Modeling Gamakās of Carnatic Music as a Synthesizer for Sparse Prescriptive Notation

Srikumar Karaikudi Subramanian

(M.Sc. by Research, NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMMUNICATIONS AND NEW MEDIA

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Srikumar Karaikudi Subramanian
6 Aug 2013

Acknowledgments

The journey of this work from a germ of a research proposal to this thesis would not have been possible without the help of teachers, family, friends and colleagues. First, I would like to express my respect and deep gratitude to my advisors Dr. Lonce Wyse and Dr. Kevin McGee, for their patient guidance, generous support and inspiration throughout this journey and for being real teachers.

I'm also deeply grateful to my father for his musical guidance during the early days of this research .. until when it got "too technical". To Dr. Martin Henz, for timely help in clarifying aspects of the research, for a clear and enjoyable introduction to constraint programming and for the fun hacking Javascript with his students. To Dr. Pamela Costes Onishi for introducing me to ethnomusicological literature and thought. To Mr. M. Subramanian, for the insightful discussions and pointers on Carnatic music and technology. To Dr. Rivera Milagros ("Mille") our former HoD who just has to walk into a space for it to fill up with her inspiring energy. Thank you Mille for your support all these years!

It is my honour to thank all the eminent musicians who participated in the study done as part of this research, who were all extremely generous with their expertise, time, space and patience. It was a privilege indeed to have their input regarding what is, as of this date, a young area.

To my cousin Shanti Mahesh and to Kamala Viswanathan-mami, themselves established *vīṇā* players, thank you for your help with references to other musicians during the evaluation study. Thanks also to Mrs. Usha Narasimhan for her help and time during the pilot phase and to Sri Tiruchy Murali for volunteering to help with contacts.

Arts and Creativity Lab and Partner Technologies Research Group offered work space and access to a studio, which was awesome, and Brhaddhvani provided the same in Chennai. I thank Norikazu Mitani-san (anclab) who went out of his way to help me with recording *vīṇā* sound samples for this work.

FASS/NUS generously provided for the opportunity to present a part of this research at ICMC 2011, for which I'm grateful. Thanks also to the CompMusic team, for the opportunity to share parts of this work and interact with a great group of researchers during their 2012 workshop and the hospitality extended.

I have to thank Ms. Retna for all her care and prompt administrative help throughout my candidature up to sending reminder emails about submission deadlines! To all my co-TAs of NM1101E, in particular Siti Nurharnani Binte Nahar, Anna Filippova, Wendy Wong and Gulizar Haciyakupoğlu, it was great working with you and the tips, tricks, gyan and laughs we've shared will stay with me.

Being away from family for a large part of this work was not easy. I thank my wife Shobana and son Uthkarsh for their patience and love, and my mother and mother-in-law for being pillars during this period. Chandra Anantharamu and my "musical friend" Divya Chandra encouraged and supported both this research and my personal musical growth. Many friends came forward with their support and I'd like to thank Anand, Aarthi, Lux Anantharaman, Chetan Rogbeer, Vivien Loong and Boey Wah Keong.

Colleagues at muvee Technologies helped by being flexible in accommodating my part-time studies and I'd like to thank Gerry Beauregard, Mafrudi bin Rubani, Terence Swee and Phil Morgan and to all of the muvee family, and especially Chetan Rogbeer, Sohrab Ali and Chua Teng Chwan.

I'm also grateful to Pete Kellock for long term friendship and mentorship, for all the amazing energizing annual mountain walks and local explorations he organized, for the great discussions on music and physics, and for general inspiration.

Finally, life in University Town wouldn't have been any fun without NEM's impromptu Kendo demonstrations and introduction to Samurai Champloo, random chats with

Maninder and watching Ganesh Iyer put on Kathakali makeup over three hours, and all those wee hours spent practicing vina with the NUS Indian Instrumental Ensemble. You guys will always be a part of me.

Srikumar Karaikudi Subramanian

To all the great vaiṇikās ...

"Boojum, huggie tha!"

- Uthkarsh S.

Publications

Aspects of this research were published in the proceedings listed here. Section 6.3 presents aspects published in [Subramanian et al., 2011]. Portions of sections 6.2.2 and 6.5 present work published in [Subramanian et al., 2012].

[Subramanian et al., 2011] Subramanian, S., Wyse, L., and McGee, K. (2011). Modeling speed doubling in carnatic music. In *Proceedings of the International Computer Music Conference*, pages 478–485, University of Huddersfield, UK

[Subramanian et al., 2012] Subramanian, S., Wyse, L., and McGee, K. (2012). A two-component representation for modeling gamakās of carnatic music. In *Proceedings of the 2nd CompMusic workshop*, pages 147–152, Bahçeşehir Üniversitesi, Istanbul, Turkey

Contents

Summary

One of the interesting and hard problems in the area of computer music synthesis is the construction of elaboration processes that translate a given sparse specification of desired musical structures into complex sound. The problem is particularly hard in genres such as Carnatic music, whose musical sophistication far exceeds that of its notation systems. In Carnatic music, compositions are communicated using a sparse "prescriptive" notation which a musician interprets using continuous pitch movements called "gamakās", with leeway for personal expressive choices. A computational model of the knowledge essential for such interpretation continues to be a challenge and open opportunity for deeper learning about the music.

Previous work can be categorized into hierarchical, constraint-based and dynamical approaches to elaboration. Hierarchical techniques include grammars used for generating melodies for Jazz chord progressions and lookup tables that map local melodic contexts to gamakā sets in Carnatic music. The traditional descriptive literature of Carnatic music provides information about permitted and forbidden melodic features that serve as constraints for composition and improvisation. A discrete optimality theoretic model of these rules as a set of ordered violable competing constraints has also been proposed by Vijayakrishnan. Dynamical models of pitch curves are common for modeling speech prosody and for vibrato and glissando effects for expressive singing synthesis.

The process of elaborating prescriptive notation in Carnatic music shows a mixture of hierarchical elements for context dependent choice of gamakās and preferences exhibited by musicians that order the set of possible gamakās over a phrase. Pure-hierarchical approaches show difficulty in modeling soft preference constraints and pure constraint-based approaches need to work with a large search space. This research goes beyond the previous work by proposing a data-derived model that combines hierarchical generation of possible gamakās with a system of soft lateral constraints for optimal phrase level selection that include adaptation of gamakās to local temporal contexts.

The method used was to first transcribe a reference performance of a sparsely specified composition into a representation that captures gamakā details and, based on the internal consistencies of the composition and the discrimination expressed by the artist in the performance, construct elaboration tables, continuity constraints on gamakās, and rules for adapting gamakās to different local melodic contexts. These were done using two different representations and the resulting elaboration systems were evaluated through interviews with expert musicians for acceptability, range of variations generated and scope of applicability.

Contributions of this research fall into two categories – computational models of the regularities of gamakās, and implications of the models for the musicology of the genre. Findings include the simplification of local melodic context necessary for elaboration and the consequent expansion of capability, constructing rules for adapting slower gamakās to higher speeds and the identification of a new representation for gamakās that separates gross movements from stylistic/ornamental movements. Some support was also found for the "competing constraints" model of elaboration in Carnatic music through the expert evaluation. The musicological consequences of the new representation and guidelines for transcription using it are also discussed.

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction and background

Contents

Honing defines musicology as “the study of formal structure in a musical form of interest” [Desain and Honing, 1992]. An important kind of musicology is the study of established musical genres through the construction of computational models that analyze and generate performances and is termed “computational musicology”. When considering genres that feature a written prescription for the music to be performed, an interesting question arises as to what musical knowledge is required to realize a performance given such a prescription, a process that we might call “elaboration”. Musical knowledge required for elaboration can include elements of what can be considered common knowledge among practitioners of the genre, as well as elements of personal style, taste and school of training. The construction of computational elaboration processes that fill the gap between prescriptive notation and performance is an interesting and challenging way to approach the knowledge that musicians bring to a performance.

Genres of music vary among and within themselves in the extent to which the music to be performed is notated. Based on the degree of notated detail and the kind of gap between notation and performance, we can identify two significant categories of elaboration namely *expressive* and *structural* elaboration . Western classical music’s staff notation system has tools for specifying a composer’s intent to a great degree of detail with variable demands on the performing musician to be expressive with timing, dynamics, timbre and some forms of pitch articulation. When computer performance systems that generate such interpretations focus on modifying the performance parameters of given melodic or rhythmic entities, they are called *expressive performance systems* or *expressive synthesis systems* . In contrast, it is common practice for a Jazz ensemble to agree on a given chord progression and improvise melodies within the harmonic structure laid down by the progression. This kind of elaboration therefore involves the creation of unprescribed melodic and rhythmic entities, which can be termed *structural elaboration*.

The elaboration of prescriptive notation[1] in Carnatic music (South Indian classical music)[2] , which is the focus of this thesis, is a combination of structural and expressive elaboration. The prescriptive notation used in the genre records melody in phrases described as sequences of notes, but the most characteristic melodic feature – continuous pitch movements called "gamakās" – are omitted from the notation. It is therefore up to the musician to interpret notated phrases using appropriate gamakās. Although the specification of a phrase is not as open ended as a chord given as part of a progression in Jazz when considered at the same time-scale, it is also not as specific as a notated work in western classical music in that it admits of multiple melodic interpretations that use tones and tone movements not explicit in the notation. Some teachers use an intermediate level of notated detail called "descriptive notation" that captures the new melodic entities introduced in an interpretation of a work given in prescriptive notation [Viswanathan, 1977].[3] The gap between a work's prescriptive notation and the descriptive notation of one of its performances is largely a *structural* gap, whereas that between a descriptive notation and its realization as a performance is largely an *expressive* gap.[4]

This chapter presents an overview of the problem of elaboration, discusses issues surrounding the study of a genre through computational means and provides background material about Carnatic music and relevant issues of culture, pedagogy and style to the extent necessary to grasp the remainder of this work. The following chapter takes up a

[1] Ethnomusicologist Charles Seeger in [Seeger, 1958] defined "prescriptive notation" as notation intended for interpretation by one or more performers which can assume as known what is considered to be common knowledge among practitioners of the genre it is intended for. In this context, the term is extended to refer to a corresponding sparse representation that serves as input to a computer program that "performs" the notated music. Though they are different entities, distinguishing between them is unnecessary for the purpose of this work.

[2] "Karnatak" is also used as an anglicized form and is closer to the pronunciation in the local languages of southern India such as the Tamil pronunciation "karnāṭaka sangītam". Some musicians prefer this spelling due to it being more phonetically accurate than "Carnatic" [Viswanathan, 1977]. This document uses "Carnatic" due its greater prevalence among recent English writings about South Indian classical music, and given that the word may be found pronounced as "karnatak" or "karnatik". The important point is that all these words and spellings refer to the same genre.

[3] The term "descriptive notation", also introduced by Seeger, stands for a notation of *a* specific performance of a prescriptive notation.

[4] In this case, the descriptive notation plays the role of a prescriptive notation, only that it provides more detail.

more detailed examination of the work relevant to the problem of elaborating the prescriptive notation of Carnatic music.

In this document, I attempt to maintain simple language and terminology in the interest of making it accessible to a broad audience who may not be familiar with Carnatic music by highlighting analogous concepts. However, suitable analogies may not be possible under all circumstances. I present genre-specific terms, concepts and clarifications either as footnotes at the appropriate points or in the glossary.

1.1 Computational musicology

In this section, we look at what makes the study of a music[5] through computational processes appealing, followed by issues of perception, modeling and knowledge representation surrounding such studies, and relates them to Carnatic music.[6]

Approaches in computational musicology, as applied to established musical genres tend to fall into two categories of means – *analysis* and *synthesis*. Analytical approaches begin with musical artifacts and attempt to develop algorithms that relate features of these artifacts to musical concepts derived from the known musicology of the genre. The active field of Music Information Retrieval (MIR) consists of analytical approaches that work with sound recordings as the starting point, with a focus on techniques for comparison, indexing and search [Typke et al., 2005]. Due to the intricacies of pitch, time, harmony, timbre, editorial, textual and bibliographic facets and the complex interactions between them that make up the problem of MIR, Downie describes MIR as "a multifaceted challenge" [Downie, 2003, p. 297]. Analytical approaches might also use symbolic representations of musical artifacts as their starting point, with the aim of developing procedures to identify structures and regularities in the music, for composition or comparative studies. The older Humdrum toolkit and the recent music21 toolkit are examples of systems built to facilitate symbolic analytical approaches [Huron, 1993, Jan, 2004, Huron, 2002, Cuthbert and Ariza, 2010]. Synthetic approaches aim to study some aspect of a music by attempting to recreate it using algorithms. As a mirror of analytical approaches, synthetic approaches might either have the actual sounds as the end point [Battey, 2004, Sundberg et al., 1983, Friberg et al.,

[5]Here, "a music" is used as short hand for "a genre of music" and subsumes the notion of "a music culture" within it. The term also lends itself to pluralization as "musics". These are common usage in ethnomusicological writings.

[6]Using computers for music composition is a much larger area of work and it is neither necessary nor possible for this document to cover the entire field. Other authors have written extensive and excellent works on the topic to which the reader is referred to [Dodge and Jerse, 1985, Roads, 1996, Leman, 1996, Rowe, 2004, Boulanger, 2000, Todd and Loy, 1991].

2006, Berndtsson, 1996], or have a symbolic intermediate representation such as Musical Instrument Digital Interface (MIDI) as the end point [Kippen and Bel, 1992, Cope, 1989, Cope, 1991b].

Though it is useful to examine an approach in terms of the above categories, goals often appear mixed – i.e. analysis might be performed with the express goal of using the result to synthesize a related musical structure, or synthesis might be attempted with the goal of discovering concepts and structures relatable to the known musicology of a genre. Cope's work on EMI ("Experiments in Musical Intelligence", pronounced "emmy") is about generating compositions in the styles of known composers such as Mozart, Bach, and Chopin. Despite the focus on composition, Cope expresses the interplay between analysis and synthesis and its value to musicology thus —

> "Research with the Experiments in Musical Intelligence program also extends my understanding of the importance of style, voice leading, hierarchy, and other compositional implications of the composer's original music." [Cope, 2000, p. 32]

A reasonable critique of Cope's statements is that they are indicative of the idiosyncratic nature of the concepts and representations embodied in EMI and Cope acknowledges the same in his writings. Furthermore, Kippen and Bel in their attempt to model the "largely intuitive knowledge of North Indian drummers" by building an expert system based on generative grammars, also conclude that "a BP [Bol Processor] grammar can be nothing other than a joint construction of the informant and the analyst". In other words, the grammar resulting from the process followed in their work is dependent on both the informant and the analyst and a different grammar may be constructed if the participants were to be different. To remedy this subjectivity, Kippen and Bel suggest that "automated learning procedures" might help bring objectivity to the task [Kippen and Bel, 1989]. This appears to justify the approach taken in the field of MIR in the application of unsupervised machine learning techniques such as self-organizing maps to the analytical task [Typke et al., 2005].

Apart from musical concepts and representations that originate in the already developed musicology of a genre, synthesis based approaches to musicological discovery serve as another source of such representations, which can inform work on MIR.[7] This input is important because research in MIR de-emphasizes the musicological relevance of the techniques used to achieve the operational goal.[8] The Humdrum toolkit, the WEDELMUSIC

[7] This comment considers only features at a higher level of music perception than those that originate in signal processing and the psycho-acoustic features close to it. A "musicologically relevant feature" can be, to a first approximation, described as psycho-acoustic features independent of timbre.

[8] "For information retrieval, we are not interested in explanation so much as we are in comparison or

format, music21 and polymetric expressions in the Bol Processor are examples of such contributions [Huron, 1993, Bellini and Nesi, 2001, Cuthbert and Ariza, 2010, Bel, 1998, Bel, 2005].

We now look at some computational techniques used to study music by means of either analysis or synthesis.

1.2 Carnatic music notation and performance

The earliest notated musical forms that can be associated with Carnatic music are the seventh century Kudumiyanmalai inscriptions [Widdess, 1979], which indicates a long though sparsely documented musical history. Despite the early history, the notation system in use has seen little attention from practitioners, possibly due to the emphasis on oral traditions, improvisation and interpretation. As Vijayakrishnan writes –

> "The tradition of notation is not as firmly entrenched in Carnatic music as it is in, say, Western music across genres. There are two diametrically opposing views on the nature and use of notation in Carnatic music among practitioners: Carnatic music cannot be notated as it is an oral tradition and that no useful purpose is served by any type of notation; and the minority view is, of course, the pursuit of honing notational skills to improve the status of notation in Carnatic music." [Vijayakrishnan, 2009]

Modern publications in Carnatic music continue to use a sparse form that does not include details of gamakās. Figure 1.1 shows an extract from the prescriptive notation of a varṇam given in appendix B. The top line provides the solfa names of the pitches to be performed, together with the time structure indicated using vertical bars. The second line provides the lyrics associated with the notes above. The use of roman letter representations of solfa is common practice in publications that intend to cross regions, though the same presentation structure as used in regional language publications in southern India is used (see table 1.1). The notation presented here is a simplified form that makes the time structure explicit – i.e. where the "," symbol indicates a time gap of one-fourth of a count[9], publications abbreviate ", ," using ";".

similarity measures. Any technique which produces features that aid the retrieval process is useful." [Pickens, 2001]

[9] A tāḷa cycle consists of a number of *counts* spaced equally in time. It can be considered equivalent to a *bar* in western classical music when the tāḷa is in a slower tempo of, say, 30 counts per minute. Such a *count* is known by the name *aksharā*.

P	,	m	,	G	,	G	m		R	,	G	R	S	,	,	,	||
ka		*ru*		*ṇim*									*pa*				
ṇ	S	R	G	R	S-	ṇ	S	|	Ḍ	P̣	ṃ	Ḍ	,	ṇ	S	R	||
i				*di*					*man*			*ci*					

Figure 1.1: A snippet of prescriptive notation

1)	C	$D_\flat$	D	$D^\sharp$	$E_{\flat\flat}$	$E_\flat$	E	F	$F^\sharp$	G	$A_\flat$	A	$A^\sharp$	$B_{\flat\flat}$	$B_\flat$	B
2)	sa	ri$_1$	ri$_2$	ri$_3$	ga$_1$	ga$_2$	ga$_3$	ma$_1$	ma$_2$	pa	da$_1$	da$_2$	da$_3$	ni$_1$	ni$_2$	ni$_3$
3)	*sa*	*ra*	*ri*	*ru*	*ga*	*gi*	*gu*	*ma*	*mi*	*pa*	*da*	*di*	*du*	*na*	*ni*	*nu*
4)	S	r	R	g	R	g	G	m	M	P	d	D	n	D	n	N

Table 1.1: Pitch class naming conventions used in Carnatic music (2..4) and their relationship to pitch classes of western music (1).

Descriptive notation[10] was introduced for the purpose of greater precision in musical communication in [Viswanathan, 1977]. It is not common practice to notate compositions at that level of detail in publications. Figure 1.2 shows an attempt to graphically describe the nuances of the music in detail [Subramanian, 1985b]. The figure shows different levels of detail of the melody including an approximate translation into staff notation. At the top is the prescriptive notation written using solfa names. It is followed by descriptive notation and a graphical notation that is referred to by the author as an "emotion graph". The difference in detail between the prescriptive notation at the top and the graphical notation captures the gap in musical features that needs to be bridged by a musician seeking to interpret the prescriptive notation.

[10] "Descriptive notation" is notation of a specific performance of a composition after the fact [Seeger, 1958].

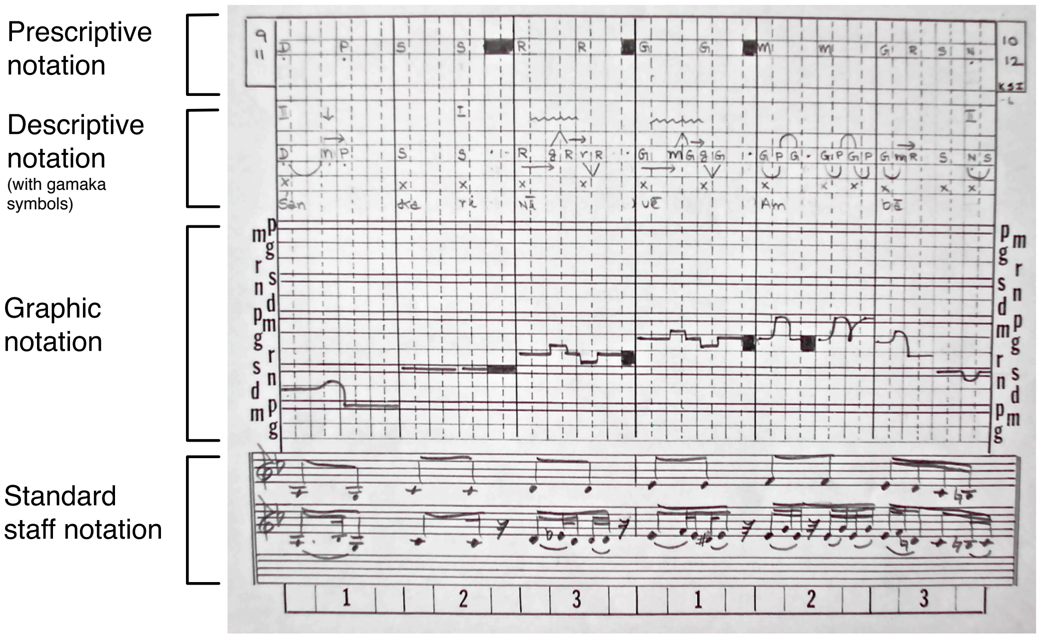


Figure 1.2: Detailed transcription of two 3-beat cycles of the composition *Śankarinīvē*. Used with author's permission from [Subramanian, 1985b].

Chapter 2

Related work

Contents

The previous chapter introduced the category of elaboration systems – processes that synthesize a performance from music given as prescriptive notation – and the sub-categories of *structural* elaboration systems and *expressive synthesis* systems and presented some theoretical frameworks used by such systems. The problem of synthesizing Carnatic music from its prescriptive notation was introduced as an elaboration problem that is a combination of structural and expressive elaboration. In this chapter, I review previous work that provides formalisms and techniques relevant to the elaboration problem in Carnatic music and other genres. The musicological literature of Carnatic music contains descriptive material about rāgās and ontologies for gamakās that, though subject to debate, provides a starting point. In contrast to formal grammars that have been applied to other genres such as Jazz and tabla improvisation, an optimality theoretic framework has been proposed for formulating the principles of Carnatic music. Techniques based on pattern matching, augmented transition networks and recombination procedures have been applied to automatic composition of western classical music from partial specifications. Rule systems for singing synthesis and speech prosody modeling deal with continuous signals that parallel gamakās. The *Gaayaka* system has an "automatic gamakam" feature for user guided interpretation of prescriptive notation that is based on expanding local melodic contexts using a phrase database.

I begin with the theoretical frameworks relevant to the elaboration problem in Carnatic music.

2.1 Music theory

Carnatic music has a rich musicological literature that has a direct bearing on the problem of elaborating prescriptive notation. The literature describes the characteristics of several formal structures which are part of the genre including composition types, systems of melodic constraints called "rāgās" and ontologies of pitch ornamentations – i.e. "gamakās". Due to the largely oral tradition of teaching and an emphasis on improvisation and variation, practitioners have written down what might be called the ground rules of the genre.

The primary musicological entity we need to examine here is the "rāgā" and the gamakā ontologies that have been developed to describe their attributes.

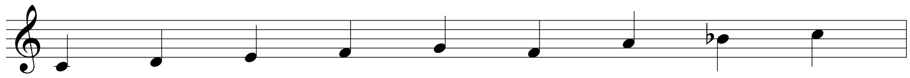
2.1.1 Rāgās and Rāga lakshaṇās

The term "rāgā" is not a precise concept in Carnatic music and yet knowledge of the rāgā of a notated composition is crucial for a musician to interpret it. It can loosely be said to encapsulate those properties that lend melodies characteristic tonal "colour".[1] Shankar, for example, describes a rāgā as a "melody-mould" [Shankar, 1983, p. 33]. From a practical perspective, a rāgā constrains the selection and sequencing of pitches that can constitute a melody. These pitches are considered relative to a tonic and are therefore better described as "pitch classes". Rāgās are typically recognized through a set of pitch classes as well as by specific phrases and gamakās.

Descriptive literature on rāgās written by established practitioners of the genre are called "*rāga lakśaṇā*-s" . Perhaps the most famous historical work in this regard is the 13th century work "Sangīta Ratnākarā" by Sāranga Dēvā. A more recent treatise specific to the Carnatic genre that continues to serve as a reference is the early 20th century work of Subbarama Dikshitar "Sangīta Sampradāya Pradarśiṇi" [Dikshitar, 1904]. As an example,

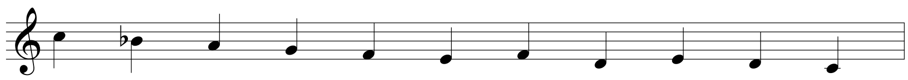
[1] "Colour" is one of the translations of the word "rāgā".

Ascent



C	*D*	*E*	*F*	*G*	*F*	*A*	$B_\flat$	$C^{8^{va}}$
sa	ri_2	ga_3	ma_1	pa	ma_1	da_2	ni_2	$\dot{sa}$
sa	*ri*	*gu*	*ma*	*pa*	*ma*	*di*	*ni*	$\dot{sa}$
S	R	G	m	P	m	D	n	$\dot{S}$

Descent



$C^{8^{va}}$	$B_\flat$	*A*	*G*	*F*	*E*	*F*	*D*	*E*	*D*	*C*
$\dot{sa}$	ni_2	da_2	pa	ma_1	ga_3	ma_1	ri_2	ga_3	ri_2	sa
$\dot{sa}$	*ni*	*di*	*pa*	*ma*	*gu*	*ma*	*ri*	*gu*	*ri*	*sa*
$\dot{S}$	n	D	P	m	G	m	R	G	R	*S*

Table 2.1: Ascent and descent pitch patterns for the rāgā “Sahānā”. Note the zigzag nature of these patterns.

the feature details of rāgā Sahānā are given in the appendix C, reproduced from [Mahesh, 2007] with the author’s permission.

The rāgā traits relevant to the problem of elaborating prescriptive notation that are described in *rāga lakśaṇā*-s are –

1. Characteristic gamakās that announce the rāgā. This involves specific movements around pitch classes that are part of the rāgā and also approximate timing information about these gamakās.

2. Out-of-scale pitches permitted or forbidden in the articulation of gamakās.

3. Precaution on use of phrases that overlap with another rāgā or minor phrase variations that would invoke another rāgā.

I now describe Dikshitar’s gamakā ontology on which later musicologists such as Viswanathan and Gopalam based their works.

2.1.2 Gamakā ontologies

Though gamakās are primarily continuous pitch movements, the notion of discrete categories for gamakās is well established in the musicological literature of the genre. Two prominent works that attempt to lay out an exhaustive ontology of gamakās used in the musical practice of their respective times are Subbarama Dikshitar's "Sangīta Sampradāya Pradarśiṇi" [Dikshitar, 1904] and Vidya Shankar's transcriptions of Śyāmā Śāstri's compositions [Shankar, 1979]. The former is a three volume treatise detailing attributes of various rāgās in the classic "rāga lakshaṇā" style in addition to providing transcribed compositions for each rāgā. To improve on the accuracy of the transcription, Dikshitar introduces and uses symbols for various categories of gamakās that feature in his transcriptions. Shankar borrows Dikshitar's terminology, categories and notation for the transcriptions and describes Dikshitar's categories in the language of contemporary practice.

In [Gopalam, 1991], Gopalam finds that although Shankar's categories reference those of Dikshitar, they also depart in some important ways due to the need for interpretation of Dikshitar's verbal descriptions as well as change in musical practice since the earlier work. The lack of audio recording facility during Dikshitar's times forces reliance on aural transmission from teacher to student over several generations. Therefore the terms introduced by Dikshitar and their descriptions are prone to error in direct interpretation as well as cumulative deviations from the original intended meanings over time. Gopalam's thesis contains a detailed account of the differences in the ontologies expressed in those two works and therefore serves here as a recent expert's view of known gamakā ontologies.[2]

In table 2.2, I present an approximate condensed visual interpretation of the verbal descriptions of these gamakā categories by the three scholars mentioned. In addition to their verbal descriptions, the examples for the types of gamakās presented in descriptive notation in Viswanathan's dissertation also helped disambiguate possible interpretations of the text [Viswanathan, 1977, p. 33-34]. Other ontologies based on Dikshitar's work include [Iyengar, 1965] and [Mallikarjuna Sharma, 2007].

2.1.2.1 Instrument as medium of definition

In their respective works, both Dikshitar and Shankar provide operational definitions for gamakās, by describing techniques for performing them on the *vīṇā*. The use of an instrument as a medium to describe gamakās raises the important issue of which gamakās are to

[2] A detailed study of the gamakās described by Dikshitar which uses Gopalam's comparative study as a key reference point can be found in [Jayalakshmi, 2002].

