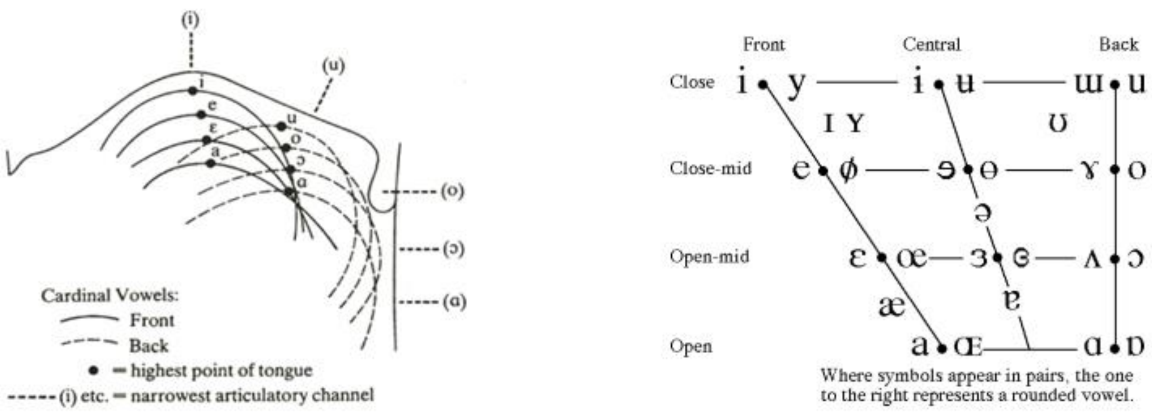
What qualifies as an open voice is perhaps not immediately obvious. According to the International Phonetic Association (IPA, 2012), the articulatory qualities of vowels are dictated by the shape and size of the oral cavity and by the shape of the lips. This means in practice that they are determined by the combination of the following three factors:

Vertical position of tongue (vowel height)	close	near-close	open
Horizontal position of tongue (vowel backness)	front	central	back
Roundness of lips (lips shape)	unrounded (neutral)	protruded	compressed

Table 7.6. IPA vowel features.

- **Vowel height:** It is achieved by the relative vertical position of the tongue in the mouth. ‘Close’ means that it is positioned as close as possible and ‘open’ means as far as possible from the roof of the mouth.
- **Vowel backness:** It is achieved by the relative horizontal position of the tongue in the mouth. ‘Front’ means as far forward as possible and ‘back’ as far backward as possible in the mouth.
- **Lips shape:** It is achieved by the roundness of the lips or degree of aperture between them. It can vary between ‘unrounded’ (with relaxed lips, as in ‘ɑ’), ‘protruded’ (when the corners of the mouth are pressed together but the lips protrude, as in ‘u’) and ‘compressed’ (when the corners as well as the lips are pressed together, as in ‘ɤ’).

The combination of these three factors is determined by the tongue position in the oral cavity, as shown in Figure 7.4.



(a) (b)

Figure 7.4. (a) Tongue position of front and back vowels (Catford, 1988) and (b) Cardinal vowel chart (Zsiga, 2012).

It should be noted here that an ‘open’ (vs. ‘closed’) sound should not be confused with the term ‘open’ (vs. ‘close’) in the cardinal vowel chart. The first is a descriptor of a desirable acoustic quality, while the second is an anatomical descriptor for the vertical distance of the tongue from the roof of the mouth. It should also be noted that the chart above refers to spoken rather than sung vowels. In singing the articulatory position of vowels can vary according to the textural needs of the melodic context and musical style.

It could be assumed that the quality of 'openness' in the voice might be associated with the quality of lax vs. tense. This quality refers to the looseness or tightness of each vowel and it is not achieved by the horizontal position of the tongue but instead by the tension of the tongue.

Table 7.7 presents the results of the association analysis between the vowels of the sung nom-tom syllables (for instance 'om', 'num', 'ta', 'ra', 'na', 'ri') and the shape of the left hand during side-glides. Vowels were annotated based on the IPA[91] standard and they include the 'o' (as in 'om'), the 'u' (as in 'num'), four variations of the 'a' ranging between quite closed to quite open (as in 'nā', 'na', 'rā', 'ra', 'ta') and two variations of the 'i' (as in 'ri').

Hand shape vs. vowel Umakant Gundecha	Close → open vowels							
	Close to mid-close vowels				Mid-open to open vowels			
	'u'	Close to mid-close 'i'		'o'	close → open 'a'			
	u	i	ɪ	o	ə	ɜ	ʌ	ɐ
open	6			2				2
half-open	3			1			1	1
half-closed	1	1	1		3	1	1	
closed	1					1		

Table 7.7. Umakant Gundecha: Vowel (open/close/mid-close) vs. left hand shape (open/closed/half-closed).

In general terms, the results point to a quite straightforward equivalence between the openness of the vowel (in terms of the vertical position of the tongue in the mouth) and the openness of the left palm. In fact, the results are in lines with both the IPA system and what the musician reported about the relationship between hand shape and articulation of vowels in approaching a stable note of the rāga (see section 4.2.3.). More specifically, the open 'a' is mainly associated with an open or half-open hand, while a closed 'a' is associated with a closed or half-closed palm. Both types of 'i' as well as the 'o' vowel are performed with a half-closed palm shape.

Apparently deviating from Umakant Gundecha's assessment in the interview is the relationship of the 'u'—supposedly a closed vowel—with an open or half-open hand shape. He observed that the syllable 'num' should be associated with a closed hand due to the closed sound of 'u'. However, in the interview the musician might be referring to a quality of the voice described as free vs. closed, which is not necessarily a matter of simple anatomical position of the tongue. A free (or open) voice is one that can also be described as round, relaxed, unrestricted, rich and resonant and it might be associated with the lax vs. tense quality of a vowel. This can be adjusted by lifting the soft palate and lowering the larynx through the ideal positioning of the articulators (dropping the jaw, protruding the lips and compressing the tongue downwards). In this way, the size of the mouth cavity is increased

[91] http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/

and a sensation of resonance is felt at a high position in the far back of the mouth. This allows the vocal tract to maintain an ‘open’ posture and prevents tension from developing in the area of the throat. By doing so, a quality of openness, flow or freedom in the voice can be produced for both vowels, ‘a’ and ‘u’, and a ‘throaty’ quality can be avoided. For example, in order to produce an open ‘u’ a common instruction by teachers is to wrap the lips ‘around’ an ‘a’ vowel. On the contrary, variations of the vowels ‘i’ and ‘e’ seem to be more tensed.

(f) Effort–gesture class & melodic phrase associations

Table 7.8 displays mean effort levels against annotated gesture and melodic phrase classes.

Mean effort levels vs. gesture class & melodic phrase Umakant Gundecha	Melodic phrase & mean effort levels						
	straight ascent	straight descent	1 octave+ pitch glide	double pitch glide	weight & release	single note	mean effort /gesture
active-passive	3						3
passive-passive	3.63		5				3.7
passive-active	6.67						6.67
stretching	5.33		7.5	7	8	3.33	5.7
pulling	5.78	4.75				3.5	4.8
collecting/taking		2.75				1.5	1.7
pushing-away	5.2					4.5	5
pushing-compressing			6.67		7.25		7
throwing						2	2
hold steady						2	2
mean effort / melodic phrase	4.94	3.75	6.39	7	7.63	2.81	3.7

Table 7.8. Umakant Gundecha: Gesture class–melodic phrase–effort level associations (numbers of appearance).

According to Table 7.8, imitations of manipulating an elastic object by acting against a variable force (stretching or pushing-compressing) are the most effortful. Next come pulling and pushing-away gestures in which a steady force is acted against (pulling forms the second part of passive-active gestures), while throwing and collecting gestures, as well as the holding of an object in one position in space involve quite low levels of effort.

In terms of melody, ascending phrases of the ‘weight & release’ type, double-slope pitch glides, as well as large ascending pitch glides over a range of an octave or more are associated with the most effortful gestures. A straight ascent, a (small) straight descent or the keeping of a note steady are associated with less effortful interactions with imagined objects.

Effort levels for the annotated movement events are also scatter-plotted in Figure 7.5 against the elapsed time of the performance to examine whether they are associated with the melodic development (progressive pitch rise) of the ālāp improvisation and the pitch height of the voice, i.e. the mechanical effort required for voice production. Scatter plots are produced separately for interactions with elastic vs. rigid objects.

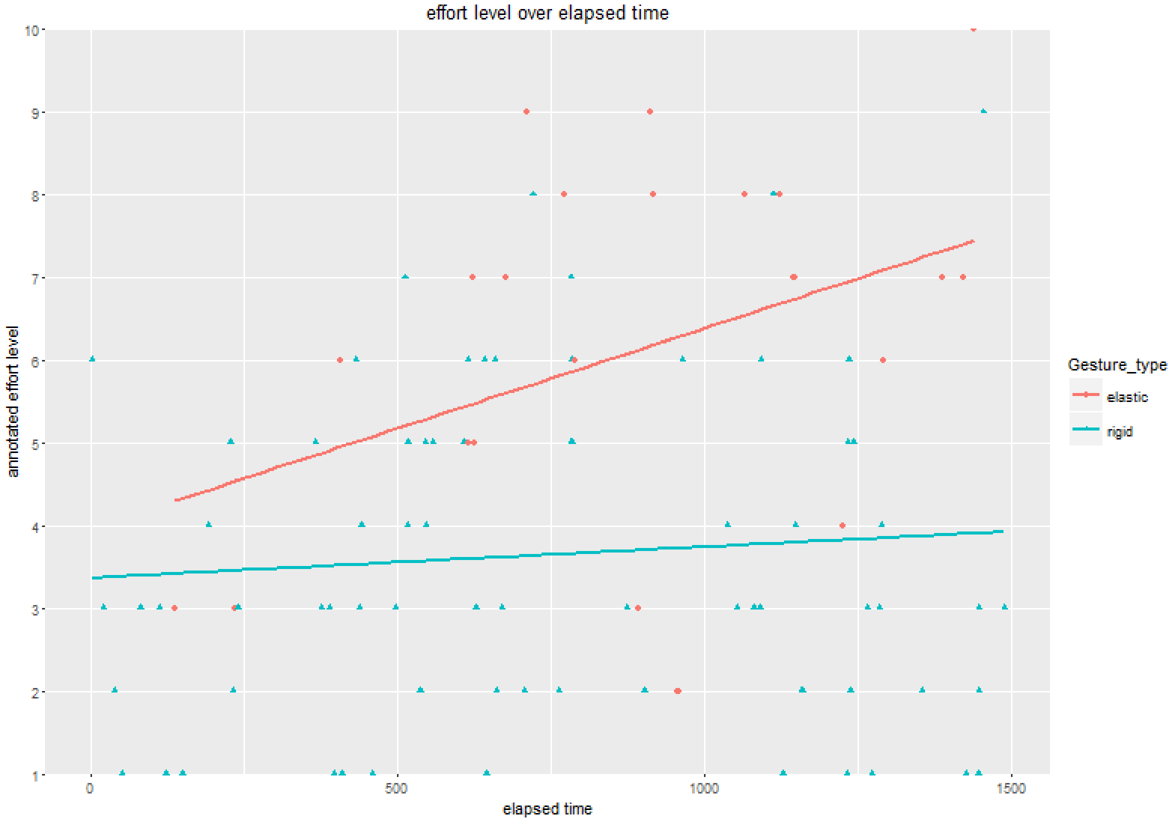


Figure 7.5. Umakant Gundecha: Effort levels vs. elapsed time according to gesture class.

For both types of gestural interaction, effort levels show a positive linear dependency upon elapsed time; however, this phenomenon is clearer for interactions with elastic objects. The trend signifies a monotonic rise of effort level with the progressive expansion of the ālāp improvisation. It may possibly also imply an association with the mechanical strain of voice production. However, in that case I would also expect a rise in effort for both the high and the low end of the pitch spectrum, which is not the case here.

7.2.2. Ramakant Gundecha

This section offers a presentation of results for Ramakant Gundecha. Particularly notable is assessing the extent to which they agree with the results for his brother, Umakant Gundecha.

(a) Effort levels

Figure 7.6 displays the histogram for Ramakant Gundecha's effort values.

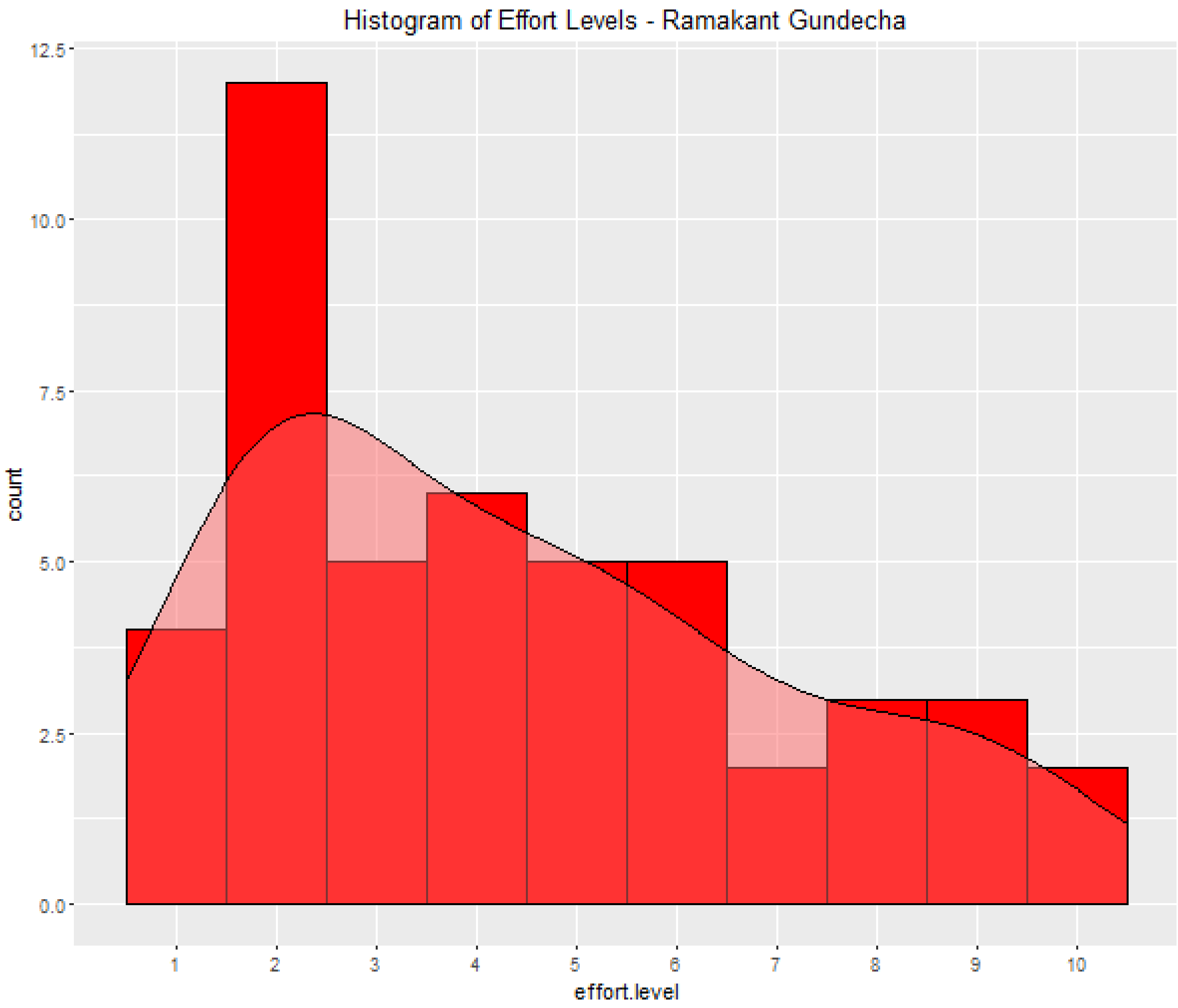


Figure 7.6. Ramakant Gundecha: Effort distribution.

The graph illustrates that the probability curve for the effort levels is skewed and deviates from a standard normal distribution, with its peak appearing in the lower range of effort values. Despite the skew, it can still be assumed that the requirement for a normal distribution of response values in linear regression methods is not violated in the quantitative analysis. The mean effort value lies at around 4.4, which means that, like his brother, the vocalist seems to use mostly the middle and lower part of his effort level range.

(b) Gesture classes

The movement annotation of the whole ālāp includes 47 gestures, which belong to six classes of MIIOs and which last between 0.28 and 14.76 sec. Table 7.9 captures the frequency of occurrence for each gesture type in the case of Ramakant Gundecha.

Gesture classes Ramakant Gundecha	number	percentage (%)
stretching	16	34.04
pulling	7	14.89
lifting	5	10.64
lifting & letting fall	9	19.15
collecting (or taking)	2	4.26
holding steady	8	17.02

Table 7.9. Ramakant Gundecha: Gesture classes.

The stretching of an elastic object, the lifting and the pulling of a rigid object are the most frequently used gestures, along with holding the object in place.

(c) Melodic phrase classes

The melodic annotation of the whole ālāp includes 47 melodic segments of 35 different melodic movement types, which are classified into six pitch classes.

<table>
<tr><th colspan="3">Melodic phrases
Ramakant Gundecha</th><th>number</th><th>percentage (%)</th><th>total</th></tr>
<tr><td rowspan="6">1. straight ascent</td><td>.../1</td><td>5/1</td><td>1</td><td rowspan="6">12.8</td><td rowspan="6">6</td></tr>
<tr><td rowspan="3">.../5</td><td>1/5</td><td>1</td></tr>
<tr><td>2/5</td><td>1</td></tr>
<tr><td>3/5</td><td>1</td></tr>
<tr><td>.../2</td><td>1/2</td><td>1</td></tr>
<tr><td>.../6</td><td>1/6</td><td>1</td></tr>
<tr><td rowspan="4">2. straight descent</td><td>...\1</td><td>2\1</td><td>1</td><td rowspan="4">8.5</td><td rowspan="4">4</td></tr>
<tr><td>...\5</td><td>6\5</td><td>1</td></tr>
<tr><td>...\3</td><td>5\3</td><td>1</td></tr>
<tr><td>...\2</td><td>3\2</td><td>1</td></tr>
<tr><td colspan="2" rowspan="2">3. 1 octave+ ascending glide</td><td>1/1’</td><td>1</td><td rowspan="2">4.3</td><td rowspan="2">2</td></tr>
<tr><td>5/6’</td><td>1</td></tr>
<tr><td colspan="2" rowspan="5">4. double pitch glide</td><td>2/5\3</td><td>2</td><td rowspan="5">42.5</td><td rowspan="5">20</td></tr>
<tr><td>1/5\3</td><td>2</td></tr>
<tr><td>1/5\1</td><td>1</td></tr>
<tr><td>1/3\2</td><td>3</td></tr>
<tr><td>2/3\2</td><td>1</td></tr>
</table>

		3/1\6	1		
		1/2'\5	1		
		1/6\5	1		
		5/6\5	1		
		1/1'\6	1		
		5/5'\3	1		
		5/1\6	3		
		6/1\5	1		
		5/3\2	1		
5. weight on one note and release on another	(...-/) ^...\...	(2-/) ^5\3	1	8.5	4
	^.../...	^5/6	1		
	(...-) ^...\...	(6-) ^6\5	1		
	(...-) ^...\...	(3-\) ^5-(5)	1		
6. single note (either short or prolonged)		1	3	23.4	11
		5	2		
		3	3		
		6	1		
		2	2		

Table 7.10. Ramakant Gundecha: Annotated and classified melodic phrases.

According to Table 7.10, Ramakant Gundecha makes a more extended use of double-sloped pitch glides than Umakant Gundecha (as is to be expected for rāga Bhūpālī, at least for the approach of the 2^{nd}, 3^{rd} and 6^{th} degrees). Single prolonged notes and straight ascents are less frequent in the former's case.

(d) Gesture class–melodic phrase associations

The results from the association analysis between gesture and melodic phrase classes performed on the 47 annotated segments are displayed in Table 7.11:

Gesture class vs. melodic phrase Ramakant Gundecha	Melodic phrase					
	straight ascent	straight descent	1 octave+ pitch glide	double pitch glide	weight & release	single note
stretching	1			11	4	
pulling	2	2				3
lifting	3		2			
lifting-letting fall				9		
collecting/taking		2				
hold steady						8

Table 7.11. Ramakant Gundecha: Gesture class–melodic phrase associations.

From this association table, the following conclusions can be drawn:

- Double-sloped pitch glides, which abound in Ramakant's performance, are performed by either interacting with an elastic object through stretching (similar to Afzal Hussain) or by lifting a solid object and letting it fall (both resemble the two opposing phases of a double-sloped pitch glide). The first, effortful phase of the gesture—when the hands extend by moving away from the body—coincides with the ascending part of the pitch glide, and the effortless recoil coincides with the descending part of the glide.

- The melodic phrases called "weight & release" are associated with stretching an elastic object (and thus the concept of elasticity), as was the case with Umakant. Ramakant Gundecha does not use pushing-compressing gestures (which are also interactions with elastic objects).

- For large pitch glides of an octave or more, Ramakant moves as if lifting an object into the air. This is in contrast to the concept of elasticity that Umakant uses with these phrases, where the first part of a stretching gesture (the extension of the object) is associated with the hand ascending and the second part (the recoil) is associated with letting the object drop down. What is shared between both cases (vertical lifting and horizontal stretching) is the principle of asymmetry in the two opposing poles, with one direction reflecting intensification and the other relaxation. The difference lies in the way effort is exerted by the performer (the agent of the gesture), which signifies the morphology of the intended melodic movement (the pitch contour) in the auditory domain.

- In a similar fashion to Umakant Gundecha, straight ascents are mostly associated with either pulling or lifting gestures (i.e. acting against a constant force in one direction in the physical space).

- Straight descents are associated with either pulling or collecting gestures (again a single direction in the physical space), as is the case with Umakant Gundecha.

- Last, a note can be kept steady by pulling (as if a resistance needs to be imagined in order to avoid any fluctuations in the voice) or through effortless collecting gestures (in which case there seems to exist a melographic relationship of the voice with the gestures).

(e) Effort–gesture class & melodic phrase associations

Table 7.12 displays mean effort levels against annotated gesture and melodic phrase classes.

Mean effort levels vs. gesture class & melodic phrase Ramakant Gundecha	Mean effort levels & melodic phrase						
	straight ascent	straight descent	1 octave+ pitch glide	double pitch glide	weight & release	single note	mean effort /gesture
stretching	4			6.6	7.3		6.56
pulling	2.5					3.3	2.86
lifting	5		5				5
lifting-letting fall				4.3			4.3
collecting/taking		1					1
hold steady						1	2
mean effort / melodic phrase	4.2	1	5	5.4	7.3	2.2	4.4

Table 7.12. Ramakant Gundecha: Gesture class–melodic phrase–effort level associations.

According to Table 7.12, stretching gestures are the most effortful, next come lifting gestures (with or without letting the object fall) and then pulling gestures, while collecting and holding gestures fall into the low range of the effort level scale.

With respect to their melodic counterparts, the most effortful stretching gestures are associated with 'weight & release' melodic phrases. The high effort level is exerted in the first part of the melodic phrase, when weight is placed upon the initial note to emphasise it. Similarly high effort levels are required when performing the first part of double-sloped pitch glides. Large pitch glides and monotonic ascents are associated with the next most effortful gestures, those of lifting an object (only one sample was associated with a stretching gesture).

Effort levels were also scatter-plotted against the elapsed time of the ālāp performance, however no dependency was found.

7.3. Discussion

The third study in this thesis focuses on a performance by two Dhrupad singers, who are brothers and performed the same rāga together. The aim of this study was to examine to what extent there exists consistency across performers during MIIOs in terms of movement, required effort and melodic context.

Before discussing the results in detail, I would like to comment on a generic observation about the gesturing manners of the two musicians. Despite their common background, the two brothers exhibit quite distinct gestural styles in rendering rāga Bhūpālī. For example, although both singers dedicate a large part of the improvisation to the presentation and establishment of the tonic through a variety of ascending glides, the associated gestures exhibit considerable diversity in every aspect, such as the form (e.g. the spatial progression, imagined as a trace in the air), the direction, the type of the grip, and the active or passive attitude of the musician in fighting against or giving in to some apparent resistance. This observation underlines the role of performer idiosyncrasy, which can be explained as both unconscious personal style and conscious, deliberate choice in embodying music. At the same time, for each of the musicians a certain degree of consistency in the movement repertoire is observed throughout the entire performance.

An overview of the most important findings is given in Table 7.13:

	Umakant Gundecha	**Ramakant Gundecha**
Straight ascent	**Associated gestures:** 1. Passive-passive 2. Opposing a constant force by 2a. Pulling 2b. Passive-active (like pulling in the 2nd part) 2c. Pushing-away (like pulling, but in the opposite direction) 3. Elasticity, i.e. opposing a varying force by 3a. Stretching 3b. Active-passive (like stretching to the side) **Effort level:** Medium (average value 4.7)	**Associated gestures:** 1. Opposing a constant force by 1a. Pulling 1b. Lifting 2. Elasticity, i.e. opposing a varying force by 2a. Stretching **Effort level:** Medium-low effort level (average value 4.2)

Weight & Release	**Associated gestures:** 1. Elasticity, i.e. opposing a varying force by 1a. Compressing 1b. Stretching **Effort level:** Most effortful = high (average value 7.6)	**Associated gestures:** 1. Elasticity, i.e. opposing a varying force by 1a. Stretching **Effort level:** Most effortful case (average value 7.3)
1 octave+ (large) pitch glides	**Associated gestures:** 1. Elasticity, i.e. opposing a varying force by 1a. Compressing 1b. Stretching **Effort level:** Medium-high effort levels (average value 6.4)	**Associated gestures:** 1. Opposing a constant force by 1a. Lifting **Effort level:** Medium effort levels (average value 5)
Double pitch glide	**Associated gestures:** 1. Elasticity, i.e. opposing a varying force by 1a. Stretching **Effort level:** Effortful (average value 7)	**Associated gestures:** 1. Elasticity, i.e. opposing a varying force by: 1a. Stretching 2. Opposing a constant force by 2a. Lifting & letting fall **Effort level:** Medium effort level = (average value 5.4)
Straight descent	**Associated gestures:** 1. Opposing a constant force by 1a. Pulling 2. Collecting **Effort level:** Mid-Low effort (average value 3.7)	**Associated gestures:** 1. Opposing a constant force by 1a. Pulling 2. Collecting **Effort level:** At the very low end of the effort scale (average value 1)

Single note	**Associated gestures:** 1. Hold steady 2. Collecting 3. Elasticity, i.e. opposing a varying force by 3a. Pulling **Effort level:** Low end of effort range (average value 2.8)	**Associated gestures:** 1. Hold steady 2. Opposing a constant force through 2a. Pulling **Effort level:** At the low end of the effort range (average value 2.2)

Table 7.13. Umakant - Ramakant Gundecha comparison.

From these findings it can be deduced that despite the idiosyncratic gestural habits and the flexibility in the way the two brothers might demonstrate MIIOs while singing the same rāga, there is a notable degree of consistency in the association of gestures, melodic phrases and effort levels. This indicates that some underlying concepts are shared between the two vocalists. The following summarises the findings that reflect these shared concepts:

7.3.1. Gesture–sound associations

Gesture classes are dictated by the common morphology that they share with their melodic counterparts; double pitch glides are performed with gestures comprising two opposing phases, while monotonic pitch glides are performed with hand gestures of a single phase. In specific:
- Gestures consisting of two opposing spatial phases (such as the stretching and releasing of an elastic object or the lifting and letting fall of a rigid object) are associated with melodic movements that, likewise, consist of two opposing phases; an increase and a decrease. In practice this means either a double-sloped pitch glide (ascent and consecutive descent) or a monotonic ascending glide in which the emphasis is placed on the first note (which coincides with the intensification phase of the gesture that is subsequently released). Sometimes stretching gestures are also performed with large ascending pitch glides (one octave or more). I assume that this is a spatial requirement; a large pitch glide would require a parting of the hands that might even go beyond the length of the arms. A compromise is a stretching gesture that consists of two opposite movements (away and then back towards the performer's body).

- Gestures performed as if acting against an imagined constant force in a single direction in space (such as moving a rigid object without a change of direction, as in pulling or pushing-away) are typically associated with a monotonic pitch glide, such as a straight ascent or descent over a small pitch interval.

- Collecting gestures have a (limited) use and are associated with descending pitch glides (as well as single notes in the case of Umakant Gundecha).

- Holding gestures are employed primarily for prolonged notes.
- Throwing gestures (not observed with Ramakant) are associated with individual, short notes.

7.3.2. Effort–sound associations

For both performers effort levels are strongly associated with sound based on the asymmetry between increase vs. decrease (pitch ascent vs. descent associated with effort intensification vs. abatement). Effort levels tend to be higher for melodic movements that include a large melodic ascent (over an interval of a whole octave or more), those of the 'weight and release' type as well as double pitch glides and lower for descents, small ascents and prolonged single notes. Thus, they possibly reflect the requirements of vocalisation and/or images of instrumental gestures. In playing the rudrā vīṇā, a larger pitch ascent is achieved by pulling a string further away from its rest position (which means raising the tension of the string and the amount of effort required by the performer), while pitch descent is achieved by simply relaxing the hands and releasing the string (which means releasing tension and letting effort drop). This is also supported in interview testimony, when Umakant Gundecha acknowledged the influence of visual engagement with instrumental (rudrā vīṇā) gestures by his former teacher Zia Mohiuddin Dagar on his own gesturing habits (see section 4.2.4.). A similar process can be imagined for the vocal folds of a performer while singing. Additionally, in the case of Umakant Gundecha there is a tendency for effort levels to rise with the elapsed time of the ālāp rendition, i.e. they reflect the progressive intensification in the macro-structure of the improvisation. This phenomenon is independent from the particular rāga.

7.3.3. Effort–gesture class associations

For both performers, gestures associated with manipulating an elastic object (stretching or pushing-compressing) by moving against an opposing varying (with distance) force are overall more effortful than those performed to transpose a rigid object in the physical space by acting against a constant force (pulling, lifting or pushing-away a rigid object, which may be associated the concept of gravity or friction), while throwing, collecting and holding are the least effortful gestures. Thus, on average interactions with elastic objects tend to require higher levels of effort than those involving rigid objects.

7.4. Summary

As in the case of Afzal Hussain, the analysis of the annotated material for the Gundecha brothers revealed that MIIOs are not performed arbitrarily throughout the performance, but that they are associated in a systematic way with the melody. Despite the idiosyncratic gestural habits of each of the brothers, they still share underlying similarities in gestural interactions with objects and in matching effort levels to their melodic counterparts. The gesture–sound associations shared between the two vocalists are mostly defined by analogous cross-modal morphologies and they possibly reflect images of instrumental gestures and the mechanical requirements of vocal production, without any findings implying an association with the pitch space organisation of the particular rāga (as was the case with Hussain).

Chapter 8 – Quantitative Studies

The current chapter aims at exploring whether a small set of movement and sound features can be found so as to enable us to infer the annotated effort levels and gesture classes computationally without the manual labour for each individual performer. Additionally, it aims to assess through comparison whether findings between various data sources and different coders converge. The quantitative methods are applied on two performers (Afzal Hussain and Lakhan Lal Sahu) and focus on:

(1) The estimation of the amount of bodily effort (numerical response) through linear regression.
(2) The classification of gestures in terms of interactions with rigid vs. elastic objects (categorical response) through logistic regression.

The chapter first presents the features and response values that were used for fitting the models. It then offers a detailed presentation of the strategy by which the best models for each individual performer were selected, as a trade-off between model accuracy, compactness and simplicity in interpretation. Due to the explorative character of this work, not only the final best model but also alternative models of comparable goodness of fit to the data are discussed, in order to provide a broad insight to cross-modal phenomena. Slightly less successful but largely overlapping cross-performer models were also devised, that include almost the same number and type of features. By ignoring individual nuances and using a small number of easy to extract features, these models do not describe MIIOs for each individual performer in the most accurate way but they are beneficial in expressing the otherwise complex phenomena during MIIOs in more generic terms. In the comparison part of this chapter both the most successful models as well as the more generic models are discussed in detail.

8.1. Analysis

8.1.1. Response data and Features

This section details the response values and the feature set that were used for the regression models. The response values were drawn from the annotations and include the effort levels and gesture classes. The features were drawn from the core movement and audio feature set that was proposed by Nymoen et al. (2013). However, novel features as well as alternative ways for computing some of the features mentioned in the paper were progressively added and tried on the models, because they were considered to be more appropriate for the specific performance context and were meant to raise the explained variance of the estimated responses.

(a) Response values (from annotations)

The response values that were used in the models as ground-truth—or in other words the correct output values—correspond to the manual annotations. The full body of my annotations was used, which is considered valid due to the cross-validation process (of 1/3 of the data) that was described in section 5.1.4. These annotations were exported from the ANVIL environment for each movement event and they included the perceived level of effort (on a scale between 0 and 10) and the gesture class, along with their time signatures (manually segmented).

However, multinomial logistic regression (response classes>2) requires careful consideration of the sample size. Due to the small size of the datasets that were available per performance (64 for Hussain and 102 for Sahu) and the high number (five) of classes that needed to be inferred, an average of only about 12-20 samples/class would be available for each performance. Even worse, in real-life situations, samples are skewed (i.e. they are not necessarily equally distributed between classes, see e.g. Table 5.4 and Table 6.2), meaning that basic requirements for the number of samples per class would easily fail in being met. For these reasons, the classes to be inferred were reduced to only two by transforming the fine gesture types into two coarse categories, distinguishing between (a) moving a rigid object in space and (b) manipulating an elastic object. Table 8.1 displays the correspondence between fine and coarse gesture classes.

Coarse vs. fine gesture classes	**Fine gesture classes**	
Coarse gesture classes	**Afzal Hussain**	**Lakhan Lal Sahu**
Elastic	stretching, stretching/pushing[92]	stretching, pushing-compressing
Rigid	pushing[93], pulling, throwing, collecting (or taking)	pushing-away, pulling, collecting (or taking)

Table 8.1. Reduction of gesture types from fine to generic classification.

[92] This is the equivalent of 'pushing-compressing' in the case of Sahu. It refers to the stretching of an elastic object away from one's body and is not to be confused with simply 'pushing-away' for Sahu, which refers to moving a rigid object away from one's body.
[93] This is the equivalent of 'pushing-away' for Sahu.

The ambiguous cases of classes and the two grip types (grappolo and kite-flying) were excluded from the datasets. An additional 20% of the data were removed in the case of Lakhan Lal Sahu due to some issues with the frame rate in migrating to a newer version of the proprietary motion capture environment of the NaturalPoint Optical Tracking system (from the older Arena to the newer Motive). This left the following dataset size for the binomial logistic regression:

Data size of generic classes/performer	Generic classes	
	Afzal Hussain	Lakhan Lal Sahu
Elastic	18	72
Rigid	46	30
Total	64	102

Table 8.2. Afzal Hussain and Lakhan Lal Sahu: Data size for binomial logistic regression.

The statistical results of the models will be used only for inference purposes; therefore the non-uniformity in the distribution of samples per class is not considered an issue.

(b) Features (extracted computationally)

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)[94], as illustrated in Figure 8.1. All feature values were first normalised prior to analysis, in order to keep the model coefficients in a limited scale range.

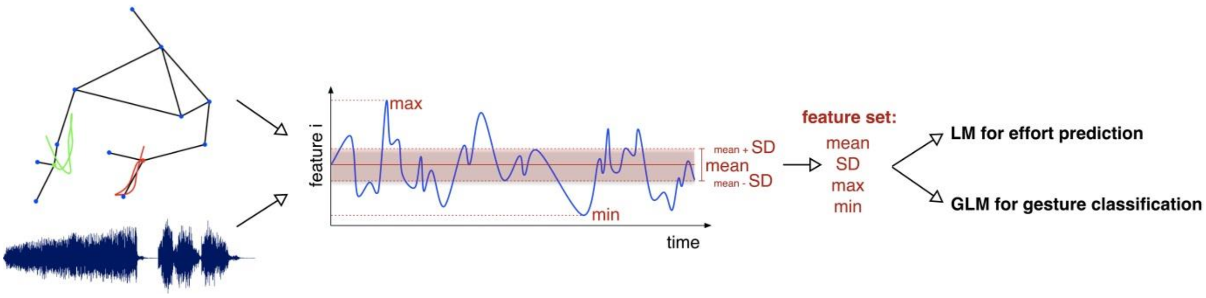


Figure 8.1. Feature extraction.

Time-varying features

Movement descriptors

There is a wide range of sophisticated high-level descriptors that can be computed from the raw (time-series) motion data for music-related experiments, such as kinematic models of the human body (Müller and Röder, 2006). Nevertheless, basic movement features that can be directly extracted from the motion data through derivatives of position coordinates (velocity and acceleration) have proven

[94] Peeters et al. (2011) differentiate between 'global' and 'time-varying' features by processing audio information at different time scales, ranging from local (sample-by-sample) to segments of some seconds, and from physical (low-level) to perceptual (high-level).

robust and pertinent in violin bow strokes (Rasamimanana et al., 2006) and in sound tracing experiments (Nymoen et al., 2012, 2011; Glette et al., 2010). They can therefore provide a sufficient representation of the movement dynamics at the level that is required for developing the regression models. These movement features were extracted using the MoCap Toolbox. As already mentioned, the variables reported by Nymoen et al. (2013) were used as the core feature set, which was progressively enriched by novel features that are being discussed in the following section.

Acoustic descriptors

Although there is a large variety of auditory descriptors, pitch, loudness and timbre are widely considered as the most salient in relation to sound (Peeters, 2004) and are also the ones that were used by Nymoen et al. (2013). These basic features were extracted primarily by using the MIR Toolbox for Matlab (Lartillot et al. 2008). The fundamental pitch was the only exception, as it and was extracted through autocorrelation in Praat (Boersma, 1993) due to the high robustness of the algorithm in detecting the pitch of the voice accurately in the ambience of the drone instrument(s). Loudness was calculated as the RMS value of the audio waveform and timbre as the spectral centroid (the barycentre of the frequency spectrum of the sound). I also tried alternative ways in the MIR Toolbox to represent timbre, such as brightness or roughness, and tried these on the models.

Core set of descriptors

The time-varying features that were exported according to the reference paper by Nymoen et al. (2013) are the following:

Time-varying features	
Motion	**Description**
Vertical position	Distance to the floor
Absolute velocity	Euclidean distance between successive position samples
Absolute acceleration	Euclidean distance between successive samples of the first derivative of position (second derivative)
Hand distance	Euclidean distance between the hands
Audio	**Description**
Pitch	Fundamental frequency (in PRAAT)
Loudness	The RMS value of the sound wave (in the MIR toolbox of Matlab)
Timbre/Brightness	The spectral centroid of the sound wave (in the MIR toolbox of Matlab)

Table 8.3. Core time-varying features, from Nymoen et al. (2013).

Global features

Global features were computed from the time-varying features by applying descriptive statistics. The next section takes the reader through these features. The explanation starts with describing the core set of features used by Nymoen et al. (2013) and moves to the novel and alternative features.

Core features

Movement features

From the time-varying movement features of the reference paper that were presented earlier, the following core movement features were computed[95]:

Core global movement features (Nymoen et al., 2013)	
Name	**Description**
VerticalVelocityMean	The mean value of the first derivative of position, calculated on the mean vertical position of the two hands.
AbsVelocityMean	The mean value of the absolute velocity, calculated on the mean three-dimensional position of the two hands.
AbsVelocitySD	The standard deviation of absolute velocity, calculated on the mean three-dimensional position of the two hands.
AbsAccelerationMean	The mean value of the absolute acceleration, calculated on the mean three-dimensional position of the two hands.
HandDistanceMean	The mean value of the distance between the hands.
HandDivergenceMean	The mean value of the first derivative of hand distance.
HandDivergenceSD	The standard deviation of the first derivative of hand distance.
VerticalEnergyUM	The sum of the signed vertical kinetic energy of a unit mass calculated as: $VerticalEnergyUM = \sum_{k=1}^{n} \lvert v_k \rvert \cdot v_k$, where v_k is the k-th sample of the vertical velocity vector of length n. This feature is associated with the mean value of velocity, but it is computed in such a way as to reflect the asymmetry between the strength of velocities in opposite directions in space. For instance, a movement that consists of a fast upward and a subsequent slow downward movement would receive a positive value, thus showing that the upward direction is executed faster.
OnsetAccelerationMean	The mean of onset acceleration is calculated from the mean position of the two hands for the onset of the movement event. In the current context, there is no designed stimulus, thus the duration of the onset cannot be defined in absolute numbers of samples. Instead, it is computed as a proportion of the whole length of each event (15% of the total duration for events of a

[95] Two features from the reference feature set were excluded, as they are considered irrelevant to the current research context, namely 'symmetry' and 'shaking'.

	duration>1.7sec and 45% for events with a smaller duration). For shorter durations than these, derivatives cannot be computed.

Table 8.4. Core global movement features

Acoustic features

Similarly, from the time-varying acoustic features, the global values of mean and standard deviation were computed. These are not included in the core set of Nymoen's reference paper, as the scope of that study was different to the one of the current thesis.

Core global acoustic features	
Name	**Description**
PitchMean	Mean of fundamental frequency in Hertz (linear scale).
PitchSD	Standard deviation of fundamental frequency in Hertz (linear scale).
LoudnessMean	Mean of loudness in rms.
LoudnessSD	Standard deviation of loudness in rms.
SpctrdMean	Mean of spectral centroid.
SpctrdSD	Standard deviation of spectral centroid.

Table 8.5. Core global acoustic features.

Novel or alternative features

Movement features

(a) In a real performance setting, singers can perform movements with similar qualities and with a number of variations in terms of direction in space and handedness (uni- or bi- manual), even for the same or similar auditory events. For instance, an interaction with an elastic object can be performed by extending it (a) between the hands while moving them apart, as in Figure 8.2a, or (b) only in one direction. In the second case, the object can be extended by either (b1) grasping the object with both hands on one side and moving them together in a single direction (bi-manual gestures), in which case the other side of the elastic object is imagined as fixed in space, as in Figure 8.2b, or (b2) keeping one hand attached to the object and fixed in space, while the other grasps its other side and extends it (uni-manual), as in Figure 8.2c,d,e. Figure 8.2 displays sequences of pictures for these three variations in stretching an imaginary object, while Figure 8.3 illustrates two-dimensional trajectories (ground plans extracted from three-dimensional data) for the same variations (plots produced from the data of Afzal Hussain).

As can be seen from these illustrations, the distance between the hands is not always a good descriptor of the interaction with the object. For example, calculating the distance between the hands for a gesture where the hands move in parallel is not a good measure for the extension by which an elastic body has been stretched from its initial position. Therefore, there is a choice to be made here about whether the 'position' of the hands (and its derivatives) should be calculated as the mean of the two hands regardless of the type of gesture or if only (what seems to be) the hand(s) participating to the movement should be taken into account and used for the calculations. Similarly, there is a decision to be taken about whether the ‘HandDistance’ and the ‘HandDivergence’ features should be calculated as a measure of the distance between the two hands regardless of the handedness of the gesture.

In (Nymoen et al., 2012), which is also supposedly followed in their later paper (Nymoen et al., 2013), the authors computed position according to the active hand participating in the movement (i.e. the position of the active hand for uni-manual gestures or the mean position of both hands for bi-manual gestures), while they calculated the distance between the hands regardless of the type of movement. In this work, it made more sense to explore all possibilities. Therefore, in the basic set of features that I used, I calculated the mean position from the two hands (as a measure of ‘position’) and the distance between the two hands (as a measure of ‘distance’) without taking into account which hand is active. Later on, I also calculated features that take into account the specific type of gesture.

Through a close visual inspection of these cases it can be suggested that there is a kind of 'reference point' in relation to which the movement is always performed (considering the fact that there is supposedly an imaginary object that is manipulated by the hands). In the case of a single active hand or both hands moving together from one point to another, this reference point can be imagined as the beginning point of the movement (single position for uni-manual gesture and mean of two hands for bi-manual gesture), while for the hands moving apart the reference point for the one hand is the position of the other. The following describes the way 'HandDistance' was calculated according to the classes of bi-manual interactions that were observed and annotated:

HandDistance calculations	
Handedness	**Description**
Single hand (right or left)	Distance between the instantaneous position of the moving hand and its initial position.
Both hands moving together in the same direction	Distance between the instantaneous mean position of the two hands and the initial mean position of the hands.
Hands moving apart	Distance between the two hands.

Table 8.6. HandDistance calculations according to handedness.

In other words, in the first two cases ‘HandDistance’ and ‘HandDivergence’ were computed in relation to the initial position of the hand(s), while in the last case they were computed as the distance between the hands (as suggested in the reference paper). All features calculated by taking into account the handedness of the gesture were named by adding the '_accord' suffix, while those calculated in the same way independently of handedness (as in the paper by Nymoen et al.) received the ‘_betw’ suffix. Similar designations were used for the first and second derivatives (velocity and acceleration). Although gesture handedness could have been computationally identified (based on features such as quantity of motion and hand direction), it was manually annotated instead and these annotations were taken into account for the computation of features (named by adding the '_accord' suffix).

(b) The calculation of 'position' features requires an understanding of how performers make use of the physical space by their hands. Interview testimonies and video observations suggest that there is a strong conceptualisation of pitch by Indian musicians as a movement (of the hands) in space that is executed in relation to the performer's body (see also section 9.3.). Therefore, depending on whether the gesture was uni-manual or bi-manual, the distance was calculated either as the position of the active hand in relation to the torso for uni-manual gestures or as the mean of the two hands in relation to the torso for bi-manual gestures. Nymoen et al. (2012) had taken a similar approach, but in reference to the (absolute position of the) ground (vertical position), which is supposedly associated with gravity applied on the hands. In the current research context, not only gravity but also other forces are imagined acting on the hands due to the imagery of objects being attached to and manipulated by the hands, which do not have any specific direction. Due to the fact that observations and interview testimonies did not suggest the vertical position to be necessarily an important feature (it might be the case for Ramakant Gundecha, but not for the two performers analysed in this chapter), instead of measuring vertical position (in relation to the ground), position was computed in relation to the performer's torso, which is the central point of reference in space for any kind of hand movement. Movement features computed in relation to the torso were named by adding the '_torso' suffix.

(a)

(b)

(c)

(d)

(e)

Figure 8.2. Snapshots taken over the course of a stretching type of gesture, illustrating variations in the way this can be performed: (a) hands moving apart from each other, (b) hands moving in parallel, (c),(d),(e) moving only one hand.

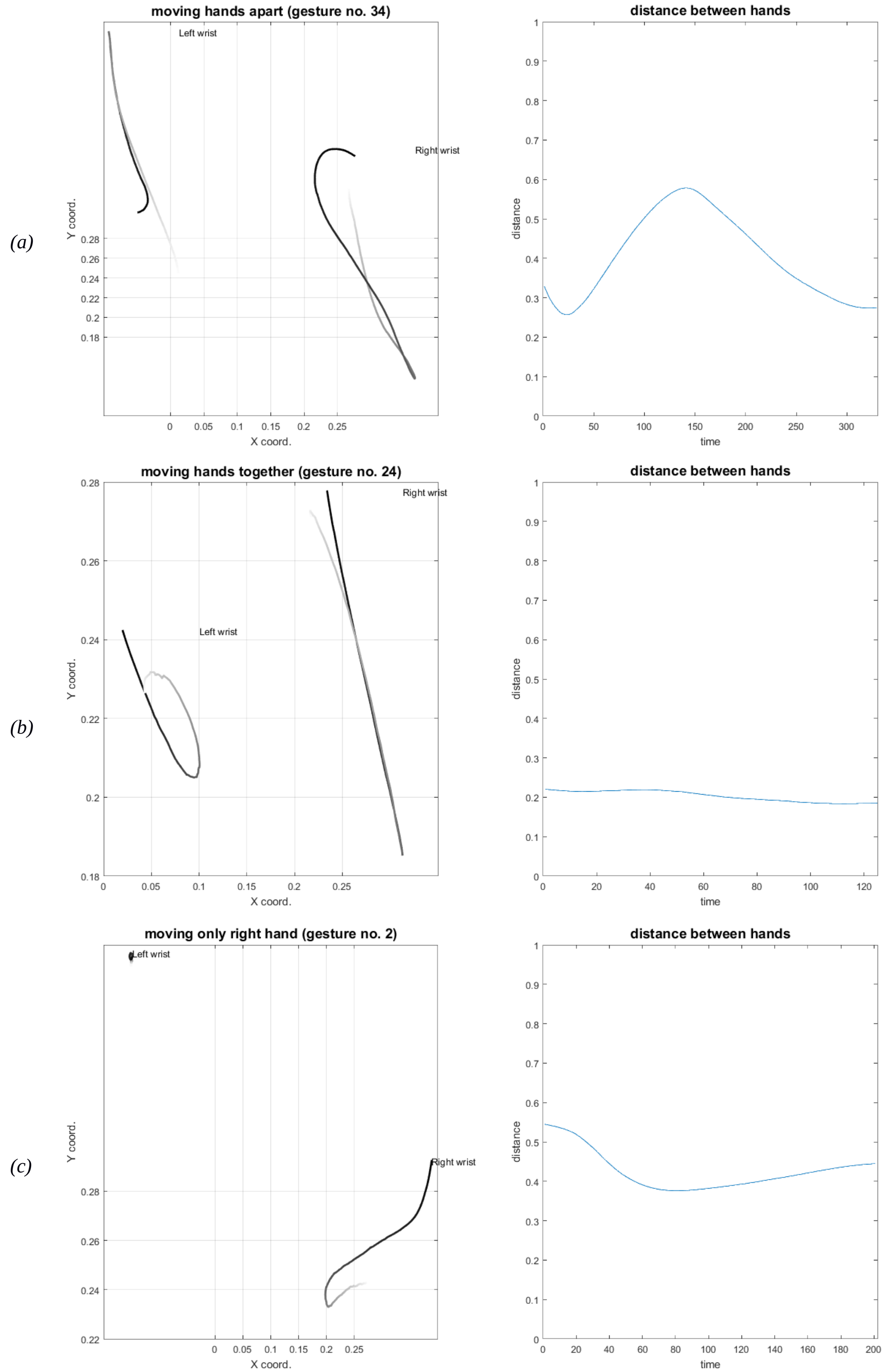


Figure 8.3. 2d trajectory (horizontal plane extracted from 3d data, with nose pointing upwards) and hand distance vs. time graph. The progression of time is indicated by a lightening of tone in the line (darker=earlier, lighter=later).

Summing up what has been discussed until now about alternative ways of computing the features of the core set, the following list gives an overview of these new movement features that were added to the list.

Novel global movement features	
VelocityMean_torso	Mean velocity, calculated as the first derivative of the mean position of the two hands in relation to the torso.
EnergyUM_torso	The sum of the signed vertical kinetic energy of a unit mass, calculated on the VelocityMean_torso.
HandDistanceMean_accord	Mean of distance, calculated between the points that are involved in the movement, either the distance between the hands or the distance of each of the hands in relation to the torso or the mean distance of the hands in relation to the torso.
HandDivergenceMean_accord	Mean of hand divergence calculated as the first derivative of HandDistanceMean_accord.
HandDivergenceSD_accord	Standard deviation of hand divergence calculated as the first derivative of HandDistanceMean_accord.
AbsAccelerationMean_accord	Mean of absolute acceleration calculated as the second derivative of HandDistanceMean_accord.
OnsetAccelerationMean_accord	Mean of acceleration calculated as the second derivative of HandDistanceMean_accord on the gesture onset (15% of total duration for events of duration>1.7sec and 45% for events with duration<=1.7sec).
energyUM_accord	The sum of the signed vertical kinetic energy of a unit mass, calculated on the HandDivergenceMean_accord.

Table 8.7. Novel global movement features.

Acoustic features

Of the three widely accepted basic attributes of sound, timbre is not easy to quantify and therefore it required alternative measures to be computed which include brightness, roughness and flatness. Additionally, while the mean and standard deviation values may be useful global measures of pitch, they cannot capture the morphology of a melodic movement. Considering the importance of the precise way by which a note is approached in Dhrupad singing, measures that reflect the shape of the melodic movement had to be computed. According to the qualitative analysis, melodic movements fall under the following three basic cases and they were manually annotated:

- ascent (from '*PitchminPre*' to '*Pitchmax*');
- descent (from '*Pitchmax*' to '*PitchminPost*');
- double pitch-slope (from '*PitchminPre*' through '*Pitchmax*' to '*PitchminPost*').

Three different ways for representing fundamental aspects of melodic movement were examined:

(a) By using the critical points of the melodic movement. These pitches were extracted in a semi-automatic way from Praat. More specifically, first the maximum pitch was automatically allocated within the boundaries of a specified time interval and then, depending on the type of movement (whether only ascent, only descent, or a double-sloped pitch glide), the minimum/a on one or either side of the maximum were also allocated by Praat. The time intervals between the minima and the maximum were also extracted.

(b) By fitting polynomial expressions of various degrees (2^{nd} - 5^{th}) to the pitch data. Their coefficients were used as features in the regression. Initially this was done on the whole melodic movement, but as this process produced shapes that diverged substantially from the original melodic contour, the polynomials were finally fit exactly between the annotated critical points of the pitch and were forced to always pass through them (three or two points according to the movement type, that was already manually annotated). The following graphs of Figure 8.4 illustrate one example for a 2^{nd} and a 5^{th} order polynomial fit in the case of an ascending, descending and double-sloped pitch glide.

(a) 2^{nd} order polynomial fit *(b) 5^{th} order polynomial fit*

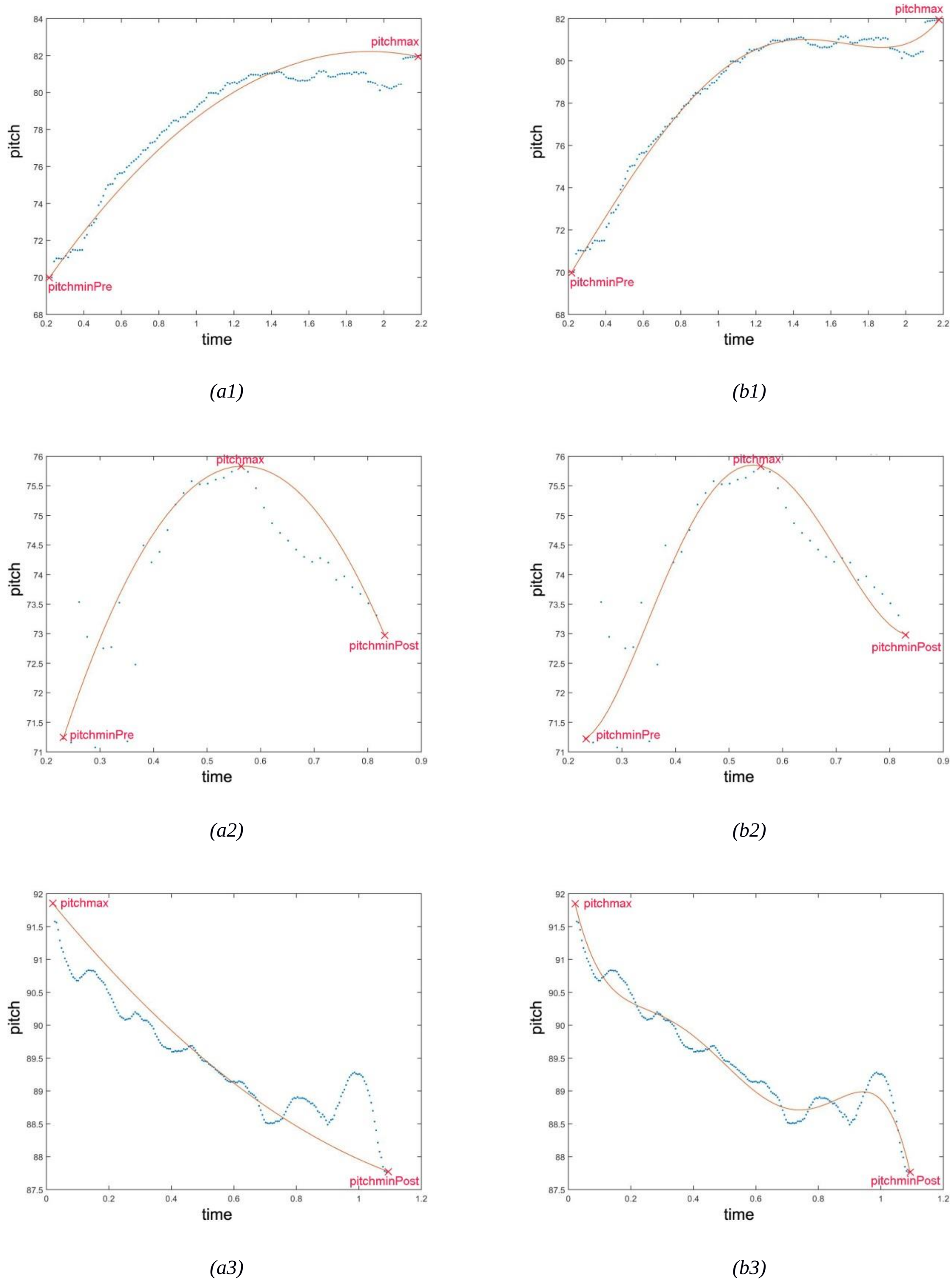


Figure 8.4. (a) 2^{nd} and (b) 5^{th} order polynomial fits on 1. ascending, 2. double and 3. descending melodic movement.

(c) By computing the strength of pitch ascent vs. descent in a compact way, as in energyUM:

$$Pitchmov = (asc\_pitch \cdot |asc\_pitch|) + (desc\_pitch \cdot |desc\_pitch|)$$

In order to also include the timing aspect, the following expression was developed, which is in practice a measure of the asymmetry between ascending and descending melodic movement speeds:

$$Pitchmov\_speed = \frac{(asc\_pitch \cdot |asc\_pitch|)}{timemax - timeminPre} + \frac{(desc\_pitch \cdot |desc\_pitch|)}{timeminPost - timemax}$$

, where *asc_pitch* is the ascending and *desc_pitch* is the descending pitch interval.

Due to the logarithmic nature of pitch, all pitch-related features were computed on three different scales. To distinguish the way they were computed, pitch features received the appropriate suffix to their name.

Pitch computation	
suffix	**scale**
lin	linear (in Hz)
abslog	absolute logarithmic scale, computed as $100 \cdot \log\left(\frac{f}{f_{ref}}\right)$, where f_{ref}=440Hz (absolute frequency value)
rellog	relative logarithmic, computed as $100 \cdot \log\left(\frac{f}{f_{ref}}\right)$, where f_{ref}=the tonic of the specific octave where the melodic movement takes place (based on the manual annotations)

Table 8.8. Pitch computation.

The absolute logarithmic values refer to the absolute pitch height and thus they reflect the mechanics of voice production. The relative logarithmic values express the pitch height of the degrees of the scale and thus they are assumed to better reflect the organisation of the specific rāga.

Table 8.9 gives an overview of all global alternative acoustic features that were included in the trial-and-error processes of determining the best models.

Novel or alternative global acoustic features				
melodic movement	critical pitch values	*PitchminPre*	minimum pitch before maximum (not for descending glides)	in 'lin', 'abslog', 'rellog' scales
		Pitchmax	maximum pitch	
		PitchminPost	minimum pitch after maximum (not for descending glides)	
		timeminPre	timestamp of *PitchminPre* from beginning of glide	
		timemax	timestamp of *Pitchmax* from beginning of glide	
		timeminPost	timestamp of *PitchminPost* from beginning of glide	
	polynomial expression (for pitch contour)	$coeff_2$, $coeff_1$, $const$	2nd order coefficients & intercept of the expression $coeff_2 \cdot x^2 + coeff_1 \cdot x + const$	
		$coeff_5$, $coeff_4$, $coeff_3$, $coeff_2$, $coeff_1$, $const$	5th order coefficients & intercept of the expression $coeff_5 \cdot x^5 + coeff_4 \cdot x^4 + coeff_3 \cdot x^3 + coeff_2 \cdot x^2 + coeff_1 \cdot x + const$	
	strength of pitch ascent vs. descent	*Pitchmov*	$(asc\_pitch \cdot \lvert asc\_pitch \rvert) + (desc\_pitch \cdot \lvert desc\_pitch \rvert)$	
		Pitchmov_speed	$\frac{(asc\_pitch \cdot \lvert asc\_pitch \rvert)}{timemax - timeminPre} + \frac{(desc\_pitch \cdot \lvert desc\_pitch \rvert)}{timeminPost - timemax}$	
timbre	brightness	*M_br*	Mean of brightness	in 'lin', 'abslog', 'rellog' scales
		SD_br	Standard deviation of brightness	
		M_br_log	Mean of log(brightness)	
		SD_br_log	Standard deviation of log(brightness)	
	spectral entropy	*M_spentropy*	Mean of spectral entropy	
		SD_spentropy	Standard deviation of spectral entropy	
	roughness	*M_roughness*	Mean of roughness	
		SD_roughness	Standard deviation of roughness	
	flatness	*M_flatness*	Mean of flatness	
		SD_flatness	Standard deviation of flatness	

Table 8.9. Novel or alternative global acoustic features.

8.1.2. Determining Best Models

The best models were determined by a trial-and-error process of finding the best trade-off between accuracy of model fit, compactness of easily interpretable features (small number of independent variables) and simplicity in feature extraction. The whole statistical analysis was performed exclusively in the R environment[96]. Due to the endless possibilities in the quest of the best combination of features, as already mentioned, this work started from an already successfully proposed set of features, as presented in Nymoen et al. (2013), and was then progressively enriched with the novel features that it was expected might raise the explained variance in the estimated response.

A large number of possible combinations of features were tried out, but only some of the most important steps that improved the goodness of fit are presented here. As the effects of single features cannot be simply extended to multiple feature regression, individual features had to be added or removed from the set one at a time and they also had to be tried on various combinations of feature sets. More specifically, the effect that a feature might have on the estimation of the response when the regression is run based on a single feature is not expected to be the same as when this feature is used in combination with a set of other features; different sets will also modify the effect of a single feature on the estimation of the response[97].

Although the analysis aimed at producing the best fitting model, alternative models of poorer but still comparable results are also included as they are useful in better understanding the complexity of MIIOs and the modelling tasks. Detailed statistical results are given only whenever it is considered important, while for less important models these are summarised by the most important statistical measures. The following symbols are used in order to signify probability values of the coefficients belonging within specified value ranges:

'***' for $0 \leq p < 0.001$

'**' for $0.001 \leq p < 0.01$

'*' for $0.01 \leq p < 0.05$

'.' for $0.05 \leq p < 0.1$

' ' for $0.1 \leq p < 1$

For an interpretation of probability values the reader can refer to section 3.3.1. For the interested reader wishing to follow the process of determining the best model step-by-step, Appendix B1 offers such an opportunity.

[96] https://www.rstudio.com/

[97] As the discussion will be mostly revolving around multi-parametric regression, it should be pointed out that the contribution of a feature is interpreted by assuming a fixed value for the other features.

8.2. Results

The current section offers a comparison between (a) the most successful models and (b) the most similar models (in terms of shared features) between the two performers. The first targets more performer/performance-specific aspects, while the second can be seen as an attempt to find more generic descriptions of MIIOs.

8.2.1. Effort Level Estimation (Linear Models)

(a) Best selected LM models

The following are the most successful LM models for each of the two performers:

Afzal Hussain: LM 4 (best)	
LM Feature	**Coefficients**
PitchminPre_rellog	-1.65***
Pitchmax_rellog	+2.42***
AbsVelocityMean	-1.28**
AbsVelocitySD	+0.71.
HandDistanceMean_accord	+0.71*
R^2adj. (p<.001)	***0.6***

Table 8.10. Afzal Hussain: Best LM (4).

Lakhan Lal Sahu: LM 2 (best)	
LM Feature	**Coefficients**
PitchSD_abslog	+0.56***
Pitchmax_abslog	+0.9***
HandDivergenceSD	+0.43*
OnsetAccelerationMean	+0.43*
R^2adj. (p<.001)	***0.44***

Table 8.11. Lakhan Lal Sahu: Best LM (2).

Afzal Hussain

According to the results for Hussain, higher bodily effort levels are required for 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. They are accompanied by hand movements that are slower but exhibit a larger speed variation while moving the hands further away from the point of reference.

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 rāga. In the most extreme case (in respect to the highest allowed degree of the scale) in rāga Jaunpurī, a glide can ascend up to the 7^{th} degree of the scale and this melodic movement is associated with the highest possible bodily effort level. Lower maximum pitches refer to glides that reach up to the 5^{th} or 4^{th} degree of the scale and they are associated with lower bodily effort levels. Thus, effort may be associated with the melodic tension defined by the character of the specific rāga.

As was already discussed in section 5.1.2. , the 7th degree is the most unstable degree of rāga Jaunpurī and it therefore requires the subsequent descent and final resolution of the melodic movement upon the 6th or 5th degree. It is associated with gestures of interacting with an elastic object; pitch ascent is associated with the stretching away of the object and pitch descent with the recoil phase of the hands to their original position. Thus, combining the results of model 4 and the qualitative findings of Chapter 5, it can be concluded that stretching an elastic object wider apart while ascending to the 7th degree of the Jaunpurī scale requires higher levels of effort. In fact, the most important feature of model 4 is the highest degree of the scale that is reached through the melodic glide and the second most important feature is the starting degree of the melodic glide, while the body movement features are less significant for the estimation of bodily effort (due to the probability values by which they appear in the model).

The boxplots of Figure 8.5 illustrate the trends of individual features relating to the bodily effort levels. Despite some confusion in the distribution of the data across the effort level values, especially for the movement features, overall association trends can be clearly visually discerned.

Lakhan Lal Sahu

According to the results for Sahu, higher bodily effort levels are required for larger melodic glides that reach up to higher maximum pitches. They are accompanied by hand movements that exhibit a larger variation of hand divergence (change of speed of moving the hands apart from or closer to each other), with a strong mean acceleration in the beginning (onset).

In this case, pitches are expressed in an absolute logarithmic scale, which better reflects aspects of voice production; a higher bodily effort is thus expected for pitches at the extremes of the performer's pitch range (extreme high and extreme low pitches). The most important features of this model are again the acoustic, while the movement features are less significant for the estimation of effort. The boxplots of Figure 8.6 depict the association between effort level and individual features of the model. Despite some confusion in the data distributions, they illustrate the overall trends that were reflected by the model, especially in the movement features (and mostly the mean of onset acceleration).

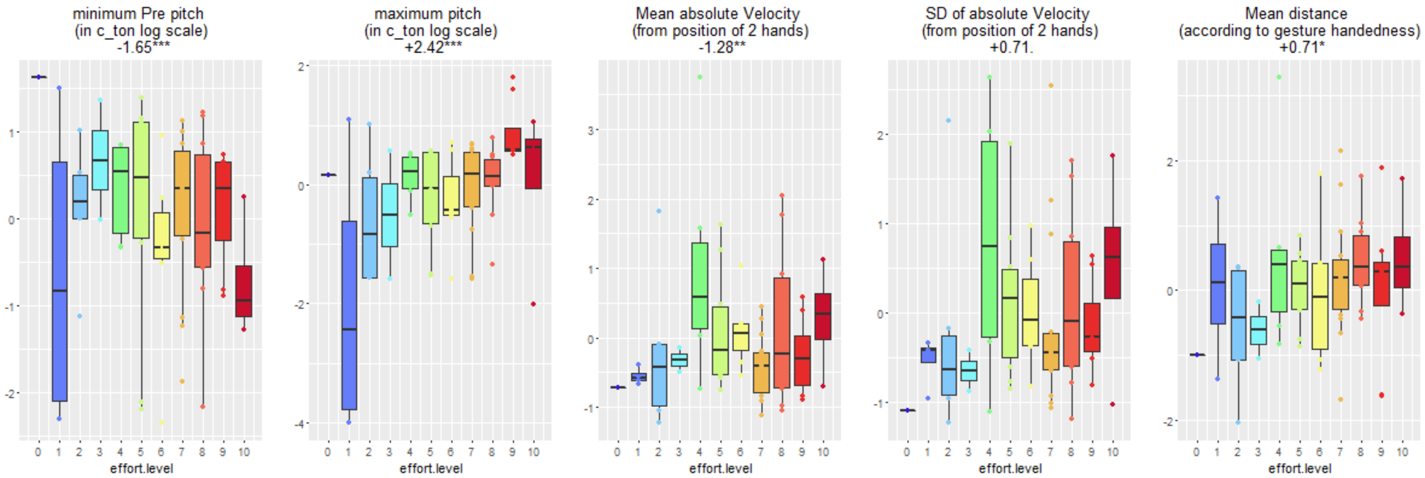


Figure 8.5. Afzal Hussain: Boxplots for LM 4.