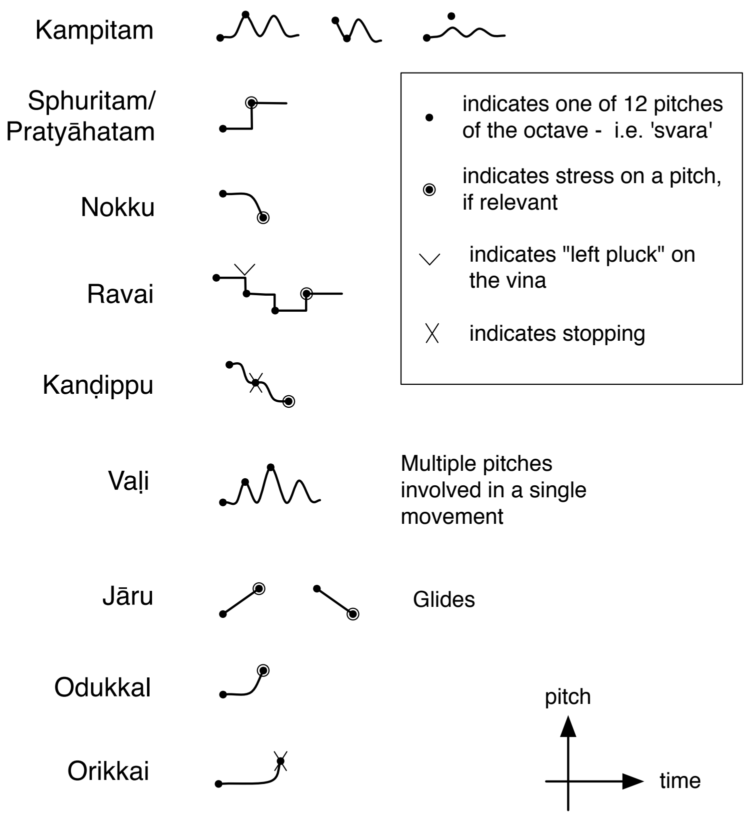


Table 2.2: A rāgā-agnostic illustration of the approximate shapes of gamakā types described in the musicological literature of Carnatic music. Some types of gamakās are specific to the *vīṇā*.

be attributed to the music and which are instrumental techniques. In a genre with repertoire common to vocal and instrumental performance, it is also questionable whether such a separation is indeed possible, given the continuous process of musical exchange among practitioners. Gopalam finds the operational definition of gamakās problematic –

> "The equating of a gamakā with its production in a particular medium [...] may have as its basis lack of understanding of the gamakā as an entity of music. A further basis for equating of the gamakā with its production in a particular medium is a lack of understanding that which is very specific to only voice or a given instrument will, by extension, be disposable to music, and therefore not a gamakā." [Gopalam, 1991, p. 67-68].

Viswanathan's use of descriptive notation does serve to abstract their form from the techniques necessary to perform them on an instrument. However, realizing a piece of descriptive notation on an instrument requires the artist to interpret the abstract description in terms of the techniques available on the instrument.[3] The necessity for interpretation implies that a given piece of descriptive notation does not unambiguously resolve a gamakā among alternatives.

The role of the instrumental medium in gamakā articulation is amplified when approached through computer models. It is common in computer music to conceive of a synthesis system in two parts – an "instrument model" that describes the sound produced and its relationship to a set of exposed "control parameters", and a component that produces a "score" consisting of the time evolution of the controls exposed by the instrument model used. CSound, for example, makes an architectural separation between an *orchestra*, which consists of a set of instrument models, and the driving *score* which describes the time sequence of instantiation and control messages to be sent to the orchestra [Vercoe, 1986]. When mapping gamakās onto such a two-component synthesizer, it is important to clarify which attributes of the music are being modeled in which component.

2.1.2.2 Attributes of gamakās

In principle, the complete description of a gamakā requires the three attributes of pitch, timing and dynamics. Yet, that is also the apparent order of their importance in the literature. Whereas pitch is the dominant feature of *rāga lakśaṇā* treatises, timing is given much less importance and dynamics even lacks representation in active vocabulary.

[3]Note that descriptive notation, when used like this, serves a prescriptive role.

Dikshitar and Shankar provide summary descriptions of timing characteristics of gamakās — whether a particular gamakā is to be used with a "long" or "short" notes, that the end point of an "*orikai*" is a "brief deflection", and so on. The descriptive notation introduced by Viswanathan articulates the timing of the movements that constitute a gamakā to a higher degree of precision by using durations that are simple fractions of a beat, such as 2/4 and 3/4 [Viswanathan, 1977, p. 33-34].

The significant part of the problem of elaboration in Carnatic music lies in modeling pitch and timing characteristics since the dynamics of gamakās finds little mention in the ontology compared to pitch and timing. As Gopalam notes –

> "We do, however, have gamakā names which are distinguished by this single factor [dynamics], i.e. *nāmita* and *humpita*, forming part of the group of fifteen gamaka-s. But these terms exist only in name and we have practically no rapport with them." [Gopalam, 1991, p. 70-71]

To explain this lack of rapport, Gopalam proposes that listeners familiar with Carnatic music understand the dynamics component of gamakās not as such but through its *emotive* effect on them [Gopalam, 1991, p. 70]. However, we also need to consider the possibility that the poor representation for dynamics in active vocabulary is indicative of its tertiary significance in traditional practice.

2.1.3 Vijayakrishnan's optimality theoretic framework

In section 2.1.1, we saw that the traditional musicological works seek to provide guidelines to practitioners by describing properties and rendering constraints for each rāgā. Recently, a formulation of the principles of Carnatic music that covers the entire musicological ground based on Prince, McCarthy and Smolensky's "Optimality Theory" has been proposed by Vijayakrishnan in [Vijayakrishnan, 2007]. Optimality Theory (OT for short), takes the stand that rule systems are too strong for modeling well-formedness of productions in a language. Instead, it proposes to model the grammar of language as a system of violable constraints, with some of them taking precedence over others in a hierarchy.

Vijayakrishnan proposes that OT is a suitable framework for modeling aspects of Carnatic music owing to its language-like properties. OT's basic architecture follows a generate-and-test approach to defining a language as opposed to the generative grammars which define a language by producing only valid instances of it. This architecture is shown in figure 2.1. The place where OT departs is that it permits some of the constraints in its set to

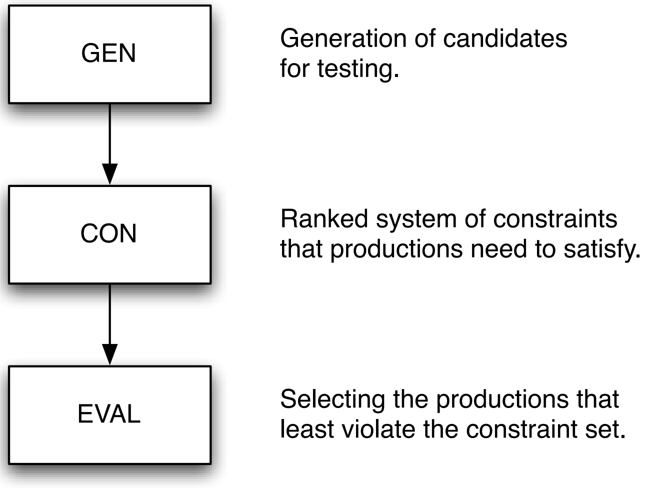


Figure 2.1: The basic architecture of Optimality Theory.

be violated. Vijayakrishnan posits therefore that Carnatic music also features constraints that can be violated under certain circumstances, provided certain other constraints are held. This approach is followed for all levels of music, from the interpretation of pitch values indicated in notation, to the "musical line" to the rāgā and higher level stylistic and performance context properties.

One of the contributions of this work that is relevant to modeling gamakās is the introduction of new discrete categories for the known twelve tones of the octave. New labels called "augmented" and "reduced" are attached to tones depending on whether they are approached from above or below in a given gamakā. For example, a "ri" (D) approached from the "sa" (C) below is labelled "red.ri". The core idea here is that these "reduced" and "augmented" tones are the appropriate interpretation of the 22 microtones[4] per octave described in traditional Carnatic musicological literature. A complete movement is described as proceeding from an "anchor" tone to a "target" tone. Constraints such as "the anchor of a glide is the note which is to be augmented" are then laid out for rendering these reduced and augmented tones [Vijayakrishnan, 2007, p. 102].

In all, the framework of soft violable constraints does appear promising as a foundation for describing the principles of Carnatic music. Although the principles have been detailed well in Vijayakrishnan's work, it is not yet clear how such a system of constraints can be assembled and realized as a computation that generates output considered valid in

[4] Called the 22 *śrutī/s*.

the genre.

2.2 Structural elaboration

2.2.1 Gaayaka

Gaayaka is a computer program developed by M. Subramanian for synthesizing a performance from a plain text notation of music in the Carnatic genre [Subramanian, 2009b]. The plain text notation supports both detailed descriptions of gamakā[5] characteristics such as microtonal inflections, as well as automatic determination of these details extracted from a database of phrases, given a skeletal description of the music that resembles the prescriptive notation used by practitioners. The latter function is representative of the class of elaboration systems and is important to this work due to its specificity to Carnatic music. Here I summarize the characteristics of Gaayaka's notation system and its mechanism for automatic phrase expansion.

2.2.1.1 Notation system

The primary components of Gaayaka's notation system are the "sa ri ga ma" solfege symbols standing for pitch classes, characters indicating temporal structure, and characters indicating microtonal positions. Table 2.3 shows a sample of such a detailed notation of the beginning of a traditional composition. [6]

Octaves are represented using letter capitalization and up to three octaves can be represented in Gaayaka's scheme. The lower octave pitch for "ga" is written as "Ga" and the higher octave pitch is written as "gA". The postfix symbols "$>$" (decrease pitch) and "$<$" (increase pitch) are used to notate pitch inflections smaller in extent than a semitone. Parentheses are used to group expressions to be performed at speed factors that are powers of two — the deeper the level of nesting of the parentheses, the higher the speed.

The key elements of the notation are words such as "sa" and "pa" which are solfege symbols, the comma indicating the continuation of a note, levels of parenthesis representing doubling of speed and the $<$ and $>$ symbols which are microtonal adjustments of the pitch classes that the solfege symbols represent.

[5]The author uses the word "gamakam" which stands for the same musical feature as what is meant by the term gamakā in this text.

[6]The composition is the well known "aṭa tāḷa varṇam ‘’ "Vanajākshi ninnē kōri".

Table 2.3: A detailed notation of one cycle of a composition in rāgā Kalyāṇi using Gaayaka's syntax including the necessary microtonal and microtemporal aspects.

Prescriptive notation

```
sa , nI dA    nI sa ri ga   |  pa ,  ma ,    ga , ri ,    ri ,  ,  ,  |
ga ma pa ga   ri sa nI dA   |  ri sa sa ,    ri dA ga ri  ga pa ma ri |
ga ma pa da   pa ma ni da   |  ni pa  , ma    ga da ma ga | ri sa nI dA nI sa ri ga
```

Transcription of expected performance

```
{T 48}

((pA sa,,)) , ((sa , sa>>> sa)) -((dA. sa. dA))   ((sa , sa>> sa))
- sa ri ((ga<< ga , ,))   |
(((ga<<. ga. , ga<<. ga. , ga<<. ga. , ga<<. ga. ,)))
(((da pa , , , pa>>> pa , , , , , , , , ,)))  |
((pa , pa>>> pa)) ga ((ga<<< ga , ga<<<)) ri    ((sa ri , ,)) ,  ,  ,  |

((ga ma ga pa pa>>> pa , da)) -(pa ((pa>>> pa , ,)) ga ((,,,ma>>>)))
 ((ri,,ga>>)) ((sa , , ga)) ((sa , sa>> sa)) ((dA sa dA,))   |
-((ri<<< ri , ri<<<)) pA ((pA , sa dA sa , sa>>> sa))   ((ri<<< ri , ri<<<))
-dA ((ga<<< ga , ga<<<)) ri   -((ga pa>>> ga pa pa>>> pa pa>>> pa ga ma ga , ri ga ri,)) |

((ga. pa. ga)) pa -pa ((da<<< da , da<<<))   pa -((pa , , pa>>> pa , , pa>>>))
((Sa -Sa da Sa, ,ni Sa))   |
-((da pa ma pa , , , da))  -((pa , pa>>> pa))    ((ga ga<<< ga ,)) ((Sa Sa>>> , Sa))
((da ,, da<<)) ((pa>>> , , pa)) ((ga , , ga<<))   |
((ri , , ri<<)) sa -((dA sa dA ,)) ((sa , sa>>> sa)) -sa ri ((ga<< ga,,))
```

2.2.1.2 Automatic gamakā expansion

Gaayaka has an "automatic gamakam" feature which provides gamakā suggestions for phrases specified in a skeletal form close to the prescriptive notation used among genre practitioners. The program provides these gamakā suggestions by looking up the melodic context of each notated pitch in a phrase database [Subramanian, 2009a].

Gaayaka interprets a given piece of notation in the context of a rāgā setting. This setting affects the meaning of the solfege symbols "sa ri ga ma pa da ni" and also selects the database to use to elaborate a given phrase using gamakās. The gamakās are therefore specific to the rāgā selected. The melodic context of each note in the given phrase consists of —

1. the note's pitch class,
2. the note's duration folded into five discrete duration categories,
3. the preceding pitch class, and
4. whether the note is part of an ascent, a descent or an inflection pattern.

Gaayaka's "automatic gamakam" mechanism serves as a *guided* elaboration system for prescriptive notation of Carnatic music. The database consists of a lookup table that maps each possible context in a rāgā to a number of phrase choices. The multiple choices, if available, are presented to the user at elaboration time to enable manual selection according to taste.

2.2.2 Bol Processor

Generative grammars are a general formalism for expressing transformations of abstract representations to move concrete forms, as well as to analyze concrete instances in terms of a pre-specified abstract set of rules. The Bol Processor system features such a grammar engine capable of both analysis and production. In order to enable a grammar to model musical transformations using string rewriting rules, the Bol Processor models temporal concatenation as textual concatenation using "polymetric expressions" [Bel, 1998].

In [Kippen and Bel, 1992], Kippen and Bel outline the process of deriving the grammar of a tabla composition given a few instances. The essence of their process is to recognize structure in the composition instances and model the structure as substitution rules in a pattern grammar. The end goal is for the grammar, when run in reverse, to be able to generate patterns similar in spirit to the original patterns. It is interesting to observe how deep and complex the rule system becomes even for the domain of rhythmic patterns where there is a good match between temporal concatenation of rhythms and textual concatenation. Changing a grammar to accommodate or describe new features becomes more difficult the more complex the grammar is. Despite the complexity, working with grammars have yielded important learnings about the construction of expert systems for musical modeling. In [Kippen and Bel, 1989], the authors conclude that "... a BP grammar can be nothing other than a joint construction of the informant and the analyst" and recommend automatic learning mechanisms as a possible solution to this problem.

2.2.3 Cope's EMI

David Cope's Experiments in Musical Intelligence (EMI) [Cope, 1987, Cope, 1989, Cope, 1991a, Cope, 1992] is an important example of an attempt to answer the question of "can computers compose music like our great masters". Though elaboration is not as open-ended a problem as automatic composition through imitation of known musical styles, some of the modeling techniques developed by Cope can be seen as constituting an elaboration sub-system, which is worth going into some detail in this context.

Cope takes the approach of developing algorithms to analyze a selection of compositions by a composer, abstracting a "style" from the developed rules and generating new compositions incorporating the stylistic elements in it. One of the unique characteristics of EMI is the fact that a "listener" is built into the system, which monitors the evolution of a composition and retrains accordingly. EMI draws on many techniques from the domain of artificial intelligence such as connectionist concept networks for the modeling of musical concepts and the relationships between them, pattern matching, statistical analysis, augmented transition networks, databases of abstracted patterns and rule systems for their "recombination". Most systems in the category of fully automatic composition limit themselves to a few styles within a genre and EMI is no exception to that. However, the success of the program in emulating the style of Chopin, for instance, lends credence and hope to the idea of using composition algorithms to model known kinds of music using algorithms.

One of Cope's important contributions has been the SPEAC system for hierarchical analysis of melodic and harmonic structures that is inspired by Schenkerian analysis. Using the SPEAC system, new compositions are generated from skeletal representations extracted from known works of classical composers through pattern matching techniques. SPEAC is an acronym that stands for (S)tatement, (P)reparation, (E)xtension, (A)ntecedent and (C)onsequent. Musical phrases are, in the analysis phase, classified into one of these roles at various levels. The idea is that the role played by harmony can depend on context, much like the fact that words can take on different meanings depending on context. Cope further splits each of these roles into multiple "levels". For example, an expression classified as S1 is a higher level and more abstract statement than one classified as S3. It appears that Cope's SPEAC system is a significant contribution to the analytical toolkit of the classical musician and student. Cope also departs from the conventional approach to western classical composition, which emphasizes harmonic structure and brings melody under its umbrella, and considers melody and harmony to be separate aspects of the composition despite their interplay and models them separately in EMI to good effect. The EMI composer works on structural constraints laid out by the SPEAC system. EMI analyzes known works to create temporal sequences labelled with the symbols S1, S2, S3, ..., P1, P2, P3, etc. The composer then works by elaborating on known SPEAC patterns by looking up a database of phrases labelled with their SPEAC analyses and stitching them together using local recombination rules. It is not uncommon to find such examples of elaboration sub-systems being used in what are otherwise fully automatic composition systems.

2.2.4 Jazz melody generation

Creating improvised Jazz melodies that harmonize with given chord progressions and the generation of variations of melodies are instances of structural elaboration problems and several systems have been developed for these purposes, usually with the goal of automatic accompaniment for practice [Ulrich, 1977, Pennycook et al., 1993, Ramalho and Ganascia, 1994, Gillick et al., 2010, Biles, 1994, Keller and Morrison, 2007].

Ulrich, being a clear precursor to the others in automatic jazz improvisation, lays down the basic approach of performing a functional analysis of a Jazz song that results in identifying "key centres" and groups of measures that move between these key centres. The generation of melodies that conform to the analyzed harmonic structure is a structural elaboration problem. Ulrich's approach is primarily grammatical, together with procedures for determining structural information that is used as input to the melody generator which comes from the author's knowledge of Jazz. The analysis is performed by searching through a space of possible key and chord assignments for the song, which are then used to generate variations of the main melody. The grammars developed by Ulrich show the use of hierarchical structure to ensure melodic continuity across harmonic boundaries and no context dependent productions are used. The grammar based approach is carried forward by Keller and Morrison who use probabilistic grammars to tackle the improvisation problem [Keller and Morrison, 2007]. These techniques are expressible within the Bol Processor grammar engine, which also supports context sensitive production in addition to purely hierarchical productions. Probabilistic grammars and the automatic determination of rule-weights from production sets are also possible [Bel and Kippen, 1992].

2.3 Expressive synthesis and speech prosody

Expressive singing synthesis systems and prosody models in text to speech synthesizers deal with pitch articulation that has semantic or stylistic value and are therefore relevant to modeling gamakās. Here, I distinguish between expressive synthesis that deals with dynamic models of continuously controlled parameters and the systems which aim for expressive performance of, typically, baroque music through modification of pitch, volume and timing of notated events. Dynamical models in systems of the former kind deal with executing expression that is only approximately notated even in western classical music, and where different performers may choose to execute them alike. Expressive MIDI piano performance of baroque music on the other hand involves generating variations on pitch, volume and timing attributes of note events already available in sheet music or MIDI form.[7] With the latter kind of expressive synthesis, the purpose is to generate different renditions or to mimic the style of a performer, usually through statistical analysis [Kirke and Miranda, 2009]. Dynamical models of vibrato and glissando, or coloratura[8] on the other hand, aim to produce acceptable renditions of notated instructions and do not focus on generating a variety of renditions. These are therefore closer to the problem of modeling gamakās where we don't yet have clear models of their musical function, without considering expression.

The work of Schwarz on expressive concatenative synthesis techniques based on corpus analysis is well known [Schwarz, 2007,Beller et al., 2005,Schwarz et al., 2006,Schwarz et al., 2000]. However, the rule based singing synthesis system called MUSSE DIG developed by Berndtsson and others at KTH is interesting to look at from a musicological perspective, since the principles behind the synthesis are explicitly coded in their system [Berndtsson, 1996]. The MUSSE DIG system is built on RULSYS, a language and engine developed for text to speech synthesis and which contains controls for a wide variety of vocal gestures such as front articulation, back tongue body and nasal production [Berndtsson, 1995, p. 7]. Of particular interest to gamakā modeling are the rules dealing with consonant and vowel

[7]Many such expressive piano performance systems compete at the annual RenCon – a "Musical Performance Rendering Contest for Computer Systems" [Hashida et al., 2012].

[8]Term "coloratura" used as referred to in Berndtson et al's work on singing synthesis.

durations, fundamental frequency or "F0" timing and "special singing techniques" such as coloratura. The consonant and vowel durations determine perceived rhythm [Sundberg, 1994] and, according to Berndtsson, pitch changes not completed at vowel onsets "sound strange" [Berndtsson, 1995, p. 15]. Coloratura combine a vibrato-like movement with rapid pitch steps and bear resemblance to some kinds of gamakās. Berndtsson models the vibrato components of coloratura with an amplitude[9] of a semi-tone around the given discrete pitches [Berndtsson, 1995, p. 16]. A related kind of overshoot with gamakās was noted by Subramanian, though not to a full semi-tone [Subramanian, 2002].

Speech intonation models deal with the generation of the F0 contour of speech signals and are related to gamakā representation as well. The most common model used for generating F0 contours for speech is the dynamical Fujisaki model which has been applied to both speech and singing [Monaghan, 2002]. According to this model, the F0 contour is generated as the response of a second order linear system to a sequence of discrete linguistic commands [Fujisaki, 1981]. When given a step input of the kind available to the KTH system, such a second order system would generate a overshoot depending on the extent of *damping*. The "tilt intonation" model is an explicit representation developed by Taylor and Black [Taylor, 1994, Taylor and Black, 1994] and views the F0 contours of speech as a series of pitch "excursions" and describe each using an extent, a duration and a "tilt" parameter which varies from 1 (a pure fall) through 0 (a rise followed by a fall) to +1 (pure rise). Portele and Heufts "maximim-based description" uses yet another parameterization that is similar to Taylors model [Portele and Heuft, 1998]. They specify a contour by identifying F0 maxima, their times and their left and right slopes [Portele and Heuft, 1998]. The minima are implicit in this model and sinusoidal interpolation of F0 is used to generate the complete contour using this information.

As seen above, multiple explicit representations of pitch contours have been proposed in the past. This raises the question of which representation is the more "natural" and what criteria might help choose one representation over another. Taylor notes in [Taylor, 1998] that "the linguistic justification for any existing intonation systems are weak". However, the Fujisaki model can be justified on physiological grounds. It therefore appears that there is considerable leeway in choice of a representation for pitch contours, which is likely to be the case for gamakās as well.

[9] "Amplitude" is also used in this document similarly to refer to the extent of pitch deviations around a reference pitch and not, for instance, to the amplitude of an audio signal.

2.4 Approaches to gamakā modeling

Gamakās have grammatical significance in Carnatic music and do not serve only an ornamental or expressive role. This suggests that a purely dynamic model of gamakās over the course of multiple notes may not be effective. The synthesis system that renders *Gaayaka*'s textual notation is therefore justified in using simple linear pitch interpolation between explicitly specified pitches [Subramanian, 1999]. Such a pitch movement is notated in *Gaayaka* syntax using the "*jāru*" symbols '/' and '\', with symbol repetition used to elongate movements. Battey adds more detail to the movement shape by modeling the gamakās in a Hindustani singing style using Bezier splines [Battey, 2004]. Battey's model chooses a best fit curve of minimal complexity by exploiting the Just Noticeable Difference (JND) interval in pitch perception.

From a broader perspective, the interesting parts of a metric-time performance in Carnatic music[10] lie not so much in the exact shapes of movements, but more with the timing of the onset and landing of movements and the dynamical and perceptual principles that dominate rapid movements. Therefore, I surmise that any of the earlier discussed explicit models of pitch contours would be acceptable as part of an elaboration system for prescriptive notation. The exact shapes might then express some of the idiosyncrasies of a performer or the training regime and tutelage that the performer passed through.

[10].. as opposed to a free-time performance such as with "ālāpanā" or "tānam" forms.

Chapter 3

Research Problem

Contents

The previous chapter presented several approaches to modeling aspects of the problem of elaborating prescriptive notation into a performance. Hierarchical structure was approached using database lookup of permitted phrases given local melodic context, generative grammars for deeper structures, theoretical models of the constraints of the genre as a set of violable soft constraints and rule systems and dynamical models for controlling continuous pitch movements in expressive speech and singing synthesis. This chapter formulates the problem of elaborating prescriptive notation in Carnatic music as a combination of hierarchical gamakā selection and the resolution of lateral gamakā sequencing constraints. A generalized version of the elaboration problem which accounts for arbitrary gamakā selection and sequencing preferences is first presented, followed by limitations such as restriction to short phrases and tāḷā independence that were imposed on the problem to enable this research to focus on gamakās.

Two kinds of structures arise when considering the problem of elaborating prescriptive notation in Carnatic music – (i) the *hierarchical* structure that controls the choice of gamakās based on local melodic context, and (ii) *lateral* constraints that control the sequencing of gamakās over the duration of a phrase. The gamakā lookup table approach of Gaayaka is strictly hierarchical. Generative grammars and constraint-based discrete optimization techniques are capable of modeling both. However, grammars turn out to be a conceptual mismatch for modeling lateral constraints and general constraint based approaches rely on incomplete known rules and therefore need to deal with a large search space of possible renditions of a phrase. Trained musicians demonstrate an ability to adapt gamakās used in slow tempo performance for rendering similar phrases at multiple speeds. We do not have models of gamakās that account for such transformations. Furthermore, few music systems attempt structural elaboration in the presence of continuous pitch control at the degree of complexity that gamakās pose.

The main contribution of this thesis is a computational model for selecting, transforming and sequencing gamakās based on an actual performance, that accounts for both hierarchical structure and intra-phrase constraints on gamakās. The model serves as an elaboration system for prescriptively notated phrases in the rāgā of the analyzed performance. Since transcription is a first step to analyzing an actual performance, a second contribution of this research has been two numerical representations for transcription and subsequent analysis of gamakās. One of the representations goes into a level of detail beyond conventional descriptive notation, while another two-component representation provides an intermediate level of detail between prescriptive notation and descriptive notation. The latter two-component representation has been shown to result in a catalog of abstract gamakā forms by which the scope of gamakās available in the transcribed performance can be extended to new melodic contexts. The model thus serves to concretize the abstract descriptions found in musicological literature presented in the previous chapter. An example of such descriptive literature is provided in Appendix C.

Besides a procedural understanding of gamakās as abstract forms and how they are transformed and combined depending on melodic context, the ability to generate valid interpretations of prescriptive notation in Carnatic music has uses in music education and in music communication across cultural boundaries. The ability to apply abstract gamakā forms to realized phrases provides a starting point for approaching the generative component of improvisation in Carnatic music. The approach used in this research may also permit us to ask and partially answer how past masters might have interpreted a particular composition that they had never performed during their lifetimes.

I now consider a generalized version of the elaboration problem in Carnatic music followed by limitations imposed for this research. Some of the terms and notations used in this document in discussing the problem are summarized in appendix A.

3.1 The generalized elaboration problem

In this research, an elaboration system for Carnatic music is modeled by a process that takes as input a prescriptive representation of a phrase or section, choice of gamakās and a set of constraints on them and produces a ranked set of renditions of the phrase using the given gamakās with implicit transformations. In addition to suggesting possible computational techniques for implementation, this framework also provides questions that would be reasonable to ask of similar systems that attempt to synthesize a performance given prescriptive notation. Factors that are known in common practice to influence gamakās choice are first accounted for in what follows before laying out the limitations imposed in this work.

The input phrase is given as a sequence of *svaras* or "notes" each of which is a tuple consisting of a pitch class, an octave value and a duration. Each *svara* is associated with a "local melodic context" around it consisting of the preceding and following *svaras*. The gamakā preferences to be applied to the input phrase are given in the form of two cost functions. The appropriateness of a gamakā for a given context and tāḷā is expressed as a cost function called the "markedness" of the gamakā, a term borrowed from optimality theory. For example, a bias against rendering the *svaras* "sa" and "pa" (the tonic and the fifth) using gamakās can be expressed by assigning a markedness value for all gamakās except the plain note rendition. A second cost function expresses gamakā sequence preferences in conjunction with the tāḷā. This function evaluates a pair of gamakās assigned to adjacent *svaras*, taking into account their melodic contexts, assigning a cost to pairs of incompatible gamakās. In general, these selection and sequencing constraints may have arbitrary look-ahead or look-behind and can include other dependencies such as appropriateness to lyrics as well. The musical knowledge in the elaboration system is encoded in these two cost functions. Phrase elaboration is thereby reduced to selecting gamakās by minimizing the total cost over the phrase as expressed by the markedness and sequencing cost functions.

For a given rāgā, a set of known constraints may be taken from common practice as well as from the well known musicology of the rāgā. For example, the ascent and descent constraints on a rāgā limit the kinds of gamakās that can be associated with a given note. Some *svaras* are appropriate for phrase beginnings and endings and therefore the corresponding gamakās also inherit that constraint. Another well known rule that is also discussed in [Vijayakrishnan, 2007] is that the pitch range of a gamakā used to interpret a *svara* must include the pitch class corresponding to the *svara*. The tone positions S and P (tonic and the fifth respectively) are also usually held fixed in most interpretations.

3.2 Imposed limitations

In this research, some additional limitations were placed on the general elaboration problem to help focus on understanding gamakās. Gamakā dependence on tāḷā is ignored, timing information is eliminated from the note trigram context and rāgā choice is constrained to ensure that pitch classes map unambiguously to tonal positions.

Dependence of gamakās on tāḷā can be ignored as a simplifying measure and is musicologically justifiable. In the general case, the position within the tāḷā certainly influences the kind of gamakā a musician would choose for it – depending, for example, on whether a particular beat of the cycle needs emphasis.[1] However, the common lack of reference to the tāḷā in *rāga lakśaṇā* literature suggests that this simplification is musicologically valid. Such a tāḷā dependency can then be studied as an independent problem. Dependency of gamakās on lyrics would be of importance to the interpretation of "krti" category of compositions, but not for *varṇam* category which features extensive solfa sections. Musical interpretation of lyrical meaning is beyond the scope of this work.