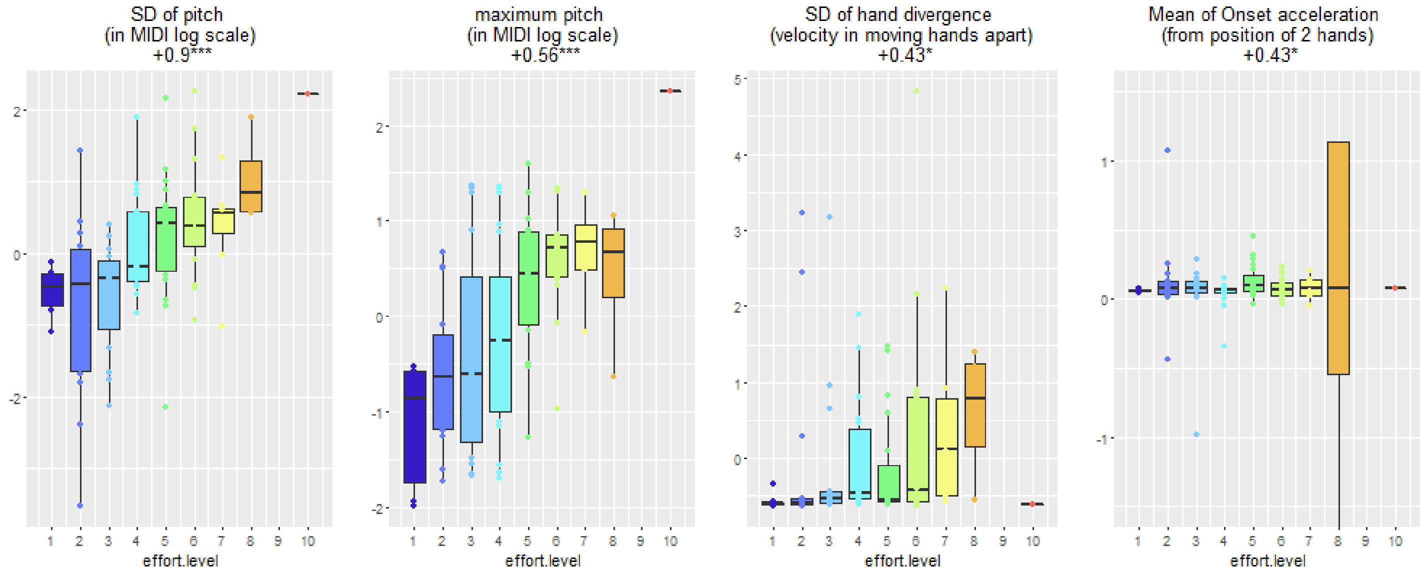


Figure 8.6. Lakhan Lal Sahu: Boxplots for LM 2.

Cross-performer comparison of best LM models

Two quite successful models were developed individually for each performer. However, these are performer/performance specific, possibly highlighting idiosyncratic aspects of the performer and/or the rāga. To examine their generic power and the possibility of using them for labour-free computations in other performances, in this section I explore the goodness of fit in applying the model of one performer to the data of the other.

First, I examine the fit of Hussain's model on Sahu's data. As Table 8.12 shows, it produces extremely poor results.

Lakhan Lal Sahu: LM 4 of Afzal Hussain	
LM Feature	**Coefficients**
PitchminPre_rellog	-0.87**
Pitchmax_rellog	+0.92***
AbsVelocityMean	-0.46
AbsVelocitySD	+0.41
HandDistanceMean_accord	-0.22
R^2adj. (p=.01)	***0.12***

Table 8.12. Lakhan Lal Sahu: LM 4 of Afzal Hussain.

In fact, from this model only two out of the five features (only the acoustic) contribute significantly to the estimation of effort. If only these two features are used, they can only estimate 13% of effort level:

Lakhan Lal Sahu: LM with only acoustic features of Afzal Hussain's LM 4	
LM Feature	**Coefficients**
PitchminPre_rellog	-0.81**
Pitchmax_rellog	+0.91***
R^2adj. (p=.002)	***0.13***

Table 8.13. Lakhan Lal Sahu: LM with most important acoustic features from Afzal Hussain's LM 4.

Instead, the use of just these two acoustic features of Hussain's model on his own dataset, yields a high goodness of fit.

Afzal Hussain: LM with only the acoustic features of model 4	
LM Feature	**Coefficients**
PitchminPre_rellog	-1.53***
Pitchmax_rellog	+2.24***
R^2adj. (p<.001)	***0.53***

Table 8.14. Afzal Hussain: LM with most important acoustic features of model 4.

Fitting Sahu's model to Hussain's data gives even poorer results.

Afzal Hussain: LM 2 of Lakhan Lal Sahu	
LM Feature	**Coefficients**
PitchSD_abslog	+1.58***
Pitchmax_abslog	-0.51.
HandDivergenceSD	-0.16
OnsetAccelerationMean	-0.09
R^2adj. (p<.001)	***0.23***

Table 8.15. Afzal Hussain: LM 2 of Lakhan Lal Sahu.

None of the movement features seems to be contributing to the effort estimation. The use of only the two most important acoustic features of Sahu's model on Hussain's data produces the following results:

Afzal Hussain: LM with only acoustic features of Lakhan Lal Sahu LM 2	
LM Feature	**Coefficients**
PitchSD_abslog	+1.56***
Pitchmax_abslog	-0.54.
R^2adj. (p<.001)	***0.24***

Table 8.16. Afzal Hussain: LM with most important acoustic features of Lakhan Lal Sahu's LM 2.

This shows that almost one fourth of Hussain's effort variance can be explained by the two strong acoustic features of Sahu's model 2.

By using these two most important acoustic features on Sahu's own data, a fit of 40% can be achieved:

Lakhan Lal Sahu: LM with only the acoustic features of model 2	
LM Feature	**Coefficients**
PitchSD_abslog	+0.63***
Pitchmax_abslog	+0.80***
R^2adj. (p<.001)	***0.4***

Table 8.17. Lakhan Lal Sahu: LM with most important acoustic features of model 2.

In summary, the above results indicate that what is shared between the two performers is mostly various qualities in the sound. Higher bodily effort levels are required for achieving higher maximum pitches in melodic glides, according to the 'best' models for both performers. Additionally, the lower starting minimum pitch that is combined with a higher value of maximum pitch in the case of Hussain can be seen as describing a similar thing to the higher value of standard deviation (the extent of variation) in the melodic movement that appears in the model of Sahu. Therefore, it can be concluded that there is a generic association of effort with some acoustic qualities that are shared between the two performers: higher bodily effort levels are required for larger melodic movements that rise to higher maximum pitches (on a logarithmic scale). The main difference is that in the case of Hussain pitches are expressed in relation to each individual tonic, while in the case of Sahu they refer to absolute values. The first case (relative logarithmic scale) reveals melodic qualities that are related to the scale of the specific rāga in terms of characteristic phrases that appear recurrently in different parts of each octave. The second case (absolute logarithmic scale) represents the incremental frequency values of the voice's pitch height. However, considering the fact that most melodic movements are performed within the boundaries of a single octave, this means also that larger glides that reach up to higher notes must have started from a lower note of the same octave. Thus, the two features should be overlapping in most of the cases.

The movement features of the best models that are associated with the estimation of effort level are quite different between the two performers. For the first performer, a higher bodily effort is perceived with slower movements that involve moving the hands further away from the reference point (between the hands or in relation to the torso or the beginning point of the movement). For the second performer, a higher bodily effort is required for movements which start with a strong acceleration and exhibit a larger change of velocity (i.e. acceleration) in moving the hands apart during the whole duration of the gesture. These observations highlight idiosyncratic factors in using the body during vocal improvisation.

Overall, in both cases we have seen extremely low probability values for the acoustic and higher probability values for the movement features. This observation may lead to the conclusion that there is a significant association between bodily effort and melodic movement (which is verified in the next section), but the way this is displayed by the performers' hands is less obvious and perhaps less consistent.

(b) Most similar LM models between the two performers

The following two LM models are the ones with the highest number of shared features between the two performers:

Afzal Hussain: LM 3 (similar)	
LM Feature	**Coefficients**
PitchminPre_abslog	-2.13***
Pitchmax_abslog	+2.3***
AbsVelocityMean	-1.34**
AbsVelocitySD	+0.93*
HandDistanceMean_accord	+0.65*
R^2adj. (p<.001)	***0.53***

Table 8.18. Afzal Hussain: Most similar LM (3).

Lakhan Lal Sahu: LM 3 (similar)	
LM Feature	**Coefficients**
PitchminPre_abslog	-0.95***
Pitchmax_abslog	+1.91***
AbsVelocityMean	-1.31**
AbsVelocitySD	+1.35**
R^2adj. (p<.001)	***0.42***

Table 8.19. Lakhan Lal Sahu: Most similar LM (3).

It was not possible to deduce the exact same model that will work equally well for both performers, as by removing HandDistanceMean_accord from Hussain's model or adding this feature to Sahu's model, the contribution of other features changes.

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). 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 from the point of reference (according to the type of gesture, distance may refer to the hands moving apart from each other or in relation to the torso or the starting point of the movement). As pitches are expressed here on an absolute logarithmic scale, they better reflect the mechanical effort required for producing the voice (higher effort for extremely high and extremely low pitches) rather than some rules of melodic organisation.

The boxplots of Figure 8.7 and Figure 8.8 illustrate the contribution of each feature on the estimation of effort levels for each of the two performers and LM models. Some confusion in the distributions can be visually identified (especially for the mean and standard deviation of the absolute velocity in the case of Sahu), but the overall trends follow the description of the models.

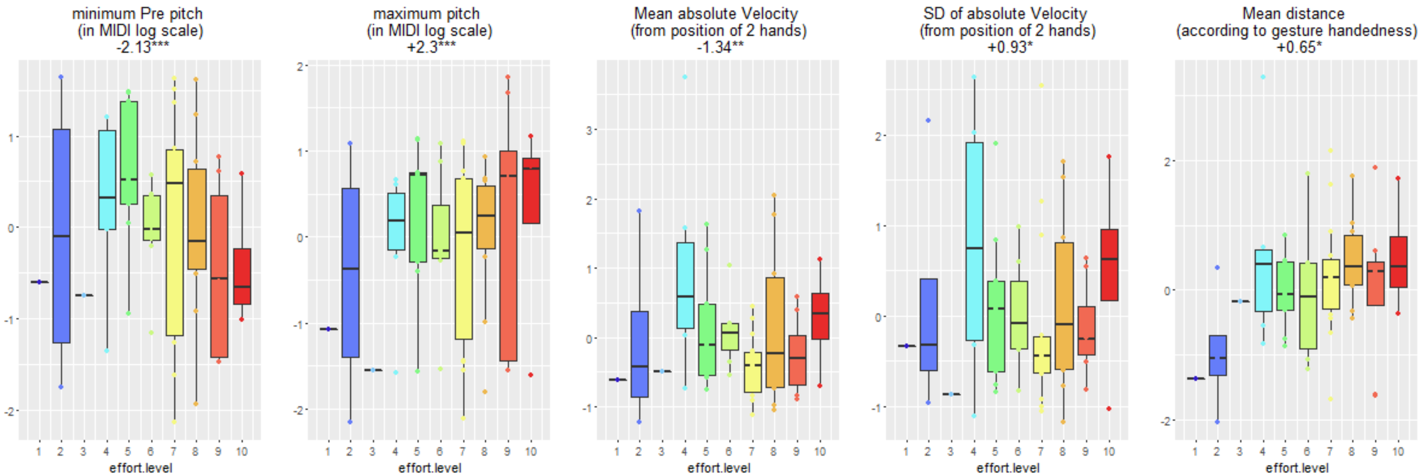


Figure 8.7. Afzal Hussain: Boxplots for LM 3.

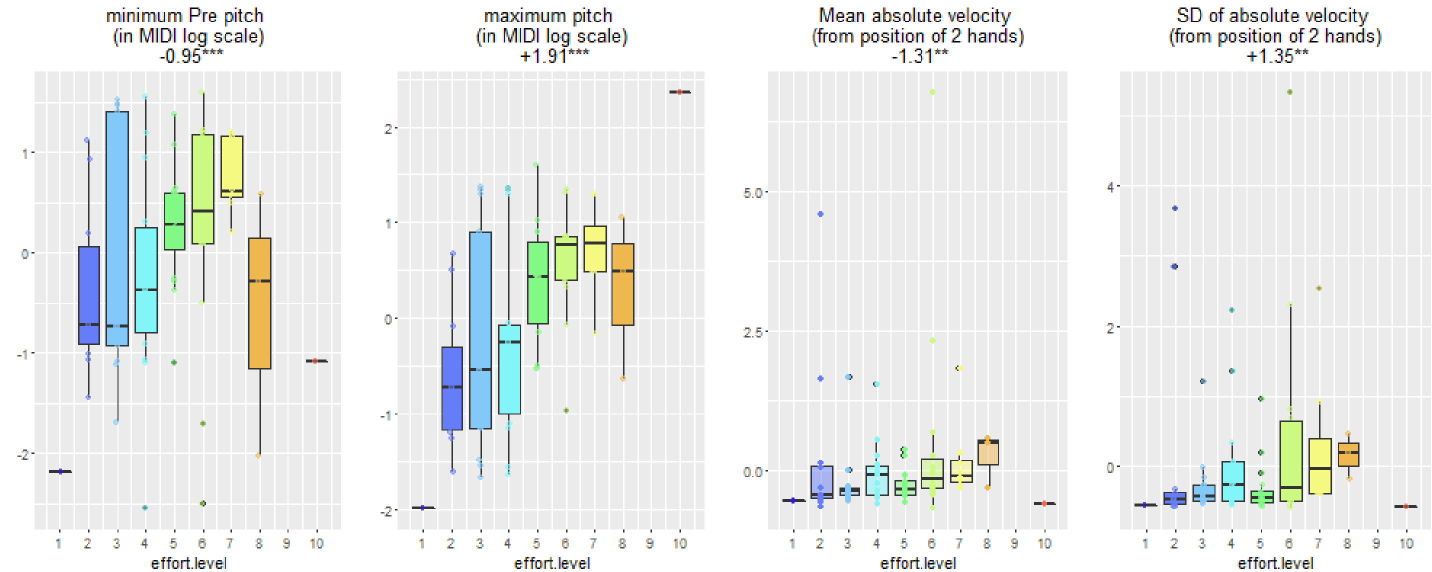


Figure 8.8. Lakhan Lal Sahu: Boxplots for LM 3.

By comparing the most overlapping models of the two performers, it can be concluded that a relatively good estimation of effort level variance can be achieved by using almost the same combination of easily extracted global features. In fact, the direction of their contribution is the same and their statistical significance is similar, meaning that acoustic and movement features shared between performers contribute in similar ways to the estimation of effort. The lower goodness of fit of these models in comparison to the best selected models discussed in the previous section reflects their more generic power in describing MIIOs for both performers.

It is interesting to note that by using only the two acoustic features of the models, a still high effort estimation rate—of 46% for Hussain and 37% for Sahu—can be achieved.

Afzal Hussain: LM on strongest pitch features	
LM Feature	**Coefficients**
PitchminPre_abslog	-2.31***
Pitchmax_abslog	+1.92***
R^2adj. (p<.001)	***0.46***

Table 8.20. Afzal Hussain: LM on strongest pitch features.

Lakhan Lal Sahu: LM on strongest pitch features	
LM Feature	**Coefficients**
PitchminPre_abslog	-0.87***
Pitchmax_abslog	+1.71***
R^2adj. (p<.001)	***0.37***

Table 8.21. Lakhan Lal Sahu: LM on strongest pitch features.

As I discussed earlier, the equivalent pitch features that are computed on a *rellog* scale (reflecting rāga-specific qualities) produce a 53% fit for Hussain and 13% for Sahu. This is an indication that while the association of bodily effort levels with the mechanical requirements of voice production is a common quality shared between the two performers/performances, the association with rāga-specific characteristics is only strong in the case of Hussain.

8.2.2. Gesture Classification (General Logistic Models)

(a) Best selected GLM models

Afzal Hussain: GLM 6 (best)		
GLM Features	**Coefficients**	**AUC**
timemax	-3.04**	.84
Pitchmax_rellog	-2.93**	0.72
PitchSD_abslog	-1.17.	0.63
AbsAccelerationMean	+3.4*	0.54
HandDivergenceSD	-3.12**	0.61
AUC		***0.95***

Table 8.22. Afzal Hussain: Best GLM (6).

Lakhan Lal Sahu GLM 3 (best)		
GLM Features	**Coefficients**	**AUC**
$coeff_2$	+0.71*	0.63
PitchSD_abslog	-0.73**	0.7
AbsVelocityMean	-1.48*	0.65
HandDistanceMean	+0.78*	0.54
AUC		***0.8***

Table 8.23. Lakhan Lal Sahu: Best GLM (3).

Afzal Hussain

According to the results for Hussain, it is more likely that interactions with elastic objects are associated with slower and larger melodic movements that ascend to a higher degree of the scale (i.e. in the upper part of the octave). They are performed by hand gestures that exhibit a lower variation in speed (the speed of the hands' average position calculated over the entire duration of the gesture) and a larger variation in the hands' divergence (the speed in moving the hands apart).

In rāga Jaunpurī the highest maximum allowed degree within an octave range is the 7^{th}, which also happens to be the most unstable pitch of the specific rāga. The instability of the 7^{th} degree reflects the strong descending character of rāga Jaunpurī and imposes a change of direction in the melodic movement (from a melodic ascent to a melodic descent), which can be seen as a reflection of the change in spatial direction during interactions with an elastic body; the first phase (extension) is necessarily followed by a second phase (recoil) towards the original position. In other words, double pitch glides (ascent-descent) are more likely associated with interactions with elastic objects that also comprise two movement phases (extension-recoil). It could therefore be suggested that the type of manual interaction is associated with the voice based on the grammatical rules of the rāga and the shared cross-modal morphologies. This finding coincides with the video analysis findings of section 5.3. , according to which double pitch glides which move to the 6^{th} degree of the scale by first going over it and ascending to the 7^{th} are more likely to be performed by an interaction with an elastic object and with higher effort levels.

The boxplots of Figure 8.9 display the contribution of individual features according to gesture classes for Hussain's model 6.

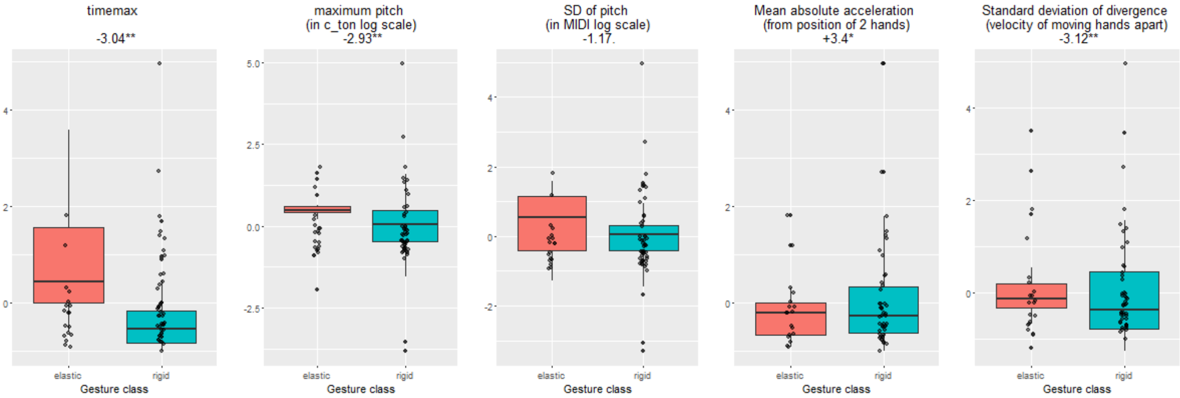


Figure 8.9. Afzal Hussain: Boxplot on features of GLM 6.

According to the boxplots, the median value of the mean absolute acceleration is lower for interactions with rigid than elastic objects and therefore contradicts the trend described by the model. However, the boxplot still shows a lot of examples with higher values for this type of gestures. The clearest case in the distinction between the two gesture classes can be seen in the boxplot for *timemax*, in which both the mean value and the whole distribution exhibit lower values for interactions with rigid rather than elastic objects.

Lakhan Lal Sahu

According to the results for Sahu, interactions with elastic objects are more likely to be performed with pitch movements of a larger size (pitch interval) and larger duration and with the hands moving faster but remaining bound to each other.

The boxplots of Figure 8.10 display the distributions of individual features according to gesture classes for Sahu's model 3.

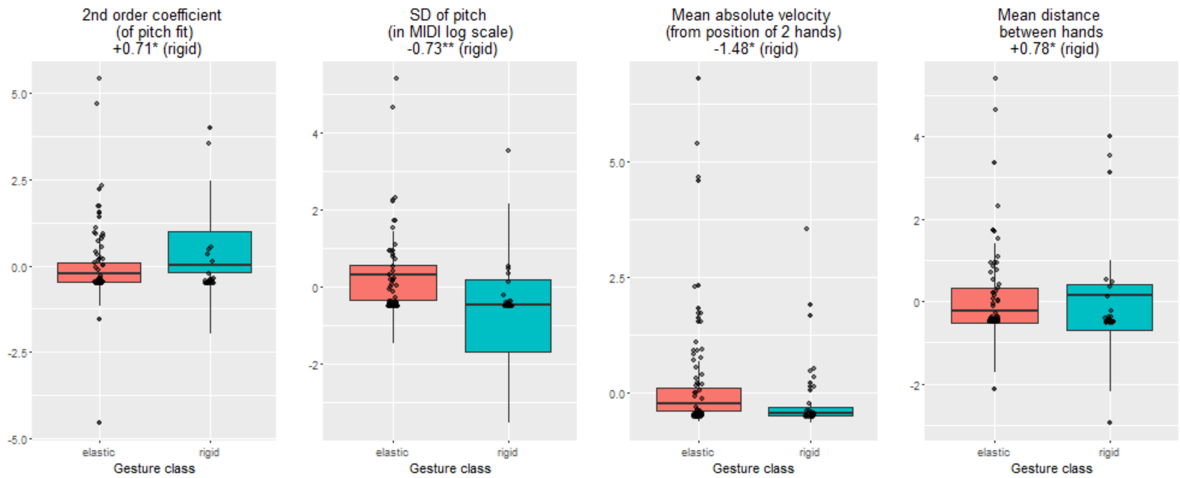


Figure 8.10. Lakhan Lal Sahu: Boxplots on features of GLM model 3.

The median values of the features are in lines with the trends described by the coefficients of the model.

Cross-performer comparison of best GLM models

Two quite successful GLM models were developed that fit each individual performer well. There is some degree of overlap in the acoustic qualities, although these are not captured by the same acoustic features. In the case of Hussain, the larger duration of the melodic movement in interactions with elastic objects refers only to the ascending part (timemax), while in the case of Sahu it refers to the whole melodic movement. Another important difference between the best GLM models of Sahu and Hussain is that the highest degree of the octave that the pitch glide reaches (on a relative logarithmic scale) is not included in Sahu's model. This means that the type of MIIO used by Sahu is independent of the scale for the specific rāga. Additionally, the movement features are different between the two performers, possibly highlighting idiosyncratic aspects in displaying MIIOs. Fitting Hussain's best GLM model to Sahu's data leads to a poor fit by a model that does not include any statistically significant features. The other way round, by applying Sahu's best GLM model to Hussain's data and removing AbsVelocityMean, which does not show a significant contribution, the following model is deduced:

Afzal Hussain: Adapted GLM 3 from Lakhan Lal Sahu		
GLM Feature	**Coefficients**	**AUC**
$coeff_2$	-2.97.	0.53
PitchSD_abslog	-0.93*	0.63
HandDistanceMean	-0.8*	0.63
AUC		***0.76***

Table 8.24. Reduced GLM 3 of Lakhan Lal Sahu applied on feature set of Afzal Hussain.

Although the gesture classification success rate is not as high as in the previously discussed models, this model highlights some of the most overlapping qualities between performers.

In summary, the best GLM models that were developed for each of the performers individually do not have a strong generic power, however, there is a degree of overlap in the melodic qualities. These qualities include the size of the melodic movement and the duration of its ascent. The movement features are different between the two performers. For Hussain, interactions with elastic objects are more likely for movements with a lower (absolute) acceleration and a larger variation in the hands' divergence, while for Sahu they are more likely for faster (absolute velocity) and more confined (hands remaining closer to each other) movements. These observations possibly highlight the idiosyncratic factor in performing MIIOs, which renders the task of describing them through a small number of movement features as non-trivial.

(b) Most similar GLM models between the two performers

The following two GLM models are the ones with the highest number of shared features between the two performers:

Afzal Hussain: GLM 3 (similar)		
GLM Features	**Coefficients**	**AUC**
PitchMean_abslog	+1.45**	0.59
PitchSD_abslog	-1.74**	0.63
HandDistanceMean	-1.1**	0.63
Overall		***0.86***

Table 8.25. Afzal Hussain: Most similar GLM (3).

Lakhan Lal Sahu: GLM 2 (similar)		
GLM Features	**Coefficients**	**AUC**
PitchMean_abslog	+0.51.	0.51
PitchSD_abslog	-1.03***	0.7
HandDistanceMean	+0.98**	0.54
AbsVelocityMean	-1.58*	0.65
Overall		***0.78***

Table 8.26. Lakhan Lal Sahu: Most similar GLM (2).

Again, it was not possible to deduce the exact same model for both performers. According to these models, interactions with elastic objects are more likely to be performed with larger melodic movements at lower pitches (calculated on an absolute logarithmic scale) and with the hands moving further apart for Hussain and remaining closer together but faster in the case of Sahu. Overall, there is an overlap between the two models in the acoustic features but no agreement in the movement features. Additionally, there are differences in the statistical significance—and therefore the relative contribution—of individual features on the models: the mean pitch value is more important for Hussain, while the standard deviation of the pitch (which expresses the size of the melodic glide) is more important for Sahu. The boxplots of Figure 8.11 and Figure 8.12 illustrate the contribution of each feature on gesture classification for each of the two performers and GLM models. Overall, the distribution of feature values coincides with the trends described by the models. Although for Sahu there is a degree of confusion in the median value of the mean pitch, there are still a number of cases that agree with the model by having higher values for interactions with rigid rather than elastic objects.

Although these models have a slightly lower goodness of fit than the best models that were developed for each performer individually, they are useful due to their generic power in classifying MIIOs for both performers. The feature that is most statistically significant and affects the classification in the same direction for both performers is the size of the melodic movement (which is larger in interactions with elastic objects). The mean pitch height is lower for interactions with elastic objects, but far less significant in the case of Hussain, while the mean distance between the hands is also important but with opposite directions for the two performers (which means that Sahu moves them further apart, while Hussain keeps them closer together when displaying interactions with elastic objects).

The extent by which the hands move away from each other is not expected to be necessarily associated with the bodily effort required to execute an action. Different distances (the effect of an action) can be achieved by changing the balance between the power of the performer's (the agent's) action (force, displayed as effort) and the sensed resistance acted by the imagined object (due to the stiffness of elastic objects in stretching or the weight of rigid objects in transposing them in space). This balance between performer's action and object's reaction seems to be different for each performer, with Hussain looking more effective in defying the imagined opposing forces. As the amount of imagined resistance by the object is not transparent to an observer, distance cannot be used as a feature reflecting the exerted effort.

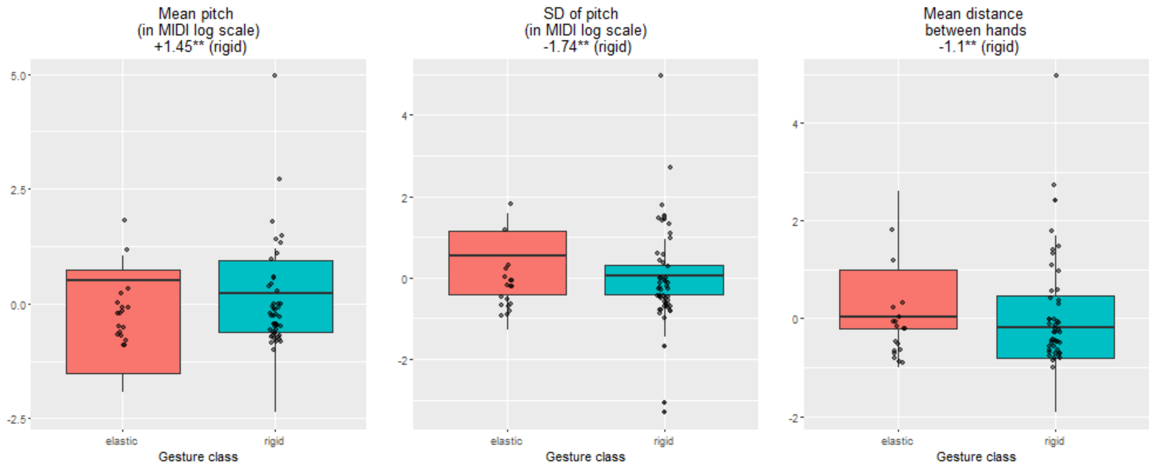


Figure 8.11. Afzal Hussain: Boxplots for GLM 3.

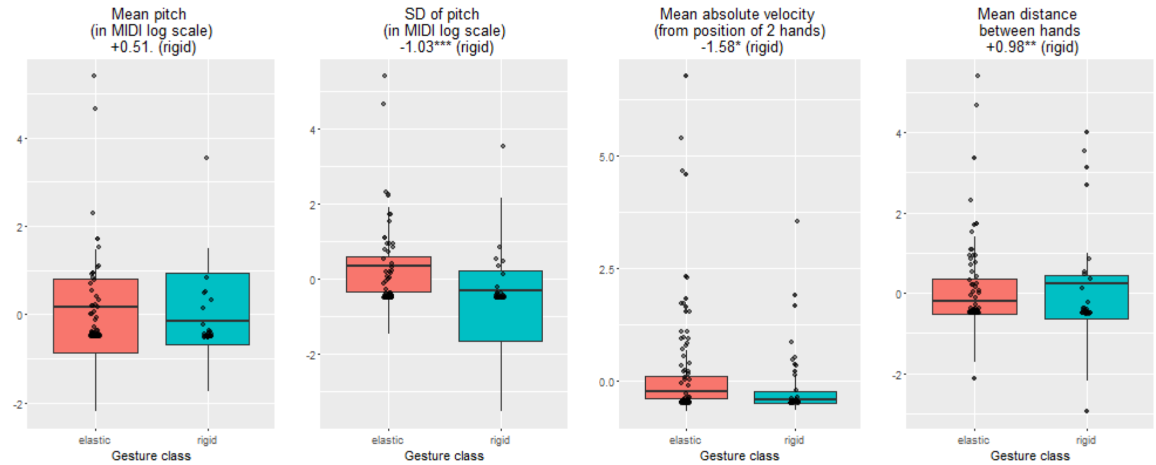


Figure 8.12. Lakhan Lal Sahu: Boxplots for GLM 2.

It is interesting to note that just the use of the strongest melodic feature (the size of the pitch glide) yields a significant classification rate that shows how important this feature is.

Afzal Hussain: GLM on the strongest pitch feature	
GLM Feature	**Coefficients**
PitchSD_abslog	-0.6.
AUC	***0.63***

Table 8.27. Afzal Hussain: GLM on strongest pitch features.

Lakhan Lal Sahu: GLM on the strongest pitch feature	
GLM Feature	**Coefficients**
PitchSD_abslog	-0.83**
AUC	***0.7***

Table 8.28. Lakhan Lal Sahu: GLM on strongest pitch features.

8.3. Discussion

The current chapter offered a detailed description of the exploration process in fitting LM and GLM models to the movement and audio data of two vocalists: Afzal Hussain and Lakhan Lal Sahu. The following points summarise the quantitative findings of this chapter.

8.3.1. Effort estimation (LM)

According to the most successful models that were fit individually to each performer, the estimation of effort relies primarily on acoustic features, while movement features contribute less, especially in the case of Hussain. Additionally there is some degree of agreement between the two performers in acoustic qualities (the way some aspect in the voice might change, even if the features by which it is captured in the models are not exactly the same) but a smaller overlap in movement features. The shared acoustic qualities include the starting note and the highest pitch reached in ascent, hence also the size of the melodic movement. However, for Hussain these pitches reflect the scale and the melodic organisation of the rāga, while for Sahu they reflect the mechanical requirements of voice production. This observation may lead to the conclusion that although there is a significant association between effort and melodic qualities, the way these qualities are displayed by each performer's hands is more idiosyncratic, perhaps less consistent and therefore more difficult to capture by the chosen movement features. However, the interpretation of the movement features makes sense in relation to the estimated output.

A more generic cross-performer estimation of effort with a slightly lower goodness of fit was also achieved by models which include features that are largely shared across the two performers. Although these models are not exactly the same (with the model for Hussain requiring one additional movement feature that is not include in Sahu's model), there is a good overlap of both acoustic and movement features. Higher effort levels are required for (larger) melodic movements that start from a lower and reach up to a higher pitch and are accompanied by hand movements that are slower but with a large speed change, and in the specific case of Hussain bigger (hands move further away from the reference point).

Overall, the results demonstrate a higher degree of cross-performer overlap for acoustic than movement features in estimating perceived effort levels. This may indicate two different things: either different idiosyncratic schemes of associating effort with the body or variations in the way the same movement quality is displayed by different performers, which cannot be so easily captured by the basic movement features that were extracted from the raw data.

Despite the different ways by which effort levels might be displayed through the distinct hand gestures of each individual performer, a good part of its variance can be estimated by using the low starting and the maximum final pitch values of a melodic glide calculated on an absolute logarithmic scale (in relation to a fixed frequency). This finding indicates a strong association between voice and bodily effort that is rooted in the mechanical requirements of voice production (higher for very high and far low pitches).

Apart from this, the models that were fit individually to each performer revealed an additional association of effort with more conceptual aspects of melodic organisation. For Hussain effort is associated with areas of potential melodic activity according to the individual rāga (higher in the upper part of the octave, when ascending to the unstable 7th degree of the Jaunpurī scale) and for Sahu with the macro-organisation of the ālāp (higher with the elapsed time of the performance and therefore also the progressive pitch ascent). This justifies a double grounding for the association of effort to the voice; both a mechanical and cognitive.

Despite its lower accuracy, I would like to briefly discuss here LM 1 for Hussain. This is the only model that reveals a link between effort levels and vocal timbre (expressed through the mean value of the spectral centroid). A "heavier" or "darker" timbre in the voice is accompanied by more powerful hand gestures (typically interactions with elastic objects, as we have seen). This finding is close to interview testimony reported in section 9.3.