The note context trigram $C(n_i)$ includes the note durations in the general case, but preliminary work indicated that note durations can be dropped from the context, provided gamakā transformation is permitted to accommodate duration changes. Therefore the complexity of the note context in this work was reduced from $C(n_i) = (n_{i-1}, n_i, n_{i+1})$ to $C(n_i) = (p_{i-1}, p_i, p_{i+1})$ and the problem of adapting gamakās to different durations separated from the general elaboration problem. This factorization also reduces the data requirement for solving the general problem.

There are 16 classes of *svara* per octave in use in Carnatic music, redundantly encoding 12 tonal positions. For many common rāgās, the distinction between the 16 pitch classes and the 12 tonal positions per octave is an unnecessary one – i.e. in many rāgās, a one to one mapping can be established between the pitch classes that feature in them, with the tones that they should be rendered with. This is, therefore, another simplification that is used in this work without introducing ambiguity and is reflected in the use of the *svarasthāna* notation to express both the prescriptive notation and the more detailed descriptive notation of gamakās.

[1] The "/samam/" (first beat) and "/arudhi/" (mid point), for example.

Chapter 4

Method

To model the melodic aspects of musical expertise involved in interpreting Carnatic music prescriptive notation, a system that performs phrase by phrase elaboration of prescriptive notation was implemented and interviews with expert musicians were conducted to evaluate the system's performance. The process involved selecting and transcribing a suitable reference performance, developing rules for adapting slow gamakās to higher speeds and iteratively determining the components of an optimization function for phrase-level gamakā selection by matching the system's output with the gamakā sequencing preferences exhibited in the reference performance. The system produced output in a detailed representation and used a simple sampling synthesis technique to generate sound from this representation. Two such detailed representations were used and the two resulting systems were compared in the expert evaluation. Figure 4.1 summarizes this work. This overview chapter presents the issues, choices, methods and performance criteria involved in the construction and evaluation of the phrase elaboration system, the details of which are presented in the chapters that follow.

A *vīṇā* performance of the*varṇam* "Karuṇimpa" in the rāgā Sahānā was chosen as the reference performance for study. The varnam form and presentation was chosen for the variety of musical structures contained in it including lyrical and solfa sections and sections rendered in two speeds. The rāgā Sahānā was chosen for the middle ground of complexity that it and its canonical *varṇam* occupy. My familiarity with the *vīṇā* is, on balance, an advantage for this research.

The reference performance was manually transcribed into two representations – Pitch Attack Sustain Release form (PASR) and Dual-PASR form (DPASR) which are described in section 6.2 – from which the performance could be re-synthesized. The transcrip-

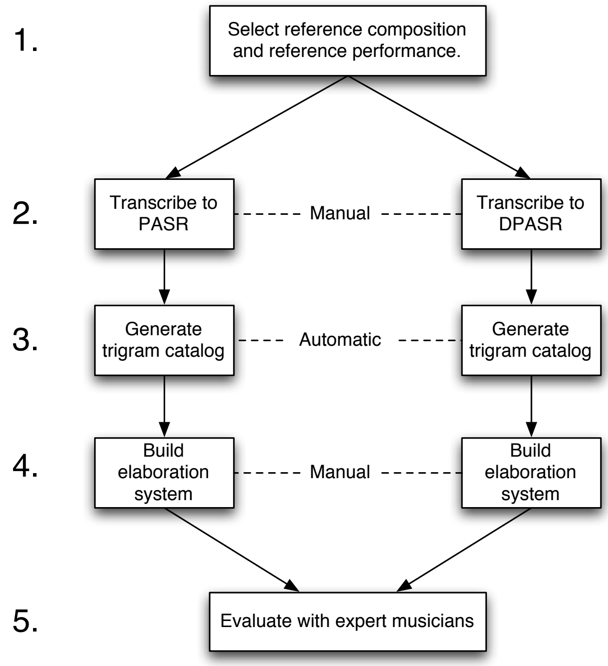


Figure 4.1: Method at a glance.

tions were verified through re-synthesis using simple instrument models. The tuning system featured in the reference performance was replaced with the equal tempered system for analytical convenience without loss of musicality and minimal quantization was applied to the timing characteristics of gamakās so that they are more amenable to being described using formal rules and preferences. The stopping and plucking techniques used in the performance were not transcribed or modeled in this work. For details of transcription, see chapter 5. Multiple pitch measurement algorithms in the Praat program [Boersma and Weenink, 2005] and the pitch preserving time stretching algorithm in Audacity [Mazzoni and Dannenberg, 2000] were used to reduce human perceptual error in the transcription process.

The gamakās identified in the transcriptions were associated with notes in the prescriptive notation and catalogued, keyed using pitch-class-trigrams to capture the local melodic context in which they feature. Though note duration is an important component of the melodic context of a gamakā, it was preferable to work with a context that didn't have the duration information because a *varṇam* would be an incomplete source of such contexts and the system would show limited generalizability to contexts in other compositions. Furthermore, a preliminary study of the two-speed renditions of sections of the selected *varṇam* yielded rules for adapting slower speed gamakās to contexts with shorter durations. This permitted the removed duration information to be restored post gamakā selection.

A phrase-optimal gamakā selection algorithm expressed as the optimal path through a Directed Acyclic Graph (DAG) was implemented to elaborate phrases using the gamakā catalog generated for each of the two representations, details of which are provided in section 6.5.4. The graph's weights were specified using a scoring function that expressed gamakā sequencing preferences. The components of this scoring function were manually determined for each representation by comparing the sequencing preferences or "discrimination" expressed by the performer in the reference performance with that of the elaboration systems and iterating until they matched.

A study with expert musicians was conducted for this research using comparative evaluation of productions of the PASR and DPASR based elaboration systems with resynthesized versions of phrases found in the reference performance. Evaluating a system which produces musical output with human listeners is a hard problem. Studies such as the listening tests reported in [Berndtsson, 1995] done to evaluate specific rules for expressive singing are rare in the field and are usually done for very narrow musical aspects. Such studies first require the system to cross a quality threshold for synthesis, and only then stand a chance if the genre of choice has a cultural common ground to support such systematic evaluation. Cope, for example, dispenses with synthesis entirely by using a human performer to play the compositions produced by his EMI program [Cope, 1987], but that comes with the methodological problem of isolating the part of the response that is associated with the performer's expressive playing. With Carnatic music, finding this common ground given its continuously diverging styles appears hard. Nevertheless, practitioners' are able to listen, appreciate and critically evaluate each other and themselves despite the divergence. With comparative evaluation, it is expected that biases brought to the evaluation of synthesized material or to the specific style of rendition in the reference performance apply equally to all the material evaluated. Differences found in the response to the various clips presented to the participants therefore likely highlight aspects of the elaboration systems alone and it would be possible to aggregate these relative scores across all the participants.

The two elaboration systems that resulted from the use of the PASR and DPASR system were evaluated for the *acceptability*, *range* and *scope* of their productions through interviews with genre experts. The interviews involved presenting two main sets of three predetermined phrases and variations produced by the two elaboration systems, interspersed with direct transcriptions of the reference performance for calibration. The variations were presented in random order. The clips were synthesized using sampled *vīṇā* sounds and gamakās were simulated by changing the playback rate of the samples. To evaluate the capability of the system to handle phrases beyond those found in the reference performance, two challenge phrases were solicited from each of the participants for which variations were generated using both elaboration systems in situ and in two speeds. Participants were asked to provide scores in the range $0 - 10$ as well as offer verbal comments on aspects of the synthetic clips played to them. The scores provided by the participants were aggregated to determine the three performance dimensions and analyzed using the verbal comments collected.

The next few chapters provide details about selecting and transcribing the reference performance, implementing the system, conducting the study, and analyzing the results.

Chapter 5

Selecting & Transcribing Reference Performance

Contents

Constructing an elaboration system for prescriptive notation requires ground data on gamakās relevant to the space of phrases to be elaborated. Towards this, a reference *vīṇā* performance of a *varṇam* in rāgā Sahānā was transcribed into two numeric representations called PASR and DPASR, each detailed enough for resynthesis to be possible using simple instrumental models. This chapter presents factors that influenced the choice of the reference performance, the techniques used in its transcription, pitch and timing normalizations applied to the transcription to aid further analytical work and aspects of the reference performance such as plucking and stopping techniques that were excluded from the transcription. A sample of these transcriptions, which form the basis for the elaboration process, is given in appendix H for reference.

5.1 Selection

The elaboration system constructed for this research is for the rāgā Sahānā and is based on a rendition of the *varṇam* "Karuṇimpa" by Smt. Rajeswari Padmanabhan, a well known *vīṇā* maestro, the details of which are given in table 5.1 and table 5.2. I now discuss the choice of this performance as the reference for this work.

A structural elaboration system that can bridge the gap between prescriptive notation of Carnatic music and performance requires two components – a set of content-indexed gamakā fragments for use in the output, and a set of rules for selecting, transforming and sequencing these fragments to form the output. The scope of the resultant system is determined by the input musical material for these stages.

The scope of this work was therefore limited to gamakās modeling based on one performance of one composition in the *varṇam* category, in one rāgā. Arriving at a set of gamakā fragments and rules that would apply in every musical context in Carnatic music is a hard problem that can only be made tractable by limiting the scope of applicability of the elaboration system to a smaller but still challenging musical context. About 500 rāgās are listed in [Pesch, 1993] and musicological literature that discusses their traits abounds in the genre. The lack of reliable automated transcription algorithms for live performance of Carnatic music and the consequent time cost placed on manual transcription necessitated that the study be limited to a single rāgā. Carnatic music features a variety of musical forms including compositional forms such as *varṇam* and *kṛti* and improvisatory forms such as *ālāpana* and *tānam*. The *varṇam* form was chosen for this study since it serves as pedagogical material for early stage students for their first introductions to several major rāgās. Though rāgā discussions among practitioners and musicologists usually happens in the context of the *ālāpana* form and *kṛti*s are known for the scope they lend to expressing rāgās, the *varṇam* form is preferable not only for its pedagogical role, but also for the variety of components that feature in it. *Varṇam*s feature lyrical sections as well as solfa sections both of which are performed in at least two speeds by convention. These sectional and speed variations provide a variety of contexts in which to study gamakā usage. Every major rāgā is associated with a canonical *varṇam*, which makes the choice of the composition simple. Furthermore, a *varṇam* is performed by an artist in a more or less consistent manner across multiple performances. However, specific musical details vary between artists even if they belong to the same *pāṇi* or "style". It is therefore reasonable to use a single *varṇam* performance as the source of gamakās and melodic contexts appropriate for the rāgā under consideration, with the understanding that further work would be involved in extending the technique to *kṛti*s and other forms.

The rāgā Sahānā occupies a middle ground of complexity between deeper gamakā-rich rāgās and scalar ones. Rāgās range in complexity from very simple to those with intricate microtonal structure. I began this research with analyzing a varṇam in the rāgā Kalyāṇi which proved to be highly challenging due to the level of detail necessary for resynthesis and modeling. The challenging nature risked obscuring what might turn out to be simple principles and therefore I chose a simpler yet idiosyncratic rāgā Sahānā. Sahānā also has a crooked scalar structure that serves to examine the amount of local melodic context necessary to capture restrictions related to such a structure. Sahānā is not claimed to be an optimal choice, but the canonical *varṇam* of Sahānā is less complex than those of other important rāgās, such as "Viribhoṇi" (rāgā Bhairavi), "Sāmi ninnē kōri" (rāgā Shankarābharaṇam), "Vanajākshi" (rāgā Kalyāṇi) and "Ērā nāpai" (rāgā Tōḍi). These rāgās are known to be "heavy weight"[1] and feature complex gamakā structures. A pilot transcription of a portion of the Kalyāṇi varṇam surfaced this complexity during the initial stages of this research. Simpler rāgās such as "Māyāmāḷavagowḷa", on the other hand, admit almost arbitrary melodic movements within the constraints of the rāgā's scale, which can make it hard to study the discrimination shown by artists in selecting gamakās for a given phrase. While Sahānā does not admit arbitrary melodic movements like the simpler rāgās due to its *vakra* or "crooked" nature, it has a distinct emotional characteristic that does not require the kind of depth of the movements necessary for the "heavy weight" rāgās in order to be perceptible to those familiar with it. This proved to be an advantage when evaluating gamakā selection for Sahānā phrases with expert musicians.

The choice of *vīṇā* as the mediating instrument is due to my own training and consequent familiarity with the instrument. My familiarity with the instrument helped greatly when transcribing the performance. It must be emphasized that the level of detail in the transcription necessary for this work is far beyond what is conventional musicological and pedagogical practice. The detail has to be high enough to permit a resynthesis of the performance that preserves the gamakās performed with high fidelity. The maximum detail found in conventional transcription is that of the descriptive notation , which is inadequate for such a resynthesis. It was also possible for me to disambiguate instrumental techniques used by the performer. My musical training also helped in decisions regarding normalization of the performance. Human performers are, for instance, never strictly metronomic in time keeping. However, it is desirable for the transcribed data to be strictly metronomic so as to not confound the study of basic gamakā rules by highlighting expressive playing that might change from one performance to another.

Familiarity with the instrument and musical training may also result in the intro-

[1] An expression in common parlance of Carnatic music which refers to musical material that is perceived to have "depth" to its tonality and the gamakās that feature in it.

duction of biases in the transcription and rule construction stages. I now discuss the tools and techniques used during the transcription phase towards reducing the biases that my own musical background may introduce and normalizations that were applied to reduce the complexity of the data for the purpose of constructing an elaboration system based on it.

Table 5.1: Details of reference performance

Composition type	*Varṇam*
Title	"Karuṇimpa"
Rāgā	Sahānā
Tāḷā	Ādi (4+2+2 beat cycle)
Composer	Tiruvottiyūr Tyāgayyar
Performed by	Smt. Rajeswari Padmanabhan on the *Vīṇā* (strings)
	Mannargudi Sri Easwaran on the Mrdangam (percussion)
Album	"Surabhi"
Performed in	Studio

5.2 Transcription

I transcribed the reference performance using manual comparison of a re-synthesis of the performance with the original, using pitch tracking and transient-preserving time stretching algorithms to clarify sections that required closer inspection.[2] The technology to eliminate or substantially reduce manual intervention in the transcription of a performance is well beyond the state of the art as of this research. Despite choosing a performance that contains only a minimal mix of instruments, initial attempts at generating a full pitch-tracker derived performance using Praat's tracking algorithms were found to be inadequate for the large scale precision transcription required for this work. On balance, the amount of human input that was necessary to compensate for the failings of pitch tracking technology in the form of octave and harmonic jumps and loss of tracking mid-tone due to slides and strumming of the side strings of the *vīṇā* and was comparable to perhaps more than a full manual transcription verified by ear. The choice of the medium of rendition of the reference

[2]An experimental multi-frequency pitch tracker using a gaussian mixture model of the power spectrum was developed for tracking short gamaka fragments and was used to determine PASR components in some difficult cases.

Table 5.2: The structure of the reference performance of *varṇam* "Karuṇimpa".

Part-1	*Pallavi*
	Anupallavi
	Muktāyisvaram
Part-1 (2x speed)	*Pallavi*
	Anupallavi
	Muktāyisvaram
Part-2 (2x speed)	*Caraṇam*
	Ciṭṭasvaram 1-4

performance requires some explanation, since a vocal rendition could have also been chosen as a reference performance. Though unaccompanied vocals may be easier for pitch tracking algorithms to deal with in some places, accurate pitch tracks are hard to obtain for fast gamakā renditions. Dynamics of the voice at the sub-gamakā level complicate the pitch tracks for such fast phrases, necessitating more frequent use of musical judgement in these cases. Other difficulties arise as well regarding the stability of the rendered pitch of a *svara*. Given that the speed of movement influences perceived pitch, a singer may not be expected to render a *svara* consistently at a fixed pitch. No known or obvious rules exist for determining whether a particular observed deviation is to be attributed to the movement or simply to instability in pitch rendition. A given deviation is in general a mixture of both. The points about pitch uncertainty and dynamics also hold for the violin. A fretted instrument such as the *vīṇā* is simpler to deal with in this regard because the pitch of a note produced at a given fret position is unlikely to vary much over the course of the performance. Any slow drift in string tuning is easy to detect and compensate for. Of fretted instruments, the *vīṇā* is a good choice for such a musicological work since the known ontology of gamakās have all been constructed with reference to renditions on the *vīṇā*.

One other important factor influencing rendition choice in this work is that using a vocal performance as the reference introduces technical difficulties in the evaluation of the resultant system. It is ideal to compare generated gamakās with the original performance by applying the same re-synthesis techniques to both. If a vocal reference rendition was chosen, this implies using a singing synthesizer, the design of which would first need to be addressed in the context of Carnatic music before such a musicological study becomes feasible. In contrast, simple techniques such as sampling synthesis, wave tables and additive synthesis are good enough to render gamakās played on a stringed instrument, placing both generated and reference gamakās on equal footing. Since the medium of rendition influences choice of gamakās in the genre, it is also problematic to transcribe using a vocal rendition and evaluate the system using resynthesis on a different medium, even if transcribing a vocal rendition were easier relative to an instrumental rendition.

The tonic and the tuning system used can vary among artists. To determine the tuning used in the reference performance, I measured the fundamental frequencies of plain tones played by the performer at various points in the performance and collected the tuning table shown in table 5.3. Though the tonic can be determined by measurement or by ear, the presence of gamakās influences the perception of the quasi-stationary pitches that constitute a melody, confounding the tuning system used. The JND[3] band for a quasi stationary pitch is known to depend on the duration of the stationary part – i.e. the sustained "tone". The perceived pitch of these "tones" also depends on the speed of the preceding and following movements as indicated by overshoots that occur during fast movements. Such overshoots have been reported in [Subramanian, 2002] as well as observed in this study. For vina performances, measuring plain notes held on frets with oscillations serves to identify the tuning system used. Since the specific tuning system of a *vīṇā* is fixed, it is orthogonal to the model construction process and can be factored out and brought back in at a later stage if deemed necessary.

Timing characteristics of gamakās can be obscured by tempo fluctuations either due to expression or drift. I compensated for these fluctuations by manually adjusting the internal time structure of gamakās where this was necessary. Tempo drift was addressed by using the symbolic duration specified in the prescriptive notation as the duration of the gamakās instead of the actual measured duration in the reference performance. Either compensation requires familiarity with the genre. The particular performance chosen for this work can be considered an "austere" or "clean" rendition of the varṇam and provides good guidance for the expected timing features of gamakās. This attribute when used in conjunction with how a phrase is rendered during repetitions helped decide which timing

[3] Just Noticeable Difference. This is the band of frequency differences within which a human ear identifies all frequencies as the same "pitch". It is a well known psychoacoustic feature.

features are important for the melody and which are “expressive” variations.

The choice of representation for gamakās influences the structure and complexity of the rule systems for working with them. For instance, modeling sequential combination of gamakās as textual concatenation complicates the modeling of gamakā sequencing constraints using generative grammars, an approach tried and abandoned early on in this research. Though the first stage transcription attempted a level of detail beyond what would be found in a descriptive notation of the performance, this was later refined into a two-component representation that added an intermediate level of detail between prescriptive notation and descriptive notation to aid modeling.

Table 5.3: Tuning table based on measurements at plain notes in the reference performance.

Pitch name	Interval	Samples	Range (Hz)	Avg tuning (Hz)	Avg tuning (cents)
da2-	-3	5	$[133.9, 134.9]$	134.3 ± 0.4	-284 ± 5
sa	0	10	$[157.6, 158.8]$	158.2 ± 0.4	0 ± 5
ri2	2	6	$[176.3, 177.9]$	177.1 ± 0.6	195 ± 6
ga3	4	10	$[197, 199.9]$	198.2 ± 1	390 ± 8
ma1	5	22	$[207.4, 212.6]$	210.7 ± 1.1	496 ± 9
pa	7	22	$[234.6, 239.1]$	237.9 ± 0.9	706 ± 6
da2	9	23	$[264.4, 267.4]$	266 ± 0.8	900 ± 5
sa+	12	7	$[316.8, 319.2]$	317.7 ± 0.7	1207 ± 4
ri2+	14	3	$[353.1, 357]$	355.3 ± 1.6	1401 ± 8

5.2.1 Verification

Each phrase transcribed was verified using A/B comparison with the corresponding snippet of the original performance. In some cases, I found that the performance speed made it difficult to describe the timing of the movement between focal pitches. In such cases, I used a factor-of-2 time stretching algorithm available in the Audacity audio editor [Mazzoni and Dannenberg, 2000] to slow down the phrase to ease transcription. I found it necessary to select the “dynamic transient sharpening” mode in order to prevent the details of the articulation from blurring into each other.

5.2.2 Normalization

In order to help further analysis, I normalized the following aspects of the performance in the transcription –

- The tuning system was factored out and replaced with the equal tempered system. This change did not significantly impact the nature of the rāgā Sahānā. Some of the controlled deviations from fret positions executed by bending the string also needed to be quantized, while leaving overshoots due to fast motion intact. These were transcribed relative to fret positions as opposed to using the tonic.

- The rhythm of the performance is not metronomic and therefore deviates from a fixed tempo slightly over the course of the performance. The reference performance was transcribed in strict local metric time, disregarding such tempo fluctuations.

- Fast gamakas in the higher speed (usually referred to among practitioners of the genre as the "second speed") were time quantized to sub-multiples of strict metric time in order to facilitate analysis. These deviations could be considered performer expression, but it is physically impossible for a performer to execute these movements precisely every time. Therefore they are more likely to be random variations in performance. The quantization did not impact the recognizable similarity of the resynthesis to the performance, although the deviations were noticeable.

- Normalization impacts the task of verifying the adequacy of the transcription of a rendered phrase. The term "smooth time" refers to musical time that moves free of a strict metric structure [Bel, 1998]. In the reference performance, some gamakās were performed in smooth time and were harder to verify than the more common metric gamakās. The smooth time gamakās were therefore compared to the original at half the normal speed using the audio editor Audacity [Mazzoni and Dannenberg, 2000].

5.2.3 Comparing re-synthesis of PASR and DPASR forms

The DPASR representation is a summation of two component movements (called "stage" and "dance") each of which is itself a pitch curve expressed in PASR form. This means that DPASR, in principle, can express a wider range of continuous movements than PASR and can therefore come closer to the performer's rendition. This difference showed up markedly in certain phrases that involve large pitch movements such as from ri2 (= D) to pa (= G) or ma1 (= F) to da2 (= A). The difference between the two is less conspicuous for small local movements.

With the DPASR representation, it is possible to use different interpolation schema for the two components. With the intention of getting as close as possible to the instrumental characteristics, I tried using an interpolation curve derived from the inverse relationship of pitch to string position for the slow "stage" component. The change in interpolation scheme was only better for the above mentioned long movements and not for shorter slides of up to 3 semitones. On the whole, the choice of interpolation scheme did not impact the transcription to such an extent that further analysis would've needed to take different directions depending on this choice.

5.2.4 Modeling *vīṇā* specific techniques

Techniques specific to the *vīṇā* such as plucking, stopping and multi-string techniques were excluded from the transcription.

In the reference performance, a brief staccato stop is applied just before the start of each pluck. These stops vary in duration and with melodic context. Modeling this stopping technique is essential to achieve a more realistic re-synthesis of the transcription since part of the purpose of this stop is to smooth transitions between notes, but this stopping technique has been ignored since it is not entirely relevant to studying phrase elaboration, though it can be significant for evaluating the resultant system with human listeners. Furthermore, it is an instrument-specific technique that is unlikely to have general implications for the genre. Nevertheless, modeling the musical role of this stylistic stopping technique is an interesting problem in itself.

The strengths of various plucks used in this performance were almost uniform. Therefore the transcriptions also do not encode information about plucking strength. Unlike the stopping technique, variation in the plucking strength, if present, may provide subtle cues to the listener about phrase boundaries, tāḷā highlights and lyrics that would be relevant had the *varṇam* been rendered using another instrument as the medium.

In *vīṇā* playing, strumming or plucking of the side strings is used for two purposes – to indicate the major divisions of the tāḷā cycle, and as filler tones during long held notes. In the chosen performance, the role of the side strings during the main sections of the *varṇam* of interest to this work were limited to indicating the tāḷā. For this reason, no attempt was made to determine the timing of these side string strums. Deviations in the timing of these strums from the tāḷā are inconsequential to parsing the performance of notated music, though they may have such a role during improvisation.

Most of the reference performance consists of melody played on only one of the main strings at a time and this was reflected in the transcription format which considers only monophonic melody. However, during some pauses, multiple main strings are strummed for effect. These strums are excluded from the transcription since they indicate sectional pauses in the composition, which information is already available in the prescriptive notation.

One *vīṇā* specific technique is a jump from the tonic (open main string) to one of the frets within a single pluck. This jump may be followed by other slides or string pulls. In the transcription, the initial jump was approximated as a slide from the tonic since that better approximates how such a gamakā would be sung. If included as is, the abrupt jump may complicate the categorization of gamakās to follow without a significant musical payoff. Modeling such expressive playing techniques is not the focus of this thesis.

The following chapter describes the implementation of the system for phrase elaboration that makes use of the musical material made available via such a transcription of the reference performance.

Chapter 6

System implementation

Contents

This chapter presents the choices made for transcription, synthesis and techniques used for phrase-level optimized gamakā selection. A simple sound model using *vīṇā* sound samples was used to verify the transcription by resynthesis. The rules for adapting gamakās to different durations and those for concatenating gamakās were isolated by studying the two-speed performances of the first half of the reference *varṇam*. This permitted the gamakā catalogs to be keyed on duration-free context information. Elaboration rules were encoded as a scoring function that evaluates the compatibility of two gamakās in sequence, which is then used for phrase-level gamakā selection by finding the optimal path through a DAG. This process is illustrated in the flow diagram shown in figure 6.1.

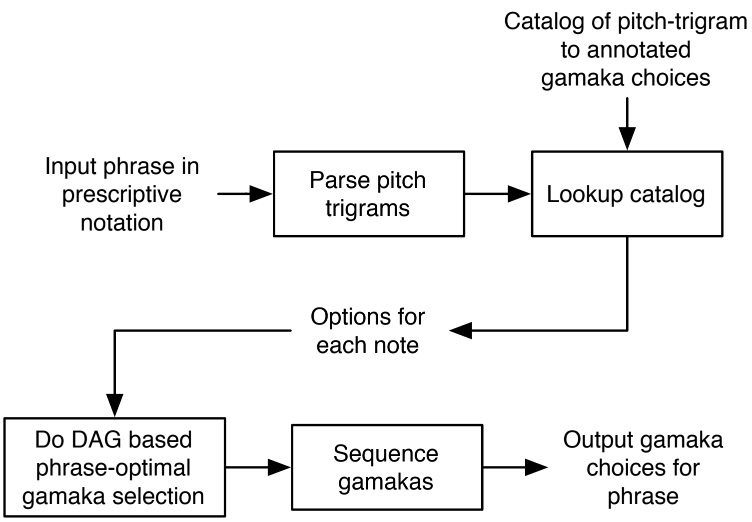


Figure 6.1: Elaborating a phrase given in prescriptive notation.

6.1 Implementation choices

Several simplifying choices were made in the construction of the elaboration system. Transcription of the reference performance was verified using a simple sound model, gamakās performed using string pulling were approximated by sinusoidal pitch bending of sampled sounds, the phrase-level optimized gamakā selection was reduced from a general combinatorial search problem to make rule construction possible to study.

The first transcription was performed using a simple instrument model for resynthesis implemented in SuperCollider [McCartney, 1996]. The synthesis system was later re-implemented to use actual *vīṇā* samples taken at various fret positions, and implementing pitch bending by changing the playback rate. The choice of instrument model for the resynthesis had negligible impact on the transcription in this case.

The artifacts of the simplistic pitch bending technique were mitigated a little by using different vina samples for each fret. Owing to the curved bridge of the *vīṇā* and the fact that the string is plucked (instead of being, say, bowed), the vibration of a *vīṇā* string is locally inharmonic - i.e. the upper partials are not in strict numeric relationship to the fundamental. Therefore change of playback rate can only maintain the string-pulling illusion over short pitch bends. For deep bends, unnatural spectral artifacts show up. By sampling each fret and performing rate changes from the nearest lower fret, these artifacts were reduced and the sound was close enough to the real instrument for evaluation purposes.

The acceptability of the sound was also checked with the experts who were interviewed to evaluate the system who unanimously felt it was good enough for them to focus on its musical aspects.

The transcription is represented as a data structure stored as text in the well known JSON format.[1] The transcription was split into a "metadata" file which provides information about the prescriptive notation as performed in the recording and a file containing the gamakā details of the phrases in both PASR and DPASR forms. One simplification done in the transcription is the use of relative durations in PASR data. The actual duration of a particular gamakā is obtained by scaling the total PASR duration values to the duration of the note specified in the prescriptive notation. For the DPASR representation, the durations of the "stage" and "dance" components of a DPASR movement must be identical and therefore they were individually auto-scaled to the duration indicated in the prescriptive notation.

In order to prepare the input for the elaboration system, a separate module was used to load the raw data and perform the annotations necessary for elaboration, including cataloguing of gamakās based on duration-free pitch trigrams and classification of dance components of gamakā fragments into the discrete categories "transient", "normal" and "sustained". This modularity helped gain clarity about the construction of the DPASR representation.

The rule system for evaluating the intra-phrase consistency of gamakās was implemented using the well known "shortest path" algorithm for directed acyclic graphs. The nodes of the graph in this case are the various possible choices of gamakās for each note in the prescription. The edges of the graph only occur between the choices for adjacent nodes and the edge weights are provided by an evaluation function that encodes the preference rules. The shortest path algorithm then selects the optimal choice of gamakās over the length of a phrase. Multiple possible elaborations were generated in the order of the indicated preference.

The final implementation of the system was done using the Javascript programming language, though Prolog was also used during a prior stage to explore techniques for encoding elaboration rules.

[1] http://www.json.org

6.2 Transcription

Transcribing a performance involves committing to a model of the musical features displayed in the performance. I chose two simple numeric representations that allowed easy re-synthesis, in place of detailed discrete representations such as "descriptive notation", so that the transcription can be verified by comparing the re-synthesized version with the performance.