8.3.2. Gesture classification (GLM)

Similarly, different schemes of gesture class association to sound were identified through logistic regression. Again, the logistic models developed exhibit a higher overlap in the contribution of acoustic rather than movement qualities on the gesture classification task between the two performers. The differing movement features reveal, as with effort, the idiosyncratic element of performing such interactions or the limitation of the chosen features in describing them in a generic way. Although the used features are not exactly the same, a closer inspection suggests that some acoustic qualities are partly shared. These include the size and the duration of the melodic movement (for Hussain duration refers only to the ascending part). Additional associations to the rāga structure were also revealed for the specific case of Hussain. No agreement in movement features is detected. This observation may either suggest that the low-level movement features that were used in the models fail to capture the individual nuances of the way these gestures are performed by each individual vocalist, or that each performer moves in a different way. The melodic qualities include the size and the duration of the melodic movement (for Hussain duration refers only to the ascending part).

A gesture classification model with a more generic power was also developed, which can still achieve a good fit (although poorer than that of the best models) by using features that are largely shared between the two performers. According to this more generic cross-performer model, interactions with elastic objects are more likely performed with larger melodic movements at lower pitches and with the hands moving further apart for Hussain and remaining closer together but faster in the case of Sahu. Thus, there is a strong overlap in acoustic features but a lower agreement in movement features. Only one movement feature is shared between the two performers, and in fact it contributes to the classification probability in the opposite direction for each of the two performers; Sahu's model includes one additional movement feature that is not included in Hussain's model.

Finally, according to Table Appendix.27 and Table Appendix.33, effort is a strong classifier of MIIOs for both vocalists, with interactions with elastic objects requiring more likely higher levels of effort than those with rigid objects.

8.3.3. Goodness of fit

The overall fit in the case of Sahu is lower than in the case of Hussain for both types of models (LM and GLM). This finding can be interpreted either as a weakness of the used features in estimating effort and classifying gestures, or as an indication that Sahu is less consistent in the way he uses hand movements accompanying the voice.

Although the goodness of fit is surprisingly high, especially in the gesture classification task, there is a limitation in the methodology that is associated with the small amount of available data. Due to the small sample size, the use of cross-validation techniques was not possible, which means that the data could not be divided into separate training and testing data sets. This means that the models may be capturing patterns that are specific to an individual performer rather than describing more generic cross-performer trends. A lower fit would be expected if these measures were computed on a new dataset. Additionally, due to the small size of the available dataset in comparison to the number of features used the models may be overfitting the actual data. Despite these concerns, the model outcomes are useful for inference purposes and provide a useful insight to tendencies that will be further discussed in comparison to the findings of the qualitative methods.

8.4. Summary

The aim of the current chapter was to take an embodied approach in developing simple, interpretable models that are able to (a) estimate perceived effort levels and (b) classify MIIOs in the most accurate and compact way. Two variations of models were presented: (a) one that best fits each individual performer and (b) one that can best describe shared cross-performer behaviours. All in all, there is a higher degree of agreement between the two performers in acoustic than movement features. Differences between models can be assumed to be caused by idiosyncratic gesturing styles, a finding that in turn suggests that each performer needs to be treated individually. It may also imply a weakness of the selected low-level features in capturing the essential qualities of the responses, a limitation caused by the small data size available, and finally a difficulty in quantifying subjective aspects of phenomena that require some level of interpretation. This is where qualitative methods can contribute to a more comprehensive picture of this study by offering a methodological approach that is grounded in human experience.

Chapter 9 – Conclusions and Discussion

The current chapter provides a summary of the whole thesis. First it summarises the main results and contributions of the work in respect to the goals that were set at the beginning. It then discusses considerations for future work and concludes with a short section on the strengths and limitations of the current approach.

9.1. Research overview

This thesis opened by acknowledging bodily effort as central to the expressiveness of musical performance and by stressing the need for more systematic approaches to its role in Hindustani vocal music. It introduced the reader to the aim of this work, which is to reveal the innate relationship between sound and physical effort upon MIIOs. For this reason, the thesis has focused on the interaction possibilities that malleable (through elasticity) vs. rigid (through weight/friction) objects can afford by accounting for the physical effort that these require. The motivation for this work was twofold: On the one hand, to gain a deeper understanding of performance practice in the Dhrupad music tradition and on the other hand, to acquire knowledge of how motor-based metaphors of familiar interactions with the real world could be related to sounds in a physically plausible way.

Chapter 2 offered a review of relevant background literature. This included an overview of the Dhrupad music tradition, the tacit transmission of movement-related knowledge and its relevance to imagery and metaphoricity, an elaboration on the concept of effort in music performance and beyond, as well as an overview of computational methods for studying cross-modal relationships.

Chapter 3 outlined and justified the combination of qualitative and quantitative methods, along with the data collection process in the field in India.

Chapter 4 presented a thematic analysis of the interview material, which offered a first overview and classification of frequently used physically-inspired linguistic descriptors and informed the later stages of the analysis. Findings clearly revealed a significant visual element in the conceptualisation of music and a frequent use of sensorial descriptors and metaphors of interactions with objects. They also offered a better insight into the role of MIIOs in the Dhrupad singing tradition, as well as a foundation for the coding scheme that was later used in the annotation process of the audio-visual material.

Chapter 5, Chapter 6 and Chapter 7 presented an observational analysis for three different case studies; Afzal Hussain: rāga Jaunpurī, Lakhan Lal Sahu: rāga Mālkauńś and the Gundecha brothers: rāga Bhūpālī. The video analysis offered a third-person perspective on the identification and classification of MIIOs and thus it provided the action-oriented ontology of MIIOs (a classification of recurrent gesture types), as well as a basic typology of the most prominent melodic movements for each of the performances. It also produced the first evidence of the notable degree of systematic associations between gesture classes, melodic phrases, melodic intention and perceived effort levels for each of the four Dhrupad vocalists. At the same time, it revealed a certain degree of flexibility in the way each of the singers might use their hands to illustrate these associations. Most importantly, it lead to the conclusion that—for limited data sets as the one discussed in the current thesis—any quantitative methods applied to the data need to be case-specific and can only follow a thorough qualitative analysis of each individual performance.

Chapter 8 detailed the quantitative method in the analysis of performances by Hussain and Sahu. By fitting compact linear models to estimate effort and classify gestures based on a small number of statistically significant movement and acoustic features which were extracted from the collected quantitative data (audio and mocap), I have achieved two things: (1) to reject the null-hypothesis that physical effort and MIIO types are unrelated to the melodic context and (2) to find statistically significant features that (a) best fit each individual performer (thus reflect idiosyncratic elements of each individual performer) and (b) describe gesture–sound links in MIIOs in the most generic way for both performers (thus reflect more generic trends).

This concluding chapter will offer an overview of results drawn from the integration of the two methodological approaches, which indicate that despite a certain degree of cross-performer divergence, there is also a notable degree of overlap and consistency.

9.2. Findings

This section brings together results from both qualitative and quantitative methods in an attempt to combine, contrast and integrate them into a coherent overview. By converging results from both methods and by comparing findings between performers/performances, this section aims to shed light on inter-performer as well as cross-performer consistency in observed cross-modal (movement–sound) relationships; an alternative result would be that idiosyncratic factors prevail to a point that findings from an individual study cannot be generalised to a wider number of performers and/or rāgas. Some generalised observations emerged already from the qualitative methods by converging outcomes drawn for different performers/performances applied to a collection of qualitative information sources, such as interviews and the non-participant observation of audio-visual materials. The quantitative methods that followed offered computational solutions for discovering recurrent patterns and cross-modal links by making use of exact measurements. Although no strong claims can be made about generalising results of cross-performer consistency on a large scale due to the small sample of performances analysed, some observed trends might be informative for future work. The specific outcomes of the analysis can now be put towards answering the initially posed questions.

9.2.1. Revisiting the Research Questions

Effort levels

[1] Is effort related to the voice in MIIOs during Dhrupad vocal improvisation in a systematic way or is it just arbitrary?

[3] If such a relationship between effort and the voice exists, is it consistent among different performers or does it apply to only an individual performer?

Some interesting tendencies of effort–sound associations were observed, which have been consistent throughout each individual performance of the performers analysed but which cannot be simply generalised for all performers and performances.

[2] Does effort simply reflect the mechanical requirements of vocal production or is it also related to cognitive aspects of melodic organisation by the performer?

[4] Which aspects of the voice is effort related to, and how can their link be best formally described (for the two performers that were used in the quantitative part of the analysis, Hussain and Sahu)?

Physical effort in MIIOs is related to both the biomechanical requirements[98] of vocal production (all performers, but less so for Ramakant Gundecha) as well as more conceptual aspects of melodic organisation and intention of the performer (all performers apart from Ramakant Gundecha).

In specific, the biomechanical aspect is reflected in:

(a) the fundamental concept of cross-modal asymmetry between intensification and abatement (all performers): higher for ascending or double-sloped vs. descending pitch glides, which may possibly

[98] Similar to the 'biomechanical effort' of speech production by Nelson, 1983.

also reflect imagery of instrumental gestures;
(b) the size of ascending pitch glides (all performers, but especially prominent for Hussain and Sahu): higher for larger vs. smaller intervals;
(c) the minimum (starting) pitch and the maximum pitch reached in ascent (only examined for Hussain and Sahu, but probably it applies to the other vocalists too): increased mechanical strain required for pitches in the extreme low and extreme high pitch regions;
(d) the average pitch height (Sahu): higher effort for higher vs. lower pitches.

The conceptual aspect is reflected in:
(a) the pitch space organisation of the rāga (Hussain): higher effort in specific pitch areas of potential melodic activity according to the individual rāga;
(b) the macro-structure of the ālāp improvisation (Sahu and Umakant Gundecha): progressively higher effort with elapsed time;
(c) the melodic intention of ascending to or towards the tonic that follows the melodic movement associated with a MIIO (Hussain): higher effort levels are required when the melodic movement fails in rising to or towards the tonic but is followed by a descent upon a lower, stable note (usually the 6^{th} or 5^{th} degree) instead, while lower effort levels are observed when the melodic movement is followed by an ascent (to the tonic if the movement happened in the upper part of the octave or to the upper part of the octave if the movement happened in the lower part of the octave).

It would be an interesting task to isolate these distinct reasons (the biomechanical and the conceptual), discern hierarchies among them, and draw conclusions about how consistently they are followed by a single performer over different performances of the same and different rāgas, or by different performers. However, to be able to generalise such findings and to cover the whole Dhrupad genre with confidence would require a larger dataset and the analysis of all possible combinations of performers and rāgas. This would far exceed the scope of a single PhD work, however the current thesis offers an original methodology in approaching the material and the analysis.

According to the linear models developed to best describe the tendencies for two of the performers (Hussain and Sahu), higher bodily effort levels are accompanied by movements which are slower but with a larger speed variation, and in the specific case of Hussain by movements during which the hands move further away from the reference point (less overlap in movement features).

Overall, Hussain has exhibited the most effortful gestures, while Lakhan has been moving in the least effortful way. This finding is a result of simple visual observation and could not be deduced from the effort annotations, due to the fact that the effort scale was adapted to the effort range of each individual performer. By comparing these and other performances of the same performers (that are available outside the context of the current thesis), it seems that such differences in the overall effort exerted by different performers lie primarily on the personality and the idiosyncratic elements of the performer. However, the possibility of effort being also associated with the character and mood of the rāga cannot be excluded.

MIIO classes

[5] Are MIIO types related to specific aspects of the voice or are they completely arbitrary?

[6] If such a relationship between MIIO types and the voice exists, is it defined by melodic modes (rāgas), is it affected by idiosyncratic factors or does it reflect more generic trends observed among various performers and rāga performances?

The analysis revealed a notable degree of cross-modal consistency throughout each performance of an individual performer. There was also a considerable amount of cross-modal consistency across performers in terms of their morphology, as patterns and contours that are shared between MIIOs and the voice. Nevertheless, the exact types of MIIOs and melodic phrases that are associated with each other are not always consistent among different performers (see sections 5.3.1. 6.3.1. and 7.3.1.) and cannot be easily generalised, as they may reflect the idiosyncratic character and the gestural habits of the individual, as well as the idiosyncrasy of the rāga.

[7] Which aspects of the voice are MIIO types related to (for all performers analysed), and how can their relationship be best formally described (for the two performers that were used in the quantitative part of the analysis, Hussain and Sahu)?

For all vocalists, associations between MIIO types and the voice may be primarily dictated by analogous cross-domain morphologies, in other words the shared way by which features in both domains (movement and sound) change over time (all performers); for instance, double pitch glides (ascent-descent) are more likely associated with interactions with elastic objects that also comprise two movement phases (extension-recoil), while monotonic pitch glides are associated with gestures performed for moving a rigid object in a single direction in space. In other words, the morphology of acoustic features may be grounded in the interaction type that the imagined object employed for the task can afford. More specifically:

- Melodic movements comprising two phases (such as gamaks, double pitch glides and glides in which the emphasis is placed on the first part (the ascent)) are more likely associated with hand movements comprising two opposing directions in space, most often due to the underlying concept of elasticity (implying a variable opposing force that leads to an initial extension of the hands from the original position and a subsequent recoil back to the original position) and are performed with the most effortful gestures.
- Monotonic ascending glides tend to be mostly (but not only) associated with hand movements of a single direction or the first part of a movement comprising two directions in space (such as pulling, pushing-away and collecting gestures (constant opposing forces)) and exhibit overall lower levels of effort than the corresponding double pitch glides.
- Larger and slower melodic glides are associated with larger and lengthier extensions of the hands during stretchings (in which case the resistance is imagined to increase).

For some performers, MIIO types may be also related to the pitch space organisation of the rāga (Hussain), meaning that specific pitch regions of melodic activity for each individual rāga are performed by specific types of MIIOs (Hussain); for instance pitch glides in the upper part of the octave in rāga Jaunpurī are performed by interactions with elastic objects (stretching or pushing-compressing an elastic object), while pitch glides in the lower part of the octave are performed by interactions with rigid objects (pulling, pushing away). In some cases, a performer may also switch between or combine these two modes of engagement (Hussain); morphological and conceptual.

According to the (linear) logistic models developed to best describe the tendencies for two of the performers (Hussain and Sahu), interactions with elastic objects are also more likely to be performed with larger melodic movements at lower pitches (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 (no overlap in movement features).

[8] Is there a systematic relationship between effort and specific MIIO types?
What has been consistent across all performers is the fact that interactions with elastic objects (stretching out or pushing-compressing) tend to require higher levels of effort than interactions with rigid objects (moving a heavy object in space).

9.2.2. General conclusion: Non-arbitrariness vs. flexibility

Results indicate that there is a good deal of overlapping findings between qualitative and quantitative methods. What is, however, even more interesting is the fact that by combining and integrating the findings from the two analysis methods, there is ample evidence for non-arbitrariness in the association of the annotated responses (effort and interaction classes) with qualities of the voice and body movements for each individual performer/performance. At the same time, while some qualities shared between performers suggest that this cross-modal consistency may be grounded in profound concepts that are common among performers, there is an additional level of flexibility in the way these are presented by each individual. Overall, there is a higher degree of overlap and agreement in acoustic than movement features (at least for the two vocalists examined in the quantitative part of the analysis), either suggesting that there is a stronger idiosyncratic gesturing character or highlighting the limited success of the extracted movement features in capturing the phenomena under study.

In the quantitative methods, non-arbitrariness has been revealed through models that were largely shared between performers and flexibility through the best models that were fit separately to each vocalist's performances. The results reflect the dual nature of the physicality of the interaction in being associated with both the mental organisation of the melodic context and the mechanical strain of vocalisation. Finally, the findings lead to the conclusion that any quantitative methods applied to the data need to be case-specific and that they require a prior thorough qualitative analysis for each individual performance. Here an overview of findings is offered as a result of combining the two types of methods.

Non-arbitrariness between performers

Effort levels

Higher effort levels are expected for larger melodic ascending movements (starting from a lower and rising to a higher pitch), with ascending (or double-sloped) melodic glides being more effortful than descending. They are most likely accompanied by hand movements that are slower but exhibit a larger variation in speed over the duration of the gesture (this was only examined for two of the four performers). For all performers, interactions of deforming elastic objects tend to be more effortful than the ones of moving rigid objects in space. Some association trends are also observed among only some of the performers analysed: For two out of the four performers, progressively higher effort levels are required with the elapsed time of the ālāp improvisation (i.e. the macro-structure of rendering the improvisation).

MIIO classes

MIIO classes correspond to melodic movement types in terms of their common morphologies as patterns and shapes; double pitch glides corresponding to MIIOs comprising two phases and monotonic pitch glides corresponding to single-phase gestures. The logistic models developed in Chapter 8 revealed some shared features among the two vocalists analysed: Interactions with elastic objects are more likely to be performed at lower pitches for larger melodic movements, and possibly also slower.

Flexibility between performers

Effort levels

In the case of Hussain, effort level is additionally associated with (a) absolute pitch height, but also with the (b) rāga pitch space and (c) the melodic intention of ascending to or towards the tonic that follows the melodic movement associated with a MIIO. In specific, for Hussain higher effort levels are required (a) with higher pitches, (b) when the melody ascends in the higher part of the octave (in specific when the melodic movement starts from a lower and moves to a higher degree of the scale in rāga Jaunpurī) and (c) when the melodic movement fails in rising to or towards the tonic but is followed by a descent upon a lower, stable note (usually the 6^{th} or 5^{th} degree) instead. While for Hussain these higher effort levels are displayed by the hand(s) moving further away from the reference point, for Sahu they exhibit a larger variation in hand divergence and stronger onset acceleration. Moreover, in the case of Sahu and the Gundecha brothers, effort levels are also associated with particular types of melodic movement: higher levels accompany gamaks, medium effort levels go together with pitch glides that include an ascent (double or only ascending), while descents and prolonged notes are the least effortful. Overall, there is a bigger overlap between in the acoustic features but lower agreement in the movement features among the vocalists.

MIIO classes

In the case of Hussain, interactions with elastic (vs. rigid) objects are associated with areas of potential melodic activity according to the rāga; in rāga Jaunpurī they are more likely to occur with melodic movements which ascend into the upper part of the octave (especially the ones that reach the unstable 7^{th} degree of the scale). These are performed in the case of Hussain with the hands moving further away from each other and exhibiting a larger acceleration over the entire duration of the gesture. In the case of Sahu, the hands move less apart but do so faster and there is no association with the rāga pitch space. Thus, no shared qualities in the hand movements have been revealed by the logistic models that were developed.

In summary, the results of the current thesis reveal that, despite the flexible character of music-movement correspondences, there is ample evidence of more generic associations that are not necessarily performer-specific or stylistic. Haga (2008) has previously argued that the flexibility of music-movement correspondences resists the use of quantitative methods for drawing cross-performer findings. Indeed, a MIIO may afford multiple (imitated) interactions and sonic results, in similar ways to the multiple affordances of a real object described by Gibson's theory (1977). However, in the current thesis it has been possible to develop compact models that capture the variation of responses (effort and gesture classes) reasonably well and that overlap to a large extent between the two performers. I argue that this has been possible due to the fact that the imagined object imposes biomechanical constraints in the (imagined) interactions that it can afford according to size, shape, material etc. These biomechanical constraints define the patterns of both gestural behaviour and sound. They restrict the possibilities of cross-domain mappings and establish a narrower and more robust action–sound palette (Jensenius, 2007) than in the case of an abstract (not goal-oriented) accompanying movement. This also makes it easier to deduce an analytical model based on a small number of cross-modal features.

MIIOs offer a special case where motor imagery is "materialised" through physical actions directed towards an imagined object. Based on the findings, I suggest that the vocalists' capacity of imagining musical sound is facilitated through the retrieval of motor programs and image schemata from well-known interactions with real objects and that this may be exactly the reason for which imaginary objects are employed. Similarly to Godøy at al. (2001), I would argue that gestures in MIIOs may be conditioned by a link between the way motor programs are involved in music production on the one hand (mechanism of vocalisation), and the imagery of motor actions (MIIOs) on the other (imagined actions for vocal production, interactions with real objects and sensorimotor contingencies from previous experience of music training).
As a closing note to this section, I would like to draw attention to the fact that in the quantitative analysis the movement data has shown quite different characteristics for the two performers, and yet qualitative judgements of the movements from different observers have been quite consistent. This does not come as a surprise. Qualitative methods rely on high-level human judgements of rich movement information, while the quantitative methods of this work used specific low-level features. These low-level features are so "literal" in describing movement qualities that they can only partly capture what a human may infer from visual information when identifying and classifying gestures. While other low-level features might better capture these shared cross-performer qualities, those used here have been successful in estimating effort and classifying gestures, even if they may not completely agree with the human judgements. In fact, it could be instructive in the future to search for higher-level features (see the corresponding discussion in section 2.3.2.) in order to capture the variations that occur when performing the same type of gesture in a more generic way. Such a development would be closer to human understanding, and might also reduce the number of low-level features even further by combining them under the umbrella of higher-level descriptors.

9.3. Musicians' responds to findings

To conclude this work, I conducted an interview over Skype with vocalist Pelva Naik and instrumentalist Mohi Bahauddin Dagar on December 27, 2016. The intention was to gather responses to the findings of this thesis from musicians I had previously recorded in India. Although I had interviewed these same musicians during my fieldwork, at that time I didn't inform them about the aims of my research in order to avoid response bias. On the contrary, with this new interview I aimed at receiving their responses to concrete findings, by clearly explaining to them what the objectives of the work are and after reporting on the outcomes. I mainly focused on effort, as I was expecting the musicians to be less aware of their own tendencies in using specific types of MIIOs in relation to specific aspects of melody while singing. I first explained what type of gestures I have been focusing on (MIIOs), I then reported my findings from section 8.2. and finally I asked them to reflect on them. The interview followed a semi-structured method of open-ended questions. All quotations of this section are excerpts from the same interview and for this reason I do not include the details of the sources each time.

Dagar starts by confirming all of the findings mentioned above:

> '*All* that you have said before… *All* of it is right. Only different people having different problems. But they all come under one thing. *EVERY* one of them is right. Effort to make the phrase, effort to do this, effort to do that, effort to push with the hand the thing…'. (Mohi Bahauddin Dagar, Palaspe, India, over Skype, December 27, 2016)

The immediate response from Pelva Naik—vocalist and student of both maestro Zia Fariduddin Dagar and his nephew, instrumentalist Mohi Bahauddin Dagar—is that while the thesis seems to focus on the powerful side of a movement, for her its goal or function is to make the voice 'effortless' or 'free'. Pelva Naik and Mohi Bahauddin Dagar comment on the concept of effort:

> 'For me... I would not... I know what you mean by effort... But I would not call it effort. It is more effortless in this context'. (Pelva Naik, Palaspe, India, over Skype, December 27, 2016)

Mohi Bahauddin Dagar comments along the same lines:

> 'I think the actions of the hands are made to make the voice effortless basically. It becomes effortless when you do this with the hand. The position of the hand [he raises his right hand straight up and points at the right shoulder with his left hand saying] is relaxing something over here [shows the 'voice' travelling in space]'. (Mohi Bahauddin Dagar, ibid.)

The 'supporting' function of movement for the singing voice is emphasised by both musicians, as in previous interviews. It is in lines with Pearson (2013), who argues that movement in the context of Hindustani singing classes may just serve as a straightforward facilitation of voice production. Dagar stresses the importance of gesture transmission from teacher to student, apparently borne out in the similarity of gesturing habits among students of Zia Fariduddin Dagar:

> 'So, while teaching, Ustad [Zia Fariduddin Dagar] has shown how to use your hand, how to *support* your voice with your hand, to make that voice or to make that particular pattern. All

> the students of Ustad, they will mostly have the same kind of gestures, because Ustad has told them: "You have to use this gesture… This supports this"'. (Mohi Bahauddin Dagar, ibid.)

Naik also stresses the importance of hand movements when learning a new rāga and acknowledges that a specific set of gestures used to be taught by Zia Fariduddin Dagar:

> 'He [Ustad Zia Fariduddin Dagar] would tell you. He used to tell. Look at my hand. And then he would show the movements through the hand. And I would follow it. And I would see. And it would come really precise and correct if I would follow his hands. In the beginning, when you don't know the rāga properly'. (Pelva Naik, ibid.)

Initially, both musicians acknowledge mechanical aspects associated with the voice timbre or 'tonal quality':

> '*Control* over the voice. Yes, the mechanics. So, you use a gesture whereas you feel the force coming from the navel. And then you bring it out like that. So, the tonal quality of the sound will change accordingly'. (Mohi Bahauddin Dagar, ibid.)

Later in the interview, this mechanical functionality is also associated with pitch and precise intonation as defined by the specific rāga:

> 'It *has* to do with the rāga, because it enhances the actual placing of the note in the rāga'. (Mohi Bahauddin Dagar, ibid.)

To my question on whether movements of different effort levels can be associated with the same …/*b*7*b*6 glide of rāga Jaunpurī, Dagar responds:

> 'Same gestures [of similar types and comparable levels of effort] can be used in different rāgas. It depends on how much intensity… I mean how much thick or thin you want a phrase. Yes, I thought of them only thinking of Jaunpurī. So, these actions can be repeated in another rāga, in a third rāga also'. (Mohi Bahauddin Dagar, ibid.)

This quotation comes in contrast to the findings for Afzal Hussain, who seems to exert higher levels of effort in the upper part of the octave in rāga Jaunpurī. This might also reflect the role of idiosyncrasy in the association between bodily effort and voice. Whereas Mohi Bahauddin Dagar would show a wider range of effort levels in his movement, Afzal Hussain might be focusing on the characteristic melodic activity of rāga Jaunpurī in the upper part of the octave. However, Dagar also acknowledges that there are specific notes in each rāga that need to be treated in a different way, even in terms of hand movements:

> 'So, if you are showing gāndhār [the 3rd degree of rāga Darbāri] which is really soft and really subtle, there are certain kinds of gestures for that [he gives examples]… if you want to show a phrase that has moved very subtly'. (Mohi Bahauddin Dagar, ibid.)

The testimony of these two performers seems to support the findings of this thesis regarding the role of movement in singing, insofar as they acknowledge its double 'supporting' functionality: the mechanical (to make the voice sound ‘effortless’ or ‘free’) and the mental (in helping the singer become more precise controlling the voice). Mohi Bahauddin Dagar talks about the role of movement:

> ‘Both [mechanical and mental] will exist together. One will be to project the voice easier. Another is to change the thickness and the thinness of the voice at the same time; controlling the voice basically. Having more control of the voice, at the same time singing it with an ease’. (Mohi Bahauddin Dagar, ibid.)

“Thickness” is a term used by Dagar to denote the “weight” of a note. What is most striking about the concept of “weight” is that “heavier” melodic movements are associated with moving the hands away or closer to the performer’s body, while melodic movements that are performed while moving the hands in the vertical axis are ‘lighter’ (‘more pointed’ to use Dagar’s own words):

> ‘So… [he sings twice a /Ni\Dha movement in rāga Jaunpurī, the first time while moving the hand up and down and the second time while moving his hand out as if against a resistance (away from his body) and again in (closer to him) and adds] this Ni [moving the hand upwards] is much more pointed in a way. And [moving away and then closer] this is much more flat; it is *thicker* over here, I can feel it. [moving again his hands upwards) I can feel it, it becomes thinner up there’. (Mohi Bahauddin Dagar, ibid.)

To my question about whether ‘thicker’ means ‘heavier’, Dagar adds:

> ‘Yes, *heavy*! You want to keep it as thick as possible. So, because Ra, Na, Na… The words have to also have weight. So, to emphasise what the weight is of the words and what the weight is of the patterns, you show it either by the finger or you show it by the hand, like this [he moves his hand as if he is pushing something downwards or as if halting something by pointing his palm away from the body]. Like āndolan [he gives two examples of an āndolan while oscillating his hands and adds] It will show the *intensity* of the oscillation’. (Mohi Bahauddin Dagar, ibid.)

It is not clear from this excerpt to which aspects of the voice the term “weight” refers; it could be the timbre of the voice or the type of melodic movement in terms of shape or contour. In fact, when Dhrupad teachers stress the need for "precise intonation”, they do not necessarily refer exclusively to pitch. Instead, the term “intonation” may function as:

> ‘a translation of uccār (or uccāran), literally 'pronunciation', a term which refers not merely to pitch, but to many different aspects of the articulation of a note, including the way it is approached, quitted, oscillated or taken as part of a glissando between two pitches, its rhythmic value, its loudness or softness etc.’
> (Sanyal & Widdess, 2004: 8).

However, what is acknowledged in the last excerpt is a strong link between a powerful quality in the movements and a powerful quality in the voice. The association of effort to timbre was not revealed through the linear models in the quantitative analysis of this thesis (exception is Hussain’s LM 1), and this could either mean that the specific measures of timbre that were used failed to capture and represent the nuances of the voice or that Dagar's terms 'thick' and 'heavy' should actually be seen as describing melodic contour, and not timbre at all.

This last excerpt also highlights an important asymmetry in the way vocalists might use space, between the absolute (i.e. in relation to the ground) direction of the vertical axis and the relative (i.e. in respect to the performer's own body) horizontal movement of the hands. While hand movements executed in space in relation to the performer's body (closer or away) are associated with a "heavier" quality in the voice, (absolute) vertical hand movements make the voice sound "lighter". According to my own observations and interview testimonies of Chapter 4, movements performed in relation to the singer's body are more frequently associated with MIIOs (the idea of an object existing between the hands that fights against changes of its physical conditions), while vertical movements are used less frequently and do not seem to involve an imagined opposing force acting on the hands (the exception was the case of Ramakant Gundecha). For instance, moving the hands away from the body against some resistance was associated with the image of pushing the hand through water in Chapter 4. The idea of moving in water is an important motor-based metaphor in Dhrupad singing.

As there is an extensive discussion on the association between vertical spatial positioning and pitch height in literature (e.g. Eitan & Granot, 2006), it is worth underlining here that according to Mohi Bahauddin Dagar, the spatial height does not necessarily reflect pitch height, but has mostly a mechanical functionality:

> 'The middle octave Sa is over here [he places his hand on top of his head]. And the upper octave Sa is over here [he shows the upper part of his chest]. So, anything about here [shows a point at the height of his face], you have to push your hand and raise the voice. It also shows the upper octave, but it doesn't actually mean upper octave. The position of the hand [he points at the shoulder] is relaxing something over here [shows the 'voice' travelling in space]'. (Mohi Bahauddin Dagar, ibid.)

During the interview I also referred to the instructions that Zia Fariduddin Dagar gave to participants during group teaching sessions I participated in Greece in 2007 ('You should think you are stretching an elastic object') and asked them to comment on how these relate to the voice.

> 'It is almost like tangible, almost like you are *sculpting* it'. (Pelva Naik, ibid.)

> 'The word tān means to stretch. It is the elasticity of the note, elasticity of imagination'. (Mohi Bahauddin Dagar, ibid.)

During the interview Dagar refers to the pulling of particular syllables ('how much do you pull the 'ri''), at which point I ask him whether it might be the case that he is using the verb 'pull' in this expression because of some association with gestures he performs on the instrument (the rudrā vīṇā) when he needs to ascend to a higher pitch through a glide. This is Dagar's response:

> 'Yeah. You are right. Because you are physically pulling the note. Because… [he gives a few examples where he imitates the pulling of a string on the rudrā vīṇā while singing various pitch glides on the syllables 'rae', 'na'] when you are playing the simple notes [here he means syllables], it is always *on* the note [meaning that the pitch should not move much]. Sometimes

> you will pull that note. But when you pull a simple syllable like Rae - ae [closed ‘a’ as in ‘ɜ’ of the IPA chart] it will be one note; it will never do that [he sings a double-sloped pitch glide], because that would be grammatically wrong. The ‘a’ [open] you will pull more [he makes a longer phrase starting with ‘ri’ and changing to ‘ɑ’, which is used for making double-sloped pitch glides]. So, ‘ae’ [‘ɜ’ of the IPA system] will go for one note. ‘A’ [‘ɑ’ of the IPA system] will go for twice the length or three times the length. And ‘i’ will go even higher. And ‘e’ will go even higher.
>
> It [rudrā vīṇā] is an instrument of measure. It measures all the smallest microtones. Which is a strong characteristic in Dhrupad. So, that’s why the rudrā vīṇā is used. Because all the different Re’s [‘e’ in IPA], all the different Sa’s [‘ɑ’ in IPA], all the different Ga’s [‘ɜ’ in IPA] you can play them on the vīṇā and then you can… even an entire phrase you can time it [he gives examples with different syllables (Ta, Na and Ri) while moving his hands]. So, which syllables to use… [is important]. Because in Hindi language there is the ‘i’ and ee, the u and the uu. And there is Ra, na. So, there are three things’.
> (Mohi Bahauddin Dagar, ibid.)

Dagar acknowledges that while in the rudrā vīṇā it is about the physical pulling of the strings, in the voice it is notes that need to be ‘pulled’. He also attests the importance of used syllables (short vs. long vowels as pronounced in Hindi) in singing and their impact on the type of melodic movements that should be executed on the instrument. In other words, according to Dagar, the extent of the pull (the size of the produced interval as well as its duration) is associated with the nature of the syllable sung, with short vowels (the ‘simple’ syllables) of the nom-tom system sung (and played) shorter than long ones.

‘Pulling’ is used here for two different purposes in the melody: the one describes the progressive pulling of a string that is associated with the tracing of an ascending pitch glide and the other reflects the positioning of the hand (and string) at a certain distance from its rest position that is required for keeping a note steady at a pitch that is higher than the one produced when the string is at its rest position. This interview testimony confirms findings of video analyses about the association between pulling gesture classes and melodic phrases. According to these findings, it is not just grappolo grips (where the hands do not move in space) that can produce single prolonged notes, but also the action of manipulating an object; pulling (as well as stretching, pushing and collecting) gestures can accompany both pitch glides and single, prolonged notes (Afzal Hussain: Table 5.19, Lakhan Lal Sahu: Table 6.4, Umakant Gundecha: Table 7.5 and Ramakant Gundecha: Table 7.11).

Bahauddin also later elaborates on stretching gestures by acknowledging the idea of a resistive force through another motor-based metaphor: grinding. However, Dagar stresses the fact that this movement quality should not enter the mechanical process of vocal production, but instead should feature as melodic tension. In other words, this tension should not be felt in the body of the performer, but it should only be conveyed to the audience:

> 'When he is doing this, his voice should not become like this [he makes a gesture showing something tensed]. It should be relaxed. And show the grinding. It is the gesture of grinding two surfaces together. Two stones surfaces together. So, when I am showing this, I can show the intensity of it. But voice should not cringe to show this gesture. […] You can show the tension, you can make the person feel the tension, but you cannot be yourself tensed for it, that's what I am trying to say'. (Mohi Bahauddin Dagar, ibid.)

Dagar makes this distinction in order to highlight examples of vocalists who make what he considers to be 'wrong' associations with the voice, such as straining the voice when performing powerful movements:

> 'So, many times when you see musicians, there is too much effort happening. Then somewhere I feel that the gestures are wrong. And they should not be like that'.
> (Mohi Bahauddin Dagar, ibid.)

One of Mohi Bahauddin Dagar's comments touches upon a possible explanation for the high degree of movement idiosyncrasy revealed through the analysis:

> 'Until the last generation—my father's generation and Ustad's [Zia Fariduddin Dagar] generation—these kinds of actions were found in a lot of musicians of their age. […] they had very similar kinds of actions and similar kind of language to explain the music. Now new generations are not using these kinds of actions. They are using some different actions, which we find a little bit wrong or they don't *fit* where they should be fitting in a way. I think that the next generation of musicians didn't listen or look carefully at these kinds of action and they don't believe in the schools of thought'. (Mohi Bahauddin Dagar, ibid.)

This interview confirmed most of the findings of the current thesis, while acknowledging that vocalist are expected to have their own ways each for associating bodily effort with the voice. Most importantly, both musicians emphasised the double supporting role of movement for the voice, the mechanical and the mental. The interview material revealed that various associations exist that reflect the timbre of the voice, the type of melodic movement and particular notes of the rāga. Finally, Mohi Bahauddin Dagar acknowledged that despite the similar gesturing habits of Zia Fariduddin Dagar's students arising as a result of his teaching of movements, nowadays vocalists tend to invest less effort into learning them. This might also partly explain the low degree of consistency in the movement features of the models developed.

9.4. Original Contributions

9.4.1. Hindustani vocal music

The current work has offered the first systematic study of the role of effort in Hindustani vocal music performance for interactions with imaginary objects. The outcomes of the current thesis regarding MIIOs are in agreement with Rahaim's (2009) findings on the 'paramparic body' in Hindustani vocal music, according to which singing movements are viewed as a result of both unconscious inheritance and personal choice. On the one hand, I have observed how several members of the Dagar gharānā use similar gestures that imitate interactions with the real world. On the other hand, the differing gesturing manners of the performers that were observed during MIIOs—and especially the distinct way by which the two Gundecha brothers gestured during the same rāga performance—are a good indication that body movements are not just a matter of the rāga but also the idiosyncrasy of performers and their distinct gesturing preferences. Drawing on these findings, I argue that the relationship between movement and voice resides in the rudimentary ecological knowledge of the interaction possibilities (imaginary) objects can afford (Gibson, 1977). The relationship is experienced by the performer and perceived by an observer through the magnitude and change of effort over time, which is determined by the imagined physicality of the object, such as viscosity, elasticity, weight, friction etc. The choice of interaction and object type, and consequently also the movement–sound relationship, highlights cognitive structures that are supposed to have emerged for a number of reasons: the recurrent, slightly varied, instances of previous sensorimotor experiences[99] with real objects (Varela et al., 1993); the tacit transmission of bodily dispositions through visual engagement with the teacher during classes (Rahaim, 2009); the repeated exercise of 'correct' movement–sound relationships learned over years' practice with the teacher (Rodger et al., 2007); the mechanical requirements of vocalisation; and possibly also the mimicking of instrumental gestures on Indian string instruments.

The opposing forces (and the required effort) associated with the imagined object are of different nature and magnitude in different pitch regions, as they can afford a dual facilitating role, mechanical or mental. Specifically, they can serve the potential needs of melodic expression in either facilitating the mechanics of voice production or highlighting potential regions of special activity in the rāga space or in the melodic structure of the ālāp improvisation. A performer can intuitively and 'naturally'—as a Hindustani musician would often report—make a cognitive shift from one class of gestural interaction (and type of imagined opposing force) to the other and can thus adjust the associated exerted effort according to the requirements of the rāga, the ālāp structure or the mechanical strain of the voice.

9.4.2. Data set

Among the most significant contributions of this thesis is the capturing of data with an unusually strong ecological validity. The current work is based on original movement and audio material, which was collected explicitly for this work in India. To the best of my knowledge, this is the first time that a

[99] Which do not include the sonic modality in the case of MIIOs.

3D motion capture system has been used in Hindustani singing and indeed with recordings that were made in India. Even more importantly, the system was set up in domestic spaces and was used to record entire performances in real conditions rather than to conduct designed experiments, in order to ensure ecological validity in the data collection process. For this, however, it was necessary to overcome various challenges so that the data would not be compromised by incompleteness or errors in its collection. These challenges are described in section 9.6.2.

9.4.3. Methodology

An original contribution to the study of movement–sound relationships has been offered by combining qualitative methods to analyse interviews and audio-visual material with quantitative methods to analyse raw mocap and audio data. The combination of qualitative and quantitative methods has potential for answering complex research questions. The qualitative methods draw upon the interpretative power of the human mind and were therefore more appropriate at the beginning of the project for developing a concrete narrative and hypothesis. The quantitative methods harness the 'cold' identification of repeated patterns in the responses and were used in the later stages for developing compact descriptions of the phenomena under study. Additionally, the combination of methods allows the use of a richer collection of data.

9.4.4. Novel features

The use of features that have been proposed previously by other scholars (Nymoen et al., 2013) did not produce any successful models for effort estimation and gesture classification of MIIOs. Therefore, I devised novel features that form part of the original contribution of this thesis. The movement features included two aspects: first, the calculation of the measure of distance by accounting for the type of handedness rather than calculating the distance between the hands in the same way for all cases; second, the calculation of hand position in relation to the torso instead of the ground as the reference point. The acoustic features included the minimum and maximum pitches to describe critical points of melodic movements, polynomial expressions to describe the pitch contour with more or less detail, as well as mathematical expressions to describe the strength and asymmetry of pitch ascent vs. descent. Additionally, all of the pitch features mentioned above were calculated on three different scales that were meant to help in differentiating between pitch height of the voice and degrees of the scale (related to the specific rāga).

9.5. Future considerations

Given the complexity of the topic, this thesis has raised various directions for future research. A non-exhaustive list of suggestions follows for those who wish to advance the findings of this study.

9.5.1. Data set

Larger datasets that include multiple performance takes of the same rāga for each performer would be beneficial for enabling a more systematic comparison between performers, performances and rāgas. They would also allow more computationally advanced and less labour-intensive methods to be applied in the selection of the best models for the effort estimation and gesture classification tasks. Finally, they would allow for the division of the data into separate training and testing sets (which was not possible with the current size of the dataset) and the use of regression analysis as a prediction method on new data.

9.5.2. Inter-person and intra-person consistency

For the moment, conclusions on the cross-modal consistency between performers (inter-person) as well as between performances of the same vocalist and rāga (intra-person) remain tentative due to the limited amount of data. It would be instructive to use a larger dataset in order to systematically examine such issues.

9.5.3. Adding the second-person perspective

As I explained in section 8.1.1. , the gesture class and effort level annotations formed the 'ground truth' for the quantitative methods that were later applied to the measured data. However, qualitative research may be viewed in the hard sciences as less rigorous as a result of the subjective interpretation involved in the observation and annotation process. Therefore, given the opportunity, spending some time with the musicians in India and asking them to reflect on their own performances and annotate their own recorded material would be beneficial in adding a second-person perspective to the methodology (Leman, 2010).

9.5.4. Continuous effort level annotation & cross-modal similarity

One interesting addition to the third-person perspective of the analysis would be to conduct an alternative annotation of effort (of the already recorded audio-visual material) by asking annotators to rate perceived effort levels on a continuous scale, with pressure or muscle tension sensors acting as the interface for them to record their ratings. In MIIOs the type of object determines not only the magnitude but also the type of the resistive force and therefore the general rules and patterns of behaving and sounding. Stretching an elastic object is associated with the motor imagery of acting against a varying force (Hooke's law) which increases with distance and causes a subsequent recoil to the original position. Pulling a heavy object (friction or gravity), on the other hand, is associated with acting against a constant force which is not necessarily succeeded by a recoil. Therefore, in manipulating an elastic object one expects to see the performer first increasing and then decreasing the effort he appears to exert on the object by following a certain pattern, while in the case of a rigid object one expects to see the performer exerting a constant amount of effort over time.

By accounting for the bell-shaped profiles by which effort changes over time—from gentle to firm or from gradual to abrupt—it is possible to systematically examine which time-dependent features of movement and sound are similar to effort (Godøy et al., 2016) in terms of the way by which they co-evolve in time. The analysis conducted in the current work was based on single effort values[100] and can be seen as a dimensionality reduction process which has highlighted the most salient cross-modal features. A next step would be to examine how these features vary in form over time to investigate the temporal congruence between different sets of features and effort (now considered as a time-series). This could be done by using Functional Data Analysis methods, such as CCA (Nymoen et al., 2011 and Caramiaux et al., 2009, 2010a) or by modelling and comparing contours of time-varying features in movement and audio (through B-spline curves, piecewise cubic Bezier etc.), such as in Canazza et al. (1998), Maestre et al. (2006), Bianco et al. (2010) and Battey (2004); the latter for pitch in Hindustani music. In this way, it might be possible to meet the challenges of addressing the continuous along with the discrete cognitive representation of sound and movement in Dhrupad singing. Hindustani singers engage in explicit discussions about pitch continua and the imagery of melody as movement in an imaginary pitch space that is transversed through physical motion. Thus, Hindustani vocal music is a good place to try to address this challenge.

9.5.5. Designed experiments

(a) Physiological measurements of muscle tension

Taking into account physiologically measureable quantities that may reflect effort would also be advisable. Given the opportunity, new material could be collected that would also include measurements from biosensing technologies for capturing the level of muscle tension or force (Gibet, 2010) in performers' hands during MIIOs. At the stage of project planning, such sensors were acquired and their use considered, but the idea was abandoned due to practical complications (a single person could not easily control the simultaneous capturing of video, audio, mocap and biosensor data).

(b) Rudrā vīṇā motor equivalence

Considering the earlier discussion on motor equivalence for the translation to pitch height (4.2.), it would be interesting to systematically study the action-oriented ontology (Leman, 2012) of the rudrā vīṇā gestures; measuring the intensity of movement in the larynx of the vocalist (Pearson, 2013) would be much harder. By using the annotated material as audio stimulus, it would be possible to measure physiological properties that are related to the bending of strings in producing various melodic glides, such as the length and tension of the string or the distance from its rest position. By doing so, it would be possible to model and develop a gesture-based typology of mīṇḍs (in a similar

[100] Global measures that were extracted from time-varying features through descriptive statistics, such as mean and standard deviation.

fashion to sliding-tones reported in Henbing & Leman (2007) for the Chinese guqin[101]) and to conduct a comparison between vocal performance and rudrā vīṇā instrumental playing in terms of the relationships between movement, sound and effort. Parametric models can be developed by using polynomial fitting techniques (linear combination of B-spline curves, piecewise cubic Bezier etc.), such as the ones used by Canazza et al. (1998), Maestre et al. (2006), Bianco et al. (2010) and Battey (2004) (the latter for modelling the articulation of sustained vowel performances in Hindustani vocal music).

(c) Assessment of effort perception and gesture classification by observers

Another aspect that future research designs might consider is an evaluation of how observers assess the effort that performers' gestures are perceived to require and how they classify MIIOs. Although it is supposed that a subject can make judgments about the effort levels they perceive from a performer, this process is not transparent in the observation analysis. Visual information is a rich source of data, and the observer's eyes could be focusing upon several different aspects of the movement. Although previous studies have shown that kinematic displays can produce similar results to videos (Scully, 1986), it might turn out that observers actually rely on secondary visual information conveyed through video for assessing the performers' effort levels; facial expressions and hand grips may also contribute along with the kinetic and dynamic aspects of the movements. It would be therefore useful to prime participants with both the video and light-point displays extracted from mocap data. Additionally, vocal timbre and quality of the voice is known to convey effort (Wallmark, 2014) and may also be implicated in the annotation strategy of an observer. It would also be interesting to examine whether one is in the position to assess how much effort each gesture is imagined to require based solely on the voice.

Hence, in order to better understand how effort is perceived and to disambiguate the abovementioned aspects, perceptual experiments should be conducted and findings should be compared by asking subjects to assess effort based on different stimuli, either of a single modality or by combining modalities. Additionally, a video from a recorded vocal performance could be combined with wrong sounds, e.g. the video of a MIIO that requires high effort levels could be paired with the audio of a MIIO with low effort levels, or the video of an interaction with a rigid object could be matched with the sound from an interaction with an elastic object, and vice versa. Musicians often refer to the fact that gestures should not be 'wrong', such as the next excerpt by Mohi Bahauddin Dagar:

> 'If you look at it, it should not seem wrong. Sometimes if you see just the actions, you don't want to listen to them sing. Or if you just hear them sing, you cannot imagine they are doing these kinds of actions'.
>
> (Mohi Bahauddin Dagar, Palaspe, India, over Skype, December 27, 2016)

Therefore, various stimuli extracted from MIIO events can be used to prime participants in such perceptual studies, such as the following:

(a) Only video material (no audio);

[101] Traditional Chinese musical instrument of the zither family.

(b) Only audio (no video);
(c) Both video and audio;
(d) Point-light displays extracted from mocap data (no audio);
(e) Both point-light displays and audio;
(f) Both video and audio, paired in wrong ways;
(g) Both point-light displays and audio, paired in wrong ways.

(d) Assessment of vocal 'weight' by musicians

Musicians during interviews (especially Mohi Bahauddin Dagar) often mentioned 'weight' as a cross-modal feature that can refer to both the body movements and the voice. To explore this observation further, it would first be necessary to examine what musicians mean by 'weight' in the voice and whether this sonic aspect can be measured. At the moment, it is unclear whether they refer to the vocal timbre, to the contour of the melodic movement or to both.

(e) Mapping validation

A systematic study for the validation of the linear multi-parametric mappings that have been deduced through this work could be conducted by using them on existing VMIs (such as those by Boem, 2014; Bennett et al., 2007 and Mulder, 1998) and evaluating them on the basis of physical plausibility.

9.5.6. Features

The approach that was followed in the regression analysis focused on rather low-level features, but it would also be beneficial to examine even more compact model alternatives. These could be achieved by either testing higher-level descriptors or applying a PCA parametric reduction method to larger datasets, in case more material is collected and analysed.

9.5.7. Non-linear modelling

Although the linear models that were explored in this thesis have produced fairly good fits to the data, an inspection of the residuals has revealed some limited but identifiable patterns that could be better described through a non-linear model. Thus although it is a valid first step to start from the assumption that the relationship is linear, further exploring non-linear models might improve the estimation of the response.

9.5.8. Movement data from video

A substantial body of video material exists for Hindustani singing that could be potentially used in case video analysis methods advance in the future at a point where effort can be safely deduced from two-dimensional data (Nister, 2004).

9.5.9. Broadening the perspective - novel interaction paradigms

The current study has reported the first ecologically valid mapping between effort, movement and voice in MIIOs. By acknowledging the dynamic properties of effort as innately linked to artistic expressive power (Luciani, 2009), I argue that artificial interactions with digital sound can be enhanced. The mappings that have been developed can contribute to the improvement of electronic musical instruments that are grounded in action-based paradigms and ‘repeated patterns of embodied experience’ (Varela et al., 1991) for digital musical expression, such as the *‘SculpTon’* by Boem (2014), the *‘DAMPER’* by Bennett et al. (2007) and the concept of *‘Sound Sculpting’* by Mulder (1998).

An important conclusion of this thesis is that movement and sound are related to each other in different ways for different performers, while some vocalists may even switch between or combine these days. Hussain follows the pitch space organisation of the particular rāga, Sahu and Hussain follow the biomechanical requirements of voice production, Sahu and Umakant Gundecha follow the progressively intensifying macro-structure of the ālāp rendition, while all performers (but mostly the Gundecha brothers) are also heavily focused on melody in a micro-timescale, in specific on morphologies that are shared cross-modally. This conclusion leads me to further suggest that in designing an electronic musical instrument a more flexible mapping scheme should be aimed for, which would enable the performer to switch between mapping modes.

The findings of this thesis allow me to propose a new hierarchical approach (Françoise et al., 2012) for the design of mappings in digital musical instruments. This new approach conforms to embodied music theories to a greater degree than those previously put forward. Instead of thinking of mappings in engineering terms as ‘gesture-input’ (from sensors) and ‘sound-output’ (to audio engine), I argue that movement and sound should be thought of as two parallel processes which emerge from a common idea. In line with the findings of the current thesis, this common idea can be represented through the amodal concept of effort and it can be thought of as a perceptual mapping layer from which both movement and acoustic features originate. In this sense, mapping is not necessarily conceived of as representing a layer that exists between movement (input) and sound (output), but as a layer that merges the two modalities into a single performance experience. By keeping the coefficients as defined by the models developed in this ecologically valid way, ‘transparency’ (Fels et al., 2002) or ‘control intimacy’ (Moore & Yamamoto, 1988) can be enhanced in the use of a digital musical instrument.

9.6. Strengths & Limitations

9.6.1. Methodology

As much as bringing the advantages of ecological validity, the approach that was followed has also posed important challenges and led to limitations. Applying a combination of qualitative and quantitative methods to real performance recordings has brought about a difficult trade-off between an ethnomusicological approach and a systematic analysis of designed and repeatable experiments. An examination of relevant work reveals that there is a broad spectrum of studies on computational approaches of music-related movements that rely on designed experiments, where subjects are asked to move in response to a source that is regarded as ground truth; either a specific audio stimulus or the score of a specific music piece. In the current work, musicians were interviewed and asked to improvise without any prior clarification on the purpose of the study and without any specific instructions; musicians were just free to perform. Although this mixed type of methodology strengthens the rigour of the analysis, a number of risks are also associated with it:

(a) Despite its ecological validity, the absence of a widely accepted ground truth predefined and shared between subjects—such as an audio stimulus or the score of a composition—imposes the need to substitute it through qualitative methods of annotating the material. Apart from the fact that this is a long and time-consuming process, it could be also criticised for its interpretative and therefore subjective character. This was nevertheless addressed through cross-validation of the coding scheme by other annotators.

(b) This work focuses on very specific gesture types, which means that the frequency by which they will appear in normal performance practice cannot be guaranteed beforehand. By avoiding typical designed experiments, where subjects are allowed to respond to specific stimuli a number of times, and instead using an ethnographic approach of recording a typical performance, there was a high risk of collecting only a small amount of data or not observing the relevant gestures at all. The dataset is indeed quite small for producing generic conclusions applicable to the entire Dagar gharānā of Dhrupad.

(c) Despite initially considering the possibility of instructing all musicians to perform a specific rāga so as to have a common ground for comparison, it was finally decided not to do so. This decision was made mainly due to the fact that I had no prior intuition as to whether MIIOs would appear at all for specific performers or in specific rāgas. This decision does not allow a distinction to be drawn between performer- and rāga- related idiosyncratic elements.

An additional limitation to the objectives of this study is related to the fact that only third-person judgments were used on effort level and gesture classification. Although ideally the material should have been annotated by the musicians themselves, for practical reasons of time and finance this was not possible.

9.6.2. Data collection

Collecting original data in the field made it necessary to undergo a challenging process of setting up the equipment in domestic spaces around India and brought about various problems not typical of laboratory work. These complications included long and unpredictable power cuts[102], ambient noise (in audio and optical data[103]), other scheduled activities taking place in the recording space, the potential for the equipment to be damaged accidentally during the normal daily uses of the space, as well as unpredicted last-minute changes of arrangements or the presence of an unexpected number of people as an audience. Extra demands in the data collection process came from the lack of assistance in carrying the equipment (weighing about 50kg) around India, as well as setting it up in a limited amount of time to use it for just few days in each city. The recording sessions themselves were equally difficult due to the requirement of controlling multiple processes in parallel (a mocap recording, two audio recordings (close-miking and ambience) and a video recording), while at the same time fulfilling the role of the audience which is essential in Hindustani music-making and then conducting the interviews (usually with video, audio and mocap being still recorded). Given the opportunity to re-do the whole process or to continue collecting data in India, it would be certainly advisable to have a second person as assistant. Additionally, given the fact that the quantitative analysis of movement data in the current work depended fundamentally upon observations of the positions of performers' hands, the use of accelerometers and gyroscopes would have been a more flexible solution to better address the peculiar demands of fieldwork in India. Due to their small size and weight, the absence of occlusion issues between markers and the fact that they can run off batteries, this might have been a more appropriate solution. However, affordable technology of this kind was limited at the time I conducted my field work (2010-2011). On the other hand, a database of multiple body positions may allow further studies requiring the computation of the whole upper-body skeleton to evolve in the future.

[102] Power cuts were normally unexpected and could vary between 2 - 7 hours per day.
[103] It included excess traffic noise from Indian highways, cooking sounds (such as mashing ginger) or other sources of domestic auditory noise and reflections in the optical data due to the small size of the recording spaces or secondary sources of light (all windows had to be covered by improvised methods and all reflective items had to be covered by cloth or removed from the recording space. Examples include a Saraswati statue in one space that had to be covered, and the entire collection of furniture in another room, including a sofa, bed and wardrobe that all had to be removed).

9.7. Closing remarks

The goal of this project was to investigate cross-modal (movement–sound) relationships in MIIOs observed in the singing practice of Hindustani music. Results indicate that despite the flexibility in the way Dhrupad vocalists might use their hands while singing, the notable degree of association between classes of MIIOs, their exerted effort levels and melody provides good evidence for non-arbitrariness.

Many alternative directions can be taken from here onwards and the material originally recorded for this study may be useful in future work. For the interested researcher wishing to carry this work forward, it is therefore my intention to upload the data to public space, such as Repovizz (2015) or Mova (Alemi, 2013).

APPENDIX

A. Chapter 5: Afzal Hussain, special cases: Ambiguous gestures and mīṇḍ slope types

(a) Ambiguous gesture class–melodic phrase associations

Table Appendix.1 displays results for the association analysis of the 41 pairs of melodic phrases and ambiguous gesture classes (stretching/pulling, stretching/pushing and pulling/collecting).

Gesture class vs. melodic phrase Hussain	Melodic phrase							
	octave range							
	lower part			5th degree	upper part		tonic	unclass.
	3-2-1 cadence	1/4\2 glide	2/5\4 glide	5th approach & hold	.../b7\b6 glide	5/2\b7\b6 glide	tonic approach & hold	other
stretching/pulling	1	5	3		13		1	
stretching/pushing	1				5	1		1
pulling/(collecting/taking)			1		5	2	1	1

Table Appendix.1. Afzal Hussain: Ambiguous gesture class vs. melodic phrase.

Overall, the results displayed here tend to follow the values that would be expected from a mix of two clear-cut gesture classes involved in each of these ambiguous annotations and thus they support the trends already discussed for the unambiguous gesture classes. More specifically:

- The ambiguous case of stretching/pulling can be performed with any of the three pitch glides, but with a bias towards the upper part of the octave, as would be expected if we combined the unambiguous cases of stretching and pulling gestures.
- Similarly, the ambiguous case of stretching/pushing displays a clear trend for being associated with pitch glides in the upper part of the octave, as would be expected if the two unambiguous cases of stretching and pushing gestures were combined.

Nevertheless, the ambiguous cases between taking/collecting and pulling gestures show an unexpected tendency towards the higher parts of the octave, which cannot be explained easily as a mix of the clear-cut cases.

(b) Special cases of gesture class (all)–mīṇḍ slope type associations

As already mentioned, the rendering of the annotated melodic phrases is never repeated in the exact same way but shows a number of variations. In the most typical case, a double-sloped pitch glide is performed with a single syllable and it is associated with a single manual gesture consisting of two parts (with the gesture performed in opposing directions for each of the ascending and descending parts of the glide). However, in a few cases glides are single-sloped, hence there are no ascending or descending parts. Also, in a few cases, each of the two parts of the melodic glide is associated with a separate gesture, one for each of the ascending and descending parts, each part performed to a different sung syllable. By looking closely into such complications or specificities in the melodic rendering, some of the findings in the previous analysis can be better explained.

Table Appendix.2 shows the number of double- and single- sloped (only ascent or only descent) pitch glides for each melodic phrase type and gesture class.

Gesture vs. melodic phrase Hussain	Melodic phrase & mīṇḍ slope type								
	1/4\2		2/5\4	…/♭7\♭6			5/2'\♭7\♭6		steady
	double slope	descent	double slope	double slope	ascent	descent	double slope	descent	tonic approach & hold
stretching				8			2		
pushing				2	2				
pulling	4	1	5	3	1	1			
throwing									
collecting/taking									2
grasping (grappolo & kit-flying)									4
stretching/pulling	5		3	10	1	2			
stretching/pushing				2	3		1		
pulling/ (collecting/taking)			1			5		2	

Table Appendix.2. Afzal Hussain: Ambiguous gesture class vs. melodic phrase & pitch slope type.

Table Appendix.3 gives an overview of associations between gesture classes, mīṇḍ slope types and melodic intention. Part of this table is produced by combining information from clear-cut and ambiguous gesture classes.

Gesture vs. melodic phrase & melodic intention Hussain	Melodic phrase, mīṇḍ slope type & melodic intention					
	lower part of octave		5th	upper part of octave		tonic	other
	1/4\2	2/5\4	5	…/♭7\♭6	5/2'\♭7\♭6	1	unclass.
stretching				double	double		
pushing			approach & hold	double / ascent			
pulling	double	double		double / descent	descent *(indirect deduction)*		
throwing							single notes
collecting/taking				descent *(indirect deduction)*	descent *(indirect deduction)*	approach & hold	
grip (kite-flying or grappolo)						approach & hold	

Table Appendix.3. Afzal Hussain: Gesture class vs. melodic phrase & mīṇḍ slope type[104].

[104] '*Indirect deduction*' indicates that the trend described in the table had to be deduced by combining information from ambiguous and other unambiguous gesture classes, as no single occurrence for the specific class was observed.

Some of the glides that appear in Table Appendix.3 as single-sloped are in practice part of a double-sloped pitch glide that is performed through two consecutive gestures (one each for the ascending and descending parts of the glide). In a few cases, a single phrase is performed by two consecutive gestures, the first of which is associated with the initial, opening note of the phrase (in the annotation process these single notes are labelled as 'other' or 'unclassified'). Table Appendix.4 displays the number and type of such peculiar phrases that are performed by using two consecutive gestures. Arrows show the paired gestures and the order of their melodic rendering, while the associated numbers indicate how many times this was observed during performance.

Gesture class vs. melodic phrase & pitch glide slope Hussain	Melodic phrase & pitch glide slope					
	1/4\2	…/b7\b6			5/2'\b7\b6	single notes (various) - extracted from 'other'
	double slope	double slope	ascent	descent	descent	
stretching						
pushing						
pulling	4					1
throwing						4
collecting/taking						1
stretching/pulling	5		1			
stretching/pushing		2	3			
pulling/(collecting/taking)				5	2	1

Table Appendix.4. Afzal Hussain: Ambiguous gesture class vs. melodic phrase performed with two consecutive gestures (numbers indicate the frequency of occurrence and arrows indicate the order of execution).

The combination of information displayed in Table Appendix.3 and Table Appendix.4 suggests the following:

1. Stretching and pushing gestures, along with ambiguous cases between the two, are typically associated with double-sloped or only ascending, but never solely descending, pitch glides in the upper part of the octave. Pulling gestures as well as their ambiguous combinations with other gestures can be associated with descending pitch glides too.

2. A few double-sloped pitch glides seem to start from a single, emphasised note that is performed with a separate 'opening' gesture.

3. Findings indicate a tendency for the ambiguous cases between pulling and collecting gestures to be associated with only the descending part of pitch glides in the upper part of the octave, which is either performed as part of a double-sloped glide (with the ascending part using another type of gesture) or as a solely descending monotonic glide. This might explain the unexpected bias for ambiguous cases of collecting/pulling gestures to occur in the upper part of the octave. Collecting gestures refer only to the descending part of a pitch glide that is performed in the upper part of the octave.

4. With respect to the as-yet-unexplained bias displayed in Table 5.16, one of the glides that does not ascend to the upper part of the octave has a single descending slope. Therefore, in practice double-sloped 1/4\2 glides performed with a pulling gesture ascend three times out of four into the upper part of the octave. Following this observation, I can now suggest that in these cases the melody tends to rise towards the tonic.

By combining the information from Table Appendix.3 on mīṇḍ slope types and from Table 5.20 on effort levels and melodic intention, it can now be argued that the perceived effort levels are not simply associated with the melodic phrase type or the melodic intention towards the tonic, but also with more fundamental concepts of ascending vs. descending movements. For example, .../b7\b6 glides of a single ascent or a double slope (stretching or pushing) are performed with higher levels of effort than when the initial ascending part is missing (as observed with some of the pulling gestures).

(c) Ambiguous gesture class–melodic phrase, mīṇḍ slope type & melodic intention associations

In this section the behaviour of ambiguous gesture classes in relation to the melody ascending toward the tonic is brought into the analysis. The purpose is to examine whether these cases follow and confirm the basic behaviour revealed from the analysis of the unambiguous gesture classes.

Table Appendix.5 refers to ambiguous gesture types for .../*b*7*b*6 pitch glides. For each of the ambiguous gestures that are performed with this glide it displays the total number of occurrence, and the number of occurrence for each separate type of melodic glide and separate melodic intention of ascending towards the tonic.

Gesture class vs. melodic intention & mīṇḍ slope type after .../*b*7\b6 Hussain	**Melodic intention & mīṇḍ slope type**			
	total number	**slope type**	**melodic intention of reaching the tonic**	
stretching/pushing	5	double slope	1	yes / rise, but after great hesitation
			1	no / fall
		ascent	3	no / fall
stretching/pulling	13	double slope	7	yes / rise, but after hesitation
			3	no / fall
		ascent	2	no / falls
			1	yes / rise, but after hesitation
pulling/(collecting/taking)	5	descent	3	no / fall
			2	yes / rise

Table Appendix.5. Afzal Hussain: .../b7\b6: Ambiguous gesture class vs. mīṇḍ slope type & melodic intention (to tonic).

A comparison between Table Appendix.5 and Table 5.13 may suggest the following:

1. In general terms, the stretching/pulling and stretching/pushing ambiguous gesture types tend to follow the values that would be expected if the unambiguous classes were combined.

2. The trend seen in the pulling/collecting gestures combined with the unambiguous cases of pulling gestures seems to imply that collecting gestures are associated exclusively with the descending part of …/*b*7*b*6 glides, with no clear intention of rising to the tonic. Nevertheless, this conclusion can only be tentative, as it is indirectly deduced from the ambiguous cases (of pulling/collecting), while no single event of collecting gestures has appeared in the upper part of the octave.

Similarly, the following applies for the ambiguous gestures of 5/2’*b*7*b*6 pitch glides:

Gesture class vs. melodic intention & mīṇḍ slope type after 5/2'\b7\b6 Hussain	Melodic intention & slope type			
	total number	slope type	melodic intention of reaching the tonic	
stretching/pushing	1	double slope	1	no / fall
pulling/(collecting/taking)	2	descent	1	yes / rise
			1	no / fall

Table Appendix.6. Afzal Hussain: 5/2 ’\b7\b6: Ambiguous gesture class vs. mīṇḍ slope type & melodic intention (to tonic).

Findings of Table Appendix.6 imply that:

1. Similar to …/*b*7*b*6 pitch glides, collecting gestures refer to the descending part of the glide (in the upper part of the octave). This is a conclusion reached indirectly.
2. Pushing gestures are not associated with an intention to reach the tonic.

Similarly, the following table refers to the ambiguous gestures of 1/4\2 glides:

Gesture class vs. melodic intention & mīṇḍ slope type after 1/4\2 Hussain	Melodic intention & slope type			
	total number	slope type	melodic intention of reaching the tonic (moving from the lower into the upper part of the octave)	
stretching/pulling	5	double slope	5	no rise to upper part of octave

Table Appendix.7. Afzal Hussain: 1/4\2: Ambiguous gesture class vs. mīṇḍ slope type & melodic intention (towards tonic).

The combination of findings from Table Appendix.7 and Table 5.16 indicates that 1/4\2 pitch glides performed with stretching gestures are not followed by the intention to rise towards the tonic.

Finally, Table Appendix.8 displays the ambiguous gesture classes in 2/5\4 pitch glides:

Gesture class vs. melodic intention & mīṇḍ slope type after 2/5\4 Hussain	Melodic intention & slope type			
	total number	slope type	melodic intention of reaching the tonic (moving from the lower into the upper part of the octave)	
stretching/pulling	3	double slope	3	no rise to upper part of octave
pulling/(collecting/taking)	1	double slope	1	no rise to upper part of octave

Table Appendix.8. Afzal Hussain: Melodic intention in rising from the lower to the upper part of the octave after the use of a 1/4\2 mīṇḍ in relation to the ambiguous gesture type used.

Similarly, findings from Table Appendix.8 combined with those from Table 5.17 indicate that 2/5\4 pitch glides performed with stretching or collecting gestures are not followed by the intention to rise in the upper part of the octave and towards the tonic.

(d) Effort–gesture class (ambiguous), melodic phrase & mīṇḍ slope type associations

In Table Appendix.9 associations between ambiguous gestures classes and effort levels are displayed:

Effort vs. gesture glass, melodic phrase & mīṇḍ slope type Hussain	Melodic phrase, mīṇḍ slope type & effort levels (mean values)									
	lower part of octave		upper part of octave					tonic	cadence	'other'
	1/4\2	2/5\4	…/b7\b6			5/2'\b7\b6		1	3-2-1	unclass.
	double slope	double slope	double slope	ascent	descent	double slope	descent	approach & hold		(single notes)
stretching/pulling	6.2	6.667	4.9	5				7	5	
stretching/pushing			8	8.33		10			7	8
pulling/ (collecting/taking)					3		1.5	2		2

Table Appendix.9. Afzal Hussain: Mean effort levels of ambiguous gesture class vs. melodic phrase & mīṇḍ slope type.

Overall, the effort values seem to follow the trends expected from the simple combination of the unambiguous classes found in Table 5.19. The only exception here is the …/b7\b6 double slope in stretching/pulling, which was expected to have a slightly higher level of effort as a combination of stretching gestures with high effort levels and pulling gestures with medium effort levels. Still, overall the ambiguous classes seem to confirm the trends of the unambiguous classes and they certainly do not contradict them.