6.2.1 PASR form

The first representation I chose is a numeric representation that is close in structure to Viswanathan's descriptive notation [Viswanathan, 1977], but permits greater detail about the timing and movement between pitches to be represented. This "PASR form" consists of describing the time spent at and moving towards and away from "focal pitches". Focal pitches correspond to the note labels used in descriptive notation, with the addition that they are numerical and can deviate from the twelve-tone system to represent microtonal variations. The key characteristic of focal pitches are that they are quasi-stationary points within a gamakā. The recorded features result in a representation of a gamakā as a sequence of focal pitches, with four numbers describing each focal pitch - its "Pitch number" which gives the equal tempered semitone value of the pitch relative to the tonic, "Attack time" which is the time taken in moving toward the focal pitch, "Sustain time" which is the time spent stationary at the focal pitch and "Release time" which is the time spent moving away from the focal pitch. Henceforth, this representation is called the "PASR form" of a gamaka.

The PASR form is one of the simplest possible representations for gamakas. It can represent most movements in the chosen performance with enough fidelity to permit further musical analysis. The previous chapter presented various pitch curve models as used in the domain of speech prosody modeling. In choosing PASR over those, I made use of the representational freedom that exists for pitch curve modeling, where there aren't clear criteria that make any one representation objectively better than another. The correspondence between focal pitches and the svarasthanas used in descriptive notation is an advantage for musicological study, making the PASR form convenient for the purposes of this research.

It is worth noting that the PASR form is not a general representation that can be used for arbitrary curve fitting tasks. This is because it places an important musical constraint on the focal pitches – that they be quasi-stationary. Though this constraint is appropriate for Carnatic music, it may not be applicable in general. For example, in certain long movements called “meends” in Hindustani classical music, a movement may slow down slightly on intermediate svaras instead of lingering long enough for the quasi-stationary condition to hold. Extending the PASR representation to account for such movements is, however, straight forward. So, for the purpose of this work, I did not consider modeling such “meend”-like movements.

The transcription into the PASR form was synthesized using two types of interpolation - sinusoidal and skew-sine which are shown in table 6.1. While sinusoidal interpolation is straight forward, it cannot model sharp rises from one focal pitch with a soft landing on the following focal pitch. The skew-sine interpolation scheme, shown in figure 6.2, was intended to handle such movements and was constructed as a two-sinusoidal fragments which are continuous and differentiable at their join point, which is known from the PASR components. The skew-sine shape influences the perceived speed of movement around the end points without sacrificing the quasi-stationary property of focal pitches. Other curves are possible which influence the speed more dramatically than the skew-sine does. However, if JND is taken into account for dynamic pitch changes, these are indistinguishable from a PASR model with longer sustains on the respective focal pitches. Furthermore, it is desirable simplification to have the speed of a gamakā movement to be computable from the simple sum of release and attack durations without taking the shape’s skew into account. The skew-sine was tried since it satisfies these properties while also providing some variety in the shape of a movement.

As expected, the skew-sine interpolation scheme captured the relationships of movements to the underlying rhythmic pulse better than the pure sine interpolation upon resynthesis. However, once the transcription was completed in PASR form, the specific shape of the interpolation curve proved to be relatively insignificant compared to the effect of the rules for selecting and sequencing gamakās. The table 6.2 shows basic statistics about the composition transcribed. It was possible to perform much of the transcription by comparing a re-synthesis to the original performance. However, some microtonal focal pitches proved to be difficult to transcribe and I took the aid of the automatic pitch extraction algorithm available in the Praat program [Boersma and Weenink, 2005] to determine the exact pitch values. The algorithm used was “sub-harmonic summation”. The “auto-correlation” pitch tracker proved to be less stable overall, but was used to measure steady pitches for which it worked well.

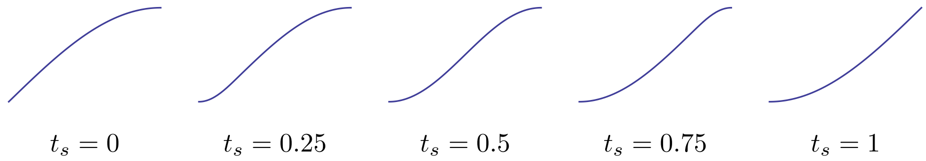


Figure 6.2: Skew-sine interpolation shapes for various skew points t_s. See table 6.1 for the formula for computing skew-sine shapes.

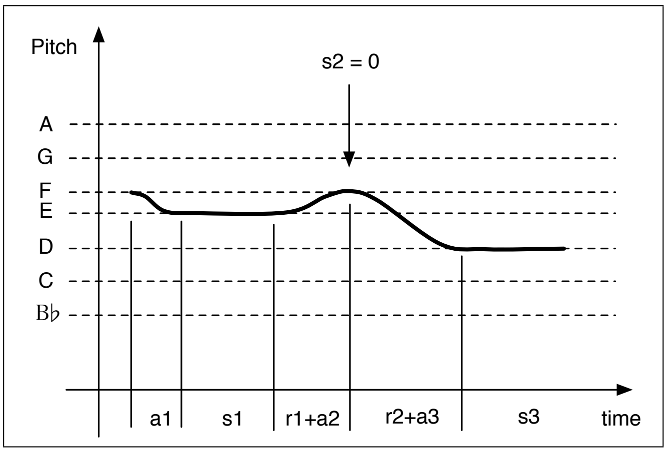


Figure 6.3: Concatenating gamaka fragments FEF and EFD of phrase $FEFD$ fuses their "attack" and "release" intervals using sinusoidal interpolation. This phrase would be expressed as $ri2$ in prescriptive notation, which is a pitch class that corresponds to D.

Table 6.1: Interpolation formulae for re-synthesis of a gamakā from its PASR representation.

Sine	$\sigma(t) = (1/2)\,(1 + \sin(\pi\,(t - 0.5)))$ where $t \in [0, 1]$
Skewed sine	$\sigma_s(t, t_s) = \begin{cases} 2t_s\sigma(t/(2t_s)) & \text{if } t \leq t_s \\ 1 - 2(1 - t_s)\sigma((1 - t)/(2(1 - t_s))) & \text{if } t > t_s \end{cases}$
Slide from semitone pitch p_1 to p_2	$\lambda(t) = p_1 - 12\log_2(1 - (1 - 2^{(p_1 - p_2)/12})\sigma(t))$

6.2.2 Dual-PASR form

The Dual-PASR form separates a unified movement representation into two components that I call the "stage" and the "dance" . The stage component captures movement at a slower pace than the inflections that comprise a gamakā and the dance component abstracts these inflections from the reference pitch from which they are performed. I call this two-component representation the "Dual-PASR" or "DPASR" form since the stage and dance components themselves are represented in PASR form.

The stage component represents movement at a level of detail closer to the prescriptive notation but contains subtle timing information necessary to connect adjacent tones. It can be thought of as the dynamic "melodic centre" of a gamakā and is itself represented in PASR form. The dance component represents movement relative to the stage, with the total movement being the given by (6.1).

$$\text{gamaka}(t) = \text{stage}(t) + \text{dance}(t) \tag{6.1}$$

The dance component is also represented in PASR form. Figure 6.4 illustrates a gamakā being decomposed into such components.

To identify the component movements of a gamakā, I first observed that the notes provided in prescriptive notation suggest a base pitch around which a gamakā is to be performed. This feature is exploited by the symbolic annotations of Dikshitar and others. However, the base pitch used can be different from the notated pitch class and depends on other factors such as the local melodic context and local continuity. Furthermore, one of the notes in prescriptive notation might correspond to two such base or "stage" pitches connected by a movement.

Table 6.2: Transcription statistics for the section of the analyzed performance of "Karuṇimpa" which occurs in two speeds.

Trait	**Speed**	**Value(s)**
Prescriptive notes	Both	296
Plucks	1x	189
	2x	100
Focal pitches	1x	626
	2x	303
Unique pitch triads	1x	56
	2x	43
Pluck duration	1x	424-1697 ms, median = 848 ms
	2x	212-1060 ms, median = 212 ms
Gamaka duration	1x	53-1697 ms, median = 178 ms
	2x	25-848 ms, median = 107 ms

I used the two-dimensional control surface of the *vīṇā* as a guide to determine these base pitches. On the *vīṇā*, a pitch movement can be performed either by sliding over the frets or by pulling on the string. The separation thus visible in a *vīṇā* performance of a gamakā is, however, not absolute and depends on the performer's style and technical facility. Also, pitch inflections can only be performed on a single fret by pulling on the string, which would constrain stage pitches to be always be the lower of the set of focal pitches involved in a gamakā. Therefore some degree of musical judgement is involved in identifying what these stage pitches are and the timing of the movements between them. The stage movement is itself represented in PASR form.

Having identified the stage pitches, the residual movement relative to the stage pitch can itself be expressed in PASR form, which I call the *dance*. On the *vīṇā*, the technique of pulling on the string to produce pitch deflections serves as an approximate guide for determining this residual *dance*. In addition to string pulls, some gamakās performed using left-hand split finger techniques, as in the movement

$$\overline{\mathrm{PmGm}}, , , ,$$

, are also included in the *dance* component. The complete gamakā is the sum of the pitch curves corresponding to these two component PASR forms. Figure 6.4 illustrates an example of such a decomposition.

The separation of a gamakā into its *stage* and *dance* components permits different interpolation techniques to be applied to the focal pitches featured in the two components. This was exploited to bring the resynthesis closer to the sound of the *vīṇā* by interpolating the *stage* component using semitone steps to simulate sliding along the fretboard of the *vīṇā*. The extra fidelity that resulted from this helped fine tune the two components in the manual transcription phase.

6.2.2.1 Refactoring gamakā amplitudes

With the straight forward decomposition of a gamakā done thus far, rāgā-specific pitch information is still split between the focal pitches of the stage component and the amplitudes of the deviations from zero-pitch in the dance components. These deviations in the dance component occur associated with the sustain portions of stage pitches and therefore the stage pitch values can be augmented with the amplitudes of the associated deviations. Each stage focal pitch f, is associated with a sequence of amplitude values α_i. With the amplitude values thus factored out of the dance component, only three values for the amplitudes remain

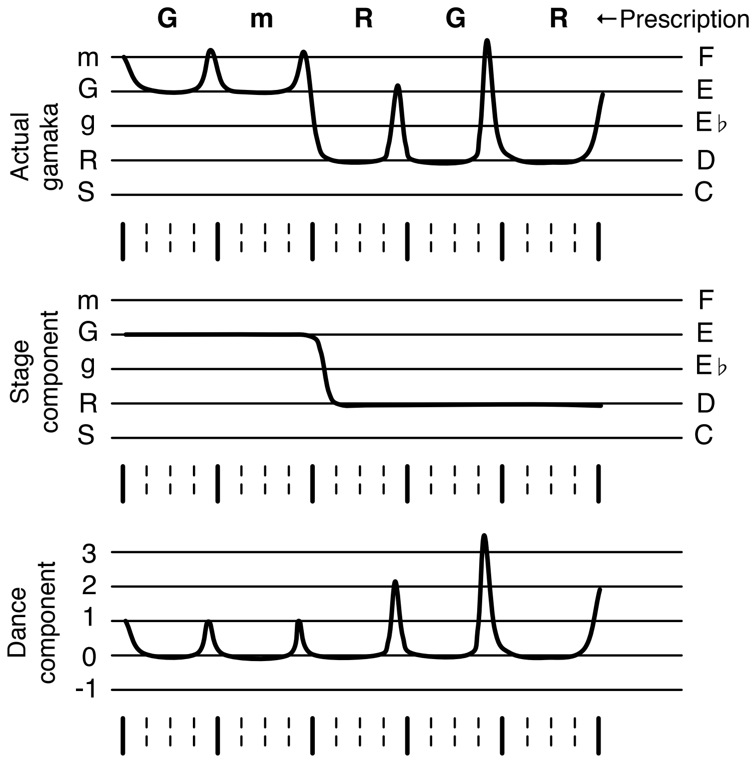


Figure 6.4: Example of decomposing a gamakā into "stage" and "dance" components.

– +1, 0 and −1 – which we denote using $^\wedge$, − and $_\vee$ for clarity.

To give an example, a gamakā for the note "m" in the local context of the movement "P,m," has the following stage-dance decomposition –

stage: $[(7, 0, 15, 0), (4, 15, 30, 0)]$

dance: $[(0, 0, 38, 7), (3, 15, 0, 0)]$

The tuples are (p, a, s, r) values where the pitch p is in units of semitones with reference to the tonic and the attack a, sustain s and release r durations are given in normalized units where the duration of the whole movement is taken to be 60 units.

The amplitude information can now be moved into the stage representation, leaving the dance component's movement normalized.

stage: $[([7, \alpha_1 = 0], 0, 15, 0), ([4, \alpha_1 = 3], 15, 30, 0)]$

dance: $[(-, 0, 38, 7), (^\wedge, 15, 0, 0)]$

This transformation resulted in a large reduction in the complexity of capturing the pitch movements constituting a gamakā. Out of 787 instances, 48.9% of the stage focal pitches were held constant, 47% featured a unique amplitude value associated with them and 4.1% featured two distinct associated amplitudes, irrespective of the number of oscillations in the corresponding dance components. Therefore most of the of stage components could be assigned a single amplitude value for the associated dance movement. In these cases, it was straightforward to assign real amplitudes to dance components when given the unique amplitudes associated with stage focal pitches.

6.2.2.2 Categorizing focal pitch shapes

Focal pitches in the "dance" component of the DPASR representation were found to fall into three categories depending on the metric shown in equation 6.2 that captures the common shapes found in the performance. In equation 6.2, $\mathbf{f} = (f_p, f_a, f_s, f_r)$ is the full PASR tuple for the focal pitch and the f_a, f_s and f_r are its attack, sustain and release durations respectively. This formula was chosen such that $\mu = -1$ corresponds to no sustain time being spent at the focal pitch and $\mu = +1$ corresponds to a pure sustained tone. The histogram of the shape parameter $\mu(\mathbf{f})$ with signed logarithmic compression applied to it is shown in figure 6.5.

$$\mu(\mathbf{f}) = \frac{f_s - (f_a + f_r)}{f_s + f_a + f_r} \tag{6.2}$$

The dance focal pitches could therefore be further simplified by classifying them into the following three categories –

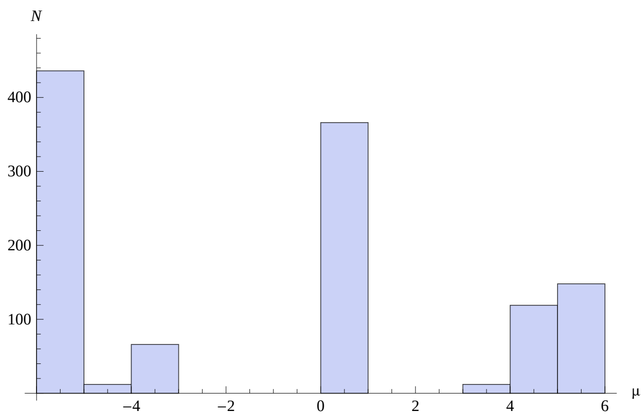


Figure 6.5: Histogram of dance component shapes. The x-axis shows $\mu(\mathbf{f})$ values with signed logarithmic compression applied.

Transient focal pitches (TFP) In the case of focal pitches with strongly negative μ value, much of the time is spent moving towards or away from the pitch. These focal pitches can therefore be labelled "transient".

Normal focal pitches (NFP) Normal focal pitches have some sustain duration in addition to time spent moving between focal pitches.

Sustained focal pitches (SFP) These focal pitches have strongly positive values for μ, which means that most of the time is spent at the focal pitch itself, with relatively little time spent reaching or moving away from it. These play an important role in the adaptation of a gamakā to a given duration, since they can be arbitrarily extended in time.

6.2.2.3 Choosing a reduced stage-dance representation

Multiple reductions based on the observations of the previous sections are possible and need to be resolved in order to proceed with further modeling. Wiggins et al. have expressed that multi-viewpoint representations "can be vital" for music and have proposed a qualitative assessment of representations based on the two axes of expressive completeness and structural generality [Wiggins et al., 1993]. The choice of representation, however, usually precedes model construction and is either based on suitability for a purpose, or is the result of pre-commitment to specific paradigms including symbolic paradigms such as note-based representations and grammars, and signal based paradigms such as the audio spectrum and its derivatives. This section presents the choices available for gamakā representation based on the simplifications described in the previous sections, identifies candidates and justifies the representation selected using a simple heuristic based on entropy estimates.

Three possible simplifications for the stage and dance components can be derived by – a) omitting either component entirely, b) forming a "minimal" reduction that omits all timing information and movement amplitudes (denoted by suffix M), and c) forming an ideal "reduced" representation that preserves all the discrete categories described in the preceding sections (denoted by suffix R). Stage_M consists of only the stage focal pitch values, whereas Stage_R includes the amplitudes of dance movements associated with these focal pitches as described in section 6.2.2.1. Dance_M similarly consists of only the dance movement directions $^\wedge/-/_\vee$, while Dance_R includes the discrete categories of section 6.2.2.2 as well. Each of these possibilities may include or exclude duration information taken from the prescriptive notation. Therefore *duration-free* and *duration-sensitive* context variations exist for each of these three simplifications.

A notion of "residual uncertainty" that measures the work that post-processing steps will need to do was used to select a simplified gamakā representation from among several possibilities. This residual uncertainty is the entropy of the possible gamakās conditional on the choice of the discrete representation . For a given local melodic context L, say the number of candidate gamakā expansions is N_L. The information required to select one of them (which is equivalent to "entropy") in the absence of any other information is then given by $\log_2 N_L$. If we choose a simplified gamakā representation R that has more information than available with L, then the remaining ambiguity is measured by how many gamakās are possible given an elaboration in terms of R. If L can be expanded into k variations in the simplified gamakā representation R, each of which have g_i (with $i \in [1, k]$) possible gamakās, then the information required to complete gamakā selection when the simplified gamakā representation i has already been selected is $\log_2(g_i)$. The mean such information required for context L is then estimated as $\sum_{i=1..k} (g_i / \sum g_i) \log_2 g_i$. Note that if the choice of representation uniquely determines gamakās, then all $g_i = 1$ and the residual uncertainty value would be 0. At the other end, if $k = 1$, then no information has been added by the representation and the uncertainty remains $\log_2 N_L$ bits. To compare two simplified gamakā representations, the mean of this residual information over all available local melodic contexts was considered. If the two representations have comparable residual uncertainties, then the simpler of the two representations was preferred. In performing this comparison, it was also important to track the maximum residual information presented for a local melodic context, which indicates the worst case performance of the representation choice.

Table 6.3 presents these entropy estimates in units of bits-per-prescribed-note and highlights in bold those options that balance generality of representation with minimizing the residual uncertainty. The smaller these bit values, the smaller the gap remaining to be bridged in order to match the original performance. The larger these bit values, the more information needed to elevate the specification of a gamakā to the detail adequate for resynthesis. The table lists both mean values and maximum values in order to keep in view the impact of the representation choice in the average case as well as the worst case. The relatively high values of the worst case residual uncertainty across the board indicates the cases for which gamakā post processing needs to do the most work. For this performance, the number of such worst case scenarios is small enough for these discrete representations to be useful. The options for our model therefore are –

1. Dance_R is determined from *duration-sensitive* local melodic context,
2. $\text{Stage}_M+\text{Dance}_R$ is determined from *duration-free* local melodic context, and
3. $\text{Stage}_R+\text{Dance}_M$ is determined from *duration-sensitive* local melodic context.

It is interesting to note that the residual uncertainty of the *duration-free* option is comparable to those that consider note durations. Choosing the *duration-free* representation would enable gamakās to be transformed for different temporal contexts. However, the simplest approach to determining Stage_M for a context is through a lookup table. To save additional steps in rendering a gamakā, the Stage_R representation can be directly selected instead through the lookup table. Henceforth, the R suffix may be dropped.

6.3 Speed doubling

The use of the duration-free pitch class trigram as the local melodic context in cataloguing gamakās for elaboration is contingent on the existence of techniques for transforming gamakās between different speeds. In other words, the timing information removed from the context needs to be inserted back into the system by other means. By studying how the double speed performance of the varnam was related to the normal speed, I worked out the following rules that enabled the slower speed gamakās to be adapted to the higher speeds. The main techniques of gamakā adaptation for this purpose are limiting the speed of movements permitted, aligning the onsets of gamakās to sub-pulses, determining which focal pitches to preserve and which to drop based on the speed limit constraint, maintaining

	Stage_R	Stage_M	Stage_{omit}
Dance_R	$\frac{0.43(3.81)}{0.31(3.58)}$	$\frac{\mathbf{0.47(3.81)}}{0.35(3.58)}$	$\frac{0.54(4.46)}{\mathbf{0.41(3.7)}}$
Dance_M	$\frac{0.84(3.81)}{\mathbf{0.57(3.58)}}$	$\frac{0.95(3.81)}{0.65(3.58)}$	$\frac{1.07(4.46)}{0.72(3.7)}$
Dance_{omit}	$\frac{1.16(4.58)}{0.71(3.58)}$	$\frac{1.66(4.86)}{0.99(4)}$	$\frac{2.03(4.95)}{1.21(4.17)}$

1. Numbers are estimates of residual entropy in bits-per-prescribed-note given in "mean(maximum)" form.
2. The "Stage_R+Dance_R" box is, for example, read as follows – "if the Stage_R+Dance_R representation can be determined given local melodic context, the remaining mean(max) uncertainty (in bits) is 0.43(3.81) if the context is *duration-free*, and 0.31(3.58) if the context is *duration-sensitive*."
3. The numbers in bold indicate choices of representation that minimize the information content in the representation while remaining effective compared to those representations with smaller residual entropy.

Table 6.3: Conditional entropy of stage and dance components given their reduced versions and local melodic contexts known from prescriptive notation.

oscillatory continuity between consecutive gamakās computed, and performing microtonal adjustments of pitch values of transient focal pitches that feature in higher speed gamakās. These rules were published in a paper titled "Modeling speed doubling in Carnatic music" at the ICMC 2011 [Subramanian et al., 2011] and this chapter details that work.

6.3.1 Movement speed limit

In the reference performance of "Karuṇimpa", the speed of continuous movement between two pitches had an upper limit of about 100ms per tone. String pulls and fret slides were treated in the same way since there is no such distinction in the vocal tradition that the genre is based on. Movements occuring in the second speed hover around this "speed limit" and therefore display a constant speed effect where more time is taken for deeper movements than for shallower movements. Pitch intervals larger than a tone take proportionately longer to span. The focal pitch preservation and dropping rules come into effect when this speed limit is reached for a movement in the first step of simple speed doubling.

Table 6.4: Summary of transformation rules for speed doubling [Subramanian et al., 2011].

Type	Description
Speed limit for gamakās	The minimum time over which a movement spanning a semi-tone may be executed was set to 50 ms.
Onset alignment of gamakās	Alignment of either the beginning or the ending of a higher speed gamakā to sub-pulses. Long range movements are aligned using their landing points and shorter movements are aligned using their starting points.
Focal pitch preservation and dropping	Reduction, due to time limits, of gamakā complexity in higher speed by pulse aligning the focal pitches and using a prioritized simplification procedure. Sustained focal pitches are preserved and transient focal pitches not conforming to the prescriptive notation are dropped in higher speeds.
Oscillatory continuity	For preserving continuous rhythmic movements in higher speed renditions. Two consecutive gamakās featuring oscillating pulse aligned movements are edited so as to extend the oscillation.
Microtonal adjustments	Adjustment of focal pitch tonal positions for transient focal pitches involved in deep movements, done for perceptual reasons.

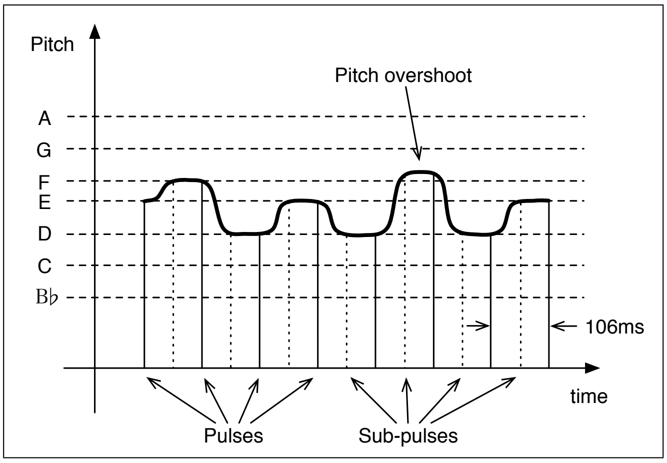


Figure 6.6: Alignment of movement onsets to pulses and landing points to sub-pulses in the gamakā $EFDEDFDE$. The prescriptive notation of this movement is D, ED.

6.3.2 Onset alignment of gamakās

The movement between two pitches were found to follow two types of pulse alignment in the slower speed - a) the onset of the movement aligns with a pulse and b) the landing point of the movement aligns with a pulse. The former dominated quicker intra-note movements and the latter occurred in slow fret slides.

In the second speed rendition, the dominant alignment is of the first kind. Therefore the transformer directly uses this information and aligns the onset of all gamakas on 1/8 count boundaries. To be precise, the onset of each gamaka fragment aligns with a 1/8 pulse and ends on the immediately following 1/16 pulse, as illustrated in figure 6.6.

A special case occurs when two notes of durations 1 count and 2 counts occur in sequence in the first speed performance. The performer, on such occasions, may choose to symmetrize it by phrasing them both to be 1.5 counts long in the first speed. Such phrases were realigned to the 1+2 pattern before transforming for the second speed.

6.3.3 Focal pitch preservation and dropping

For the purpose of this section, a gamakā is seen as a sequence of focal pitches - for example $FEFDF$. Gamakā complexity is reduced by dropping certain focal pitches of a phrase to accommodate others that need to be preserved. The following rules were found to be adequate for this purpose. A pre-processing step for these rules is the removal of extra plucks in the slower speed. A pluck is considered extra if it features in the middle of a syllable of the lyrics. Extra plucks are inserted by *vīṇā* artists for audibility of long notes since the sound of the vibrating string decays over time.

6.3.3.1 Pulse assignment

Assign each focal pitch to an integer number of pulses. The sustain part of a focal pitch is to begin on a 1/16 sub-pulse and end on a 1/8 pulse, except if the focal pitch occurs at the start of a pluck, in which case the sustain part also starts on a 1/8 pulse. Movement is to last for half a pulse, unless overridden by the "speed limit" rule for large intervals. If more time is available, distribute pulses to the focal pitches which have longer sustain times in the slow speed gamakā. If less time is available, apply one of the dropping rules and try again. One way to understand this transformation is by analogy to text to speech synthesis systems which time stretch vowels while preserving consonants. Focal pitches with relatively long sustains (within a pluck) seem analogous to vowels.

6.3.3.2 Stress preservation

For focal pitches articulated with a pluck on the *vīṇā*, the previous movement's ending focal pitch needs to be preserved in any transformation. One reason why this works is perhaps because a pluck on a focal pitch acts as a stress marker on it, and dropping the preceding focal pitch may result in the stress being altered considerably. A more sophisticated approach would be to model stress directly, but this simple rule was adequate to cover the ground for this performance.

6.3.3.3 Accommodation

To accommodate the focal pitches that need to be preserved, some transient and non-salient focal pitches need to be dropped due to the non-availability of pulses during pulse assignment.

1. The first focal pitch of a pluck in the slower speed is dropped in the double speed rendition if it is a moving focal pitch - i.e. if it has zero sustain.

2. The first focal pitch of a pluck in the slower speed is also dropped in the double speed rendition if it has the same pitch value as the ending focal pitch of the preceding pluck. This pluck is then a "continuity pluck". Note that this rule applies even if the starting focal pitch has a non-zero sustain duration.

3. If a prescribed pitch is assigned two focal pitches in the slow speed rendition and the time scaled movement is too fast in 2x speed, then the two focal pitches can be replaced with an stationary focal pitch (attack = release = 0) that is the same as the prescribed pitch.

4. An oscillatory pattern $xyxyxy$ can be reduced to $xyxy$ in the double speed version if not enough pulses are available to accommodate all the focal pitches and if it occurs in the middle of a gamaka.

6.3.4 Oscillatory continuity

When two successive notes in the second speed are such that at least one of them features an oscillatory gamaka and the adjacent note also has a movement, then additional movements continuing from the oscillation are added to the adjacent note in the second speed rendition, creating a feeling of continuity between them.

For example, the connected movement $\underline{DED}EF$ in the slower speed, where the DED is of the same duration as the E and F, is transformed into $DEDFEF$ where the extra oscillation DFE has been added.

6.3.5 Microtonal adjustments

In addition to the above rules, microtonal adjustments to the focal pitch values of some movements performed by deflecting the string were necessary for perceptual reasons. In these cases, without an overshoot, the target focal pitch sounds flatter than it actually is. This observation is consistent with vibrato studies which indicate that the perceived frequency of a note with vibrato is an average of the extreme frequencies [Prame, 1994, Horii, 1989]. The occurrence of such overshoots in Carnatic music has been studied by Subramanian [Subramanian, 2002] and Krishnaswamy [Krishnaswamy, 2003]. Subramanian also suggests that the intended pitch be approximated by a sliding window average. Figure 6.6 also illustrates one such overshoot occurring on the second F of the gamakā $EFDE\underline{DFD}E$ which occurs in the middle of the deep oscillation DFD.

Apart from perception, another reason for such overshoots could be the difficulty of precisely reaching pitches in fast oscillatory phrases using string pulling on the *vīṇā*. These two factors didn't need to be separated for this work because the overshoots are perceptually resilient to small variations ($\sim \pm 10\%$) when evaluated in the context of a phrases several seconds long. Therefore the effect of the skill dependent physical precision constraint is not significant for the purpose of resynthesis.