The following summarises what the findings for the special and ambiguous gesture classes added to the analysis of the clear-cut cases:

1. The ambiguous cases of gestures do not contradict, but further support the trends observed by the unambiguous gesture classes.

2. It is mainly the first (ascending) part of each glide that defines the association with the gesture class. In fact, whenever no ascending part is included in the melodic glide, the behaviour of the ambiguous classes cannot be simply explained as the combination of the unambiguous (double-sloped) classes (e.g. the pulling/collecting ambiguous class).

3. This last observation, combined with the annotated effort levels, seems to suggest that judgements on the perceived levels of effort are primarily based on the first part of a double-sloped pitch glide, i.e. with the ascent. Additionally, double-sloped or ascending glides are associated with higher levels of effort, while descending glides are perceived as having lower levels of effort. For example, the effort annotations for glides in the upper part of the octave, which are either double-sloped or ascending, are much higher than those for descending glides.

B. Chapter 8: Determining Best Models

B1. Linear Model (LM) for Effort Level Estimation

(a) Afzal Hussain: LM

First a model was fit on the core feature set (Table 8.4 for movement features and Table 8.5 for acoustic features). From this set only two features seem to be strongly contributing to the model and they can estimate 1/4 of the effort level variance:

Afzal Hussain: LM 1	
LM Feature	**Coefficients**
PitchSD_lin	+1.36***
SpctrdMean	-0.51*
R^2adj. (p<.001)	***0.23***

Table Appendix.10. Afzal Hussain: LM 1.

According to LM 1, with just two audio features it is possible to estimate 23% of effort level variance. In short, the model can be written as:

$$effort = 1.36 \cdot pitchSD - 0.51 \cdot spctrdMean$$

This means that, the higher the standard deviation of pitch (which is indicative of the melodic movement interval) and the lower the mean of the spectral centroid (which represents the average value of the voice's brightness), the higher the perceived effort level. In other words, the larger the melodic movement and the less bright the voice, the higher the perceived effort. Due to its low probability value (***), the feature that contributes most strongly to the estimation of effort level is the standard deviation of pitch. The next most important feature (slightly higher probability value, but still significant) is the mean value of the spectral centroid. In fact, if we compute an alternative model that uses only the standard deviation of pitch on a linear scale, 20% of effort level variance can still be explained. The fit increases slightly when pitch is expressed on an absolute logarithmic scale.

A significant improvement in the fit is achieved by (a) expressing pitch in terms of the critical pitches of the melodic movement on logarithmic scales and (b) taking into account the handedness of the gesture for the computation of position, distance and its first derivatives. From the three critical pitches, *pitchminPost* is excluded due to its strong correlation (over 0.9) to *pitchmax*. A considerable tendency for AbsVelocityMean to be associated with AbsVelocitySD (cor=0.81) is also detected and should be considered in the final selection of the best model. The following four models of similar goodness of fit (between 0.53 and 0.6) are produced by using the critical pitch values in the two types of logarithmic scales:

Afzal Hussain: LM 2-5 FOR PITCHES EXPRESSED IN ABSOLUTE AND RELATIVE LOGARITHMIC SCALES				
LM Feature	**Absolute logarithmic scale**		**Relative logarithmic scale**	
	2	**3**	**4**	**5**
	Coefficients	**Coefficients**	**Coefficients**	**Coefficients**
PitchminPre	-2.17***	-2.13***	-1.65***	-1.62***
Pitchmax	+2.22***	+2.3***	+2.42***	+2.48***
AbsVelocityMean	-1.25*	-1.34**	-1.28**	-0.65*
AbsVelocitySD	+0.99*	+0.93*	+0.71.	
HandDistanceMean_accord	+1.02**	+0.65*	+0.71*	+0.59*
VerticalEnergyUM	-0.51.			
R^2adj.	***0.55 (p<.001)***	***0.53 (p<.001)***	***0.6 (p<.001)***	***0.59 (p<.001)***

Table Appendix.11. Afzal Hussain: LM 2-5.

When pitches are expressed on an absolute logarithmic scale, up to 55% of effort level variance can be estimated with the use of six features. According to model 2, high bodily effort levels are required by Afzal Hussain for melodic movements which start from a low and reach up to a high maximum pitch (on an absolute logarithmic scale). They are performed through slow movements of extending the hands further away from the reference point, which are slower in the first part of the gesture (moving away from the reference point) than in the opposite (recoil or retract) direction and exhibit a large overall variation of speed. If the compactness of the model is of higher priority, then by removing VerticalEnergyUM (due to its higher probability value) still a 53% estimation rate can be achieved with just five features (model 3).

When pitches are expressed in relation to the tonic of each octave, the overall fit can reach up to 60% in the estimation of effort level variance with the use of just five features. Pitches refer this time to areas within an octave, or in other words to the critical pitch values of glides (mīṇḍs) associated with the characteristic organisation of the specific rāga. According to model 4, high bodily effort levels are required by Afzal Hussain for melodic movements which start from a low degree (closer to the tonic) and reach up to a high degree (further away from the tonic) of the scale within the boundaries of each octave. They are accompanied by movements which are slow on average but with a large variation of speed, where the hands move further away from the point of reference (according to the type of gesture, this may mean hands moving apart from each other or from the torso).

In case the compactness of the model (the number of used features) is of higher importance, by removing AbsVelocitySD it is still possible to achieve a 59% estimation accuracy with just four features. In fact, if it is more important to avoid the long manual annotation process for the handedness of the gesture that is required for the computation of the HandDistanceMean_accord feature, by excluding this feature a 55% estimation accuracy can still be achieved, this time with just three features.

According to the information of Table Appendix.11, the most important are two acoustic features, so by only keeping those a very strong model is achieved that can explain 53% of Hussain's effort variance with just two features. This means that the three movement features that are removed contribute by only 7%.

Afzal Hussain: LM 6	
LM Feature	**Coefficients**
pitchminPre_rellog	-1.53***
pitchmax_rellog	+2.24***
R^2adj. (p<.001)	***0.53***

Table Appendix.12. Afzal Hussain: LM 6.

Best model selection

Models 2-6 are all good candidates with a similar goodness of fit that ranges between 53-60%. They all include pitch features calculated on either an absolute or relative logarithmic scale. The absolute logarithmic calculations for models 2 and 3 produce a fit of 53-55%, while the relative logarithmic calculations (in relation to the individual tonic) for models 4 and 5 produce a slightly higher fit of 59-60%. Interestingly, only two critical pitch values calculated on a relative logarithmic scale suffice for an estimation of 53% of effort variance (model 6). This means that there is a strong dependence of bodily effort levels on rāga-related (Jaunpurī in this case) aspects of the melody rather than on the mechanical requirements of vocal production expressed by the absolute pitch height. Although model 4 produces a slightly higher fit than model 5 (by 1%), the latter model (5) seems like a good trade-off due to the lower number of features (four instead of five) and therefore it is the one chosen as the best for Afzal Hussain. Although movement features differ between most of the models that were explored, there are two basic movement features that are shared among the strongest model candidates and these are the mean absolute value of velocity (calculated as the mean of the two hands) and the mean distance between the moving parts and the point of reference (which depends on the type of gesture). This leads to the conclusion that slow gestures of moving the hands further away from the point of reference while accompanying ascending melodic movements (or the ascending parts of melodic movements) that rise from a lower minimum to a higher maximum pitch are the most effortful ones.

(b) *Lakhan Lal Sahu: LM*

The same process was repeated for the second performer. First the model was fit on the initial feature set and then additions or substitutions of alternative and novel features were made with the aim to raise the fit while keeping the model compact.

By excluding those features from the core feature set that did not contribute significantly to the estimation of effort or that were highly co-linear (AbsAccelerationMean with HandDivergenceSD), the following model was deduced:

Lakhan Lal Sahu: LM 1	
LM Feature	**Coefficients**
PitchMean_lin	+0.69***
PitchSD_lin	+0.85***
HandDivergenceSD	+0.39*
OnsetAccelerationMean	+0.55**
R^2adj. (p<.001)	***0.38***

Table Appendix.13. Lakhan Lal Sahu: LM 1.

According to this model, with the use of just four non-correlated features one can estimate about 38% of the effort level variance for Sahu. The substitution of the linear pitch features by their absolute logarithmic equivalent does not lead to any significant improvement in the fit, but by using only the acoustic features (calculated on the absolute logarithmic scale) a 36% of effort level estimation can still be achieved.

The fit is raised by expressing melodic movement in terms of its critical pitch values calculated on an absolute logarithmic scale, from which only the maximum pitch proves to be significant:

Lakhan Lal Sahu: LM 2	
LM Feature	**Coefficients**
PitchSD_abslog	+0.56***
Pitchmax_abslog	+0.9***
HandDivergenceSD	+0.43*
OnsetAccelerationMean	+0.43*
R^2adj. (p<.001)	***0.44***

Table Appendix.14. Lakhan Lal Sahu: LM 2.

By using the critical pitch values (*minPre* and *max*) as in the case of Afzal Hussain, a model of similar success rate is produced:

Lakhan Lal Sahu: LM 3	
LM Feature	Coefficients
PitchminPre_abslog	-0.95***
Pitchmax_abslog	+1.91***
AbsVelocityMean	-1.31**
AbsVelocitySD	+1.35**
R^2adj. (p<.001)	***0.42***

Table Appendix.15. Lakhan Lal Sahu: LM 3.

It is interesting, that the use of just the two most important pitch features from each of the last two models can explain 37-40% of bodily effort level.

Lakhan Lal Sahu: LM 4		Lakhan Lal Sahu: LM 5	
LM Feature	Coefficients	LM Feature	Coefficients
PitchSD_abslog	+0.63***	***PitchminPre_abslog***	-0.87***
Pitchmax_abslog	+0.8***	***Pitchmax_abslog***	+1.71***
R^2adj. (p<.001)	***0.4***	***R^2adj. (p<.001)***	***0.37***

Table Appendix.16. Lakhan Lal Sahu: LM 4 & 5.

Last, the relationship between effort and elapsed time is examined for Sahu too. The progressive intensification of the rāga seems to have an effect on the level of power by which Sahu performs his gestures. As the ālāp improvisation progresses, the gestures' effort level rises and almost ¼ of the effort level variance can be estimated just by the elapsed time.

Lakhan Lal Sahu: LM 6	
LM Feature	Coefficients
Elapsed time	+0.03***
R^2adj. (p<.001)	***0.24***

Table Appendix.17. Lakhan Lal Sahu: LM 6.

Nevertheless, the combination of this feature with other features does not lead to a model with a better fit than model 2.

Best model selection

Model 2 is the most promising due to the highest adjusted R^2 value and the smallest number of features, while model 3 ranks second in goodness of fit. The combination of the two acoustic features

(*PitchSD_abslog* & *Pitchmax_abslog*) in model 2 can explain the biggest part of effort level variance, while the addition of the two movement features (HandDivergenceSD & OnsetAccelerationMean) raises the overall fit by only 4%. Although model 4 has the strong advantage of the small number of features, model 2 is chosen as the best model for Sahu due to its slightly better fit (by 4%) and the fact that it also includes movement features. According to model 2, a higher bodily effort level is required by Sahu for larger melodic movements (large glides) that reach up to a higher maximum pitch (on an absolute logarithmic scale)[105]. They are accompanied by movements with a larger initial (onset) acceleration and a larger variation of divergence (speed of moving the hands apart from or closer to each other) throughout the entire duration of the gesture.

[105] As the size of the pitch glides is expressed on a logarithmic scale in reference to a specific frequency, it means that naturally glides of higher pitches (frequencies) will be bigger in size (even when compared to equivalent glides of a lower octave). Nevertheless, no co-linearities were found.

B2. General Logistic Model (GLM) for Gesture classification

The two response classes (rigid and elastic) are inferred by creating a dummy variable, which takes the value of 0 for elastic gestures and 1 for rigid gestures. In other words, the probability values presented here correspond to the probability of the gestures falling under the category of interacting with rigid objects rather than elastic. In specific, the value and sign of each feature's coefficient describe the estimated amount and direction respectively by which the likelihood of a MIIO being classified as an interaction with a rigid object would increase in case the specific feature was increased by one unit. The opposite sign would apply for the likelihood of classifying a MIIO as an interaction with an elastic object.

(a) *Afzal Hussain: GLM*

First a model is fit on the core feature set, from which only four features prove to be significant and produce a good fit. Pitch features are computed on either a linear or an absolute logarithmic scale.

Afzal Hussain: GLM 1		
GLM Features	**Coefficients**	**AUC**
PitchMean_lin	+1.22*	0.59
PitchSD_lin	-1.62**	0.63
AbsAccelerationMean	3.68**	0.54
SD_vel_betw	-3.04**	0.61
Overall		***0.89***

Table Appendix.18. Afzal Hussain: GLM 1.

Afzal Hussain: GLM 2		
GLM Features	**Coefficients**	**AUC**
PitchMean_abslog	+1.18*	0.59
PitchSD_abslog	-1.79**	0.63
AbsAccelerationMean	+3.4**	0.54
SD_vel_betw	-2.8**	0.61
Overall		***0.87***

Table Appendix.19. Afzal Hussain: GLM 2.

According to GLM model 1, with only four basic movement and acoustic features a success rate of close to 90% can be achieved in classifying gestures as interactions with rigid vs. elastic objects; pitches are calculated here on a linear scale. By expressing the pitch features on an absolute logarithmic scale, a good fit—although slightly lower—can still be achieved. Interactions with rigid objects are more likely to be performed at higher pitches for smaller melodic movements, while the hands exhibit a high (absolute) mean acceleration (computed on the mean position of the two hands) and a low variation of divergence (variation of speed in moving the hands apart). It should be noted here that the contribution of *PitchMean* is less significant than that of the other features (higher p-value: *).

A comparable fit is achieved by the following model:

Afzal Hussain: GLM 3		
GLM Features	**Coefficients**	**AUC**
PitchMean_abslog	+1.45**	0.59
PitchSD_abslog	-1.74**	0.63
HandDistanceMean	-1.1**	0.63
Overall		***0.86***

Table Appendix.20. Afzal Hussain: GLM 3.

According to this model, interactions with rigid objects are more likely to be performed at higher pitches for smaller melodic movements (calculated on an absolute logarithmic scale) and for gestures where the hands remain confined, i.e. they remain close to each other.

The fit is further raised by substituting some of the movement features of the core set by their equivalents that are calculated according to the handedness of the gesture:

Afzal Hussain: GLM 4		
GLM Features	**Coefficients**	**AUC**
PitchSD_linear	-1.16**	0.63
OnsetAccelerationMean_accord	-1.09*	0.57
AbsAccelerationMean_accord	+4.01***	0.66
HandDivergenceSD	-3.1***	0.61
Overall		***0.9***

Table Appendix.21. Afzal Hussain: GLM 4.

According to this model, interactions with rigid objects are more likely to accompany smaller melodic movements. They are performed by hand gestures that exhibit a low acceleration value at the onset and a high (absolute) overall acceleration value in extending the hand(s) from the reference point, as well as a low variation in the speed of moving the hands apart. Although there is a tendency for AbsAccelerationMean_accord and HandDivergenceSD to be correlated with each other (cor=0.8). This is not considered prohibitive for the validity of the model.

Some promising models are produced by using alternative ways for expressing melodic movement. The use of critical pitches calculated on a relative logarithmic scale lead to the following model:

Afzal Hussain: GLM 5		
GLM Features	**Coefficients**	**AUC**
PitchminPre_rellog	+1.08.	0.52
Pitchmax_rellog	-2.00**	0.72
AbsAccelerationMean	+4.29**	0.54
HandDivergenceSD	-3.27**	0.61
Overall		***0.89***

Table Appendix.22. Afzal Hussain: GLM 5.

Through this model a classification success rate of more than 89% can be achieved by using four features. According to this model, interactions with rigid objects are more likely to be performed when the hands exhibit a high absolute mean acceleration value (computed from the mean position of the two hands calculated over the entire duration of the gesture) and a low variation in the hands' divergence. These are performed while accompanying melodic movements which start at a relatively high degree and ascend to a relatively low degree of the scale within each octave (the target always being higher than the starting pitch), in other words for smaller melodic movements.

By adding the ascending time from the minimum to the maximum pitch, the following model of a higher classification rate is deduced:

Afzal Hussain: GLM 6		
GLM Features	**Coefficients**	**AUC**
timemax	-3.04**	0.84
Pitchmax_rellog	-2.93**	0.72
PitchSD_abslog	-1.17.	0.63
AbsAccelerationMean	+3.4*	0.54
HandDivergenceSD	-3.12**	0.61
Overall		***0.95***

Table Appendix.23. Afzal Hussain: GLM 6.

According to this model, interactions with rigid objects are more likely to be performed for small and quick melodic movements which ascend to a relative low degree of the scale within each octave. These are performed by hand gestures that exhibit a high absolute mean acceleration value (computed from the mean position of the two hands calculated over the entire duration of the gesture) and a low variation in the hands' divergence.

Using the coefficients of a 2^{nd} order polynomial expression fit to pitch values (on a linear scale), the following model is deduced:

Afzal Hussain: GLM 7		
GLM Features	**Coefficients**	**AUC**
coeff₂	-18.82*	0.53
coeff₁	-7.87**	0.46
intercept	-4.22*	0.66
OnsetAccelerationMean_accord	-1.42*	0.57
AbsAccelerationMean_accord	+3.83***	0.66
HandDivergenceSD	-3.38***	0.61
Overall		***0.9***

Table Appendix.24. Afzal Hussain: GLM 7.

According to this model, interactions with rigid objects are more likely to be performed for lower values of the 2nd order polynomial coefficients and intercept (i.e. for wider melodic movements that are executed during a longer stretch of time and start from a lower pitch), with hand gestures that exhibit a low acceleration value at the onset and a high (absolute) overall acceleration value in extending the hand(s) from the reference point, as well as a low variation in the hands' divergence. There is a certain degree of correlation between the 1st order coefficient and the intercept (cor=0.89), which could be considered as just acceptable (< 0.9). However, the fit rises only slightly in comparison to model 2 (from 90.22% to 90.34%) and it requires two more features.

By describing melodic movement through the innovative feature of *Pitchmov* on a linear scale the following model is also devised:

Afzal Hussain: GLM 8		
GLM Features	**Coefficients**	**AUC**
Pitchmov_lin	-1.35**	0.71
OnsetAccelerationMean_accord	-1.05*	0.57
AbsAccelerationMean_accord	+3.74***	0.66
HandDivergenceSD	-3.04***	0.61
Overall		***0.89***

Table Appendix.25. Afzal Hussain: GLM 8.

According to this model, an almost 90% classification success rate can be achieved by using just four features. Interactions with rigid objects are more likely to be performed for melodic movements in which the ascending part does not intend not to be much larger than the descending part, and with hand gestures that exhibit a low acceleration value at the onset and a high (absolute) overall acceleration value in extending the hand(s) from the reference point, as well as a low variation in the hands' divergence.

What is also examined here is the significance of pitch features (p-value) and the direction of their effect (sign of coefficients) when used individually in classifying gestures. The relationship between the classified gesture and an individual feature is not necessarily the same as when used in combination with other features. Various pitch descriptors are tried here, however only the following shows a strong contribution in gesture classification when used individually:

Afzal Hussain: GLM models 9-12 on INDIVIDUAL features		
GLM Features	**Coefficient**	***AUC***
PitchSD_lin	-0.61*	***0.75***
Pitchmax_rellog	-0.95*	***0.74***
Pitchmov_lin	-1.00**	***0.7***
Pitchmov_rellog	-1.06**	***0.69***

Table Appendix.26. Afzal Hussain: GLM 9-12.

Due to the negative signs of all the above mentioned pitch descriptors, it is deduced that interactions with rigid objects are more likely performed with smaller melodic movements in which the pitch does not ascend into the higher degrees of the scale and in which the ascent is not much larger than the descent. According to Table Appendix.26, the relationships when features are used individually remain similar (coefficients of similar value and same sign) to those found when the features are used in combination with other features (of the previous models).

Last, the classification power of the perceived effort level in discerning between interactions with elastic and rigid objects is also examined here. This can be seen as an attempt to better understand whether one type of gesture (e.g. interaction with elastic object) tends to require more bodily effort (as perceived by the observer) than the other. For this, a GLM model is fit to the single feature of effort level.

Afzal Hussain: GLM 13		
GLM Feature	**Coefficients**	**AUC**
Effort level	-0.82***	***0.89***

Table Appendix.27. Afzal Hussain: GLM 13.

The results indicate that the (perceived) effort level is a strong classifier of interactions with elastic vs. rigid objects, with the former requiring a higher amount of bodily effort than the latter. This is also illustrated in the boxplot of Figure Appendix.1.

Figure Appendix.1. Afzal Hussain: Boxplot for effort levels vs. gesture classes.

The confusion matrix of Table Appendix.28 shows that the success rate of classifying gestures based simply on their perceived effort level can reach 78%.

Afzal Hussain: confusion matrix (classification on effort)				
Logitpred/class		**actual class**		
		elastic	**rigid**	**total**
inferred class	**elastic**	12	8	20
	rigid	6	38	44
total		18	46	64
success Mean		***0.78***		

Table Appendix.28. Afzal Hussain: Confusion matrix for classification success based on effort level.

Best model selection

Models 1-8 are all promising model candidates due to the small number of used features (three to six) and the high classification success rates (between 87%-95%). However, model 6 is selected is the most successful due to its highest AUC. In the graph of Figure Appendix.2 one can identify one feature (timemax) having a large impact on the model due its large AUC value, as its ROC curve ascends fast towards the top.

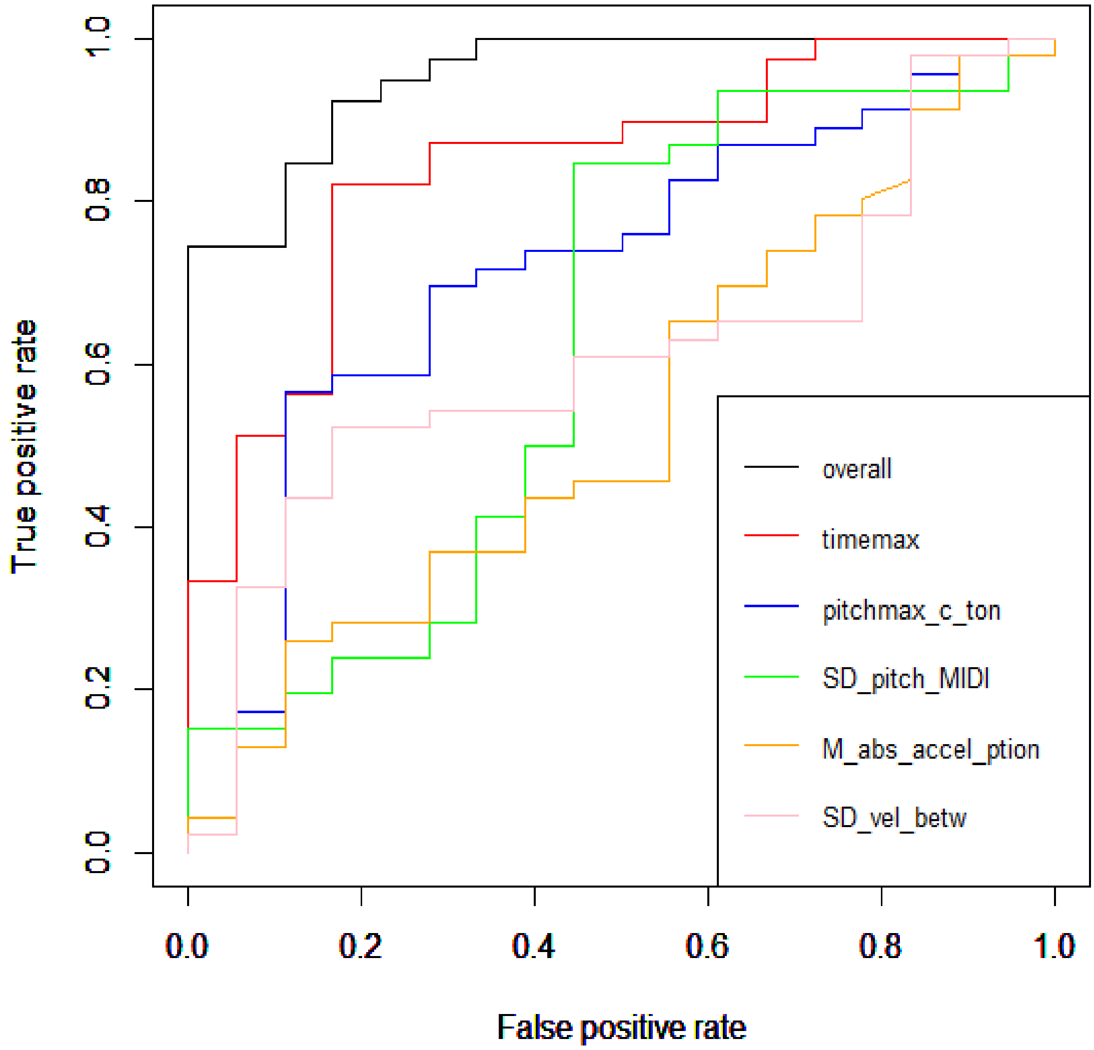


Figure Appendix.2. Afzal Hussain: ROC for GLM 6 and its individual features.

Table Appendix.29 displays the confusion matrix for GLM 6. The mean classification success rate reaches almost 90% and is expressed as the proportion of correct classifications over the entire set of responses:

Afzal Hussain: confusion matrix (GLM 6)					
Logitpred/class		actual class		
		elastic	rigid	total
inferred class	elastic	14	2	16
	rigid	4	37	41
total		18	39	57
success Mean		***0.89***		

Table Appendix.29. Afzal Hussain: Confusion matrix for classification success of GLM 6.

(b) *Lakhan Lal Sahu: GLM*

First a model is fit on the core feature set. By removing all non-significant as well as co-linear features, the following model is developed, which consists of only two movement features:

Lakhan Lal Sahu: GLM 1		
GLM Features	**Coefficients**	**AUC**
AbsVelocityMean	-1.21*	0.62
OnsetAccelerationMean	-0.98.	0.52
Overall		***0.68***

Table Appendix.30. Lakhan Lal Sahu: GLM 1.

These two features are not significantly correlated (abs(cor)=0.54) and can correctly classify 68% of the gestures.

The calculation of the pitch features on a relative logarithmic scale results in a promising model that consists of four (non-correlated) features that can classify almost 78% of the gesture types.

Lakhan Lal Sahu: GLM 2		
GLM Features	**Coefficients**	**AUC**
PitchMean_abslog	+0.51.	0.51
PitchSD_abslog	-1.03***	0.7
AbsVelocityMean	-1.58*	0.65
HandDistanceMean	+0.98**	0.54
Overall		***0.78***

Table Appendix.31. Lakhan Lal Sahu: GLM 2.

According to model 2, interactions with rigid objects are more likely to be performed for smaller melodic movements at higher pitches, while the hands move slower but further apart from each other. Due to the far lower statistical significance of mean pitch in the classification, the use of the voice at higher pitches is considered less important, while the use of small melodic glides is the most contributing feature.

By expressing the shape of the melodic movement through a 2nd order polynomial fit, a compact model is produced, which consists of four non-co-linear features and has a high classification rate:

Lakhan Lal Sahu: GLM 3		
GLM Features	**Coefficients**	**AUC**
PitchSD_abslog	-0.73**	0.7
$Coeff_2$	+0.71*	0.63
AbsVelocityMean	-1.48*	0.65
HandDistanceMean	+0.78*	0.54
Overall		***0.8***

Table Appendix.32. Lakhan Lal Sahu: GLM 3.

The 2^{nd} order coefficient expresses the sharpness of the curve (i.e. the rate of change in pitch over time) and according to model 3 it is more likely larger in interactions with rigid objects. In other words, in interactions with rigid objects, $coeff_2$ can be interpreted as either a melodic movement of a smaller duration (if we imagine the melodic movement going through three fixed critical pitches) or as a larger melodic movement (a larger pitch interval for a fixed duration). Due to the likeliness of interactions with rigid objects in being associated with smaller melodic movements (negative coefficient of *PitchSD_abslog* in interactions with rigid objects), it can be assumed here that $coeff_2$ reflects melodic movements of a smaller duration in interactions with rigid objects. Thus, according to model 3, interactions with rigid objects are more likely to be performed with melodic movements of a smaller size (pitch interval) and duration, and with the hands moving slower and further apart from each other.

The classification power of the perceived effort level is also examined here. The gesture classification is based here only on the perceived effort levels.

Lakhan Lal Sahu: GLM 4		
GLM Feature	**Coefficients**	**AUC**
Effort level	-0.65***	0.77

Table Appendix.33. Lakhan Lal Sahu: GLM 4.

The results suggest that—as in the case of Hussain—the effort level is a strong classifier of gestures between elastic and rigid object interactions, with the former requiring higher bodily effort levels than the latter. This is also illustrated in the boxplot of Figure Appendix.3:

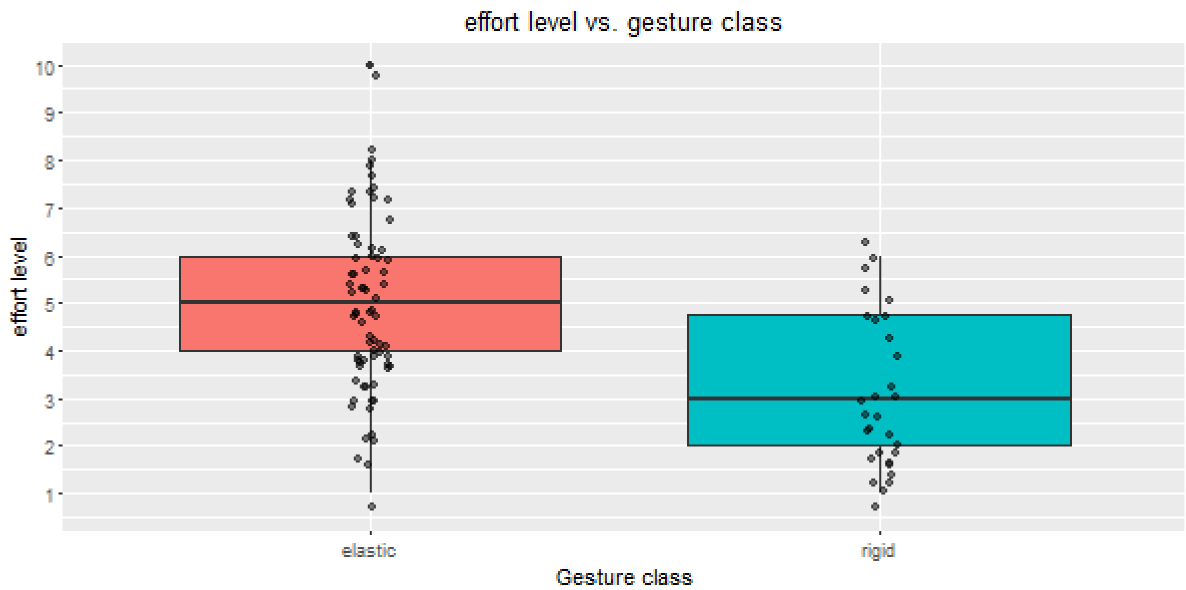


Figure Appendix.3. Lakhan Lal Sahu: Boxplot for effort levels vs. gesture classes.

The confusion matrix of Table Appendix.34 shows that the success rate of classifying gestures based simply on their perceived effort level can reach more than 78%, which is the same as with Hussain.

Lakhan Lal Sahu: confusion matrix (classification on effort)				
Logitpred/class		actual class		
		elastic	rigid	total
inferred class	elastic	66	16	82
	rigid	6	14	20
total		72	30	102
success Mean		***0.78***		

Table Appendix.34. Lakhan Lal Sahu: Confusion matrix for classification success based on effort level.

Best model selection

Model 3 is the most promising, with a good combination of number of features and classification success. Table Appendix.35 displays the confusion matrix and classification success rate for model 3.

Lakhan Lal Sahu: confusion matrix (GLM 3)				
Logitpred/class		actual class		
		elastic	rigid	total
inferred class	elastic	59	23	82
	rigid	13	7	30
total		72	30	102
success Mean		***0.65***		

Table Appendix.35. Lakhan Lal Sahu: Confusion matrix for classification success of GLM 3.

Additionally, the graph of Figure Appendix.4 illustrates the ROC curves for model 3 as well as for models that have been fit on the individual features that constitute this model.

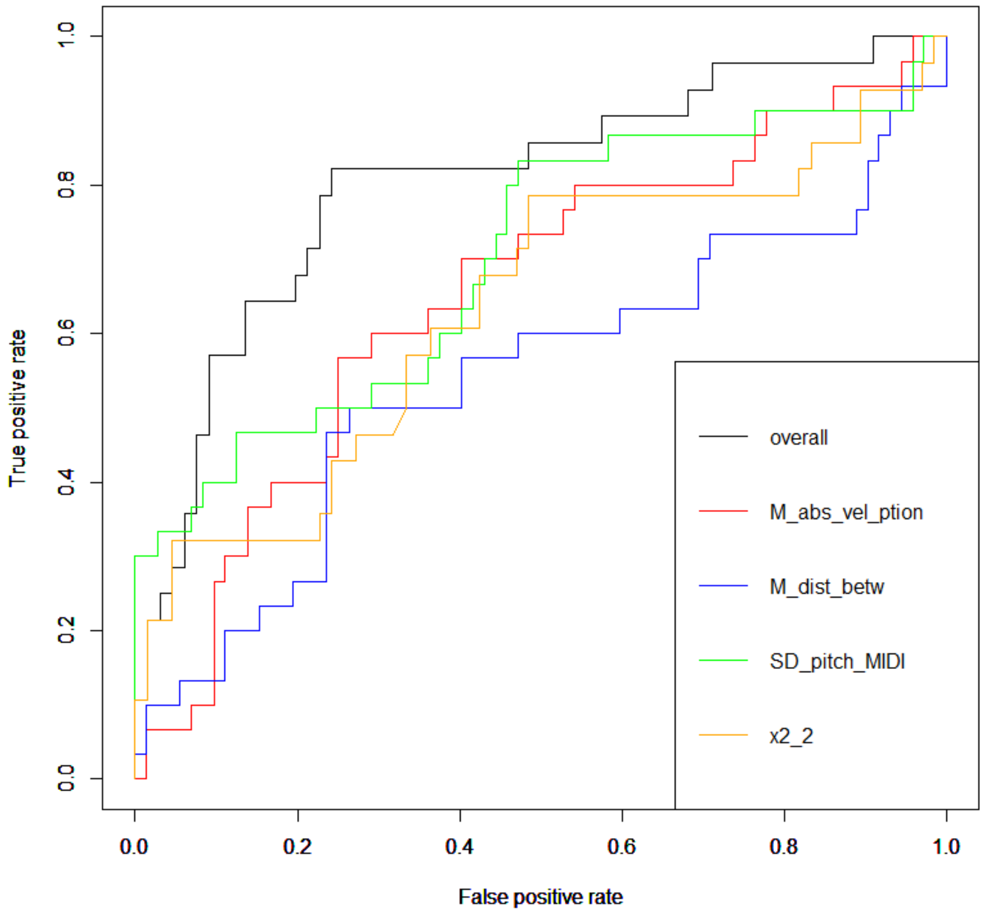


Figure Appendix.4. Lakhan Lal Sahu: ROC for GLM 3 and all individual features.

Despite the relatively high overall accuracy (AUC value), none of the individual features seems to exhibit a high sensitivity (i.e. a ROC curve that is close to the upper left corner).

C. Consent and Recording Agreement Release Forms

C1. Interview Consent Form

I,		(print name)

agree to a videotaped/recorded interview for this research project.

I have had the purposes of the research project explained to me.

I have been informed that I may refuse to participate at any point by simply saying so.

I have been assured that my confidentiality will be protected, and my name will not be published unless I have specified below that I wish it to be published.

I agree that the information that I provide can be used for educational or research purposes, including publication.

I understand that if I have any concerns or difficulties I can contact prof. Martin Clayton, Chair in the Department of Music, Durham University, Palace Green, Durham, DH1 3RL, United Kingdom (tel: 0044 191 3343151).

If I wish to complain about any aspect of my participation in this project, I can contact Panagiota-Styliani Paschalidou or prof. Martin Clayton both at the address above, or at their respective e-mail addresses: panagiota-styli.paschalidou@durham.ac.uk or stepadou@hotmail.com and martin.clayton@durham.ac.uk

I	would like to be identified by name	do not wish to be identified by name

(please delete one) in any publication of this research.

I assign the copyright for my contribution to the Open University for use in education, research and publication.

Signature	Date

C2. Recording Agreement and Release Form

I Panagiota-Styliani Paschalidou request your permission to make an audio/video/mocap recording of your music classes, performances and interviews at ___________________________ (place) on ___________________________ (date).

I ______________________________ (artist name) grant Panagiota-Styliani Paschalidou non-exclusive permission to make this videotape and to use it for the purposes of:

- Academic research
- Publication on CD or DVD for educational and not-for-profit purposes
- Publication on the Internet for educational and not-for-profit purposes
- Deposit in an archive to preserve and disseminate by all means including the internet or in any other medium not known or to be discovered (only educational purposes and not-for-profit)

I ______________________________ (artist name) am the owner of the copyright of all material recorded. I understand that all uses of my performance(s), compositions, name or likeness, as they appear on the videotape you are making, will be for non profit, educational purposes. I understand that all rights may be transferred to others without limit; are for the entire world; and shall last the entire duration of any copyright or other protection.

I ______________________________ (artist name) agree that I have received sufficient consideration and that no other royalties shall be paid to me.

Signed ....Date ....

Name (block letters) ....

Permanent address ....

....

....

I Panagiota-Styliani Paschalidou acknowledge your legal and moral rights in the recording and guarantee no use will be made of it other than those agreed to above without your written permission.

Signed ....Date ....

Name: PANAGIOTA-STYLIANI PASCHALIDOU

Address: Department of Music, Durham University,

Palace Green, Durham, DH1 3RL,

United Kingdom

Email: panagiota-styli.paschalidou@durham.ac.uk

BIBLIOGRAPHY

Alaoui, S. F., Caramiaux, B., Serrano, M., & Bevilacqua, F. (2012). Movement qualities as interaction modality. In *Proceedings of the International Conference on Designing Interactive Systems Conference* (pp. 761-769). Newcastle, UK.

Alemi, O. (2013). *Mova: Movement Analytics Platform*. Retrieved from http://www.sfu.ca/~oalemi/mova/ (last accessed December 15, 2016).

Allbeck, J., & Badler, N. (2004). Representing and parameterizing agent behaviors. In *Life-Like Characters* (pp. 19-38). Springer Berlin Heidelberg.

Anthony, G. T. (2005). Applying thematic analysis theory to practice: A researcher's experience. *Contemporary Nurse, 19,* pp. 75-87.

Autrimncpa (2012a). Glossary. Retrieved from https://autrimncpa.wordpress.com/glossary/ (last accessed December 15, 2016).

Autrimncpa (2012b). Jaunpuri. Retrieved from https://autrimncpa.wordpress.com/jaunpuri/ (last accessed December 15, 2016).

Autrimncpa (2012c). Malkauns. Retrieved from https://autrimncpa.wordpress.com/malkauns/ (last accessed December 15, 2016).

Autrimncpa (2012d). Bhupali. Retrieved from https://autrimncpa.wordpress.com/bhupali/ (last accessed December 15, 2016).

Autrimncpa (2012e). Melody. Retrieved from https://autrimncpa.wordpress.com/bhupali/ (last accessed December 15, 2016).

Bakker, S., Antle, A. N., & Van Den Hoven, E. (2009). Identifying embodied metaphors in children's sound-action mappings. In *Proceedings of the 8th International Conference on Interaction Design and Children* (pp. 140-149), Como, Italy.

Bartenieff, I., & Lewis, D. (1980). *Body movement: Coping with the environment*. New York: Gordon and Breach Science Publishers.

Battey, B. (2004). Bézier Spline Modeling of Pitch-Continuous Melodic Expression and Ornamentation. *Computer Music Journal, 28,* 4, pp. 25-39.

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.

Bernhardt, D., & Robinson, P. (2007). Detecting Affect from Non-stylised Body Motions. *Lecture Notes in Computer Science,* 4738, pp. 59-70.

Berthoz, A. (2000). *The brain's sense of movement*. Cambridge, Mass: Harvard University Press.

Bianco, T., Freour, V., Rasamimanana, N., Bevilacqua, F., & Causse, R. (2010). On Gestural Variation and Coarticulation Effects in Sound Control. *Lecture Notes in Computer Science,* 5934, pp. 134-145.

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research : an International Journal of Perception, Attention, Memory and Action, 62,* 4, pp. 237-254.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion, 19,* 8, pp. 1113-1139.

Boem, A. (2014). Sculpton: A malleable tangible interface for sound sculpting, In *Proceedings of the 2014 Joint Conference: International Computer Music Conference and Sound and Music Computing Conference*, Athens, Greece.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences*, *17,* 1193 (pp. 97-110), Amsterdam, The Netherlands.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (version 6.0.06) [computer program]. Retrieved from http://www.praat.org/ (last accessed December 15, 2016).

Bonini F., Rodà A. (2001). Expressive content analysis of musical gesture: an experiment on piano improvisation, *Workshop on Current Research Directions in Computer Music*, Barcelona.

Bor, J. (Ed.). (1999). *The raga guide: A survey of 74 Hindustani ragas*. Nimbus Records with Rotterdam Conservatory of Music.

Bowling, A. (2014). *Research methods in health*. McGraw-Hill Education (UK).

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage Publications.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3,* 2, pp. 77-101.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Bradford Books.

Bresin R., Friberg A. (2000). Rule-Based Emotional Colouring of Music Performance. In *Proceedings of the 2000 International Computer Music Conference* (pp. 364-367), San Francisco, USA.

Britten, N. (1995). Qualitative research: qualitative interviews in medical research. *Bmj*, *311*, 6999, pp. 251-253.

Broughton, M. C., & Davidson, J. W. (2014). Action and familiarity effects on self and other expert musicians' Laban effort-shape analyses of expressive bodily behaviors in instrumental music performance: a case study approach. *Frontiers in Psychology, 5.*

Brower, C. (2000). A Cognitive Theory of Musical Meaning. *Journal of Music Theory, 44,* 2, pp. 323-379.

Brunkan, M. C. (2015). The Effects of Three Singer Gestures on Acoustic and Perceptual Measures of Solo Singing. *International Journal of Music and Performing Arts, 3,* 1, pp. 35-45.

Brunkan, M. C. (2015a). Relationships of a Circular Singer Arm Gesture to Acoustical and Perceptual Measures of Singing: A Motion Capture Study. *Update: Applications of Research in Music Education, 34,* 3, pp. 56-62.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done?. *Qualitative research*, *6,* 1, pp. 97-113.

Burger, B., & Toiviainen, P. (2013). MoCap Toolbox-A Matlab toolbox for computational analysis of movement data. In *Proceedings of the 10th Sound and Music Computing Conference* (pp. 172-178), Stockholm, Sweden.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: a reply to McNeill. *Psychological Review, 96,* 1, pp. 168-174.

Buxton, W., & Myers, B. (1986). A study in two-handed input. In *ACM SIGCHI Bulletin, 17*, 4, pp. 321-326.

Cadoz, C. (1988). Instrumental gesture and musical composition. In *Proceedings of the 1988 International Computer Music Conference* (pp. 1-12), Kologne, Germany.

Cadoz, C. (1994). *Les réalités virtuelles: Un exposé pour comprendre, un essai pour réfléchir*. Paris: Flammarion.

Cadoz, C., & Wanderley, M. M. (2000). Gesture and Music. In: Wanderley, M., Battier, M. (eds.), *Trends in Gestural Control of Music (CDROM)*. Ircam/Centre Pompidou, Paris.

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends in Cognitive Sciences, 2,* 7, pp. 247-253.

Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its application in the study of expressive timing. In *Proceedings of the 2001 International Computer Music Conference* (pp. 232-235), Havana, Cuba.

Camurri, A., & Trocca, R. (2000). Movement and gesture in intelligent interactive music systems. In: Wanderley, M., Battier, M. (eds.), *Trends in Gestural Control of Music (CDROM)*. Ircam/Centre Pompidou, Paris.

Camurri, A., De Poli, G., Leman, M., & Volpe, G. (2001). A multi-layered Conceptual Framework for Expressive Gesture Applications. In *Proceedings of the Workshop on Current Research Directions in Computer Music* (pp. 29-34), Barcelona, Spain.

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. *International Journal of Human Computer Studies, 59*, 1-2, pp. 213-225.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004a). Multimodal analysis of expressive gesture in music and dance performances. In Camurri A., Volpe G. (Eds.), *Gesture-based Communication in Human-Computer Interaction*, LNAI 2915 (pp. 20-39), Springer Verlag.

Camurri, A., Mazzarino, B., & Volpe, G. (2004b). Analysis of expressive gesture: The EyesWeb expressive gesture processing library. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science),* 2915, pp. 460-467.

Camurri, A., Volpe, G., Poli, G. D., & Leman, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia Magazine*, *12*, 1, pp. 43-53.

Camurri, A., Canepa, C., Ghisio, S., & Volpe, G. (2009). Automatic Classification of Expressive Hand Gestures on Tangible Acoustic Interfaces According to Laban's Theory of Effort. *Lecture Notes in Computer Science,* 5085, pp. 151-162.

Canazza, S., De Poli, G., Di Sanzo, G., & Vidolin, A. (1998). A model to add expressiveness to automatic musical performance. In *Proceedings of the 1998 International Computer Music Conference* (pp. 163-169), San Francisco, USA.

Canazza, S., De Poli, G., Rodà, A., & Vidolin, A. (2003). An Abstract Control Space for Communication of Sensory Expressive Intentions in Music Performance. *Journal of New Music Research, 32,* 3, pp. 281-294.

Canazza, S., De Poli, G., Drioli, C., Rodà, A., & Vidolin, A. (2004). Modeling and control of expressiveness in music performance. In *Proceedings of the IEEE, 92,* 4, pp. 686-701.

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2010a). Study on Gesture-Sound Similarity. In the 3rd Music and Gesture Conference, Montreal, Canada.

Caramiaux, B., Bevilacqua, F., & Schnell, N. (2010b). *Mimicking sound with gesture as interaction paradigm*. Tech. rep., IRCAM-Centre Pompidou.

Caramiaux, B., Bevilacqua, F., & Schnell, N. (2010c). Analysing gesture and sound similarities with a hmm-based divergence measure. In *Proceedings of the 7th Sound and Music Computing Conference*, 9, Barcelona, Spain.

Caramiaux, B., Bevilacqua, F., Schnell, N. (2009). Towards a gesture-sound cross-modal analysis. *Lecture Notes in Computer Science* (pp. 158-170). Springer Berlin Heidelberg.

Caramiaux, B., Susini, P., Bianco, T., Bevilacqua, F., Houix, O., Schnell, N., & Misdariis, N. (2011). Gestural Embodiment of Environmental Sounds: an Experimental Study. In *Proceedings of the 11th international conference on New Interfaces for Musical Expression* (pp. 144-148), Oslo, Norway.

Caramiaux, B., Wanderley, M. M., & Bevilacqua, F. (2012). Segmenting and Parsing Instrumentalists' Gestures. *Journal of New Music Research, 41,* 1, pp. 13-29.

Caramiaux, B., & Tanaka, A. (2013). Machine Learning of Musical Gestures. In *Proceedings of the 13th International Conference on New Interfaces for Musical Expression* (pp. 513-518), Seoul, Korea.

Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O., & Susini, P. (2014). The role of sound source perception in gestural sound description. *ACM Transactions on Applied Perception, 11,* 1.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., & Pelachaud, C. (2007). Virtual agent multimodal mimicry of humans. *Language Resources and Evaluation, 41,* 3-4, pp. 367-388.

Casciato, C., Jensenius, A. R., & Wanderley, M. M. (2005). Studying free dance movement to music. In *Proceedings of ESCOM 2005 Performance Matters! Conference*, Porto, Portugal.

Castagné, N., & Cadoz, C. (2005). A goals-based review of Physical modelling. In *Proceedings of the 2005 International Computer Music Conference* (pp. 343-346), Barcelona, Spain.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *Journal of Experimental Psychology. General, 113,* 3, pp. 394-412.

Castellano, G., Villalba, S. D., & Camurri, A. (2007). Recognising Human Emotions from Body Movement and Gesture Dynamics. *Lecture Notes in Computer Science,* 4738, pp. 71-82.

Castellano, G., & Mancini, M. (May, 2009). Analysis of emotional gestures for the generation of expressive copying behaviour in an embodied agent. In *Proceedings of the* 7th *International Gesture Workshop: Gesture-Based Human-Computer Interaction and Simulation* (pp. 193-198). Springer Berlin Heidelberg.

Catford, J. C. (1988). *A practical introduction to phonetics* (p. 183). Oxford: Clarendon Press.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage Publications.

Chi, D. (1999). *A motion control scheme for animating expressive arm movements* (Doctoral thesis, University of Pennsylvania, Philadelphia, USA). Retrieved from http://repository.upenn.edu/cgi/viewcontent.cgi?article=1046&context=ircs_reports (last accessed December 15, 2016).

Chi, D., Costa, M., Zhao, L., & Badler, N. (2000). The EMOTE model for effort and shape. In *Proceedings of the 27th ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques* (pp. 173-182), New Orleans, Louisiana, USA.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chordia P., & Rae, A. (2007). Raag recognition using pitch-class and pitch-class dyad distributions. In *Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 431–436), Vienna, Austria.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In Howell, P., Cross, I., & West, R. (Eds.), *Musical structure and cognition* (pp. 209-236), London: Orlando.

Clarke, E. F. (1993). Generativity, mimesis and the human body in music performance. *Music and the Cognitive Sciences 1990,* pp. 207-219.

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. Oxford: Oxford University Press.

Clayton, M. (2005). Communication in Indian rāga performance. In Miell D., Hargreaves D., & MacDonald R. (eds.), *Musical Communication* (pp.361-381*)*, Oxford: Oxford University Press.

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

Clayton, M. & Leante, L. (2013). Embodiment in Music Performance. In Clayton, M., Dueck, B., & Leante, L. (Eds.), *Experience and Meaning in Music Performance* (pp. 188-207), Oxford: Oxford University Press.

Clynes, M. (1973). Sentography: Dynamic forms of communication of emotion and qualities. *Computers in Biology and Medicine, 3,* 2, pp. 119-130.

Clynes, M., & Nettheim, N. (1982). The living quality of music. In Clynes, M. (Ed.). *Music, mind, and brain: The neuropsychology of music* (pp. 47-82). New York: Plenum Press.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. In *Educational and Psychological Measurement, 20, 1,* pp. 37-46.

Coker, W. (1972). *Music & meaning: A theoretical introduction to musical aesthetics*. New York: Free Press.

Cook, N. (2001). Theorizing musical meaning. *Music Theory Spectrum, 23,* 2, pp. 170-195.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion Recognition in Human-Computer Interaction. *IEEE Signal Processing Magazine, 18,* pp. 32-80.

Cox, A. (2006). Hearing, feeling, grasping gestures. In Gritten, A., & King, E. (Eds.). *Music and gesture* (pp. 45-60). Aldershot, England: Ashgate.

Cox, A. (2011). Embodying music: Principles of the mimetic hypothesis. *Music Theory Online*, 17(2).

Cox, A. (2016). *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking*. Indiana University Press.

Dael, N., Goudbeek, M., & Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception, 42,* 6, pp. 642-57.

Dahl, S. (2006). Movements and analysis of drumming. In Altenmüller, E., Wiesendanger, M., & Kesselring, J. (Eds.), *Music, Motor Control and the Brain* (pp.125 -138). Oxford University Press, New York.

Dahl, S., & Friberg, A. (2007). Visual Perception of Expressiveness in Musicians' Body Movements. *Music Perception, 24,* 5, pp. 433-454.

Dahl S., Bevilacqua F., Bresin R., Clayton M., Leante L., Poggy I. & Rasamimanana N. (2010). Gestures in Performance. In Leman M. & Godøy R. I. (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 36-68), New York: Routledge.

Davidson, J. (1994). Expressive movements in musical performance. In *Proceedings of the 3rd International Conference on Music Cognition* (pp. 327–329), Liege, Belgium.

Davidson, J. W. (2001). The Role of the Body in the Production and Perception of Solo Vocal Performance: A Case Study of Annie Lennox. *Musicae Scientiae, 5,* 2, pp. 235-256.

Davidson, J. W., & Correia, J. S. (2002). Body movement. In Parncutt, R., & McPherson, G. (Eds.), *The science & psychology of music performance: Creative strategies for teaching and learning* (pp. 237-250). Oxford: Oxford University Press.

Davidson, J. W. (2005). Bodily communication in musical performance. In Miell, D., MacDonald, R. A. R., & Hargreaves, D. J. (Eds.), *Musical communication* (pp. 215-237). Oxford: Oxford University Press.

De Poli, G., Piccialli, A., & Roads, C. (1991). *Representations of musical signals*. Cambridge, Mass: MIT Press.

De Poli, G. (2004). Methodologies for Expressiveness Modelling of and for Music Performance. *Journal of New Music Research, 33,* 3, pp. 189-202.

De Poli, G., Avanzini, F., Roda, A., Mion, L., D'Inca, G., Trestino, C., & Castagné, N. (2005). Towards a multi-layer architecture for multi-modal rendering of expressive actions. In *Proceedings of the 2nd International Conference on Enactive Interfaces,* 1, 7, Genova, Italy.

De Poli, G., Mion, L., & Rodà, A. (2009). Toward an Action Based Metaphor for Gestural Interaction with Musical Contents. *Journal of New Music Research, 38,* 3, pp. 295-307.

Delalande, F. (1988). *La gestique de gould: Élements pour une sémiologie du geste musical*. In G. Guertin (Ed.), Glenn Gould Pluriel (pp. 85–111). Québec: Louise Courteau.

Denscombe, R. (2003). *The Good Research Guide* (2nd edition). Edn. Maidenhead.

Desain, P., Honing H. (1991). Generalized Time Functions. In *Proceedings of the 1991 International Computer Music Conference* (pp. 348-348), San Francisco, USA.

d'Escriván, J. (2006). To sing the body electric: Instruments and effort in the performance of electronic music. *Contemporary Music Review, 25,* 1, pp. 183-191.

Dixon, S. (2000). Extraction of musical performance parameters from audio data. In *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia* (pp. 42-45), Sydney, Australia.

dos Santos, L. C. G. F. (2013). *Laban Movement Analysis: A Bayesian Computational Approach to Hierarchical Motion Analysis and Learning* (Doctoral thesis, Universidade de Coimbra, Coimbra, Portugal). Retrieved from https://estudogeral.sib.uc.pt/jspui/bitstream/10316/24291/1/Laban%20Movement%20Analysis.pdf (last accessed December 15, 2016).

Easterbrook, S., Singer, J., Storey, M. A., & Damian, D. (2008). Selecting empirical methods for software engineering research. In *Guide to advanced empirical software engineering* (pp. 285-311). Springer London.

Ebeling, K. (1973). *Ragamala painting*. Basel: Ravi Kumar.

Eitan, Z. (2007). Intensity and Cross-Dimensional Interaction in Music: Recent Research and its Implications for Performance Studies. *Orbis Musicae*, *14*, pp. 141-166.

Eitan, Z., & Granot, R. Y. (2004). Musical parameters and images of motion. In *Proceedings of the conference on interdisciplinary musicology* (pp. 15-18), Graz, Austria.

Eitan, Z., & Granot, R. Y. (2006). How Music Moves. *Music Perception, 23,* 3, pp. 221-248.

Eitan, Z., & Granot, R. Y. (2007). pp. Intensity changes and perceived similarity: Inter-parametric analogies. *Musicae Scientiae, 11,* pp. 39-75.

Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114,* 3, pp. 405-422.

Ekman, P., & Friesen, W. V. (1972). Hand Movements. *Journal of Communication, 22,* 4, pp. 353-374.

Enoka, R. M., & Stuart, D. G. (1992). Neurobiology of muscle fatigue. *Journal of Applied Physiology (bethesda, Md. : 1985), 72,* 5, pp. 1631-48.

Erickson, D., Baer, T., & Harris, K. S. (1983). The role of the strap muscles in pitch lowering. In Bless, D. M., & Abbs, J. H. (Eds.), *Vocal fold physiology: Contemporary research and clinical issue* (pp. 279-285), College-Hill Press, San Diego, CA.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience, 15,* 2, pp. 399-402.

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.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22,* 2, pp. 133-187.

Feld, S. (1981). 'Flow like a Waterfall': The Metaphors of Kaluli Musical Theory. *Yearbook for Traditional Music, 13,* pp. 22-47.

Feldman, J., Epstein, D., & Richards, W. (1992). Force Dynamics of Tempo Change in Music. *Music Perception: an Interdisciplinary Journal, 10,* 2, pp. 185-203.

Fels, S., Gadd, A., & Mulder, A. (2002). Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound, 7,* 2, pp. 109-126.

Fenza, D., Mion, L., Canazza, S., & Roda, A. (2005). Physical movement and musical gestures: a multilevel mapping strategy. In *Proceedings of the 2nd Sound and Music Computing*, *5*, Salerno, Italy.

Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13,* 3, pp. 361-377.

Françoise, J. (2015). *Motion-Sound Mapping by Demonstration* (Doctoral thesis, Université Pierre et Marie Curie, Ircam, Paris, France). Retrieved from https://tel.archives-ouvertes.fr/tel-01161965/document (last accessed December 15, 2016).

Françoise, J., Caramiaux, B., & Bevilacqua, F. (2012). A hierarchical approach for the design of gesture-to-sound mappings. In *Proceedings of the 9th Sound and Music Computing Conference* (pp. 233-240), Copenhagen, Denmark.

Franinović, K., & Serafin, S. (2013). *Sonic interaction design*. Mit Press.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America, 87,* 1, pp. 311-22.

Frego, R. J. D. (1999). Effects of Aural and Visual Conditions on Response to Perceived Artistic Tension in Music and Dance. *Journal of Research in Music Education, 47,* 1, pp. 31-43.

Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion?: A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America (print), 1999,* pp. 1469-1484.

Gabbard, C. P. (2011). *Lifelong motor development*. Pearson Higher Ed.

Gabrielsson, A. (1973). Similarity ratings and dimension analyses of auditory rhythm patterns. 1. *Scandinavian Journal of Psychology*, *14*, 1, pp. 138-160.

Gabrielsson, A. (1995). Expressive intention and performance. *Music and the Mind Machine*, pp. 35-47.

Gabrielsson, A. (1997). The performance of music. In Deutsch, D. (Eds), *The psychology of music, 2* (pp. 501-602), San Diego: Academic Press.

Gabrielsson, A., & Juslin, P. N. (2003). *Emotional expression in music*. Oxford University Press.

Gallagher, S. (2011). Interpretations of embodied cognition. In W. Tschacher and C. Bergomi (Eds.), *The Implications of Embodiment: Cognition and Communication* (pp. 59-70). United Kingdom: Imprint Academic.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain : a Journal of Neurology, 119*, pp. 593-609.

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22, pp.* 455-479.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13,* 3, pp. 361-377.

Gibbon, D. (2009). Gesture Theory is Linguistics: On Modelling Multimodality as Prosody. In *PACLIC*, pp. 9-18.

Gibet, S., Kamp, J.-F., & Poirier, F. (2004). Gesture Analysis: Invariant Laws in Movement. *Lecture Notes in Computer Science,* 2915, pp. 1-9.

Gibet, S. (2010). Sensorimotor control of sound-producing gestures. In Leman M. & Godøy R. I. (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 212-237), New York: Routledge

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.

Gibson, J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gjerdingen, R. O. (1994). Apparent motion in music?. *Music Perception: An Interdisciplinary Journal*, *11*, 4, pp. 335-370.

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.

Glowinski, D., Camurri, A., Volpe, G., Dael, N., Mortillaro, M., & Scherer, K. (2011). Toward a minimal representation of affective gestures. *Ieee Transactions on Affective Computing, 2,* 2, pp. 106-118.

Godøy, R. I. (1997). *Formalization and epistemology*. Scandinavian University Press.

Godøy, R. I. (1999). Cross-modality and conceptual shapes and spaces in music theory. *Music and Signs*, pp. 85-98.

Godøy, R. I., & Jørgensen, H. (2001). *Musical imagery*. Lisse, the Netherlands: Swets & Zeitlinger Publishers.

Godøy, R. I. (2003). Motor-Mimetic Music Cognition. *Leonardo, 36,* 4, pp. 317-319.

Godøy, R. I. (2004). Gestural Imagery in the Service of Musical Imagery. *Lecture Notes in Computer Science,* 2915, pp. 55-62.

Godøy, R. I. (2006). Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, *11*(02), pp. 149-157.

Godøy, R. I., Haga, E., & Jensenius, A. R. (2006a). Exploring music-related gestures by sound-tracing: A preliminary study. In K. Ng (Ed.), *Proceedings of the COST287-ConGAS 2nd International Symposium on Gesture Interfaces for Multimedia Systems* (pp. 27-33), Leeds, UK.

Godøy, R. I., Haga, E., & Jensenius, A. R. (2006b). Playing “air instruments”: mimicry of sound-producing gestures by novices and experts. In Gibet S., Courty N., & Kamp J.-F. (Eds.), *Gesture in*

Human-Computer Interaction and Simulation, GW 2005, Volume LNAI 3881 (pp. 256-267). Springer Berlin Heidelberg.

Godøy, R. I. (2008a). Chunking sound for musical analysis. In *International Symposium on Computer Music Modeling and Retrieval* (pp. 67-80), Springer Berlin Heidelberg.

Godøy, R. I. (2008b). Reflections on chunking in music. *Systematic and Comparative Musicology,* pp. 117-131.

Godøy, R. I., Jensenius, A. R., & Nymoen, K. (2008c). Production and perception of goal-points and coarticulations in music. *The Journal of the Acoustical Society of America, 123,* 5, pp. 3657.

Godøy, R. I. (2009). Geometry and Effort in Gestural Renderings of Musical Sound. *Lecture Notes in Computer Science,* 5085, pp. 205-215.

Godøy, R. I. (2010). Gestural affordances of musical sound. In Leman M. & Godøy R. I. (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 103-125), New York: Routledge.

Godøy, R. I., & Leman, M. (2010). *Musical gestures: Sound, movement, and meaning*. New York: Routledge.

Godøy, R. I., Jensenius, A. R., & Nymoen, K. (2010). Chunking in music by coarticulation. *Acta Acustica United with Acustica, 96,* 4, pp. 690-700.

Godøy, R.I., Jensenius, A.R., Voldsund, A., Glette, K., Høvin, M., Nymoen, K., Skogstad, S., & Tørresen, J. (2012). Classifying Music-Related Actions. In *Proceedings of the ICMPC-ESCOM 2012 Joint Conference: 12th International Conference on Music Perception and Cognition & the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 352 – 357), Thessaloniki, Greece.

Godøy, R. I. (2013). Thinking Sound and Body-Motion Shapes in Music: Public Peer Review of "Gesture and the Sonic Event in Karnatak Music" by Lara Pearson. *Empirical Musicology Review, 8*, pp. 1 – 15.

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.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, Mass: Belknap Press of Harvard University Press.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences, 9,* 5, pp. 234-41.

Goldin-Meadow, S. (2006). Talking and Thinking With Our Hands. *Current Directions in Psychological Science, 15,* 1, pp. 34-39.

Goldin-Meadow, S. (2014). How gesture works to change our minds. *Trends in Neuroscience and Education, 3,* 1, pp. 4-6.

Goldman, A. I. (2012). A Moderate Approach to Embodied Cognitive Science. *Review of Philosophy and Psychology, 3,* 1, pp. 71-88.

Graham, R., & Bridges, B. (2014). Gesture and Embodied Metaphor in Spatial Music Performance Systems Design. In *Proceedings of the 14th international conference on New Interfaces for Musical Expression* (pp. 581-584), London, UK.

Granot, R. Y., & Eitan, Z. (2011). Musical Tension and the Interaction of Dynamic Auditory Parameters. *Music Perception, 28,* 3, pp. 219-246.

Greenhalgh, T. (2014). *How to read a paper: The basics of evidence-based medicine*. John Wiley & Sons.

Gritten, A., & King, E. (2006). *Music and gesture*. Aldershot, England: Ashgate.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Los Angeles: Sage Publications.

Hachimura, K., Takashina, K., & Yoshimura, M. (2005). Analysis and evaluation of dancing movement based on LMA. In *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication* (pp. 294-299), Nashville, Tennessee, USA.

Hackney, P. (2002). *Making connections: Total body integration through Bartenieff fundamentals*. New York: Routledge.

Haga, E. (2008). *Correspondences between music and body movement*. (Doctoral thesis, University of Oslo, Oslo, Norway). Retrieved from https://www.duo.uio.no/handle/10852/26916 (last accessed December 15, 2016).

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143,* 1, pp. 29-36.

Hartmann, B., Mancini, M., Buisine, S., & Pelachaud, C. (2005). Design and evaluation of expressive gesture synthesis for embodied conversational agents. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multi Agent Systems* (pp. 1095-1096), Utrecht, The Netherlands.

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert*. Bloomington: Indiana University Press.

Hatten, R. S. (2015). Melodic forces and agential energies: An integrative approach to the analysis and expressive interpretation of tonal melodies. *Music, Analysis, Experience,* pp. 315-330.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Henbing, L., & Leman, M. (2007). A Gesture-based Typology of Sliding-tones in Guqin Music. *Journal of New Music Research, 36,* 2, pp. 61-82.

Hermann, T., & Hunt, A. (2005). An introduction to interactive sonification. *IEEE multimedia*, pp. 20-24.

Hogan, N., & Sternad, D. (2007). On rhythmic and discrete movements: reflections, definitions and implications for motor control. *Experimental Brain Research, 181,* 1, pp. 13-30.

Honing, H. (2003). The Final Ritard: On Music, Motion, and Kinematic Models. *Computer Music Journal, 27,* 3, pp. 66-72.

Hsieh, C. M., & Luciani, A. (2005). Physically-based particle modeling for dance verbs. In *Proceedings of the 15th International Conference on Computer Graphics and Applications* (pp. 85-92), Novosibirsk, Russia.

Hunt A., Kirk R. (2000). Mapping strategies for musical performance. In: Wanderley, M., Battier, M. (eds.), *Trends in Gestural Control of* Music (CDROM). Ircam/Centre Pompidou, Paris.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, Mass: MIT Press.

Hutchins, E. (2010). Cognitive ecology. *Topics in cognitive science*, *2*(4), pp. 705-715.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science (new York, N.y.), 286,* 5449, pp. 2526-2528.

Imberty, M. (1981). *Les écritures du temps: Sémantique psychologique de la musique (tome 2). Paris: Dunod*.

IPA (2012). Full IPA Chart. Retrieved from https://www.internationalphoneticassociation.org/content/full-ipa-chart#ipachartkiel (last accessed December 15, 2016).

Jackson, D. L. (2007). The effect of the number of observations per parameter in misspecified confirmatory factor analytic models. *Structural Equation Modeling*, *14, 1*, pp. 48-76.

Jacobson, E. (1951). Muscular tension and the estimation of effort. *The American Journal of Psychology, 64,* 1, pp. 112-24.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 6). New York: Springer.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage, 14,* 1, pp. 103-109.

Jennifer, F., & Eimear, M.-C. (2006). Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods, 5,* 1, pp. 80-92.

Jensenius, A. R. (2007). *Action-sound: Developing methods and tools to study music-related body movement* (Doctoral thesis, University of Oslo, Oslo, Norway). Retrieved from https://www.duo.uio.no/handle/10852/27149 (last accessed December 15, 2016).

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical Gestures: Concepts and Methods in Research. In Leman M. & Godøy R. I. (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 13-35), New York: Routledge.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, *14*, 2, pp. 201-211.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Johnson, M. L., & Larson, S. (2003). "Something in the Way She Moves"-Metaphors of Musical Motion. *Metaphor and Symbol, 18,* 2, pp. 63-84.

Juslin, P. N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in music performance: The GERM model. *Current Trends in the Study of Music and Emotion,* pp. 63-122.

Juslin, P. N. (2003). Five facets of musical expression: A psychologist's perspective on music performance. *Psychology of Music*, *31,* 3, pp. 273-302.

Kapadia, M., Chiang, I. K., Thomas, T., Badler, N. I., & Kider Jr, J. T. (2013). Efficient motion retrieval in large motion databases. In *Proceedings of the 17th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (pp. 19-28), Orlando, FL, USA.

Karwoski, T. F., Odbert, H. S., & Osgood, C. E. (1942). Studies in synesthetic thinking: II. The role of form in visual responses to music. *The Journal of general psychology*, *26, 2*, pp. 199-222.