These findings were incorporated into the following rules -

1. Only overshoots occur, no "undershoots". It is likely that this is a consequence of the use of the *vīṇā* in the performance. The *vīṇā* being a fretted stringed instrument, it is only possible to increase the pitch by pulling on the string from a particular fret. In other performance modes such as singing or violin playing, undershoots could occur.

2. Only focal pitches with sustains of 1/16 of a count - i.e. of the duration of a sub-pulse - are given non-zero overshoots. Those with sustains of 1/8 or longer are not assigned any overshoots.

3. A "depth" is assigned to an oscillation of the form xyz, where y is the highest pitch of the three, that is equal to one less than the number of semitones of the smaller of the two intervals xy and yz. [2] For all other types of xyz movements, the depth of y is set to zero.

$$\text{depth}(xyz) = \max(0, \min(3, y - x, y - z) - 1) \tag{6.3}$$

4. Applied overshoot = depth × 25 cents.

The above rules were adequate for most of the overshoots found. An unavoidable ambiguity arose with one phrase whose slower speed rendition was transcribed with an overshoot of 80 cents. The phrase is $GAGAG$ and its execution is closer to $GB\flat GB\flat G$. This deep overshoot, however, disappears in the double speed rendition where the depth rule accounts for the performance. The strangeness of the slower speed rendition could be because the performer spends more time on the first and last G in the phrase, causing the movements in the middle to be, ironically, faster than in the pulse aligned double speed rendition. Though this suggests that the overshoot depends on the slope, the above interval rule was adequate to generate a comparable double speed performance.

6.4 Focal pitch adaptation rules

Section 6.2.2.2 reduced the variety of focal pitch shapes to three categories labelled "transient", "normal" and "sustained". These categories simplify the rules for adapting gamakās to different durations, as given below -

[2] Due to the way we've defined "focal pitch", two consecutive focal pitches within a single *gamaka* cannot be the same.

- The given note duration is divided into a number of pulses according to the timing structure of the composition. Usually this involves dividing a tala count into 4 pulses and each in turn into 4 sub-pulses.
- The sub-pulses are allocated to the various focal pitches of the gamakā, with preference to the Sustained Focal Pitchs (SFPs) and Normal Focal Pitchs (NFPs).
- If the duration of the note is longer than needed for the gamakā, and the gamakā contains only one SFP, then duration extension by repetition is preferred over time stretching.
- If the duration of the note is shorter than needed for the gamakā, the gamakā is replaced by a flat tone consisting of the last SFP and the note allocation is re-run. The note, in this case, is preferred to be held plain.
- Transient Focal Pitch (TFP) values can be inserted or removed from the ends of gamakās depending on continuity with their neighbours. Abstract gamakā forms are described only in terms of movement direction descriptors 0, 1 or -1 for flat, upward deflections (toward higher pitch) and downward deflections (towards lower pitch). This simplification and the rule that two consecutive focal pitches with the same pitch values can be merged, results in a set of variations that can be used to adapt a gamakā to different note durations.

6.5 Rule derivation

The preferences exhibited by the performer in the reference performance were constructed by iteratively matching a discrimination measure calculated from a structured representation of the performance and its prescriptive notation with that shown by the elaboration system. The complete set of procedures involves rules for gamakā selection, sequencing and smoothing. Gamakā selection involves enumerating the choices available for each local melodic context in the given phrase. Gamakā sequencing is where a set of gamakās are chosen for the input phrase optimized according to a set of local preferences expressed as a scoring function for pairs of gamakās. Smoothing refers to a simple step whereby the boundaries between gamakās are made compatible in a rāgā independent manner using the concatenative properties of the PASR representation.

6.5.1 Structuring the representation

The manual derivation of rules for elaboration for step 4 of figure 4.1 requires a structured representation of the transcribed composition which captures all the contextual information necessary for the task. The prescriptive notation of "Karuṇimpa" shows the composition partitioned into sections labelled "*pallavi*", "*anupallavi*", "*muktāyisvaram*", "*caraṇam*" and many "*ciṭṭasvaram*"s (see appendix B). These sections are further divided into phrases indicated by hyphens in the published notation. The performer often indicates these phrase boundaries with a pluck, but plucks are also used to accent the notes corresponding to syllables of the lyrics. Continuity plucks were also used to offset the decaying vibrations of the plucked string. I captured both phrase boundaries and plucks independently in the transcription. Extracts from the transcription are shown in appendix H.

For the *muktāyisvaram* and *ciṭṭasvaram* solfa sections, plucks occur on every note given in the prescriptive notation since the note names (solfege) serve as the lyrics in a sung performance of the composition.

6.5.2 Selecting gamakās for local melodic contexts

The first step of the elaboration process is selecting a number of gamakās as choices for each note specified in the prescriptive notation. No special rules are necessary to perform this step for the cases where all the local melodic contexts that feature in the input phrase are readily available in the reference performance and an enumeration of all the gamakās corresponding to all direct matches in the reference performance's transcription will suffice for such cases. Though a varṇam contains pitch triads important and characteristic of a rāgā, it cannot be expected to be exhaustive. For example, the varṇam used for this study contains about one third of the triads possible with Sahānā. For input phrases featuring contexts for which a direct match cannot be found in the reference performance, the following matching preference order expressed as a penalty score in the range $[0 - 1]$ was calculated for each of the contexts featuring in the input phrase as follows –

1. If a note in the input prescription cannot match any of the notes found in the reference performance even after considering octave differences, the input prescription is declared invalid and the elaboration process is aborted.

2. If a context is available at a different octave than the context in the input, where all three pitches match, then it is declared to be an exact match. Though this rule is broadly applicable to many rāgās of Carnatic music including Sahānā, it would be

incorrect for a few of the rāgās which have an octave range constraint. Therefore this should be considered a rāgā-specific rule. For some rāgās with symmetric gamakā structures in the lower and upper part of the scale, it may even be possible to extend this rule to match contexts between the two parts of the scale.

3. A mismatch of the preceding note gets a penalty of 0.5 and a mismatch of the following note gets a penalty of 0.4, both applied multiplicatively.

4. A mismatch of the direction of movement from the preceding note gets a penalty of 0.6 and a mismatch of movement direction to the following note incurs a penalty of 0.4.

The penalties thus accumulated are passed on to the selected gamakās for use during phrase-optimal selection.

6.5.3 Matching the performer's discrimination

The mappings formed thus far between local melodic contexts and choices of gamakās indicate the space of valid choices - the validity having been established by their use in an actual performance in the rāgā. However, on examination of the actual choices used in the performance for a given pair of consecutive local melodic contexts, we find a reduction from the space of possibilities that exceeds what one would expect from a mere increase in the size of the context. I call this reduction the "discrimination" shown in the performance and it gives an important clue to constructing rules out of the transcription data. A suggested measure of this discrimination is shown below –

$$d(c_1, c_2) = \log_2 \frac{n(c_1)n(c_2)}{n(c_1, c_2)} \tag{6.4}$$

where c_1 and c_2 are local melodic contexts and $n(c)$ stands for the number of choices present in the performance for context c.

We need to consider three kinds of discrimination ordered by increasing amount of context information.

1. Single pitch context, where each pitch mentioned in the prescriptive notation is elaborated in isolation from its neighbours. The number of choices per pitch in this case is a very large space that results in a vast increase in the combinatorial complexity of choosing an optimal set of gamakās for a phrase. I therefore argue that this is an inappropriate amount of context information for gamakā choice.

2. Pitch digram context, where we only consider a pitch in conjunction with the one following it. This provides a reduced space of choices compared to the single pitch context and the notion of "discrimination" as described above begins to show. However, it is inadequate for *vakra* rāgās such as Sahānā which have constraints about inflection points in melodic movements. A digram context pair would incorrectly conflate movements involving the inflection points of a rāgā.

3. Pitch trigram context, where a notated pitch is always considered in relation to the pitch that precedes it and the one that succeeds it. Pitch trigrams are adequate to encode a rāgā's constraints about inflection points in movements.

I used the pitch trigram context since it provided minimally complete information for selecting gamakās for Sahānā. Table 6.2 presents the transcription statistics for the reference performance.

6.5.4 Optimizing gamakā selection over a phrase

The first step towards a phrase interpreter based on the performance transcription is to create a catalog of Stage and Dance components keyed by pitch-trigram contexts. The "Karuṇimpa" composition was first divided into "notes" as specified in its prescriptive notation – i.e. each mention of a *svara* in the prescriptive notation was taken as a "note", regardless of the length of the gamakā that the note was a part of. A pitch-class trigram context was derived for each note, to which a set of Stage and Dance components were associated. This context is similar to the approach taken in Gaayaka, except that note timing information is discarded in constructing the context. Constructing such a catalog discards the discrimination expressed by the performer in choosing gamakās for a phrase, which scoring functions used in the phrase-level optimization algorithm restored.

6.5.4.1 Algorithm

To select a preferred set of gamakās over the duration of a phrase, local continuity preferences were represented as a scoring function $w(g_1, g_2)$ derived directly from the pattern of occurrences in the reference performance of "Karuṇimpa" that evaluates whether two gamakās are compatible when used in sequence. (216 such bigrams featured in the reference performance.) The choice of gamakās over a phrase is then taken to be the sequence g_i that maximizes the phrase score $\sum_i w(g_i, g_{i+1})$. The optimization is done by the well known "shortest path" or equivalently the "longest path" algorithm for directed acyclic graphs,

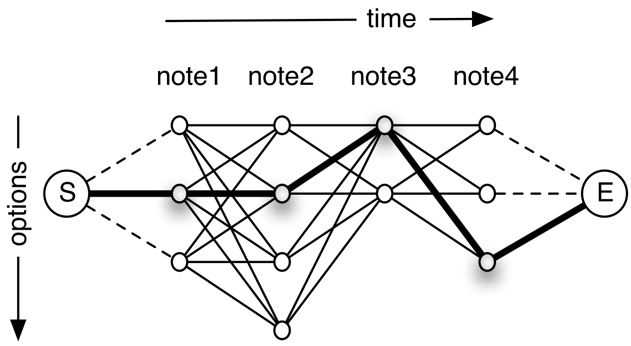


Figure 6.7: Finding the optimal choice of gamakās over a phrase as the optimal path through a directed acyclic graph. The directions on the edges are aligned with the direction of time.

illustrated in figure 6.7. The dummy start and end nodes labelled S and E are connected to the gamakā options for the first and last notes of the phrase through zero-weighted edges. The weights on the other edges are given by w. Eppstein's k-paths algorithm may also be used to explore multiple options [Eppstein, 1998].

This architecture can be seen as the fusion of a "grammar"-based approach using string rewriting rules, and a constraint-satisfaction approach. Expressing the constraint satisfaction as the optimal satisfaction of a set of potentially conflicting rules was suggested by Vijayakrishnan's proposed formulation of the "grammar of Carnatic music" based on Prince and Smolensky's Optimality Theory [Vijayakrishnan, 2007, Prince and Smolensky, 2004].

To select gamakās for a given phrase, the phrase is divided into its constituent notes and the duration-free note trigram is used as the associated local melodic context. The gamakās in the analyzed performance for corresponding note-trigrams are collected as options for each note of the given phrase, expressed in the Stage_R + Dance_R representation. We bias two consecutive Stage_R components to be continuous and also express a preference towards matching "kampita" gamakās by introducing another factor for the Dance_R component. Note that the table lookups are duration-free, but the scoring functions for the optimization passes are sensitive to the duration featured in the target prescriptive notation. Listings I.1 and I.2 give the calculations used to get a score for two gamakās being placed in sequence, using the PASR and DPASR representations. The PASR scores were used with the "longest path" algorithm and the DPASR scores were used with the "shortest path" algorithm.

kampita$(-, -, n)$	kampita$(-, \wedge, n)$
kampita$(\wedge, -, n)$	kampita$(\wedge, \wedge, n)$
$_\vee -$	$_\vee \wedge$
$- _\vee -$	$- _\vee \wedge$
$\wedge - _\vee -$ (i.e. *ravai*)	$_\vee \wedge - \wedge$

Table 6.5: Simplified dance movement catalog. kampita$(start, end, n)$ denotes sequences such as $[\wedge, -, \wedge, -, ...]$ or $[-, \wedge, -, \wedge, ..]$ The word *kampita* used is suggestive of the traditional term, but generalizes to include *odukkal* $([-, \wedge])$ and *orikai* $([\wedge, -])$ in the $n = 0$ case.

6.5.4.2 Scoring continuity of the Stage component

The Stage continuity score $w_s(S_i, S_{i+1})$ was determined by extending S_i and S_{i+1} over the joint duration of the two consecutive notes and taking the squared pitch difference between the corresponding Stage$_M$ pitch values. In addition to the continuity score, a speed bias factor that preferred less movement for shorter notes and another that preferred a match between the stage pitch and prescribed note's pitch were also used.

6.5.4.3 Scoring continuity of the Dance component

The Dance$_M$ catalog could be simplified to the categories shown in table 6.5. The Dance continuity score w_D was determined using the following –

1. A ceiling parameter for number of movements permitted per time count eliminates faster than admissible movements.
2. If the Stage movement has no associated Dance amplitudes, the dance component is fixed as flat - i.e. with a deflection of 0 and in the "sustain" category.
3. A Stage movement with both positive and negative amplitudes admits only dance movements with both $\wedge$ and $_\vee$.
4. If D_i ends in a kampita$(n > 0)$ and D_{i+1} starts with one, the score is biased such that the two kampitas are compatible - i.e. they can be represented as a single kampita when considering the two gamakās together.
5. The score for gamakās with kampitas is boosted by a factor of 2 when constructing phrases to be rendered at 2x speed.

6.5.5 Determining the scoring function through iteration

The process of rule construction, which is the same irrespective of whether the transcription is represented in the PASR or DPASR form, is described below –

- In the initial state, the system will show a discrimination of 0 because all gamakā combinations are permitted.
- Select the context sequence pair for which the difference between the discrimination expressed in the performance and that expressed by the rule system is the maximum. Discrimination is calculated as per equation (6.4).
- Add the simplest, most general rule that results in the rule system showing discrimination similar to that in the performance. Note that it is not sufficient to just match the discrimination score, but the choice made must also conform to the set of gamakās actually used in the performance.
- Re-evaluate the discrimination shown by the rule system and iterate.

On one hand, it is possible to make simple rule systems that embody the entire content of the material that they train on. One such simple rule system would involve using rich enough melodic contexts enumerated from the transcribed reference performance so that unique gamakās would be associated with each context. This is clearly not what we're after because no new knowledge about the patterns in the performance will be obtained, apart from the performance itself. In fact, such a rule set would only be a re-encoding of the transcription. On the other hand, deriving a rule system by studying a given reference performance requires that no extra-performance information is used. This means that the rules included must be based on simple categorical boundaries around concepts supported by the evidence in the performance.

The above approach is not guaranteed to converge on such a compact rule system and selecting rules is partly an art, much as selecting a hypothesis in science from the vast space of possible hypotheses is an art. However, this method of discrimination matching systematizes the process to the extent that the PASR based rule system and the DPASR based rule system are comparable.

Though the above expression of the process was iterative, it turned out to be not strictly so, and involved going back and forth on adjusting the strength of the rules as I tried to match the discrimination expressed in the performance. An important point to note is that the constructed rules should not be expected to produce the same gamakās at the phrase level as found in the reference performance. If that were a constraint, the rules would merely encode the data found in the reference performance and will not display contextual flexibility. In practice, I found exact discrimination matching to be hard anyway because the rules reach a point of diminishing returns where a change to a component to improve a match in one case would result in the deterioration of another match.

6.6 Gamakā grafting procedure

A stage movement augmented with information about the amplitudes of the focal pitches involved, as described in section 6.2.2.1, provides complete information about rāgā compatibility of the resultant gamakā. This feature of the DPASR representation permits us to take any abstract gamakā form in the extracted catalog and adapt it for a given stage movement, without loss of conformance to the rāgā. The gamakā is said to be "grafted" onto the stage movement.

Though rāgā conformance is not lost, the performer does show preferences to executing gamakās and these expressed preferences can be used to provide a score for matching gamakās to stage pitches, which combines with the score for gamakā sequencing to generate an optimal set of possible renditions of a phrase.

- When matching a gamakā to a stage movement, at most one dance focal pitch with non-zero amplitude can overlap a sustained stage pitch.
- If a stage focal pitch indicates a non-zero amplitude, then a gamakā with a movement that matches in the sign of the amplitude is preferred over one with an opposite sign or zero. In other words, if a direction of gamakā is preferred for a given stage focal pitch, then a gamakā with a dance component that matches that preference must be selected.
- If two dance focal pitches are assigned to two different stage pitches, then a TFP or NFP must feature between them during the period when the stage pitch transition occurs.

Chapter 7

Evaluation

Contents

A study with experts of the genre was conducted to evaluate the performance of the PASR and DPASR based elaboration systems relative to the resynthesis of the reference performance from its transcription. For an elaboration system that can generate multiple ranked interpretations of phrases, it is of interest to know how acceptable these interpretations are, how many of these variations are acceptable for each phrase and the space of phrases for which such acceptable variations can be generated by the system. These performance parameters are referred to in this document as *acceptability*, *range* and *scope* respectively. Expert musicians interviewed in the study were presented with a series of clips for evaluation and critical commentary. Cultural conventions and common music listening situations were taken into account in designing the study. Test sets which included phrases in two speeds as well as renditions of solicited challenge phrases were given a numerical score by participants. The scores were aggregated to estimate the performance parameters and the critical comments by the participants were used to reflect on aspects of the system. This chapter details considerations for the design of the expert evaluation of the system in terms of the performance parameters, the setup for conducting the evaluation, the structure and purpose of the test sets, and the method used to analyze the results. The format of the interview and the sets of phrases used during the evaluation is given in appendix D.

7.1 Experiment design considerations

Several issues surround the acquisition, aggregation and analysis of auditory evaluations and reflective discourse of specific musical renderings by practitioners of Carnatic music. Analytical approaches to the basic entities that characterize the genre are rare among practitioners. The microtonal structure of Carnatic music is highly debated, as noted in [Vijayakrishnan, 2007]. The role of synthesized sounds in the teaching, learning and performance of Carnatic music stops with the use of an electronic tanpura, or "sruti box" as it is known. Although Carnatic music practice defers to the vocal tradition, instrumental techniques do inform and borrow from vocal music. There is no canonical form for even the basic scalar exercises[1] taught to beginners, resulting in gamakās being avoided in early stage instruction in many schools. Learning by listening to a teacher either sing or play on an instrument and imitating the teacher is the most common form of pedagogy. Switching teachers is discouraged since considerable unlearning might be involved due to stylistic and repertoire differences. Since compositions are subject to interpretation by performers, schools of musical training ("*pāṇi*s") evolve distinct idiosyncratic repertoire ("*pāṭhāntra*") over time. The existence of common terminology does not imply shared understanding. For example, the notion of "rāgā" has many facets. Although compositions constitute a significant part of the culture, performance and pedagogy, discourse on rāgās leans towards the *ālāpana* form rather than compositions. These music-cultural aspects therefore need to be factored into the design of any study involving musicians. I now describe in greater detail some of the issues, how they may confound a study and how to account for them.

[1]"Scalar exercises" is a common term used in music teaching to refer to exercises involving patterns that move up and down a musical scale.

7.1.1 Analysis in Carnatic music

Analytical approaches to the fundamental musical forms used in Carnatic music are relatively rare. The traditional musicology of Carnatic music includes several facets of study – cultural practices, categorical studies such as ontologies, interpretation of historical discourse, composers and their works, rāgā aspects, *laya* aspects, melodic structure and gamaka and tonality theories. Despite the existence of discourse on such levels, a common cultural meme is that music is "beyond analysis". The dominant pedagogical practice involves a listen-and-imitate loop between a teacher and student, with the teacher offering corrections or alternatives wherever necessary. Teachers, or "gurus", are revered in the tradition and such one to one interaction is the canonical way to learn Carnatic music.

Some teachers may depart from this tradition or augment it with teaching techniques they've developed. For instance, the late flute maestro T. Viswanathan is well known for his use of descriptive notation in articulating the details of gamakas that would otherwise have to be learnt only through repetition. Such level of detail is an exception rather than the rule. The most used form of notation in Carnatic music remains the prescriptive notation, both when considering the written form as well as the form used in musical communication.

The strategy used to mitigate the effect of the bias against detailed musical analysis is to situate the participating experts in a hypothetical, yet familiar condition and then seek their responses. Since teaching is a situation familiar to experienced practitioners such as those interviewed, asking a participant to evaluate a variation played to them as though they're evaluating a student and provide corrections as they would in that situation, the analysis bias could be overcome.

7.1.2 Discourse

The lack of common analytical techniques and vocabulary makes opinion aggregation a hard problem in the genre. This problem surfaces when a particular interpretation of a phrase is presented to a musician who expresses his or her disapproval of it. A musician will usually be able to offer alternatives in the case of such disapproval, but is unlikely to be able to articulate the reason for their disapproval.

The many reasons for such disapproval include –

1. unacceptability of the rendition in the school that the participant belongs to,
2. unfamiliarity with instrumental or vocal techniques and limitations for rendering a particular phrase,

3. biases against specific forms that are common practice,
4. biases against specific musical styles if the rendition evokes one,
5. unfamiliarity with the sound of the rendition (such as a different tonic, and voice or instrumental quality),
6. deep familiarity with the phrase and consequent preference for the interpretations that they've learnt from their teachers,
7. a consequence of analysis they've conducted in the past in which they've considered alternatives and rejected some while keeping some.

Of the above, only in the last case can one expect to get a well articulated description of the problems with the interpretation of the phrase presented. To help establish a common ground on which the responses of multiple participants can be compared, participants were asked to provide a numerical rating on the scale of 0-10 for the snippets presented to them. To provide a common musical reference point for the rating received, phrase renditions present in the reference performance were included in the rating task without being identified as such. Different participants were expected to provide different ratings even for the renditions taken from the reference performance and this expectation held on the field.

7.1.3 Instrumental techniques

Carnatic music is considered to be based on a tradition of vocal music. The instrumental music performed draws on the same repertoire as vocal music and features the same performance structure with regard to compositions and improvisatory forms. Though instrumental performers would consider closeness to a "singing style" as desirable, it is not clear what that entails. Not only do instrumental performers adapt techniques from singing for their instrument, but vocal music practitioners also adopt instrument-inspired ornamentations. This dialogue between instrumentalists and vocalists implies that musicians may show an affinity for certain instruments, which would then influence their judgement.

Instrumentalists may choose to either limit themselves to a small range of techniques or to explore and adopt techniques used in other genres as well. For example, right finger techniques on the vina include alternating plucks with the fore and middle fingers, with or without a brief stopping at the start of a pluck, with differing degrees of strength and at different relative positions on the string. Left hand techniques include sliding, pulling on the string, finger slaps, slap and pluck, string jumps, playing overtones, etc. One artist's choice to be close to vocal renditions might constrain the techniques chosen while another's choice to go for speed may result in a simplification of the techniques most often used. Such choices can be expected to create biases in listening tests as well.

Instrumentalists may also have certain habits of performance that they attach special significance to. For instance, strumming the side strings of the vina is used as a way to show the tāḷā during performance. Not all traditions practice this, but for those who do, presenting material with or without such a tāḷā strumming may be significant.

7.1.4 Rāgā in alapana and compositions

A common understanding of a rāgā is as a set of melodic constraints and characteristics that apply to the performance of compositions and improvisation. Compositions, especially varnams, are used in the genre as sources from which to learn the nuances of a rāgā, with the shorter "krti" forms serving to highlight the feeling or "bhāva" aspect of a rāgā. However, discourse on rāgās tends to revolve around the free-time improvisatory form known as "*ālāpana*". This suggests a connection between the metrical and non-metrical forms in which a raga may be expressed. While most of the melodic constraints of a raga are common between the two forms of expression, a strong bias exists towards discourse on ragas centering on the *ālāpana* form. Since this research restricts itself to gamakās set in metrical time, such *ālāpana* centric discourses may not translate well. It is therefore desirable in expert interactions to steer conversation towards discussing metrical gamakā forms by focusing on, say, *svara kalpana*.

7.1.5 Synthesized gamakas

Synthesized gamakas are unfamiliar to the Carnatic ear and I refer to both positive and negative biases originating from this unfamiliarity as "synthesis bias". Some artists perform Carnatic music on electronic keyboards using the pitchbend wheel to execute gamakas. Though such adaptations do exist, the sound of Carnatic music continues to be dominated by voice, violin, vina and "goṭṭu vādyam" (a.k.a. "chitra vīṇā"). Since the ear is only exposed to human performance on instruments even in the case of electronic keyboards, any gap in the expression between a synthesized gamaka and a human performance is likely to be immediately noticed. The significance of this expression gap may be expressed as either disproportionately large if a musician's biases align with that view, or disproportionately small if the musician is favourable to such experimental work.

In this research, I focused on the most important left hand techniques for executing gamakās on a vina – the left finger slide and pull – and normalize all right finger plucks to be uniform. Though in preliminary tests this synthesis was acceptable to several musicians, it was still possible for musicians to hold strong opinions about it which influence their ability to comment in listening tests. For this reason, the first clip presented to the participants was a long section from the reference composition and consisted of two speeds. The purpose of this first set was to orient the participants and to determine early in the interview whether they had any objections to or difficulties with the sound that they will be presented with during the rest of the evaluation.

7.2 Logistics and setup

The logistics for conducting such a study will vary for each attempt. For this study, participants were recruited through the "Chennai musicians-dancers directory" [Ramani, 2001] and recommendations of other musicians. The participant group consisted of 13 musicians and musicologists all of whom have 12 to 50 years of teaching and/or performance experience. The participants were interviewed in a location of their choice, usually their residence, and in accordance with cultural norms. Participants required to be able to converse either in English or Tamil or a mixture of both since I served as the principal investigator for the interviews and those are the languages that I'm fluent in.

In order to keep the strain placed on the participants by the listening tests manageable, the interviews were constrained to be within approximately 90 minutes. This placed rather severe constraints on what could be tested within a single interview. In particular, it necessitated that the number of variations presented for each phrase be limited to three. Therefore the variations presented needed to be chosen such that both the DPASR and PASR systems get equal opportunity - i.e. the phrases and variations were chosen such that the two systems rated the reference performance the highest, but differed between them about two other top rated variations.

In order to minimize situational variations among the interviews, synthesis output for the test sets that evaluate a fixed set of variations were pre-recorded to mp3 format (sets 1, 2, 3 and 5 as described in section 7.4). A computer ("MacBook Air", 1.7GHz Intel Core i5 processor, with 4GB memory) was used the play back these pre-recorded as well as to present the synthesized output for the set involving challenge phrases given by the participant (set 4). The sound setup for the study was designed to mimic familiar situations in which such music is listened to. This meant playing all sound through speakers at a volume level comparable to television or radio at the participants' residence. A pair of portable external USB powered speakers were used to present the audio. The computer's audio output was set to maximum volume for the all the interviews and the speaker volume was adjusted to be acceptable for each participant during playback of the calibration set (set 1). All the pre-recorded synthesis output snippets were played back using MacOS 10.8's playback facility known as "Quick View". This made rapid switch between the snippets and seeking within a snippet quick to perform. The output of the synthesis interface and the pre-recorded snippets were ensured to be at the same level (9 dB below peak) by recording the digital audio output of the synthesis interface running within the "Google Chrome" web browser application and converting the captured samples to MP3 format.[2] The same application was then used during the challenge set (set 4).

7.3 Study overview

A study was designed to evaluate the output of the system for selecting gamakās for short phrases, in which synthesized musical examples were presented to the participants for evaluation. Musicians who were above 30 years of age and had at least 5 years of experience performing in about 50 concerts as the main artist in standard concert formats were eligible to participate. Music teachers not focused on performance, but who have trained performing students in improvisatory forms of alapana, tanam, niraval and svara kalpana were also considered. A typical interview session lasted around 90 minutes.

The interview format consisted of five sets of synthesized phrases played to the participants on a computer. The synthesized phrases were generated in the PASR and DPASR representations and synthesized using sampling synthesis where samples of plain *vīṇā* notes were played back with continuous rate variation to simulate gamakās. A tanpura (drone) accompaniment track was mixed in to indicate the tonic. Some reverb was also added to the mix to approximate standard presentation and listening situations.

Participants were asked to rate multiple interpretations of each phrase on a scale of 0 (for "fail") to 10 (for "excellent") with 5 standing for "ok". Participants were encouraged to talk about the reasons they gave the ratings they did. For the case of low ratings offered, they were asked to consider the clip to be that presented by a student and asked to suggest corrections to this "student rendition". For the case of high ratings, they were asked to

[2]The application used to do the audio capture was "Audio Hijack Pro" version 2.10.5

describe those aspects that they felt contributed to the high score.

The phrases chosen for synthesis and the interpretations presented were constructed to measure the *acceptability* of the gamakas selected by the system, the *range* of acceptable variations generated and the *scope* of phrases for which the system produced acceptable variations. A description of these variables now follows.

7.3.1 Acceptability

The acceptability of a particular synthesized rendition may be taken to be directly related to the rating given by the participant for each interpretation. Calibrating for synthesis bias requires that the phrases themselves be familiar to the participants and consistent in style. For this, the set of phrase interpretations also included resynthesized versions of interpretations of the phrase used in actual performances. If participants reject or rate low a transcribed performance, that is indicative of negative synthesis bias. Scores for other interpretations may be expected to be lower. Therefore the score for a resynthesized performance of a phrase is a normalizing factor in evaluating the other interpretations. Normalizing against evaluations of these resynthesized performances also helps account for differing degrees of "strictness" of the expressed musical judgment of the participant – i.e. some participants may be more flexible in what they consider acceptable while others may be more rigid about it.