Kelso, J. A. S., Fuchs, A., Lancaster, R., Holroyd, T., Cheyne, D., & Weinberg, H. (1998). Dynamic cortical activity in the human brain reveals motor equivalence. *Nature, 392,* 6678, pp. 814-818.

Kendon, A. (1996). An agenda for gesture studies. *Semiotic review of books*, *7*(3), pp. 8-12.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kikhia, B., Gomez, M., Jiménez, L. L., Hallberg, J., Karvonen, N., & Synnes, K. (2014). Analyzing body movements within the Laban Effort Framework using a single accelerometer. *Sensors (basel, Switzerland), 14,* 3, pp. 5725-5741.

Kipp, M. (2010). Multimedia annotation, querying and analysis in ANVIL. *Multimedia information extraction*, *19*, pp. 351-367.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly: Management Information Systems, 23*(1), pp. 67-94.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kochanek, D. H. U., & Bartels, R. H. (1984). Interpolating splines with local tension, continuity, and bias control. *Acm Siggraph Computer Graphics, 18,* 3, pp. 33-41.

Kohler, E., & Kaysers, C. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297,* 5582, pp. 846-848.

Kövecses, Z. (2002). Emotion concepts: Social constructionism and cognitive linguistics. *The Verbal Communication of Emotions. Interdisciplinary Perspectives. Mahwah, New Jersey: Lawrence Erlbaum Associates*, pp. 109-124.

Krefeld, V., & Waisvisz, M. (1990). The Hand in the Web: An Interview with Michel Waisvisz. *Computer Music Journal, 14,* 2, pp. 28-33.

Kronman, U., & Sundberg, J. (1987). Is the musical ritard an allusion to physical motion. *Action and perception in rhythm and music*, *55*, pp. 57-68.

Krueger, J. (2014). Affordances and the musically extended mind. *Frontiers in Psychology, 4.*

Krumhansl, C. L., & Schenck, D. L. (1997). Can Dance Reflect the Structural and Expressive Qualities of Music? A Perceptual Experiment on Balanchine's Choreography of Mozart's Divertimento No. 15. *Musicae Scientiae, 1,* 1, pp. 63-85.

Kurtenbach, G. & Hulteen, E. (1990). Gestures in Human-Computer Communications. In Laurel B. (Ed.), *The Art of Human Computer Interface Design* (pp. 309-317). Addison-Wesley.

Kurth, E. (1922). *Grundlagen des linearen kontrapunkts: Bachs melodische polyphonie*. Berlin: M. Hesse.

Kussner, M. (2014). *Shape, drawing and gesture: Cross-modal mappings of sound and music* (Doctoral thesis, King's College London (University of London), London, UK). Retrieved from https://kclpure.kcl.ac.uk/portal/en/theses/shape-drawing-and-gesture(2b4cf820-956f-4da9-a29e-1053870d656b).html (last accessed December 15, 2016).

Küssner, M. B., & Leech-Wilkinson, D. (2014). Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing paradigm. *Psychology of Music, 42,* 3, pp. 448-469.

Küssner, M. B., Tidhar, D., Prior, H. M., & Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in Psychology, 5.*

Laban, R., & Lawrence, F. C. (1974). *Effort: Economy of body movement*. Boston: Plays, Inc.

Laban, R., & Ullmann, L. (1971). *The mastery of movement*. Boston: Plays, Inc.

Labuschagne, A. (2003). Qualitative research-airy fairy or fundamental?. *The qualitative report*, *8*, 1, pp. 100-103.

Lagerlöf, I., & Djerf, M. (2001). On cue utilization for emotion expression in dance movements. *Manuscript in preparation, Department of Psychology, University of Uppsala*.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience, 27,* 2, pp. 308-314.

Laios S. (2008). *Rāga Mālakaunśa*. Retrieved from http://www.musicking.gr/articles/malkauns/?lang=en (last accessed December 15, 2016).

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4,* 2, pp. 195-208.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago ; London: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh : The embodied mind and its challenge to Western thought*. New York: Basic Books.

Lapadat, J. C., & Lindsay, A. C. (1999). Transcription in Research and Practice: From Standardization of Technique to Interpretive Positionings. *Qualitative Inquiry, 5,* 1, pp. 64-86.

Larson, S. (1997). Musical forces and melodic patterns. *Theory and practice*, *22*, pp. 55-71.

Larson, S., & Vanhandel, L. (2005). Measuring Musical Forces. *Music Perception, 23,* 2, pp. 119-136.

Larson, S. (2006). Musical gestures and musical forces: Evidence from music-theoretical misunderstandings. *Music and gesture*, pp. 61-74.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. In *Data analysis, machine learning and applications* (pp. 261-268). Springer Berlin Heidelberg.

Leante, L. (2009). The Lotus and the King: Imagery, Gesture and Meaning in a Hindustani Rāg. *Ethnomusicology Forum, 18,* 2, pp. 185-206.

Leman, M. (1997). *Music, gestalt, and computing: Studies in cognitive and systematic musicology*. Berlin: Springer.

Leman, M. (2008). *Embodied music cognition and mediation technology*. Cambridge, Mass. ; London: MIT.

Leman, M. (2010). Music, Gesture, and the Formation of Embodied Meaning. In Leman M. & Godøy R. I. (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 126-153), New York: Routledge.

Leman, M., & Godøy, R. I. (2010). Why study musical gestures. In Leman M. & Godøy R. I. (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 3-11), New York: Routledge.

Leman, M. (2012). Musical gestures and embodied cognition. In Dutoit T., Todoroff T., & d'Alessandro N. (Eds.), *Journées d'informatique musicale* (pp. 5-7). Université de Mons/numediart.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception, 24,* 4, pp. 329-366.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental psychology*, *16,* 6, pp. 597-607.

Liao, M. Y., & Davidson, J. W. (2007). The use of gesture techniques in children's singing. *International Journal of Music Education*, *25*(1), pp. 82-94.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1, pp. 1-36.

Luciani, A. (2004). Dynamics as common criterion to enhance the sense of presence in virtual environments. In *Proceedings of the 7th Annual International Workshop Presence 2004* (pp. 96-103), New Orleans, Louisiana, USA.

Luciani, A. (2007). Ergotic/epistemic/semiotic functions. *Enaction and enactive interfaces: a handbook of terms*, pp. 96-97.

Luciani, A., Florens, J. L., Couroussé, D., & Castet, J. (2009). Ergotic Sounds: A New Way to Improve Playability, Believability and Presence of Virtual Musical Instruments. *Journal of New Music Research, 38,* 3, pp. 309-323.

Luck, G., & Toiviainen, P. (2008). Exploring relationships between the kinematics of a singer's body movement and the quality of their voice. *Journal of interdisciplinary music studies*, *2*(1-2), pp. 173-186.

Maestre, E., Bonada, J., & Mayor, O. (2006). Modeling musical articulation gestures in singing voice performances. In Proceedings of the *Audio Engineering Society - 121st Convention Papers 2006, 1,* pp. 471-479, San Francisco, California.

Maletic, V. (1987). *Body, space, expression: The development of Rudolf Laban's movement and dance concepts (Vol. 75)*. Berlin: Mouton de Gruyter.

Mancini, M., & Castellano, G. (2007). Real-time analysis and synthesis of emotional gesture expressivity. In *Proceedings of the Doctoral Consortium of the 2nd International Conference on Affective Computing and Intelligent Interaction*, Lisbon, Portugal.

Marcora, S. (2009). Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs. *Journal of Applied Physiology (bethesda, Md. : 1985), 106,* 6, pp. 2060-2062.

Margulis, E. H. (2005). A Model of Melodic Expectation. *Music Perception: an Interdisciplinary Journal, 22,* 4, pp. 663-714.

Martin, T., & Schwartz, D. L. (2005). Physically Distributed Learning: Adapting and Reinterpreting Physical Environments in the Development of Fraction Concepts. *Cognitive Science, 29,* 4, pp. 587-625.

Maurer, D. (1993). Neonatal synesthesia: Implications for the processing of speech and faces. *Nato Asi Series D Behavioural and Social Sciences, 69,* pp. 109-124.

Mazzarino, B., Peinado, M., Boulic, R., Volpe, G., & Wanderley, M. M. (2009). Improving the Believability of Virtual Characters Using Qualitative Gesture Analysis. *Lecture Notes in Computer Science,* 5085, pp. 48-56.

McAdams, S. (1984). *Spectral fusion, spectral parsing and the formation of auditory images* (Doctoral thesis, Stanford University, Stanford, USA). Retrieved from https://ccrma.stanford.edu/files/papers/stanm22.pdf (last accessed December 15, 2016).

Mcneill, D., & Levy, E. (1980). *Conceptual Representations in Language Activity and Gesture*.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D., & Duncan, S. D. (1998). *Growth points in thinking-for-speaking*. Southampton: Cogprints.org.

McNeill, D. (2000). *Language and gesture*. Cambridge: Cambridge University Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Menin, D., & Schiavio, A. (2012). *Rethinking musical affordances*. *Avant*, *3*, 2, pp. 202-215.

Mentis, H. M., & Johansson, C. (2013). Seeing movement qualities. In *Proceedings of the 31st SIGCHI Conference on Human Factors in Computing Systems* (pp. 3375-3384), Paris, France.

Merleau-Ponty, M. (1945/1962). *Phenomenology of perception* (International library of philosophy and scientific method). London: Routledge & Kegan Paul.

Merleau-Ponty, M., Lefort, Claude, & Lingis, Alphonso. (1968). *The visible and the invisible: Followed by working notes* (Northwestern University studies in phenomenology & existential philosophy). Evanston: Northwestern University Press.

Métois, E. (1997). *Musical sound information: Musical gestures and embedding synthesis* (Doctoral thesis, Massachusetts Institute of Technology, Boston, USA). Retrieved from https://dspace.mit.edu/handle/1721.1/29125 (last accessed December 15, 2016).

Mikumo, M. (1998). Encoding strategies for pitch information. *Japanese psychological monographs*, *27*, pp. 41-48.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage Publications.

Miner, A. (1997). *Sitar and sarod in the 18th and 19th centuries* (1st ed., Performing arts series; vol. 7). Delhi: Motilal Banarsidass.

Mion L. (2003). Application of Bayesian Networks to automatic recognition of expressive content of piano improvisations. In *Proceedings of the SMAC03 Music Acoustics Conference* (pp. 557-560), Stockholm, Sweden.

Mion, L., & De Poli, G. (2007). Music Expression Understanding Based on a Joint Semantic Space. *Lecture Notes in Computer Science,* 4733, pp. 614-625.

Mion, L., & De, P. G. (2008). Score-independent audio features for description of music expression. *Ieee Transactions on Audio, Speech and Language Processing, 16,* 2, pp. 458-466.

Mion, L., de Poli, G. D., & Rapanà, E. (2010). Perceptual organization of affective and sensorial expressive intentions in music performance. *ACM Transactions on Applied Perception*, *7,* 2, pp. 1-21.

Miranda, E. R., & Wanderley, M. (2006). *New digital musical instruments: Control and interaction beyond the keyboard*. Middleton: A-R Editions.

Moore, C.-L., & Yamamoto, K. (1988). *Beyond words: Movement observation and analysis*. New York: Gordon and Breach.

Moore, D. S., & McCabe, G. P. (1989). *Introduction to the Practice of Statistics* (5th ed.). New York: W.H.Freeman and Company.

Moran, N. (2007). *Measuring Musical Interaction: analysing communication in embodied musical behaviour*. (Doctoral thesis, Open University, Milton Keynes, UK). Retrieved from http://oro.open.ac.uk/39118/ (last accessed December 15, 2016).

Morita, J., Nagai, Y., & Moritsu, T. (2013). Relations between body motion and emotion: analysis based on Laban Movement Analysis. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1026-1031), Berlin, Germany.

Mulder, A. (1998). *Design of Virtual Three-dimensional Instruments for Sound Control* (Doctoral thesis, Rijks Universiteit Groningen, Groningen, The Netherlands). Retrieved from http://www.xspasm.com/x/sfu/vmi/AM98-thesis.pdf (last accessed December 15, 2016).

Müller, M., & Röder, T. (2006). Motion templates for automatic classification and retrieval of motion capture data. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation* (pp. 137-146), Vienna, Austria.

Nakata, T., Mori, T., & Sato, T. (2002). Analysis of impression of robot bodily expression. *Journal of Robotics and Mechatronics*, *14, 1,* pp. 27-36.

Naveda, L., & Leman, M. (2010). The Spatiotemporal Representation of Dance and Music Gestures using Topological Gesture Analysis (TGA). *Music Perception, 28,* 1, pp. 93-111.

Neff, M., & Fiume, E. (2002). Modeling tension and relaxation for computer animation. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation* (pp. 81-88), San Antonio, TX, USA.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological cybernetics*, *46,* 2, pp. 135-147.

Neuhoff, J. G., & McBeath, M. K. (1996). The Doppler illusion: the influence of dynamic intensity change on perceived pitch. *Journal of Experimental Psychology. Human Perception and Performance, 22,* 4, pp. 970-985.

Neuman, D. (2004). *A House of Music: The Hindustani musician and the crafting of traditions* (Doctoral thesis, Columbia University, NYC, USA).

Niewiadomski, R., Obaid, M., Bevacqua, E., Looser, J., Anh, L. Q., & Pelachaud, C. (2011). Cross-media agent platform. In *Proceedings of the 16th international conference on 3D web technology* (pp. 11-19), Paris, France.

Niewiadomski, R., Mancini, M., & Piana S. (2013). Human and virtual agent expressive gesture quality analysis and synthesis. In Rojc M. and Campbell, N. (Eds.), *Coverbal Synchrony in Human-Machine Interaction* (pp. 269-292). CRC Press.

Nister, D. (2004). Automatic passive recovery of 3D from images and video. In *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission* (pp. 438-445), Thessaloniki, Greece.

Noë, A. (2004). *Action in perception* (Representation and mind). Massachusetts ; London: MIT Press.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Novack, M., & Goldin-Meadow, S. (2015). Learning from Gesture: How Our Hands Change Our Minds. *Educational Psychology Review, 27,* 3, pp. 405-412.

Noyce, G. L., Küssner, M. B., & Sollich, P. (2013). Quantifying shapes: Mathematical techniques for analysing visual representations of sound and music. *Empirical Musicology Review*, *8*(2), pp. 128-154.

Nussbaum, C. O. (2007). *The musical representation: Meaning, ontology, and emotion*. Cambridge, Mass: MIT Press.

Nusseck, M., & Wanderley, M. M. (2009). Music and Motion-How Music-Related Ancillary Body Movements Contribute to the Experience of Music. *Music Perception, 26,* 4, pp. 335-353.

Nymoen, K., Caramiaux, B., Kozak, M., & Torresen, J. (2011). Analyzing sound tracings: a multimodal approach to music information retrieval. In *Proceedings of the 1st International ACM workshop on Music information retrieval with user-centered and multimodal strategies* (pp. 39-44), Scottsdale, AZ, USA.

Nymoen, K., Jensenius, A. R., Tørresen, J., Glette, K. H., & Skogstad, S. A. V. D. (2010). Searching for cross-individual relationships between sound and movement features using an SVM classifier. In *Proceedings of the 10th International Conference on New Interfaces for Musical Expression* (pp. 259-262), Sydney, Australia.

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.

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.

Ohala, J. J. (1972). How is pitch lowered?. *The Journal of the Acoustical Society of America*, *52,* 1A, pp. 124-124.

O'Regan, J. K., & Noë, A. (2001). Acting out our sensory experience. *Behavioral and Brain Sciences, 24,* 5, pp. 1011-1021.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology. Human Perception and Performance, 15,* 2, pp. 331-346.

Parreren, C. F. ., Carpay, J. A. M., & Schouten-van, P. C. (1972). *Sovjetpsychologen aan het woord*. Groningen: Wolters-Noordhoff.

Pearson, L. (2013). Gesture and the Sonic Event in Karnatak Music. *Empirical Musicology Review, 8,* 1, 2.

Pearson, L. (2016). *Gesture in Karnatak Music: Pedagogy and Musical Structure in South India* (Doctoral thesis, Durham University, Durham, UK). Retrieved from http://etheses.dur.ac.uk/11782/ (last accessed December 15, 2016).

Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification) in the CUIDADO project*, CUIDADO I.S.T. Tech. Rep, IRCAM, Paris, France.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America, 130,* 5, pp. 2902-2916.

Petersen, D. (2008). Space, Time, Weight, and Flow: suggestions for enhancing assessment of creative movement†. *Physical Education & Sport Pedagogy, 13,* 2, pp. 191-198.

Pfordresher, P. Q., Demorest, S. M., Dalla, B. S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015a). Theoretical perspectives on singing accuracy: An introduction to the special issue on singing accuracy (part 1). *Music Perception, 32, 3,* pp. 227-231.

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015b). A Mechanism for Sensorimotor Translation in Singing. *Music Perception: an Interdisciplinary Journal, 32, 3,* pp. 242-253.

Piana, S., Mancini, M., Camurri, A., Varni, G., & Volpe, G. (2013). Automated analysis of non-verbal expressive gesture. In *Proceedings of the 6th International Workshop on Human Aspects in Ambient Intelligence* (pp. 41-54), Pisa, Italy.

Picard, R. (1997). *Affective computing* (Vol. 252). Cambridge: MIT press.

Pierce, A., & Pierce, R. (1989). *Expressive movement: Posture and action in daily life, sports, and the performing arts*. New York: Plenum Press.

Powers, H. S., & Widdess, R. (2001). India, sub-continent of, III. 2 Rāga. *The New Grove Dictionary of Music,(London: Macmillan)*.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K. E., Ansari, R. (2002). Multimodal Human Discourse: Gesture and Speech. *Acm Transactions on Computer Human Interaction, 9*, pp. 171-193.

Rahaim, M. (2009). *Gesture, Melody, and the Paramparic 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).

Rasamimanana, N. H., Flety, E., & Bevilacqua, F. (2006). Gesture Analysis of Violin Bow Strokes. *Lecture Notes in Computer Science,* 3881, pp. 145-155.

Rasch, R., & Plomp, R. (1999). The perception of musical tones. *The psychology of music*, *2*, pp. 89-112.

RepoVizz (2015). *RepoVizz: Multimodal online database and visualization tool*. Retrieved from http://repovizz.upf.edu/repo/Home (last accessed December 15, 2016).

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *The Journal of the Acoustical Society of America, 88,* 2, pp. 622-641.

Repp, B. H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's "Träumerei". *The Journal of the Acoustical Society of America, 92,* 5, pp. 2546-2568.

Repp, B. H. (1993). Musical motion: Some historical and contemporary perspectives. In *Proceedings of the Stockholm Music Acoustics Conference* (pp. 128-135). Stockholm, Sweden.

Repp, B. H. (1995). Expressive timing in Schumann's "Träumerei:" an analysis of performances by graduate student pianists. *The Journal of the Acoustical Society of America, 98,* 5, pp. 2413-2427.

Reybrouck, M. (2005). A biosemiotic and ecological approach to music cognition: event perception between auditory listening and cognitive economy. *Axiomathes*, *15,* 2, pp. 229-266.

Reybrouck, M. (2012). Musical sense-making and the concept of affordance: an ecosemiotic and experiential approach. *Biosemiotics*, *5,* 3, pp. 391-409.

Rieser, J. J. (1994). Imagery, Action, and Young Children's Spatial Orientation: It's Not Being There That Counts, It's What One Has in Mind. *Child Development, 65,* 5, pp. 1262-1278.

Rojc, M., & Campbell, N. (Eds.). (2013). *Coverbal Synchrony in Human-Machine Interaction*. CRC Press.

Rosenbaum, D. A. (2009). *Human motor control*. Academic press.

Rowell, L. E. (1992). *Music and musical thought in early India*. Chicago: University of Chicago Press.

Sanyal, W., & Widdess, Richard. (2003). *Dhrupad : Tradition and performance in Indian music* (SOAS musicology series). Aldershot: Ashgate.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology, 210*, pp. 419-421.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21,* 5, pp. 188-194.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: How our minds share actions and emotions*. Oxford: Oxford University Press.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*. Oxford, UK: Blackwell Publishers.

Rocchesso, D., Lemaitre, G., Susini, P., Ternström, S., & Boussard, P. (2015). Sketching sound with voice and gesture. *Interactions, 22,* 1, pp. 38-41.

Rodger, M., Issartel, J., & O'Modhrain, S. (2007). Performer as perceiver: perceiver as performer. In *Proceedings of the 4th International Conference on Enactive Interfaces* (pp. 237-240), Grenoble, France.

Rolfe, G. (2006). Validity, trustworthiness and rigour: quality and the idea of qualitative research. *Journal of Advanced Nursing, 53,* 3, pp. 304-310.

Roubeau, B., Chevrie-Muller, C., & Guily, J. L. S. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Oto-Laryngologica, 117,* 3, pp. 459-464.

Roychaudhuri, B. (2000). *The dictionary of Hindustani classical music* (Vol. 8). Delhi: Motilal Banarsidass Publ.

Ruckert, G. (2004). *Music in North India: Experiencing music, expressing culture*. New York: Oxford University Press.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112,* 4, pp. 585-615.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology,* 39, pp. 1161–1178.

Ryan, J. (1991). Some remarks on musical instrument design at STEIM€. *Contemporary Music Review, 6,* 1, pp. 3-17.

Samadani, A. A., Burton, S., Gorbet, R., & Kulic, D. (2013). Laban effort and shape analysis of affective hand and arm movements. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 343-348). IEEE.

Sanyal, R., & Widdess, R. (2004). *Dhrupad: Tradition and performance in Indian music*. Aldershot, Hants, England: Ashgate.

Saslaw, J. (1996). Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music. *Journal of Music Theory, 40,* 2, pp. 217-243.

Schaeffer, P. (1952). *A la recherche d'une musique concrète*. Paris: Éditions du Seuil.

Schaeffer, P. (1966). *Traité des objets musicaux*. Paris: Editions du Seuil.

Scherer, K. R. (1994). Affect bursts. In Goozen, S. H. M., Poll, N. E., & Sergeant, J. A. (Eds.), *Emotions: Essays on emotion theory* (pp. 161-196). Hillsdale, N.J: L. Erlbaum.

Schiavio, A. (2014). *Music in (en)action. Sense-making and neurophenomenology of musical experience* (Doctoral thesis, University of Sheffield, Sheffield, UK). Retrieved from http://etheses.whiterose.ac.uk/6313/ (last accessed December 15, 2016).

Schiel P., Heloir, A. (2008). *Expressive Gesture Synthesis for ECA and the EMOTE model for Effort and Shape*, Seminar in „Character Animation 2008" Michael Kipp / Alexis Heloir Universität des Saarlandes, SS 2008. Retrieved from http://embots.dfki.de/doc/seminar_ca/Schiel_EMOTE.pdf (last accessed December 15, 2016).

Schiphorst, T. (2009). Soft(n): Toward a somaesthetics of touch. In *Proceedings of the Conference on Human Factors in Computing Systems*, pp. 2427-2438, Boston, MA, USA.

Schmidt, M. (2012). *Prinzipien des Improvisierens in der nordindischen Kunstmusik* (Doctoral thesis, Freie Universität Berlin, Berlin, Germany). Retrieved from http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000037369 (last accessed December 15, 2016).

Schwarz, D., & Caramiaux, B. (2013). Interactive Sound Texture Synthesis through Semi-Automatic User Annotations. In *Proceedings of the 10th International Symposium on Computer Music Modeling and Retrieval* (pp. 372-392), Marseille, France.

Scruton, R. (1997). *The aesthetics of music*. Oxford: Clarendon Press.

Scully, D. M. (1986). Visual perception of technical execution and aesthetic quality in biological motion. *Human Movement Science, 5,* 2, pp. 185-206.

Seaman, C. B. (2008). Qualitative methods. In *Guide to advanced empirical software engineering* (pp. 35-62). Springer London.

Serrà J., Koduri G. K., Miron M., Serrà X. (2011). Assessing the tuning of sung Indian classical music. In *Proceedings of the 12th International Society for Music Information Retrieval Conference* (pp. 157-162), Miami, USA.

Serra, X. (1997). Musical sound modeling with sinusoids plus noise. In Roads, C., Pope, S. T., Piccialli, A., & Poli, G. D. (Eds.), *Musical signal processing* (pp. 91-122). Hoboken: Taylor and Francis.

Shapiro, L. (2011). *Embodied cognition* (New problems of philosophy). London: Routledge.

Sheets-Johnstone, M. (1966). *The phenomenology of dance*. Madison: University of Wisconsin Press.

Shove, P., & Repp, B. H. (1995). Musical motion and performance: theoretical and empirical perspectives. *The practice of performance: Studies in musical interpretation*, pp. 55-83.

Slawek, S. (1987). *Sitār technique in nibaddh forms* (Vol. 6). Delhi: Motilal Banarsidass Publ.

Snyder, B. (2000). *Music and memory: An introduction*. Cambridge, Mass: MIT Press.

Stern, D. N., Hofer, L., Haft, W., & Dore, J. (1985). Affect attunement: The sharing of feeling states between mother and infant by means of inter-modal fluency. *Social perception in infants*, pp. 249-268.

Stern, D. (2010). *Forms of vitality : Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford ; New York: Oxford University Press.

Stevens, S. S. (1975). Laws That Govern Behavior. (Book Reviews: Psychophysics. Introduction to Its Perceptual, Neural, and Social Prospects). *Science*, *188*, pp. 827-829.

Sundberg, J., & Friberg, A. (1996). Stopping running and stopping a piece of music: Comparing locomotion and music performance. *Proceedings / Editors: Klaus Riederer, Tapio Lahti,* pp. 351-358.

Sundberg, J., & Verrillo, V. (1980). On the anatomy of the retard: A study of timing in music. *The Journal of the Acoustical Society of America*, *68,* 3, pp. 772-779.

Sundberg, J., Leanderson, R., & von, E. C. (1989). Activity relationship between diaphragm and cricothyroid muscles. *Journal of Voice, 3, 3,* pp. 225-232.

Sundberg, J., Frydén, L., & Friberg, A. (1995). Expressive Aspects of Instrumental and Vocal Perfomance. In Steinberg R. (Ed.), *Music and the Mind Machine. The Psychophysiology and Psychopathology of the Sense of Music* (pp. 49-62). Springer Berlin Heidelberg.

Suter, W. N. (2011). *Introduction to educational research: A critical thinking approach*. Sage Publications.

Suzuki, K., & Hahimoto, S. (2004). Robotic Interface for Embodied Interaction via Dance and Musical Performance. In *Proceedings of the IEEE, 92,* 4, pp. 656-671.

Tagg P. (1999). *Introductory notes to the semiotics of music*. Retrieved from http://tagg.org/xpdfs/semiotug.pdf (last accessed December 15, 2016).

Tanaka, A., & Knapp, R. B. (2002). Multimodal interaction in music using the electromyogram and relative position sensing. In *Proceedings of the 2002 conference on New Interfaces for Musical Expression* (pp. 1-6), Dublin, Ireland.

Tanaka, A., Altavilla, A., & Spowage, N. (2012). Gestural Musical Affordances. *In Proceedings of the 9th Sound and Music Computing Conference* (pp. 318-325), Copenhagen, Denmark.

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

Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the Cognitive Sciences, 4,* 4, pp. 407-427.

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Perception: An Interdisciplinary Journal*, *3,* 1, pp. 33-57.

Todd, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, *91,* 6, pp. 3540-3550.

Todd, N. P. M. (1995). The kinematics of musical expression. *The Journal of the Acoustical Society of America*, *97, 3*, pp. 1940-1949.

Tubul, Z. E. N. (2010). Musical Parameters and Children's Images of motion. *Musicae Scientiae, 14,* pp. 89-111.

Van der Meer W. (2001/1999). Theory and Practice of Intonation in Hindustani Music. In Barlow, C. (Ed.), *The Ratio Book*, Köln: Feedback Papers, *43,* pp. 50-71.

Van der Meer, W., & Rao, S. (2006). What you hear isn't what you see: the representation and cognition of fast movements in Hindustani music. In *Proceedings of the International Symposium Frontiers of Research on Speech and Music* (pp. 12-20), Baripada, India.

Van Noorden, L., & Moelants, D. (1999). Resonance in the Perception of Musical Pulse. *Journal of New Music Research, 28,* 1, pp. 43-66.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind : Cognitive science and human experience*. Cambridge, Mass.: MIT Press.

Varela, F., Thompson, E., & Rosch, E. (1993). *The embodied mind : Cognitive science and human experience* (New ed.). Cambridge, Mass. ; London: M.I.T. Press.

Vertegaal, R., Kieslinger, M., & Ungvary, T. (1996). Towards a musician's cockpit: Transducers, feedback and musical function. *Quarterly Progress and Status Report / Speech, Music and Hearing*, pp. 29-32.

Vines, B. W., Wanderley, M. M., Krumhansl, C. L., Nuzzo, R. L., & Levitin, D. J. (2004). Performance Gestures of Musicians: What Structural and Emotional Information Do They Convey?. *Lecture Notes in Computer Science,* 2915, pp. 468-478.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101,* 1, pp. 80-113.

Visi, F., Miranda, E., Schramm, R., (2014). Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies. *In Proceedings of the 1st International Workshop on Movement and Computing,* pp. 100-105.

Vogt, S., & Thomaschke, R. (2007). From visuo-motor interactions to imitation learning: behavioural and brain imaging studies. *Journal of Sports Sciences, 25,* 5, pp. 497-517.

Volpe, G. (2003). Computational models of expressive gesture in multimedia systems (Doctoral thesis, University of Genova, Genova, Italy). Retrieved from http://theses.eurasip.org/media/theses/documents/volpe-gualtiero-computational-models-of-expressive-gesture-in-multimedia-systems.pdf (last accessed December 15, 2016).

Wade, B. C. (1975). *Khyāl: A study in Hindustani classical vocal music* (Doctoral thesis, University of California, UCLA, USA). Ann Arbor, Mich: University Microfilms.

Wade, B. C. (1998). Hindustani Vocal Music. *The Garland Encyclopedia of World Music. Vol. 5: South Asia: The Indian Subcontinent*, pp. 162-187.

Walker, L., & Walker, P. (2016). Cross-sensory mapping of feature values in the size–brightness correspondence can be more relative than absolute. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 1, pp. 138-150.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28,* 6, pp. 879-896.

Wallmark, Z. T. (2014). *Appraising Timbre: Embodiment and Affect at the Threshold of Music and Noise* (Doctoral thesis, University of California, UCLA, USA). Retrieved from https://escholarship.org/uc/item/99t2t939 (last accessed December 15, 2016).

Wanderley, M. M., & Battier, M. (2000). *Trends in gestural control of music* (CDROM). Paris: Ircam.

Wanderley, M. M. (2002). Quantitative Analysis of Non-obvious Performer Gestures. *Lecture Notes in Computer Science,* 2298, pp. 241-253.

Wanderley, M. M., & Depalle, P. (2004). Gestural Control of Sound Synthesis. *Proceedings of the IEEE, 92,* 4, pp. 632-644.

Wanderley, M., Vines, B., Middleton, N., McKay, C., & Hatch, W. (2005). The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *Journal of New Music Research, 34,* 1, pp. 97-113.

Ward, N. (2013). *Effortful Interaction: A New Paradigm for the Design of Digital Musical Instruments* (Doctoral thesis, Queens University Belfast, Belfast, Northern Ireland, UK). Retrieved from http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.602967 (last accessed December 15, 2016).

Warren, W. H. J., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology. Human Perception and Performance, 10,* 5, pp. 704-712.

Watson, A. H. D. (2006). What can studying musicians tell us about motor control of the hand?. *Journal of Anatomy, 208,* 4, pp. 527-542.

Widdess, R. (1995). *The rāgas of early Indian music: Modes, melodies, and musical notations from the Gupta period to c. 1250*. Oxford: Clarendon Press.

Widdess, R. (2011). *Dynamics of melodic discourse in Indian music: Budhaditya Mukherjee's ālāp in rāg Pūriyā-Kalyān* In Roeder, J., & Tenzer, M. (2011). *Analytical and Cross-Cultural Studies in World Music* (pp. 187-224). Oxford: Oxford University Press.

Widmer, G., & Goebl, W. (2004). Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research, 33,* 3, pp. 203-216.

Wilkie, K., Holland, S., & Mulholland, P. (2010). What Can the Language of Musicians Tell Us about Music Interaction Design? *Computer Music Journal, 34*(4), pp. 34-48.

Wilkie, K., Holland, S., & Mulholland, P. (2009). Evaluating musical software using conceptual metaphors. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (pp. 232-237), Cambridge, UK.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131,* 3, pp. 460-473.

Windsor, W. L. (2012). Gestures in music-making: Action, information and perception. In Gritten, A., & King, E. (Eds.), *New perspectives on music and gesture* (pp. 45-66), Farnham, Surrey: Ashgate Pub.

Winkler, T. (1995). Making motion musical: Gesture mapping strategies for interactive computer music. In *Proceedings of the 1995 International Computer Music Conference* (pp. 261-264), Banff, Alberta, Canada.

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: an instance of the ideomotor principle. *Philosophical Transactions: Biological Sciences, 358,* 1431, pp. 501-515.

Yin, R. K. (2003). *Case study research: Design and methods*. Thousand Oaks, Calif: Sage Publications.

Young, K. (2002). The Memory of the Flesh: The Family Body in Somatic Psychology. *Body & Society, 8,* 3, pp. 25-47.

Zannos, I. (Ed.). (1999). *Music and signs: semiotic and cognitive studies in music* (Vol. 2). ASCO Art and Science.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron, 47,* 1, pp. 9-12.

Zbikowski, L. M. (1997). Conceptual Models and Cross-Domain Mapping: New Perspectives on Theories of Music and Hierarchy. *Journal of Music Theory, 41,* 2, pp. 193-225.

Zbikowski, L. M. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford: Oxford University Press.

Zbikowski, L. M. (2012). Musical gesture and musical grammar: A cognitive approach. *New Perspectives on Music and Gesture*, pp. 83-98.

Zhao, L. (2001). *Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures* (Doctoral thesis, University of Pennsylvania, Philadelphia, USA). Retrieved from https://pdfs.semanticscholar.org/f146/2519c9473867afa914674ce49796b7e4d3e5.pdf (last accessed December 15, 2016).

Zhao, L., & Badler, N. I. (2005). Acquiring and validating motion qualities from live limb gestures. *Graphical Models, 67,* 1, pp. 1-16.

Zsiga, E. C. (2012). *The sounds of language: An introduction to phonetics and phonology*. John Wiley & Sons.