7.3.2 Range

I define the range of the system as the space of possible acceptable interpretations of a phrase that the system can produce. The rating received by the system-generated interpretations of phrases, normalized against the ratings received for the resynthesized reference performances indicates the range of the system. The problematic component in this is the determination of which set of variations to present to the participant. Nearby rated productions of the system may vary by little and a measure of melodic proximity is required to be able to provide sufficiently different variations to evaluate the range. This implies that it is inadequate to select, say, the top five interpretations generated by the system according to its scoring scheme since they may be too similar. A manual evaluation of the generated top scoring phrases was used and a selection of phrases that spanned a range of scores (i.e. not just the top scoring ones) were chosen for presentation to participants. In particular, those phrases for which the DPASR system and the PASR system ranked differently were chosen in order to help the evaluation discriminate between these two representations.

The range of the system can be numerically indicated using the formula in equation 7.1 –

$$\text{Range} = \frac{1.5 - R_{RP} + 0.5(R_{PASR} + R_{DPASR})}{1.5} \tag{7.1}$$

where R_{RP} is the rank of the reference performance and R_{PASR} and R_{DPASR} are the ranks for the variations preferred by the PASR and DPASR systems respectively.[3] On this scale, a value of 0 indicates that only a single interpretation is acceptable for the phrase, in which case there ought to be no "range of variations" to rank or select from. On the other end, a value of 1 indicates that all variations were acceptable without discrimination. Values close to either limit indicate that the problem of elaboration may not be a hard one since unique elaborations exist in the case of 0, which can therefore be modeled using generative grammars, and many equally acceptable variations exist in the case 1, in which case filtered random selection of gamakās may also work well enough. Since only three variations per phrase could be tested due to limits on the length of the interviews, such extreme conclusions cannot be drawn from this study if the range turns out to be close to either end. Nevertheless, such a tendency would be an important indicator of other techniques that these systems might need to be compared against.

7.3.3 Scope

I define the scope of the system as the space of prescriptively notated phrases for which the system is capable of generating acceptable interpretations. The primary gauge of the scope of the system is to solicit phrases from the expert participants for which the system is used to generate variations on the spot for scoring by participants, in two speeds. The scores normalized to those given for resynthesized reference performance versions of the phrases in the fixed sections are used as indicators of the scope of the system. Participants were asked to provide phrases as though they were to perform "*svara kalpana*" or "niraval" improvisations in Sahānā instead of "*ālāpana*" phrases. This is because this research is limited to metric gamakās by design.

7.4 Test sets

The format of the interview consisted of asking the participant to evaluate synthesized phrase snippets presented in five sets.

[3] Calculation of these Spearman ranks and a "pseudo rank" is described in section 7.5.1.

Set 1 This set consists of a single snippet presented for the purpose of familiarizing the participant with the synthesized sound, the rating scale and the musical context of the study which includes the raga and the varnam.

Set 2 "First speed" phrases and variations.

Set 3 "Second speed" phrases and variations.

Set 4 Referred to as the "challenge set" here, participants are asked for two phrases of their own imagination for which variations are generated and presented to them for evaluation on the spot.

Set 5 This set consists of one multi-phrase snippet for which gamakas are chosen by the system.

Participants were asked to rate the phrase variations presented in each set on a scale of 0 to 10 where 0 stands for "fail" and 10 for "excellent" with the middle 5 standing for "ok". The purpose of using such a numeric rating scale is to compare the preferences of the participants.

I describe the sets in detail in the following sections.

7.4.1 Set 1 - Familiarization

One 108 second long resynthesized version of the pallavi of the "Karuṇimpa" varṇam was played. The audio consisted of the synthesized melody played with a tanpura background to establish the tonic. A reverb effect was also used to bring the sound closer to what the participants might be familiar with in a concert setting. These parameters are chosen to be consistent across all the sets. Therefore the usability of the subsequent sets can be inferred from the response of the participant to this one clip.

The primary goal of this set is to ensure that the participant is able to pay attention to the musical aspects of the snippet without being distracted by the synthetic nature of the sound. Although real vina string samples were used as the basis for the sound, the models used to render a melody expressed in PASR and DPASR forms were limited in their physical accuracy. Dynamics of string plucks were not modeled and simple interpolation curves were used to model pulling and sliding techniques. The phrases rendered also featured strictly metronomic timing, which is not the norm in Carnatic music performance. This rendition also reflects the quality of my transcription. Having the participant comment on this section first therefore helps set a reference point for the other evaluations they provide.

The snippet presented included a first speed rendition of the pallavi as well as a second speed rendition. This was expected to acquaint the participant with the tempos at which all the synthesized snippets will be presented to them in the subsequent sets.

Other parameters that can affect the participant's perception of the melodic details include the volume level of playback, the amount of reverberation and the mix between the tanpura and the synthesized melody. These need to be set to help the participant attend to the melodic aspects.

Asking the participant to rate this "performance" on the 0-10 scale that will be used for all the other phrase renditions that will be presented to them help familiarize them with the rating scale and the criteria they may bring to arrive at the evaluation.

7.4.2 Set 2 - First speed phrases

This set consists of three "first speed" phrases played at a tempo of 70 beats per minute. For each phrase, three variations were presented to the participant for evaluation. One of the variations was a direct transcription from a performance and the other two variations consisted of two different selections made by the system. These variations were presented in random order to the participants and they were not informed of the origin of the variations they heard.

Participants were asked to evaluate each of these variations by providing a numeric rating and verbal comments. Some criteria suggestions were given, but they were not limited by those criteria. One suggestion given was to treat each snippet as though one of their students were performing it and rate the quality of the rendition according to their own aesthetics, and provide any corrections they would make. A second suggestion was to ask them to provide the prescriptive *svaras* for what they heard so that it can be compared with what was elaborated by the system. A third suggestion was to have the participant provide their own interpretation, in case the earlier ones proved to be insufficient.

7.4.3 Set 3 - Second speed phrases

The structure of this set is similar to that of set 2, except that the phrases are rendered in "second speed" – i.e. at a tempo of 140 beats per minute. This doubling of speed necessitates different choices for gamakas and the purpose of this set is to evaluate whether speed appropriate choices are being made by the system and where problems might lie.

Similar to set 2, one resynthesized performance and two system generated variations were used and the same suggestions given for set 2 applied to this set too.

7.4.4 Set 4 - Challenge phrases

The purpose of this set is to identify the limitations of this approach to gamaka selection and combination. Participants were asked to provide two metrical "challenge" phrases in svara form for which two variations were generated by the system in two speeds for evaluation. The phrases given by the participants were transcribed into a plain text notation accepted by the system as input.

It was anticipated that participants will lean towards providing phrases in smooth time "raga alapana" form rather than metric svara form. Therefore the metric constraint was given to them first. I avoided the "raga alapana" form since we do not yet know how to assign durations to svaras in such expressions. If the participant had any difficulty moving away from the "alapana" mode, they were asked to think of what they would do for niraval or svara kalpana in the raga. The phrases had to be of their own imagination, however there may be similarities to phrases used in the performance from which the gamaka preferences were inferred to construct the system.

It is conceivable that there exist phrases for which the system cannot select any gamakas because its component movements did not feature in the reference performance. In such cases, it may not be possible to get an evaluation from the participant, though the existence of such phrases is itself valuable knowledge.

Sahānā admits one "controversial" phrase – RGmPDnS – that does not conform to the known ascent/descent melodic constraints (the *ārōhaṇa* and *avarōhaṇa*). Phrase choices that border on these violations were also expected to pose problems for the system.

7.4.5 Set 5 - Multi-phrase section

In this set, one system-generated interpretation of one 15 second section from "Karuṇimpa" consisting of multiple phrases is presented for evaluation in the form of a rating and commentary. The purpose of this set is to evaluate the significance of phrase boundaries to gamaka selection. Participants were briefed about the purpose of this section and requested to comment on whether they felt that the phrasal structure was perceptible and to comment on the failure points. The system does not perform automatic phrase segmentation, which limits its ability to account for significant phrase boundaries. The feedback from participants on this section was therefore expected to shed light on areas of future work.

7.5 Analysis method

Analysis consists of two parts – aggregating the reported scores for the purpose of comparing system-computed rankings with scores given by the participants and qualitative analysis of the comments offered.

7.5.1 Score aggregation

The scores given by the participants were first normalized according to the maximum score given by each participant for the phrase variations taken from the actual performance. The purpose of this normalization step is to make the participants' opinions about the variations presented comparable. Phrases in sets 1, 2, and 3 were included in the normalization. Set 5 is excluded from this normalization because that would entangle phrase boundary considerations with the interpretation of individual phrases. The normalized maximum score for each participant is therefore 10 and each participant has a different normalization factor. The interpretation of the results was not sensitive to the choice of normalization factor.

The normalized scores were aggregated across all participants, and described using two parameters – mean and variance. Using these parameters, an aggregate "pseudo rank" in the range 1.0 to 3.0 was derived for the variations of phrases presented in sets 2 and 3. The pseudo rank maximizes the correlation between the aggregate score order and the rank number. Equation 7.2 gives the relative weight between two score aggregates modeled using a gaussian distribution. To determine the rank numbers corresponding to the scores for each of the three variations (DPASR, PASR and RP), all six possible rank assignments are first calculated using connection weights α_{ij} according to the formula in equation 7.2, where (μ_1, σ_1) and (μ_2, σ_2) are the parameters of the two gaussian distributions. For each of the six assignments, the correlation coefficient with the scores is then calculated. The assignment that results in the maximal value for the correlation coefficient is taken to be the assigned "pseudo ranking" of the variations. Unlike the conventional Spearman ranking correlation coefficient, this ranking assignment scheme can account for value spread when aggregation needs to be done in two stages – once across all participants and a second time across the categories that need to be ranked.

$$\alpha_{12} = e^{-\frac{1}{2}(\mu_1-\mu_2)^2(\frac{1}{{\sigma_1}^2}+\frac{1}{{\sigma_2}^2})} \tag{7.2}$$

Two rankings were calculated for each system-generated variation - one variation generated by the system using the DPASR representation and one generated by the system using the PASR representation. Both systems consider all possible gamaka selections based on note-trigrams and can therefore rate an arbitrary set of gamaka choices for each phrase. To compare these system-generated ranks with the pseudo rank, a simple Pearson correlation coefficient was computed. Positive correlation coefficients indicate that the system predicts the rank well.

7.5.2 Qualitative analysis

Opinions regarding gamaka aesthetics, speed appropriateness, raga appropriateness and stylistic preferences were sought from participating experts. These comments were used to determine the limits of the approach of determining raga rules from a single performance of a varnam.

The challenge phrases provided by the participants were evaluated against known *rāga lakśaṇā* literature for their typicality with respect to the raga Sahana. Typicality is judged based on strict adherence to *ārōhaṇa* and *avarōhaṇa* and the inclusion of characteristic sub-phrases indicated in the literature. It is possible for arohana/avarohana adherence to be ambiguous, in which case it is judged to be atypical of Sahana. Exceptional usages admitted in Sahana were avoided during the interviews and were not expected to impact analysis of the challenge section. The main difficulty in analyzing this section arises when participants provide non-metric phrases. In such cases, the provided phrases may either be discarded, or a close metrical approximation considered for use.

The descriptive words and phrases used by the participants to describe the clips that they were presented are tabulated. These were used to – a) reflect on possible confounding factors the participant may bring to the study, b) cross check the ratings given by the participants and c) annotate musical input provided by the participant as part of the interview.

Chapter 8

Results

Contents

Two elaboration systems, one built using the PASR representation and the other using DPASR representation, were evaluated through interviews with experts for *acceptability*, *range* and *scope* defined in sections 7.3.1-7.3.3. The two systems were found to perform comparably, with the expert rankings of variations correlating with either system to a fair extent. A measure of the range of variations produced by either system was found to fall in the region that suggests that grammar based deterministic approaches and constrained random generation or a combination thereof may not be adequate to tackle the phrase-level elaboration problem in Carnatic music, and several competing constraints are at play. For phrases provided by the experts that were not from the reference performance, the choice of timing-independent local melodic context resulted in either system generating acceptable variations. Furthermore, since the complexity of the encoded rules in the DPASR based system is much lower than that in the PASR based system, it can be said that the DPASR representation holds promise as a candidate for studying Carnatic music through computational means. Expert comments given during the interviews indicated that dynamics may play a greater role in rāgā definition than anticipated, and that there is much scope for improvement in our formal understanding of gamakās. This chapter details the results of the study as pertaining to the elaboration system and includes specific comments by participants on aspects of elaboration and expression raised by the clips presented to them.

8.1 Comparing PASR and DPASR

From table 8.2, we find that the phrases generated by the DPASR representation rank slightly higher than those generated using the PASR based algorithm. Due to the proximity of the ranks based on the two representations, it is fair to say that they are roughly similar in their modeling abilities. However, we also need to factor in the relative simplicity of the DPASR based gamaka selector compared to the PASR based selector. The good performance despite the simplification suggests that the DPASR representation may have some bearing on gamaka aesthetics.

Pearson's correlation coefficient was computed between the system's rankings and the observed pseudo rankings of the variations. The coefficient was 0.69 for the DPASR algorithm and 0.62 for the PASR based algorithm. The Pearson coefficients thus indicate that the DPASR representation based system is marginally better correlated with the participants rating. The comparable scores for both these representations is expressive of similar extents of musical knowledge embodied in the two systems. While the PASR based rule sets are more elaborate and consider many small cases, the DPASR representation makes some sweeping simplifications that are about as effective as the PASR representation. It can therefore be conjectured that incorporating some of the complementary musical preferences expressed in the PASR system into the DPASR based system would further improve the outcome.

Participating experts did not show a marked preference for either the DPASR based variations or the PASR based variations, sometimes preferring one and sometimes the other. Both the systems, by construction, rank the gamaka choices of the original performer the highest. However, we also see cases where the experts do not show much discrimination between the variations presented, such as with phrase-3 in set-3 (see appendix D), contrary to the discrimination demonstrated by the original performer whose work was transcribed (labelled "RP" in tables 8.1 and 8.2). This could be attributed to the space of permissible interpretations in the raga, given that many participants indicated that their evaluations would express their own musical sensibilities that come from their stylistic lineage.

8.2 Acceptability

The mean ratings of 6.91 for DPASR and 6.72 for PASR and the comments offered by the participants, taking into account that 5 was described as an "ok" score, indicates that there is considerable room for improvement in the productions of either system. However, the reference performance itself scored only a normalized 8.20 in the aggregate and the systems may be said to have generated acceptable variations beyond what was actually found in the performance. On some occasions, the variations generated by the system scored higher than the reference performance. As surprising as that was, stylistic opinion is known to vary a lot among practitioners of Carnatic music and it is a hard problem of interview design to eliminate such opinion in its entirety in such studies. It is also debatable as to whether such opinion ought to be eliminated since it is an integral component of music culture.

The comments given by the participants for the reference performance snippets ranged from "excellent" to qualified "ok"s. The synthesis was also criticized on several counts. In a couple of places, two participants felt that the pitch of G was "a bit flat for Sahānā", suggesting that a microtonal adjustment may be necessary. However, the reference performance featured unadjusted pitches for the G as well, so in that sense the systems remained faithful to the original material. The lack of dynamics was felt to be important for some gamakās, particularly those involving G and m, which were required to be "more delicate" and "subtle". Some participants did not agree with the musical decisions found in the reference performance such as the use of the glide $\overline{mD}$ to express the third (D) in ($DnD\mathbf{D}P$). On the whole, the most significant factors that lowered the non-normalized aggregate score to an unexpected 6.7 for the reference performance's resynthesis were stylistic divergence among artists, absence of dynamics in the string plucking style used in the resynthesis, and timing quantization done for some gamakās during the transcription.

8.3 Range

When considered in the aggregate, the equal pseudo ranks for the raw scores given in table 8.2 suggests that the variations generated by both these systems have comparable acceptability. Though it turned out so in the aggregate, participating experts did individually discriminate between the system-generated variations and the reference performance. This discrimination is visible in the pseudo ranks for the *normalized* scores which placed the reference performance around 0.75 above the other variations generated by the system. In the case of maximal discrimination, this difference would be 1.5, with expected ranks for

"DPASR", "PASR" and "RP" being 1.5, 1.5 and 3.0 respectively.

The measure of the range of the system according to equation 7.1 is $(1.5 - 2.48 + (1.79 + 1.72)/2)/1.5 = 0.52$. This indicates that the elaboration problem likely cannot be solved by approaches that generate unique solutions since competing constraints are at play. The range would be closer to 0 if that were to be the case. Furthermore, it also indicates that randomized or constrained random approaches are also unlikely to be satisfactory, in which case the range can be expected to be closer to 1.0. The range score of 1.0 for an elaboration system would mean that any variation it generates would be acceptable. If a more extensive study on the range of possible elaborations in Carnatic music were to be done and such a high range were found, it would imply that the task of elaboration in Carnatic music is a simpler problem that doesn't warrant the approaches used in this work. The observed range of 0.52 therefore suggests that competing constraints are at play in selecting gamakās, that need to be resolved through an optimization procedure. However, it would be incorrect to single out the specific optimization procedure used here (i.e. shortest distance in a DAG) based on this study, since several approaches that model the discrimination expressed by a performer in selecting gamakās are possible.

8.4 Scope

The primary measure of the scope of the system in this study is the scores given by participating experts for renditions of phrases of their own imagination in the "challenge phrases" set 4. Most participants provided phrases that are not present in the reference performance.

The relatively high ratings of 7.7, 8.1, 7.5 and 7.8 received for the challenge phrases section in the aggregate (average 7.76 indicated in table 8.3) compared to the rating of 8.2 received for the reference performance from sets 2 and 3 suggests that the system is fairly capable of extrapolating the material available in a reference performance to melodic contexts that were not present in the reference performance. This extensibility is possible due to the use of reduced local melodic context in the gamakā catalog which does not take note duration into account. This conclusion can be drawn particularly because the measured range of 0.52 indicates that elaborating new phrases based on a given catalog is not a trivial problem – one that is neither deterministic nor arbitrary.

In the aggregate, the second speed variations scored marginally higher (8.1 and 7.8) than the first speed variations (7.7 and 7.5). This can be expected because the possibilities in higher speeds are considerably reduced compared to the lower speeds and the speed transformation rules adequately cover the space. Another factor that influences this difference is that many participating experts already had an interpretation in mind by the time they gave the svara patterns for synthesis. During the slower speed renditions, many participants showed initial disappointment when the system did not generate the variation they expected, but then proceeded to score what they heard based on whether the variations presented were themselves acceptable. This process in the interviews was aided by the general awareness of the participants regarding the range of stylistic variability in the musical culture. This shows promise for the cultural acceptability of computational methods to elucidate musicological issues.

Using a rating threshold of 6.0, table 8.5 lists the phrases which scored about the threshold and table 8.4 lists those that scored below or equal to the threshold. The latter set of phrases, due to the their low score, are considered problematic for the system and I examine these individually below.

1. The phrase "R G **m P D n D** P m G m R" features an exceptional phrasing of Sahānā, where the common *ārōhaṇā* movement is "m P m D n". These kinds of exceptional movements will need to be treated as special cases and validated separately since the reference performance does not contain even a single instance of such an *ārōhaṇā* movement. It should be noted that in the published version of the reference composition, the svara sequence "R G m P D n Ṡ ," features in the muktayisvaram section. However, the artist whose performance was used as the reference had altered that movement and used the svara sequence "R G m P m D n Ṡ" instead.

2. The phrase "R G **m P D n , D** P m G m R" is similar to the previous phrase except for a longer *n*, but has the same arohana problems.

3. The participant who suggested the phrase "R n D P , P" expected a *jāru* (slide) to be used to connect the fragment "R n", which the system did not perform. Again, this particular movement is not present in the reference composition and therefore the system does not have it in its vocabulary. It rendered the movement with *R* and *n* separately which did not satisfy the participant. This case can be addressed, perhaps, by adding a rule to the effect that svaras that are separated by a wide gap need to be connected by a "jaru" when enough time is available to perform the movement.

4. In the phrase " **R , , R** G m P , m G m R , G R S", the problematic part was the long held "R , ,". This was a case where the duration-independent rules of the system

selected an inappropriate gamakā that needed to be rendered with some delicacy, given that *R*, *G* and *m* are central to the rāgā.

5. For "Ṡ , Ṡ P , P R P m G m R , G R S", the problematic movements were "Ṡ , Ṡ" and its connection to "P , P". The expected movement here is a general aesthetic common to many ragas in Carnatic music, but did not feature in the reference performance. This forced the system to generate time transformations of other gamakās taken from second speed phrases. The general rule behind this idiomatic usage is unclear, but it can be added as a special case.

6. The phrase "Ṡ Ṡ , P P , R R , ṇ S R G m" is literally found in the reference composition, but is present only in the second speed form. Adapting first speed gamakā choices to the higher second speed phrasings was found to be feasible, but the other way around leads to aesthetic problems. This edge case demonstrated such an aesthetic violation where the participant scored the rendition low only for the first speed variation presented. A satisfactory rendition would include "jaru"s connecting the first of the "Ṡ", "P" and "R", with "jantai" used to render the second repeat svara.

7. For "D n Ṡ **Ṙ m** D n **Ṡ R** G m P", the highlighted gaps straddle the fragment boundaries. The first speed rendition using the PASR algorithm was the most objectionable in this case, and the participant described the rendition as "like mercury". The pitch triads "Ṡ Ṙ m" and "Ṡ R G" are difficult exceptions to handle for the PASR algorithm, whereas the DPASR is able to do better by relying on the stage components alone in this case. One way to handle these large jumps is to treat them as discontinuities, effectively splitting the phrase into three independent parts - "D n Ṡ Ṙ", "m D n Ṡ" and "R G m P". Though that would be acceptable for practical purposes, it would preclude certain kinds of long *jāru* movements that would also work if adequate time for such a movement is available.

8.5 Expert comments

Participating experts provided musical and verbal comments on the snippets presented to them on dynamics, the rāgā and speed appropriateness of gamakas, clarity of articulation, instrumental characteristics, phrase characteristics requiring emphasis and handling of symmetric phrases. In this section I compile the comments provided by the expert participants on the above mentioned aspects. Though many of these comments are not directly relevant to evaluating the system and provide detail beyond what is possible to infer from the chosen reference performance, I present them here because they offer suggestions for future work in this area.

For brevity of notation, the single letters "SRGmPDn" will be used to unambiguously refer to the pitch values "sa ri_2 ga_3 ma_1 pa da_2 ni_2" ("C D E F G A $B_\flat$" respectively) of the raga Sahānā for the sake of brevity. When discussing the internal movements within a phrase, descriptive notation as found in [Viswanathan, 1977] is used below and makes use of the single letter *svara* identifiers.

8.5.1 Consensus

The comments offered by the participating experts showed some divergence of opinion, however consensus prevailed regarding a few central traits of the rāgā. Most of critical and detailed comments offered were regarding the handling of G, m and D when followed by n and R.

A common comment was to produce G with an intonation closer to m. In *avarōhaṇa* phrase fragments such as "GmR", the G was required to be approached from m. "Scalar" renditions of "GmR" such as in Set-3/Phrase-2 were not appreciated. In Set-1, two different renditions of "GmR" were used. In the first instance, G was approached from m and in the second instance G was given more prominence. One participant pointed out these two different usages and judged the second interpretation (i.e. giving G prominence) as a rāgā error. Other participants pointed out this aspect in Set-2 and Set-3.

In interpretations of D from P in the phrase fragment "PDP", gamakas originating on P were expected to extend up to n. This movement is executed as a rapid deep pull on the *vīṇā* on the P fret. In terms of the DPASR representation, the D was encoded using P as the stage and a dance component amplitude of 2 semitones, but it needs to be extended to 3 according to the comments.

Rendering the note P with gamakas was not appreciated in both first and second speeds. "Plain" renditions were preferred as opposed to introducing "anusvaras". This is a commonly stated rule. However, the general conditions under which this needs to be strictly obeyed seemed unclear since movements landing on S and P are common in performance. For example, one of the challenge phrases ($\dot{S}$,$\dot{S}$P,PR,R) was considered "pleasing" if the ($\dot{S}$,) were rendered as a slide from P and the $(P,)$ as a slide from $\dot{S}$.

Three participants pointed out the role of string plucks in emphasizing the right svaras of a phrase and to establish phrase boundaries. The plucks needed to be softer or "hrsva" in the Set-3/Phrase-1/PASR and Set-3/Phrase-1/RP variations.

Four of the participants interviewed played the *vīṇā*. Among them, there was consensus that fewer gamakās should be used if more plucks are used and vice versa.

8.5.2 Divergences

Musical preferences can vary among practitioners of Carnatic music and this showed up in the interviews sometimes as divergent view points, which are recorded in this section.

One participant said that the Set-2/Phrase-1/SD variation "gives the feeling of Sahānā" whereas another said that the variation "lacked the essence of Sahānā". Though my own evaluation of this phrase would lean towards the latter critical comment, it is interesting to note that opinions regarding a single phrase can be so opposite.

Despite critical comments given by some participants for Set-3/Phrase-1/PASR, one participant gave a raw score of 10 and another commented that the "tone" is good and that the "progression is acceptable".

Some participants preferred certain phrase fragments to be rendered "plainer" – i.e. without gamakās. The reference performance for the phrase Set-3/Phrase-2 presents the terminal phrase fragment "GmP" in plain form, but one participant preferred the same fragment to be presented as a "rounded gamaka".

In Set-3/Phrase-3/PASR, one participant rejected the rendition of m in "PmD" using the movement $\overline{PmPmD}$ as "not ok" whereas another "liked" it.

Two participants perceived the way the reference performance starts for Set-3/Phrase-3 as "GGm" instead of "RGm". In the same variation, which a participant declared as "excellent", another participant found the "mD" movement executed using a "jaru" (slide) "not acceptable" and another found "mPm" rendered "too plain".

In Set-2/Phrase-1, one participant said that the middle n and D in "DnD" and "PDP" needed to be stressed whereas another preferred the two symmetric parts of the phrase to be highlighted by stressing the first svaras D and P of the two parts "DnDDP" and "PDPPm" respectively. This is not a contradiction but indicates divergence in the perception of accents within a phrase.

In the reference performance variation of Set-2/Phrase-1, four participants disagreed with the reference version where the phrase fragment "PDP" as $\overline{P,,DP,,DP,,,}$, calling it "not appropriate".

In the SD variation of Set-2/Phrase-2, one participant commented that the rendition of "nṠṘ" needed "more differentiation", but another required that it continue from the previous gamakā.

8.5.3 Dynamics

Four participants commented on the importance of dynamics in expressing a rāgā well. In particular, the rendition of "GmR" in the upper register needed to be "more subtle". As noted in section 2.1.2.2, dynamics is among the least discussed aspects of gamakās and rāgās in the traditional literature of Carnatic music. Therefore the significant number of comments by the participants on the dynamic aspects of rendering certain gamakās in Sahānā is an interesting development. It is conceivable that the aesthetics of the genre have evolved from the period of the canonical literature on gamakās through present day practice to include dynamics as a discriminating factor.

One kind of dynamics that was brought up repeatedly by several participants is the "stressing" of svaras by playing them with a combination of a plain note and a string pluck. In Set-3/Phrase-3, for example, one participant mentioned that the svaras m, P and D "needed stress" in the renditions. In Set-5, the role of such stress in delineating phrase boundaries was also brought up by three participants.

In Set-2/Phrase-1, one participant commented that the first "DP" movement in "DnDDP" needed to be "more delicate". From a representation point of view, it is unclear to me how this extra "delicacy" can be encoded and I interpreted this comment as involving some dynamics to show the "delicacy". In all, modelling such dynamics was not considered in scope for this research and more work is necessary to understand the origin, evolution and role of dynamics in Carnatic music.

8.5.4 Symmetry

Set-2/Phrase-1 (DnDDP-PDPPm), Set-3/Phrase-1 (PmGmR,-DPmGmR,) and Set-3/Phrase-3 (RGmP,P-mPmD,D) are such that the phrases can be split into two non-identical symmetric parts. Three participants preferred that this symmetry be recognized and the phrase fragments treated accordingly. All the presented renditions including the variation selected from the reference performance treated the two parts differently.

8.5.5 Gamaka shapes

Several comments were given describing gamakā shapes and how they needed to be changed in certain situations, although such shape preferences were acknowledged to be stylistic in nature. When gamaka shape suggestions were made, they were qualitative in nature. Words such as "rounded", "smooth", "glide" and "subtle" were used, from which we can gather only that the shapes presented were not completely satisfactory to the artists. It is not common in the musical culture to communicate shapes explicitly. In some cases, though it was possible to get some clarification by asking for timing aspects of a gamakā and these have already been described in the earlier sections.

Three participants objected to rendering D in the phrase fragment "DP" using the movement $\overline{mDP}$. One participant indicated that the movement feels inappropriate, but would be "acceptable if there was a pause before m".

In the SD variation of Set-2/Phrase-1, three participants noted that the ending gamaka that might be described as $\overline{P,,mP}$ (or $\overline{P,,GP}$) implies that the svara that follows the gamakā is D.

Table 8.1: Ratings given by participants for the various sets

Set	Phrase	Variation	DPASR rank	PASR rank	Rating mean(stdev)	Rank mean(stdev)	Pseudo rank
Set 1					8.8(1.9)		
Set 2	phrase 1	DPASR	2.0	1.0	6.4(2.3)	2.0(0.6)	2.07
		PASR	1.0	2.0	4.9(1.8)	1.3(0.4)	1.43
		RP	3.0	3.0	7.4(1.4)	2.7(0.3)	2.51
	phrase 2	DPASR	2.0	1.0	7.0(2.4)	2.1(0.5)	2.01
		PASR	1.0	2.0	5.8(1.5)	1.5(0.7)	1.49
		RP	3.0	3.0	8.3(1.8)	2.5(0.7)	2.50
	phrase 3	DPASR	2.0	1.0	5.5(1.1)	1.3(0.4)	1.32
		PASR	1.0	2.0	6.9(1.9)	2.1(0.5)	2.22
		RP	3.0	3.0	7.9(2.2)	2.6(0.4)	2.51
Set 3	phrase 1	DPASR	2.0	1.0	7.5(1.2)	1.5(0.6)	1.50
		PASR	1.0	2.0	7.9(2.0)	1.8(0.6)	1.63
		RP	3.0	3.0	9.2(0.8)	2.7(0.5)	2.82
	phrase 2	DPASR	2.0	1.0	7.3(1.6)	2.0(0.4)	1.91
		PASR	1.0	2.0	6.2(1.5)	1.3(0.6)	1.42
		RP	3.0	3.0	8.8(1.5)	2.7(0.6)	2.64
	phrase 3	DPASR	2.0	1.0	7.9(1.8)	1.9(0.5)	1.93
		PASR	1.0	2.0	8.7(1.9)	2.3(0.6)	2.10
		RP	3.0	3.0	7.7(2.2)	1.8(0.6)	1.91
Set 4	DPASR	1x			7.7(1.8)		
		2x			8.1(1.6)		
	PASR	1x			7.5(2.1)		
		2x			7.8(1.8)		
Set 5					7.3(3.2)		

Table 8.2: Summary of ratings

Variations	Rating mean(stdev)	Rank mean(stdev)	Pseudo rank (normalized)	Raw rating mean(stdev)	Pseudo rank (raw scores)
DPASR	6.91(1.9)	1.80(0.61)	1.79	5.64(1.77)	2.00
PASR	6.72(2.3)	1.70(0.69)	1.72	5.56(2.26)	2.00
RP	8.20(1.8)	2.50(0.62)	2.48	6.70(1.9)	2.01

Table 8.3: Evaluation parameters

Acceptability (DPASR/PASR) ($[0-10]$)	Range ($[0-1]$)	Scope mean(stdev) ($[0-10]$)	Ranking correlation (DPASR/PASR) ($[-1 to 1]$)
6.91/6.72	0.52	7.76(1.8)	0.62/0.62

Table 8.4: Challenge phrases with normalized score <= 6.0

1. R G m P D n D P m G m R
2. R G m P D n , D P m G m R
3. R n D P , P
4. R , , R G m P , m G m R , G R S
5. Ṡ , Ṡ P , P R P m G m R , G R S
6. Ṡ Ṡ , P P , R R , ṇ S R G m
7. D n Ṡ Ṙ m D n Ṡ R G m P

Table 8.5: Challenge phrases with normalized score > 6.0

1. D Ṙ n , D P m D D n Ṡ Ṙ ,
2. P m D n Ṡ P , P R , , G m P m G m R ,
3. R n D D P P m G m R G R S ṇ S
4. Ṡ Ṙ n Ṡ , Ṡ P D m P , P
5. n Ṡ Ṙ Ġ ṁ n , Ṡ D , n D P ,‘
6. P R , G m P ,
7. D , D m D , D R G m P m D , D
8. D , n Ṡ , , D n Ṡ n , D P m D , D
9. R G R G m G m P m P D n , D P m G m R ,
10. P m D n Ṡ Ṙ Ġ ṁ Ṙ Ġ Ṙ Ṡ
11. D m D , m D n Ṡ Ṙ Ṡ n Ṡ D ,
12. Ṡ , Ṡ P , P R , R
13. Ṡ D , Ṡ n D P D m P , m
14. m D m D n Ṡ Ṙ n Ṡ

Chapter 9

Discussion

Contents

One of the main contributions of this work is the identification of the DPASR representation for gamakās which captures two kinds of movements that are superposed to construct the final gamakā form. Unlike the PASR form, transcription into DPASR requires familiarity with the genre. A few indicators such as performance on a *vīṇā* and the prescriptive notation helped reduce the dependency on musical expertise on the part of the transcriber. This chapter presents these guidelines for DPASR transcription and discusses the relevance of the representation to musicological analysis and pedagogy.

9.1 Guidelines for DPASR transcription

The DPASR representation requires more familiarity with the genre to use as a transcription target when compared to the PASR representation. Transcription into the PASR form can be handled mechanically, requiring little domain knowledge. Even if the transcriber does not have enough musical training to deal with the speed of the movements encountered in the performance being analyzed, time stretching tools that preserve pitch and transients can be used to slow the performance down to a speed at which it can be analyzed accurately.

Pitch estimation tools can help with identifying and measuring difficult focal pitches such as transients. Determining the "stage" component of the DPASR form of a gamakā is not as straight forward. Once the stage component is known, though, determining the dance component is as straight forward as transcribing into PASR. Below are some guidelines for determining the stage and dance components based on the experience gained in transcribing the reference performance.

9.1.1 The *vīṇā* as a guide

Vīṇā performances offer a closer view on the DPASR representation and if a different instrument is being used, it is perhaps helpful to consider how a particular movement might be performed on the *vīṇā*. This is because there are two kinds of movements possible on the *vīṇā* – sliding along the fret board and pulling on a string. In my transcription attempt, I found that the sliding movement corresponded in most circumstances to what my musical judgement told me was the stage pitch involved. Telling apart sliding movements from pulling movements takes familiarity with the instrument, but that is much easier than working without such a guiding principle. I did not have a video recording of the transcribed performance, but in cases where familiarity with the instrument cannot be assumed, a video recording with adequate closeups can help.

A first approximation of the stage component is, therefore, the movement executed by sliding on a *vīṇā* fret board. This heuristic fails when the performer uses the "*sphuritam*" and "*pratyāhatam*" techniques which need to be interpreted as dance components since the gamakā's melodic centre does not move in these cases, though the fret position changes. Furthermore, they might also involve pitch classes that are not permitted as part of the scale, but can occur as transients for the purpose of stresses.

The envelope of minima of continuous gamakā movements was also useful to consider in transcribing the stage component. This heuristic is related to the fact that pitch bending on the *vīṇā* can only be achieved from a lower to a higher pitch value, but knowledge of the instrument's techniques can help disambiguate cases in which a performer simulates a movement in the other direction using a lower fret as the base.

A first approximation of the dance component is the movement performed by pulling on the string, but for the exceptions mentioned above. On occasion, however, the performer may execute a tonal centre shift entirely by pulling on the string alone, as demonstrated in the movement $\dot{R}\dot{G}\dot{m}\dot{R}$ with the whole movement executed on the $\dot{R}$ fret. On such occasions, it can be useful to look for alternative renditions of the same phrase within the performance, perhaps in a different octave, and check whether the ambiguity is resolved.

One disadvantage of this *vīṇā* heuristic is that dance components tend to be biased towards positive amplitudes. Therefore it is important to make corrections to the first approximation using the *vīṇā* if the implied tone is not the one performed on a fret. For instance, a m might be rendered by pulling on the G fret, but sustaining on the m tone. In this case, m is the intended pitch and the dance component for the m must feature a negative amplitude that indicates that the movement around m involves G. The prescriptive notation can be of guidance in such cases, since the fact that the note is m would be indicated in it.

9.1.2 Prescriptive notation as a guide

The stage movements and the prescriptive notation were often correlated, indicating that the prescriptive notation may be useful to disambiguate stage movements where necessary. A couple of exceptions are in order though.

Certain phrasings characteristic of the rāgā may cause deviation from the prescription. In the case of Sahānā, we find R often being used as the base from which a prescription of G is realized through a movement. (Vijayakrishnan also describes gamakās of this kind in [Vijayakrishnan, 2007].) One may also consider the converse of this rule – that one characteristic of a rāgā is the various "stages" from which certain notes in the rāgā are rendered.

A second characteristic of stage movements corresponding to a prescribed note is that at most two focal pitches are necessary to describe it. That is, the correspondence between stage pitches and prescribed notes are either one-to-one or a movement between two pitches. This feature may be helpful in those circumstances where it appears that more than two melodic centres are involved. No exception to this rule was necessary in the transcription of the reference performance used for this work.

9.1.3 Transient pitches in dance movements

Finding dance movements with uniform amplitudes aids with the simplicity of model construction. I encountered many instances where a gamakā would need to be transcribed with a transient at the end. If the purpose of the transcription is fidelity, then this would need to be preserved. However, if the transcription is being done for the purpose of constructing an elaboration system that already accounts for inserting and removal of transient focal pitches in dance movements, then these can be omitted during transcription. For example, R may be rendered as $\overline{mG,,mR,,G}$. The core of this movement is $\overline{mG,,mR,,}$ and the last gesture towards G may be added in contexts where S (or such a pitch class lower than the G) would follow R. This transient G can be added as a post processing step. Additionally, dance movements with uniform amplitudes help clarify rules that extend a gamakā over longer period by repetition.

9.2 DPASR and musicology

This work was motivated by its potential applications to pedagogy, but the techniques developed may be useful for some kinds of musicological analysis as well.

9.2.1 Musicological analysis

The stage component of the DPASR representation is of an intermediate level of complexity between the prescriptive notation and descriptive notation used among musicologists studying the genre. The descriptive notation provides a linear temporal description of all the movements constituting a gamakā, not unlike the PASR representation which is even more detailed than the description notation for the sake of synthesis. The transformation of prescriptive notation into descriptive notation can itself be viewed as an elaboration problem. On the surface, this problem seems amenable to systems based on deterministic grammars, which has therefore been the default direction of research on gamakās thus far. However, the complexity of the resulting grammars for such systems obscure the simpler principles based on competing violable constraints on continuous gamakās. This suggests that the stage component may capture the music at a level of detail that not only makes interpretation of the music less ambiguous but can also provide a style-independent scaffolding to describe the movements involved. Expressive style can be delegated to the dance component.

9.2.2 Pedagogy

The pitch positions that feature in the stage component may not always correspond to the *svara*s declared in the prescriptive notation. This can be a problem for communicating the stage component through singing in normal teaching situations since the uttered *svara*s can “feel wrong”. However, this limitation may not apply to cases where the student is musically adept in a different genre. In such cases, the ability to describe the movements to a degree of detail beyond that given in the prescriptive notation, but still less than that offered by the descriptive notation can be a useful stepping stone towards communicating the rules for applying gamakās in various contexts.

Chapter 10

Conclusion

Contents

This work studies the principles underlying gamakās of Carnatic music using an analysis-by-synthesis approach. The method used was to build a computational model encoding the expertise required to interpret sparse prescriptive notation of the genre using appropriate gamakās, based on a reference composition and a reference performance of it. The optimality theoretic (OT) model of least violation of competing constraints was found to be a good fit to describe gamakā selection and sequencing within phrases. A new two-component representation for gamakās, one of whose components encodes an intermediate level of detail between prescriptive and descriptive notation, was shown to be useful in simplifying the rules behind gamakā selection and sequencing. Interviews conducted with experts showed that the elaboration systems built for this work could adapt to phrases not found in the reference performance, a feature attributable to procedures developed for adapting gamakās to different speeds.

10.1 Review

The term "elaboration" was introduced to refer to the process of interpreting each notated entity (*svara*) in terms of gamakās, and systems that performed such elaboration were terms "elaboration systems". Two types of elaboration in music systems were identified namely *expressive* and *structural* elaboration. Singing synthesis systems and F0 contour modeling in speech synthesis were presented as examples of expressive elaboration. Jazz melody generation systems and the "automatic gamakam" system of Gaayaka were presented as examples of structural elaboration systems.

The analysis-by-synthesis approach followed in this work is similar to how rules for singing synthesis were iteratively determined in [Berndtsson, 1995]. The common steps are that of defining rules controlling the synthesis of a performance, generating performances and refining the rules to improve the result. The additional step necessary for this work is that of using a transcription of a reference performance of a prescriptive "score" as the starting point, since the culture admits multiple interpretations of notated music. Re-synthesis is used in this work to reduce the subjectivity involved in the transcription process, which is otherwise a traditional approach in ethnomusicological research.

The DPASR representation introduced in this work separates gamakā movements into a slow moving "stage" component and an expressive "dance" component. Such an additional level of detail where the slow moving component has been useful in speech intonation modeling as well as Jazz melody generation. Intonation in speech is described using "F0 mean" and "F0 shape" parameters, and melody generation for Jazz is constrained by a moving "melodic centre", itself determined to fit a given harmonic context. By comparison with the "F0 mean" and the concept of "melodic centre", the "stage" component may be expected to play an analogous analytical role for Carnatic music.

Furthering the analogy with speech where a given spoken form can be adapted to different speeds while being perceived as "the same", slower speed gamakās were shown to be adaptable to higher speed contexts. This also had a practical implication that duration could be factored out of the local melodic context necessary for elaboration, which resulted in the elaboration systems being applicable to phrases not part of the reference composition that they were derived from. This was borne out in the expert evaluation where the system interpretations of challenge phrases given by the participants scored comparably to the reference performance clips, in the aggregate. The kind of gamakā combination necessary to achieve this extension is analogous to Cope's work on recombinant music where known compositional material is repurposed for new contexts [Cope, 1989].

Hierarchical selection and competing lateral constraints on gamakās sequencing were both identified as components required for elaborating prescriptive notation in Carnatic music. Gaayaka features hierarchical selection of possible gamakās based on a rich description of local melodic context, but delegates phrase-level gamakā selection to the user. Phrase-optimal gamakā selection by resolving the lateral constraints using a DAG satisfied the computational task implied in Vijayakrishnan's optimality theoretic (OT) formulation of

Carnatic music by simpler means [Vijayakrishnan, 2007]. While the source of phrase-level constraints on gamakās to use in a strict OT formulation is unspecified, it was possible in this work to use a method of "discrimination matching" to identifying candidate lateral constraints from the reference performance.

In the evaluation studies with experts, only synthesized material was presented for the purpose of comparing the musical choices made by two different elaboration systems. This comparative approach is similar to the method used by Berndtsson where individual rules are turned off and on and audience response to the results are observed, except that individual interviews were conducted so that more detailed comments and clarifications can be obtained. Cope, on the other hand, uses human performers to play sheet music generated by EMI.

On the whole, this work has shown that an analysis-by-synthesis approach to studying the foundations of Carnatic music is a productive line of research. In particular, the transcription and discrimination matching techniques developed for building the elaboration system for Sahānā can be repeated for other rāgās using other reference performances towards producing more general models. The representations developed for gamakās also hold potential for reuse across the genre.

I now discuss some limitations of this research and how they can be addressed in future work in this area.

10.2 Future work

The methods developed in this research to build an elaboration system for short metric phrases in the rāgā Sahānā come with several limitations that suggest ways to extend and improve on this work in the future.

The rules and preferences encoded in the systems developed for this research may be unique to the specific musical context of the *varṇam* "Karuṇimpa" or Sahānā and may not be applicable to other rāgās as is. This is so even if it can be argued that a principle such as "bias towards continuity of the stage component" which does not refer to specific gamakās suggests generality. One way to determine the generality of the encoded rules and preferences is to repeat this for other rāgās, while still restricting to the *varṇam* category of compositions.

The *varṇam* category of compositions serves as raw material for characteristic movements of a rāgā, but experts of the genre hold that the lyrical *kṛti* forms offer much more scope for expression of "rāgatva" (translation: "rāgā's nature"). The range of variations that an artist may show in a particular melodic context in a *krti* is higher than in a *varṇam*, particularly if the variations are aggregated across multiple performances, artists and styles. The study of *krti* forms along these lines will therefore require dealing with a potentially much larger space of gamakā variations, necessitating different computational techniques. Furthermore, phonological information about the lyrics of *krti*s may be a necessary component of such systems to function. Beyond *krti*s, another major aspect of rāgā is free-time *ālāpana* improvisation. The only result from this work that suggests that a unified rāgā model for *ālāpana* and compositions is possible is the speed adaptation rules for gamakās. Therefore more work is required to cover the melodic space of even the one rāgā considered in this work.

The technique of treating gamakā sequencing constraints as the optimal path selection through a DAG is limited in temporal extent. In this work, the optimization was applied to short phrases of around 10 notes or fewer. The effectiveness of the algorithm will reduce as the phrase length increases. In longer phrases, gamakā preferences will no longer be strictly local since the tāḷā becomes relevant. Introducing tāḷā dependence in the function that scores gamakā sequences is possible, but adds a new dimension to every component therein, making it difficult to perform manual iterations to develop models and requiring a larger body of reference material to draw on. If automatic precise transcription were possible in the large, this problem will become amenable to machine learning techniques.

The results of this work are strictly about gamakās pertaining to the *vīṇā*. Since instruments have different performance constraints which influence the choice of gamakās and speed of playing, not all of the findings of this work may apply to other instruments or vocal music. Furthermore, other styles of *vīṇā* playing exist which make use of more ornamental gamakā techniques such as rapid string jumping and vibrato which are not included in the models developed. Some of these techniques require improvements to the PASR and DPASR representations so that they can be modeled in a similar fashion.

The evaluation studies pointed out that modeling stopping and plucking techniques on the *vīṇā* is an important area to improve on. The use of uniform plucking strength in the resynthesized examples did not play well with the expectation of appropriate dynamics in certain phrases of Sahānā for some of the participants. Though this is itself an interesting result, given that dynamics is one of the least discussed aspects in the musicological literature of Carnatic music, the importance of dynamics cannot now be denied. Similarly, the simplistic proportional stopping model used in the elaboration systems did not satisfy some of the participants who were more familiar with the instrument. Progress in modeling these aspects will therefore improve the musical quality of the output and hence the acceptance level of the system.

The evaluation study involved playing a series of snippets for the participants to score and comment on. The time taken and the concentration required on their behalf for this prevented a more thorough evaluation of the *range* and *scope* of the system. Perhaps a different study design that focused on only one aspect of the evaluation such as challenge phrases, or choosing more variations generated by each of the PASR and DPASR based systems can be used in the future to collect more reliable feedback.

Quoting computer scientist Alan J. Perlis,

> "The only constructive theory connecting neuroscience and psychology will arise from the study of software."

The similar motivating belief behind this work is that an important, if not the "only", way to understand musical cultures and the music they produce is through the construction of software models of the processes of producing them. The belief also expresses a hope that a musicology founded on such modeling may add knowledge about aspects that even those intimate with the genre may be unaware of. This work is a small step in that direction for Carnatic music.

Bibliography

[Battey, 2004] Battey, B. (2004). Bzier spline modeling of pitch-continuous melodic expression and ornamentation. *Computer Music Journal*, 28(4):25–39. ArticleType: research-article / Full publication date: Winter, 2004 / Copyright 2004 The MIT Press.

[Bel, 1998] Bel, B. (1998). Migrating musical concepts: An overview of the bol processor. *Computer Music Journal*, 22(2):5664.

[Bel, 2005] Bel, B. (2005). Two algorithms for the instanciation of structures of musical objects.

[Bel and Kippen, 1992] Bel, B. and Kippen, J. (1992). Bol processor grammars. *Understanding Music with AIPerspectives on Music Cognition*, page 366401.

[Beller et al., 2005] Beller, G., Schwarz, D., Hueber, T., and Rodet, X. (2005). A hybrid concatenative synthesis system on the intersection of music and speech. *Journes dInformatique Musicale (JIM)*, page 4145.

[Bellini and Nesi, 2001] Bellini, P. and Nesi, P. (2001). WEDELMUSIC format: an XML music notation format for emerging applications. *Web Delivering of Music, 2001. Proceedings. First International Conference on*, pages 79–86.

[Berndtsson, 1995] Berndtsson, G. (1995). The KTH rule system for singing synthesis. *STL-QPSR*, 36(1):1–22.

[Berndtsson, 1996] Berndtsson, G. (1996). The KTH rule system for singing synthesis. *Computer Music Journal*, 20(1):7691.

[Biles, 1994] Biles, J. (1994). GenJam: a genetic algorithm for generating jazz solos. In *Proceedings of the International Computer Music Conference*, page 131131.

[Boersma and Weenink, 2005] Boersma, P. and Weenink, D. (2005). Praat: doing phonetics by computer [Computer program]. *Version*, 5:21.

[Boulanger, 2000] Boulanger, R. C. (2000). *The Csound book: perspectives in software synthesis, sound design, signal processing, and programming.* The MIT Press.

[Cope, 1987] Cope, D. (1987). An expert system for computer-assisted composition. *Computer Music Journal*, 11(4):30–46. ArticleType: research-article / Full publication date: Winter, 1987 / Copyright 1987 The MIT Press.

[Cope, 1989] Cope, D. (1989). Experiments in musical intelligence (EMI): non-linear linguistic-based composition. *Journal of New Music Research*, 18(1):117139.

[Cope, 1991a] Cope, D. (1991a). *Computers and Musical Style.* A-R Editions.

[Cope, 1991b] Cope, D. (1991b). Recombinant music: using the computer to explore musical style. *Computer*, 24(7):2228.

[Cope, 1992] Cope, D. (1992). Computer modeling of musical intelligence in EMI. *Computer Music Journal*, page 6983.

[Cope, 2000] Cope, D. (2000). *The Algorithmic Composer.* A-R Editions, Inc.

[Cuthbert and Ariza, 2010] Cuthbert, M. S. and Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In *Int. Society for Music Information Retrieval Conf.(ISMIR 2010).*

[Desain and Honing, 1992] Desain, P. and Honing, H. (1992). *Music, Mind, and Machine: Studies in Computer Music, Music Cognition, and Artificial Intelligence.* Thesis Pub.

[Dikshitar, 1904] Dikshitar, S. (1904). *Sangita Sampradaya Pradarsini.* Ettayapuram Samasthanam.

[Dodge and Jerse, 1985] Dodge, C. and Jerse, T. A. (1985). *Computer music: synthesis, composition, and performance.* Macmillan Library Reference.

[Downie, 2003] Downie, J. S. (2003). Music information retrieval. *Annual review of information science and technology*, 37(1):295340.

[Eppstein, 1998] Eppstein, D. (1998). Finding the k shortest paths. *SIAM J. Comput.*, 28(2):652673.

[Friberg et al., 2006] Friberg, A., Bresin, R., and Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, 2(2):145161.

[Fujisaki, 1981] Fujisaki, H. (1981). Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations. In *Proceedings of the 4th FASE Symposium on Acoustics and Speech*, volume 2, page 5770.

[Gillick et al., 2010] Gillick, J., Tang, K., and Keller, R. M. (2010). Machine learning of jazz grammars. *Computer Music Journal*, 34(3):5666.

[Gopalam, 1991] Gopalam, S. (1991). *Facets of Notation in South Indian Music*. Sundeep Prakashan, Delhi.

[Hashida et al., 2012] Hashida, M., Hirata, K., and Katayose, H. (2012). Renconmusic - music performance rendering contest for computer systems. http://renconmusic.org/.

[Horii, 1989] Horii, Y. (1989). Acoustic analysis of vocal vibrato: A theoretical interpretation of data. *Journal of voice*, 3(1):3643.

[Huron, 1993] Huron, D. (1993). The humdrum toolkit. *Available online at music-cog. ohio-state. edu/Humdrum/index. html*.

[Huron, 2002] Huron, D. (2002). Music information processing using the humdrum toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26(2):11–26. ArticleType: research-article / Full publication date: Summer, 2002 / Copyright 2002 The MIT Press.

[Iyengar, 1965] Iyengar, R. (1965). *Sri Kriti Mani Malai*. Sabarmati, Madras, 2 edition.

[Jan, 2004] Jan, S. (2004). Meme hunting with the humdrum toolkit: Principles, problems, and prospects. *Computer Music Journal*, 28(4):68–84.

[Jayalakshmi, 2002] Jayalakshmi, R. S. (2002). *Gamakas explained in Sangita-sampradaya-pradarsini of Subbarama Diksitar*. PhD thesis, Madras University.

[Keller and Morrison, 2007] Keller, R. and Morrison, D. (2007). A grammatical approach to automatic improvisation. In *Proceedings, Fourth Sound and Music Conference, Lefkada, Greece, July.Most of the soloists at Birdland had to wait for Parkers next record in order to find out what to play next. What will they do now.*

[Kippen and Bel, 1989] Kippen, J. and Bel, B. (1989). Can a computer help resolve the problem of ethnographic description? *Anthropological Quarterly*, 62(3):131–144. Article-Type: primary_article / Issue Title: Expert Systems Applications in Anthropology, Part

2 / Full publication date: Jul., 1989 / Copyright 1989 The George Washington University Institute for Ethnographic Research.

[Kippen and Bel, 1992] Kippen, J. and Bel, B. (1992). Modelling music with grammars: formal language representation in the bol processor. *Computer Representations and Models in Music, Ac. Press ltd*, page 207232.

[Kirke and Miranda, 2009] Kirke, A. and Miranda, E. R. (2009). A survey of computer systems for expressive music performance. *ACM Computing Surveys (CSUR)*, 42(1):3.

[Krishnaswamy, 2003] Krishnaswamy, A. (2003). Application of pitch tracking to south indian classical music. In *Proc IEEE ICASSP*.

[Leman, 1996] Leman, M. (1996). *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*. Springer.

[Mahesh, 2007] Mahesh, S. (2007). *Raga, an Insight: abstract from my Ph.D. thesis, "An exploration of the concept of raga in Karnatak music"*. Madurai.

[Mallikarjuna Sharma, 2007] Mallikarjuna Sharma, A. (2007). Kampitas and gamakas. In *AMS Easy Methods*, pages 25–31. AMS Foundation.

[Mazzoni and Dannenberg, 2000] Mazzoni, D. and Dannenberg, R. (2000). Audacity: Free audio editor and recorder. http://audacity.sourceforge.net/.

[McCartney, 1996] McCartney, J. (1996). SuperCollider, a new real time synthesis language. In *Proceedings of the International Computer Music Conference*, page 257258.

[Monaghan, 2002] Monaghan, A. (2002). State-of-the-art summary of european synthetic prosody R&D. In Keller, E., Bailly, G., Monaghan, A., Terken, J., and Huckvale, M., editors, *Improvements in speech synthesis: COST 258: the naturalness of synthetic speech*, pages 93–103. Wiley.

[Pennycook et al., 1993] Pennycook, B., Stammen, D. R., and Reynolds, D. (1993). Toward a computer model of a jazz improvisor. In *PROCEEDINGS OF THE INTERNATIONAL COMPUTER MUSIC CONFERENCE*, page 228228.

[Pesch, 1993] Pesch, L. (1993). *Raga Dhana: An Alpha-Numerical Directory of Ragas*. Natanakairali, Irinjalakuda (Kerala), 2 edition.

[Pickens, 2001] Pickens, J. (2001). A survey of feature selection techniques for music information retrieval. In *Proceedings of the 2nd International Symposium on Music Information Retrieval (ISMIR)*.

[Portele and Heuft, 1998] Portele, T. and Heuft, B. (1998). The maximum-based description of f0 contours and its application to english. In *Fifth International Conference on Spoken Language Processing*.

[Prame, 1994] Prame, E. (1994). Measurements of the vibrato rate of ten singers. *Journal of the Acoustical Society of America*, 96(4):19791984.

[Prince and Smolensky, 2004] Prince, A. and Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Wiley Online Library.

[Ramalho and Ganascia, 1994] Ramalho, G. and Ganascia, J. G. (1994). Simulating creativity in jazz performance. In *Proceedings of the National Conference on Artificial Intelligence*, page 108108.

[Ramani, 2001] Ramani, N. (2001). *Chennai Musicians-dancers Directory: Classical*. Columbus Publications.

[Roads, 1996] Roads, C. (1996). *The computer music tutorial*. The MIT press.

[Rowe, 2004] Rowe, R. (2004). *Machine Musicianship*. The MIT Press.

[Schwarz, 2007] Schwarz, D. (2007). Corpus-based concatenative synthesis. *Signal Processing Magazine, IEEE*, 24(2):92104.

[Schwarz et al., 2006] Schwarz, D., Beller, G., Verbrugghe, B., Britton, S., et al. (2006). Real-time corpus-based concatenative synthesis with catart. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06),(Montreal, Quebec, Canada)*, page 279282.

[Schwarz et al., 2000] Schwarz, D. et al. (2000). A system for data-driven concatenative sound synthesis. In *Digital Audio Effects (DAFx)*, page 97102.

[Seeger, 1958] Seeger, C. (1958). Prescriptive and descriptive music-writing. *The Musical Quarterly*, 44(2):184–195.

[Shankar, 1979] Shankar, V. (1979). *Shyama Sastry's Compositions*. C. S. Ayyar, Madras, 2 edition.

[Shankar, 1983] Shankar, V. (1983). *The Art and Science of Carnatic Music*. The Music Academy Madras.

[Subramanian, 1985a] Subramanian, K. S. (1985a). An introduction to the vina. *Asian Music*, 16(2):7–82. ArticleType: research-article / Full publication date: Spring - Summer, 1985 / Copyright 1985 University of Texas Press.

[Subramanian, 1985b] Subramanian, K. S. (1985b). *South Indian Vina Tradition and Individual Style*. Ph.D. thesis, Wesleyan University, Connecticut, USA.

[Subramanian, 1999] Subramanian, M. (1999). Carnatic music and the computer. *Journal of Sangeet Natak Akademi, New Delhi*, pages 16–24.

[Subramanian, 2002] Subramanian, M. (2002). Analysis of gamakams of carnatic music using the computer. *Sangeet Natak*, XXXVII(1):26–47.

[Subramanian, 2009a] Subramanian, M. (2009a). Carnatic music - automatic computer synthesis of gamakams. *Sangeet Natak*, XLIII(3).

[Subramanian, 2009b] Subramanian, M. (2009b). GAAYAKA - carnatic music notation player. http://carnatic2000.tripod.com/gaayaka6.htm.

[Subramanian et al., 2011] Subramanian, S., Wyse, L., and McGee, K. (2011). Modeling speed doubling in carnatic music. In *Proceedings of the International Computer Music Conference*, pages 478–485, University of Huddersfield, UK.

[Subramanian et al., 2012] Subramanian, S., Wyse, L., and McGee, K. (2012). A two-component representation for modeling gamakās of carnatic music. In *Proceedings of the 2nd CompMusic workshop*, pages 147–152, Bahçeşehir Üniversitesi, Istanbul, Turkey.

[Sundberg, 1994] Sundberg, J. (1994). Perceptual aspects of singing. *Journal of Voice*, 8(2):106122.

[Sundberg et al., 1983] Sundberg, J., Askenfelt, A., and Fryden, L. (1983). Musical performance: A synthesis-by-rule approach. *Computer Music Journal*, 7(1):37–43.

[Taylor, 1994] Taylor, P. (1994). The rise/fall/connection model of intonation. *Speech Communication*, 15(1-2):169186.

[Taylor, 1998] Taylor, P. (1998). The tilt intonation model. In *Fifth International Conference on Spoken Language Processing*.

[Taylor and Black, 1994] Taylor, P. and Black, A. W. (1994). Synthesizing conversational intonation from a linguistically rich input. In *The Second ESCA/IEEE Workshop on Speech Synthesis*.

[Todd and Loy, 1991] Todd, P. M. and Loy, D. G. (1991). *Music and connectionism*. The MIT Press.

[Typke et al., 2005] Typke, R., Wiering, F., and Veltkamp, R. C. (2005). A survey of music information retrieval systems.

[Ulrich, 1977] Ulrich, J. W. (1977). The analysis and synthesis of jazz by computer. In *Fifth International Joint Conference on Artificial Intelligence*, page 865872.

[Vercoe, 1986] Vercoe, B. (1986). Csound: A manual for the audio processing system and supporting programs. *Program Documentation. Cambridge, Massachusetts: MIT Media Lab*.

[Vijayakrishnan, 2007] Vijayakrishnan, K. G. (2007). *The Grammar of Carnatic Music*. Mouton De Gruyter, Har/MP3 edition.

[Vijayakrishnan, 2009] Vijayakrishnan, K. G. (2009). The function and scope of notation in carnatic music. *Journal of the Indian Musicological Society*, (40):140–145, 272.

[Viswanathan, 1977] Viswanathan, T. (1977). The analysis of rāga ālāpana in south indian music. *Asian Music*, 9(1):13–71. ArticleType: research-article / Issue Title: Second India Issue / Full publication date: 1977 / Copyright 1977 University of Texas Press.

[Widdess, 1979] Widdess, D. R. (1979). The kudumiyamalai inscription: a source of early indian music in notation. *Musica Asiatica*, pages 115–50.

[Wiggins et al., 1993] Wiggins, G., Miranda, E., Smaill, A., and Harris, M. (1993). A framework for the evaluation of music representation systems. *Computer Music Journal*, 17(3):31–42. ArticleType: research-article / Full publication date: Autumn, 1993 / Copyright 1993 The MIT Press.

Appendix A

Formal definitions and notations

Where precision is required in the interpretation of a term, the following definitions may be used for the purpose of this work.

Pitch An integer in the range 0–48 to cover 3 octaves. There are 16 pitch classes per octave redundantly representing 12 tones per octave. $r(p) \in [0, 36)$ denotes the physical (or "real") pitch corresponding to p.

Duration An integer in the range 1–16 (upper limit is not fixed).

Note A pitch-duration pair written as $n = (p, d)$. In Carnatic music, the term "*svara*" would better fit this and the term "note" is usually used to denote a group of *svara*s rendered in a single movement.

Phrase A sequence of k notes, written as $[n_1, n_2, ..., n_k]$. Phrases constituting the input for this system don't usually exceed 15 notes in length.

Gamaka Connective pitch movements between tonal positions constituting a scale. This includes both continuous smooth movements between tonal positions, as well as discrete step movements. In the elaboration system, a gamakā, written g_i, is represented as an integer indexing into a table of continous pitch forms extracted from a given reference performance and its prescriptive notation.

Tāḷā A cyclic time structure of integer period T imposed on a performance. A typical value for T is 32 beats per cycle.

Rāgā Ascent and descent constraints on pitch patterns. Usually includes characteristic and prohibited gamakas as well.

Prescriptive notation The discrete description of melody used in the conventional published notation of Carnatic music. This document uses the *svarasthāna* letters "S r R g G m M P d D n N" to represent the 12 tones constituting an octave. Higher octave tones are indicated by a dot above the corresponding letter, such as $\dot{S}\dot{R}\dot{G}$. Lower octave tones are indicated by a dot below the corresponding letter, such as $\underset{\cdot}{P}\underset{\cdot}{D}\underset{\cdot}{n}$. A fragment of prescriptive notation is presented as, for example, $(RGmPmDn\dot{S})$.

Descriptive notation Notation that approximates the internal movements of gamakās in terms of discrete movements between intermediate tonal positions. An over-line is used in this case. For example, the G in the prescriptive notation fragment (GRS) can be rendered using the movement $\overline{GR,m\ \ R,,G\ \ S,,,}$.

Focal pitch Quasi-stationary tonal position that occurs within a gamakā. Though such tonal positions often correspond to one of the 12 tones of the octave, this is not necessarily so. Therefore focal pitch values are represented numerically in units of semitones.

PASR representation Representation of a gamakā as a sequence of (p, a, s, r) tuples where p is a focal pitch in semitones, a is the "attack time", which is the time spent moving towards this focal pitch from the preceding one, s is the "sustain time", which is the time spent at the focal pitch and r is the "release time", which is the time spent moving away from this focal pitch towards the following one. Note that the interpolation curve necessary to perform a gamakā expressed thus is abstracted away.

DPASR representation Representation of a gamakā as the sum of two pitch curves each represented in PASR form. The component PASR curves are referred to as "stage" and "dance". "Stage" is a slow moving component compared to "dance". While "stage" may have at most two focal pitches for each note of the prescriptive notation, the "dance" component many have an arbitrary number of them.

Trigram context The local prescriptive melodic context of a *svara*. For example, the *svaras* in the phrase fragment $(DPmGmR)$, have the trigram contexts $(-DP)$, (DPm), (PmG), (mGm), (GmR), and $(mR-)$ respectively, with "$-$" standing for the phrase boundary. The generalized trigram context for a note n_i is $c(n_i) = (n_{i-1}, n_i, n_{i+1})$. The duration free trigram consists only of pitch information – i.e. $c_p(i) = (p_{i-1}, p_i, p_{i+1})$.

Appendix B

Varṇam: "karuṇimpa"

Composer:	Tiruvotriyūr Tyāgayyar
Rāgā:	Sahānā
Tāḷā:	Ādi
Ārōhaṇa:	S R G m P m D n Ṡ
Avarōhaṇa:	Ṡ n D P m G m R G R S

Pallavi:	*karuṇimpa idi manci taruṇamu sāmi*
Anupallavi:	*parula vēḍalēnu nā pāli śrī vēṇugōpāla dēva*
Caraṇam:	*kṛpa jūḍumi ī vēḷa*

Pallavi

P	,	m	,	G	,	G	m		R	,	G	R	S	,	,	,	||
ka		*ru*		*ṇim*									*pa*				
ṇ	S	R	G	R	S-	ṇ	S	|	Ḍ	P̣	ṃ	Ḍ	,	ṇ	S	R	||
i				*di*					*man*			*ci*					
P	m	G	m	R	G	R	S		,-	R	G	m	P	m	D	,	||
ta							*ru*								*ṇa*		
n	,	Ṡ—	P	,	,-	R	G	|	m	D	P-	P	m	G	m	R	||
			mu			*sā*								*mi*			

Anupallavi

D	n	D	-D	P-	P	D	P		P	m-	D	P	m-	G	m	R	||
pa			*ru*		*la*						*vē*			*ḍa*			
G	m	P-	ṇ	S	R	G	m	|	P	,	m	,	D	n	Ṡ	,	||
lē			*nu*												*nā*		
D	n	Ṡ	Ṙ	,	Ṙ-	n	Ṡ		Ṙ	Ġ	ṁ	Ṙ	,	Ġ	Ṙ	Ṡ	||
pā			*li*			*śrī*				*vē*				*ṇu*			
Ṙ	n	,-	Ṡ	D	,-	n	D	|	P-	D	P	P	m-	G	m	R	||
gō			*pā*			*la*				*dē*				*va*			

Muktāyi svaram

P	,	,	P	m	G	m	,		,	R	G	m	P	m	G	m	||
R	,	,	G	R	S	ṇ	S	|	R	G	m	P	m	D	n	Ṡ	||
P	,	-D	n	Ṡ	n	Ṡ	Ṙ		Ġ	ṁ	Ṙ	,	Ġ	Ṙ	Ṡ	n	||
Ṡ	Ṙ	n	Ṡ	,	Ṡ	P	D	|	m	P	,	ṇ	S	R	G	m	||

(*karuṇimpa*)

Caraṇam

D	,	D	,	,	,-	n	,		,	,	n	Ṡ	D	,	n	D	||
kṛ		*pa*				*jū*									*du*		
P	,	,	,	,	,-	D	P	|	m	G	m	R	G	m	P	m	||
mi						*ī*				*vē*		*ḷa*					

Ciṭṭa svaram

(1)	P , , ,	, , m ,		, , D ,	P , m ,	||
	G , , m	, , R ,	|	, G , m	, P , m	||

(*kṛpa*)

(2)	D , P m	G m R ,		G m P- ṇ	S R G m	||
	P , m G	, m R ,	|	G R S- R	G m P m	||

(*kṛpa*)

(3)	P D m P	G m- R G		m P m G	m R G R-	||
	S R ṇ S	Ḍ ṇ S R	|	G m- S R	G m P D	||
	P m- D n	Ṡ R S R		n Ṣ- Ṙ Ġ	ṁ Ṙ Ġ Ṙ	||
	Ṡ Ṙ- n Ṙ	Ṡ D n Ṡ	|	D n D- P	D m P m	||

(*kṛpa*)

(4)	n , , Ṡ	Ṙ Ṡ- n Ṙ		Ṡ n Ṡ n	D , n D	||
	P , , R	, , G m	|	P m- G m	R G R S	||
	R G m P	, P m P		m D , D	P D n D	||
	P D m P	G m- R G	|	R S , R	G m P ,	||
	P m G m	R ,- D P		m G m R	,- G m P	||
	ṇ S , R	G m P m	|	D , n Ṡ	Ṙ , Ṙ ,	||
	Ġ ṁ Ṙ Ġ	Ṙ Ṡ- n Ṙ		Ṡ- n Ṡ D	P m- D n	||
	Ṡ Ṡ ,- P	P ,- R R	|	,- ṇ S R	G m P m	||

(*kṛpa*)

References

1. Smt. Rajeswari Padmanabhan's *vīṇā* rendition in the album "Surabhi" (primary)
2. Shivkumar Kalyanaraman's notation
   `http://www.shivkumar.org/music/varnams/karunimpa-sahana-adi-varnam.pdf`

3. http://www.karnatik.com/c3007.shtml

Appendix C

Sahānā

Author's note: This appendix is included to provide a sample of typical *rāga lakśaṇā* presentations in the musicology of the genre. The use of descriptive notation (in tiny font above regular notation) is unusual for such presentations. The content is an excerpt adapted, with permission, from [Mahesh, 2007, p. 42-43]. Minor edits were made to the text for clarity and some formatting errors in the publication were corrected. Music notation was reformatted for use here, but not altered otherwise. The section on Sahānā in the source also tabulates 14 phrases along with their prescriptive and descriptive notation, which are not included here.

		SGR		mGm				ṠSDDṠD		
ārōhaṇa:	S	R	G	m	P	m	D	n	Ṡ	
		D,ṠD,Ṡ		DP		m,,G,	mGm,	m,GR	mRm	
avarōhaṇa:	Ṡ	n ,	D	P	m	G ,	m	R	G	R S

Sahānā is a rāgā comprising of just a few phrases and it is defined by them. Each phrase is pregnant with rāgatva. Almost all compositions in Sahānā have similar tunes or "varṇameṭṭu". Sahānā is almost synonymous with the phrases analyzed, but one could be creative, not by creating new phrases, but in the variation in their alignment.

For example, the phrase (S, P, m, D) could be either an end phrase or a germinating phrase for succeeding phrases.

P,m Gm, R GGR,	-	SP, m D
R GmPm Gm RGGR	-	SP, m D
Ṙ n, DPm Gm RGGR	-	SP, m D
n, n, DPm Gm RGGR	-	SP, m D
ṇSRGm	-	P, m D

In Sahānā, phrases proceed as –

PmDnḊ, mPmD,
mPmGmR, mPDm GmR

Descending phrases figure as –

Ṡ n D P	-	(RGm Pm.....)
DPm	-	GmRGRS
PmGm	-	RGRS
n, DPmGm	-	RGRS

The phrase $(RGmPDn, DPm)$ figures at least once, and at the most twice in every *ālāpana*. This phrase, however, does not figure in the basic tune of any composition. It is incorporated in some Padam-s – ex: *Mogaḍuci* – as a variation of the previous line, and does not seem to have been part of the basic varṇameṭṭu. Some phrases, conceived in the *ālāpana*, do not lend themselves to be rendered as *svaras*. The *svaras* have been designated to such phrases only for the purpose of notation.

Ex.	R	G	G	R		
	ṇ	m,	G	R	S	Ḍ

End phrases

R GG R S
RG, R R, R,
P m D
n D P ... R
RSṇS .

Appendix D

Evaluation format

D.1 Biographical information

Age:________ Years of experience:________

D.2 Set 1 - Pallavi of "Karuṇimpa"

Sound: Is the snippet audible? Is the sound of adequate quality for commenting on? Is the synthetic instrument used an adequate medium for expressing the music?

Identification: Can participant identify the composition and performance style?

Rating: Rate the overall quality of the synthesis on a scale of $[0 - 10]$, where –

0	fail
2-3	poor
5	ok
8-9	good
10	excellent

What criteria did the participant use to rate? Specific comments?

D.3 Set 2 - First speed phrases

Instructions for participant: In this set, three "first speed" phrases will be presented in synthesized form. Three variant interpretations will be presented for each phrase. Rate each interpretation on the scale of $[0 - 10]$ presented earlier.

Instructions for interviewer: Repeat scale description. Present variations of each phrase in random order. Ask participants to evaluate each interpretation as though it were performed by a student. Tell participant that they may evaluate using their own musical sensibilities without regard to those of other musicians. If they ask to repeat a particular snippet, do so. Ask for corrections or improvement suggestions. Use questions from Appendix E where appropriate.

Phrase 1: `^da2 ni2 da2 ^da2 pa ^pa da2 pa ^pa ma1`

DPASR: [ ] PASR: [ ] RP: [ ]

RESPONSES:

Phrase 2: `^da2 ^ni2 ^sa+ ^ni2 ^sa+ ^ri2+ ^ga3+ ^ma1+ ^ri2+:2 ^ga3+ ^ri2+ ^sa+`

DPASR: [ ] PASR: [ ] RP: [ ]

RESPONSES:

Phrase 3: `^da2 pa ma1 ^ga3 ma1 ri2 ^ga3 ma1 ^pa`

DPASR: [ ] PASR: [ ] RP: [ ]

RESPONSES:

D.4 Set 3 - Second speed phrases

Instructions for participant: In this set, three "second speed" phrases will be presented in synthesized form. Three variant interpretations will be presented for each phrase. Rate each interpretation on the scale of $[0 - 10]$ presented earlier.

Instructions for interviewer: Repeat scale description if necessary. Otherwise same as for the "first speed" phrases.

Phrase 1: `^pa ^ma1 ^ga3 ^ma1 ^ri2:2 ^da2 ^pa ^ma1 ^ga3 ^ma1 ^ri2:2`

DPASR: [ ] PASR: [ ] RP: [ ]

RESPONSES:

Phrase 2: `^da2:2 ^pa ^ma1 ^ga3 ^ma1 ^ri2:2 ^ga3 ^ma1 ^pa`

DPASR: [ ] PASR: [ ] RP: [ ]

RESPONSES:

Phrase 3: `^ri2 ^ga3 ^ma1 ^pa:2 ^pa ^ma1 ^pa ^ma1 ^da2:2 ^da2`

DPASR: [ ] PASR: [ ] RP: [ ]

RESPONSES:

D.5 Set 4 - Challenge phrases

(The interface shown in appendix F was used for this section.)

Instructions for participant: In this set, you will have to provide two phrases in Sahana for each of which four variations will be generated and played to you. Two of these variations will be in “first speed” and two will be in “second speed”. The phrases must be metrical. You can think of these as “svara kalpana” or “niraval” phrases instead of as “alapana”. I will translate your phrases into a notation that can be input into the computer. Rate the variations presented to you on the same scale of $[0 - 10]$ used in earlier sets.

Instructions for interviewer: Present the variations for each speed in random order. Speak out (not sing) the phrases given by the participant as svaras and ask participant to confirm that what you've notated is correct. The synthesis interface can be shared with them only after obtaining their responses. Ask for participant's own interpretation(s) of the phrases they provide. Otherwise same as for the previous two sets.

D.5.1 Phrase 1

Ratings	Speed 1x	Speed 2x
PASR		
DPASR		

RESPONSES:

D.5.2 Phrase 2

Ratings	Speed 1x	Speed 2x
PASR		
DPASR		

RESPONSES:

D.6 Set 5 - Continuous section

Instructions for participant: This set involves one section with multiple phrases. Its purpose is to gauge the influence of phrase structure on gamakas. Only one interpretation of the section will be presented. Rate the interpretation on the scale of $[0-10]$ and comment on it.

Instructions for interviewer: Play the whole section before asking questions. Focus questions and discussion on phrase structure on the connecting gamakas used across phrase boundaries.

```
^da2   ni2   da2  ^da2  pa    ^pa    da2   pa    ^pa    ma1
^da2   pa    ma1  ^ga3  ma1    ri2  ^ga3   ma1   ^pa
^ni2-  sa   ^ri2   ga3  ma1   ^pa    ma1  ^da2:2       ^ni2:2    ^sa+:2
^da2   ni2   sa+  ^ri2+:2     ^ri2
^ni2   sa+   ri2+ ^ga3+ ma1+  ^ri2+:2     ^ga3+   ri2+ ^sa+
^ri2+ ^ni2:2      ^sa+ ^da2:2      ^ni2   da2    pa
^da2   pa   ^pa    ma1 ^ga3   ma1   ri2
```

DPASR: [ ]

RESPONSES:

Appendix E

Questions for use during evaluation interviews

Below is a list of some clarifying questions to be asked of participants depending on the situation. Words referring to the clip played such as "clip", "rendition", "variation", "version" or their vernacular equivalents need to be used appropriate to the context of the question. Addressing the participant during the session is to be according to cultural norms.

1. Is the clip audible? Is the synthesis clear and of adequate quality for further comments?
2. What rating would you give this clip on a scale of zero to ten? Assume this clip was performed by a student.
3. If you were to make one correction to improve how this phrase was rendered, what would it be?
4. This rendition was not acceptable to you. Which part of it did you find unacceptable?
5. You said this gamakā needs to be more "subtle". Are you referring to the shape of the movements or to the volume or "dynamics" with which the gamakā was rendered?
6. When you say "dynamics", are you referring to the loudness, volume or brightness of the sound at that point, or were you referring to the movement? (The term "dynamics" has several meanings in the culture.)
7. You said this variation is "unusual". By that do you mean you yourself won't use it, but others might? (Clarification of stylistic preference.)

8. In this rendition, were you able to perceive the various individual phrases?
9. When you said "this note needs to be sharper", are you referring to the pitch used for the note, or, perhaps, how to the movement with which the note was approached?
10. When you said "this note needs more importance", do mean the note must be held for longer than featured in this clip?
11. When you said "this movement is unnecessary", do you find it unacceptable to use the movement in this context? In your opinion, how should this note be handled?
12. You found the ending gamakā "odd". Do you consider it "odd" independent of what follows? If, say, the note is followed by *CHOICE1* or *CHOICE2*, would the gamakā be less or more appropriate?
13. You said that note *X* was played "too prominently"? Are you referring to the pluck on the note or perhaps the time spent on it?
14. When you said that note *X* needs to be "stressed", were you referring to just making it louder, or perhaps attack the note using a jantai-like technique?
15. You said the rendition sounded "robotic". Is there some particular aspect of the rendition that you found to be robotic? (The intention is to determine whether the timing of the rendition was the target of the comment.)
16. You said the previous rendition was not acceptable to you. By that did you mean that it might be considered acceptable in some other styles? If so can you identify a style where it might be acceptable?
17. You said you heard a different phrase than what was just mentioned. Can you describe the *svara*s for how you heard it? (The intention is to get alternative prescriptive *svara*s for the phrase.) In order for it be the *svara*s mentioned, how would you change the rendition?

Appendix F

Synthesis interface

Figure F.1 presents the interface to the elaborator used during the challenge phrase section of the evaluation (set 4, see appendix D.5).

- The input prescriptive notation is keyed into the box labelled "prescriptive notation" at the top of the interface. The prescriptive notation is to conform to the formal syntax described in appendix G.
- Once the prescriptive notation is entered into the box, clicking on the "Play" button will result in the notation being rendered using gamakās.
- A *tanpura* drone continuously plays in the background to reinforce the tonic for the participant.
- Checking the "SD" checkbox will enable elaboration based on the DPASR representation. "SD" is used here as an abbreviation of "stage-dance". When unchecked, the PASR based elaboration system is used.
- The "Tempo" slider is used to change speeds between 75 bpm and 150 bpm. Changing the tempo can, depending on the prescriptive notation given, result in different gamakās being selected.
- Participants were shown this interface so that they can check the correct entry of the prescriptive phrase they dictated.

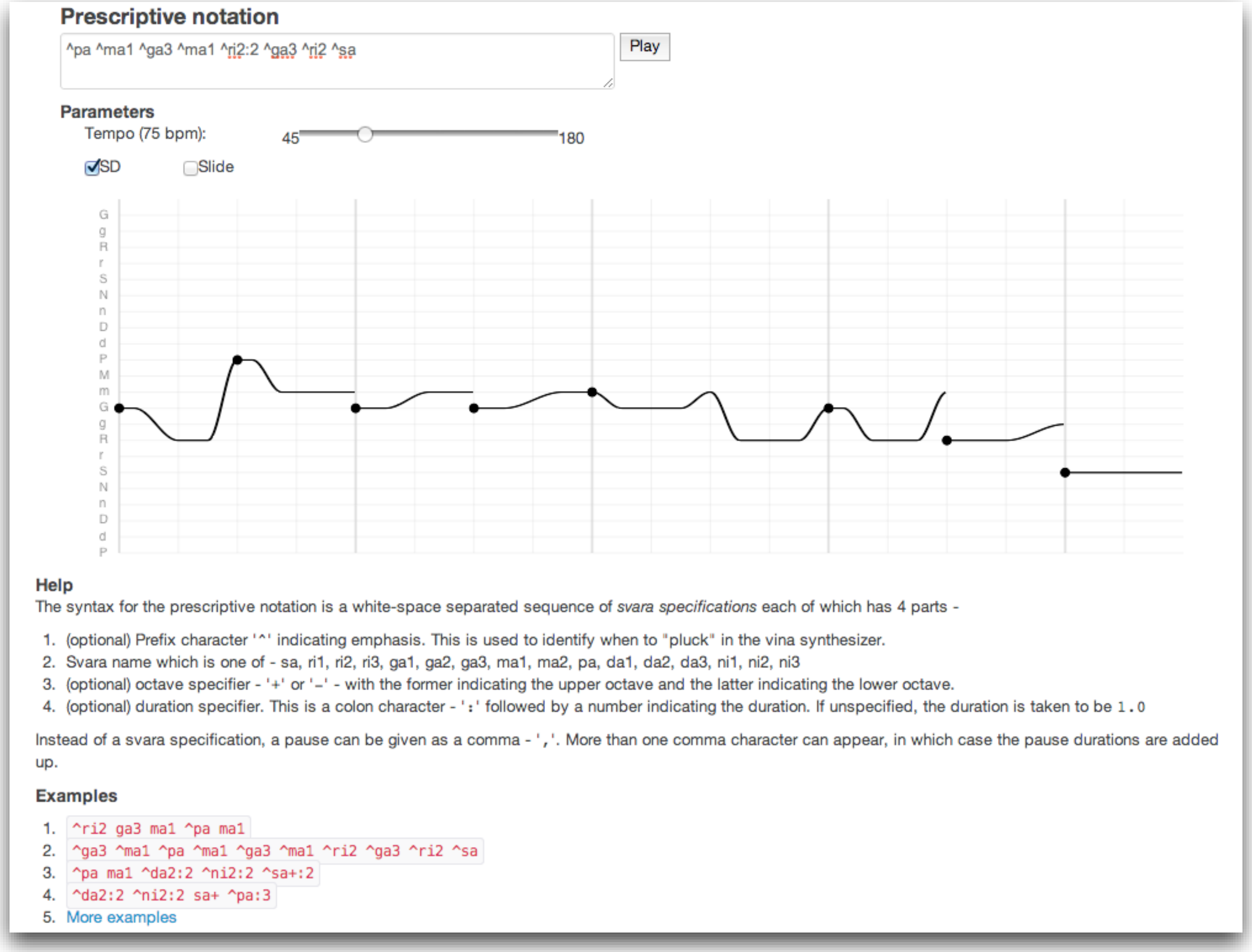


Figure F.1: Screenshot of synthesis interface.

Appendix G

Plain text prescriptive syntax

The formal syntax accepted as the "prescriptive notation" by the elaborator is given below as a Backus-Naur Form (BNF) grammar.

```
<expression> ::= <term> { <whitespace> <term> }
<term>       ::= <pause> | <svara>
<whitespace> ::= <wschar> { <wschar> }
<wschar>     ::= " " | "\t" | "\n"
<pause>      ::= "," { "," }
<svara>      ::= [<pluck>] <pitchclass> [<octave>] [<duration>]
<pluck>      ::= "^"
<pitchclass> ::=   "sa"  | "ri1" | "ri2" | "ri3" | "ga1" | "ga2" | "ga3"
                 | "ma1" | "ma2" | "pa"  | "da1" | "da2" | "da3" | "ni1"
                 | "ni2" | "ni3"
<octave>     ::= <oct_up> | <oct_down>
<oct_up>     ::= "+"  { "+" }
<oct_down>   ::= "-"  { "-" }
<duration>   ::= ":" <digit>
<digit>      ::= "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8"
```

expression The expression accepted by the elaboration system consists of a sequence of one or more terms separated by whitespace.

term A term specifies a **svara**, or indicates a **pause**.

pause Pause is indicated by a sequence of one or more "," characters, with each "," representing one symbolic time unit.

svara Pitch and time information regarding one *svara* in a prescriptive notation.

pluck The pluck marker is an optional synthesis aid that indicates where to insert a *vīṇā* pluck. This information is not used during the elaboration phase, but used only by the synthesizer. If plucks are omitted for all the notes in an expression, the system assumes that each given note is to be synthesized with a pluck. This is a useful shorthand that helped speed up typing during evaluation interviews.

pitchclass One of the 16 pitch class names that redundantly encode the 12 tones of an octave.

octave Higher octaves are indicated by one or more "+" symbols and lower octaves are indicated by one or more "-" symbols. The absence of an octave marker indicates that the *svara* is in the middle range.

duration Duration of a *svara* is indicated by a ":" followed by a digit giving the number of symbolic time units the *svara* should take. If duration is omitted, the *svara* is assumed to have a duration of one symbolic time unit.

Appendix H

Transcriptions

Listing H.1 shows a sample from the transcription database of the "Karuṇimpa" *varṇam* [1] and listing H.2 shows a sample of the encoded information about the prescriptive representation.[2] The database is a single JavaScript Object Notation (JSON) formatted data structure which consists of a sequence of sections, each comprising a sequence of phrases, each phrase comprising a sequence of *svaras*, for each of which a set of numerical gamakā transcriptions are given. The *svaras* are given in the syntax according to appendix G. Each of the "stage", "dance" and "PASR" components is given as an array of tuples of the form – $[[p_1, a_1, s_1, r_1], [p_2, a_2, s_2, r_2], ...]$. The p_i are focal pitch values expressed in semitones relative to the tonic. The a_i, s_i and r_i are respectively attack, sustain and release durations of focal pitches p_i. The durations of the attack, sustain and release components are considered to be normalized such that the total corresponds to the duration indicated in the *svara* specification. Note that the "stage", "dance" and "PASR" are all expressed as such PASR tuple arrays.

Listing H.1: Extract from unified transcription of ""Karuṇimpa"".

```
{
    "info": "sahana_db_meta",
    "performance": [ // Array of sections
    {
        "meta": "pallavi.line1",
        "speed": 1,
        "pasr": [ // Array of phrases
            [["^pa:2", /* ... */],  // One entry for each svara.
```

[1] Complete transcription data available from `http://sriku.org/dpasr/sahana_db.js`.
[2] Complete prescription data available from `http://sriku.org/dpasr/sahana_db_meta.js`.

```
            ["^ma1:2", /* ... */],
            ["^ga3:2", /* [stage], [dance], [PASR] */],
            ["^ga3", [[4,0,4,0]], [[1,0,0,0.5],[0,1,1,1],[1,0.5,0,0]],
                [[5,0,0,0.5],[4,1,1,1],[5,0.5,0,0]]],
            ["ma1", [[4,0,8,0]], [[1,0,0,0.5],[0,1,1,1],[1,0.5,0,0]],
                [[5,0,0,0.5],[4,1,1,1],[5,0.5,0,0]]],
            ["^ri2:2", [[4,0,4,0],[2,1,3,0]],
                [[1,0,0,0],[0,1,2,1],[1,0,0,0],[0,1,2,1],[2,0,0,0]],
                [[5,0,0,0.5],[4,1,1,1],[5,0.5,0,0.5],[2,1,1,1],[4,0.5,0,0]]],
            ["^ga3", /* ... */],
            ["ri2", /* ... */],
            ["^sa:4", /* ... */]],
           [/* ... next phrase ... */]
        ]
    },
    //...
}
```

Listing H.2: Extract from prescriptive representation of ““Karuṇimpa””.

```
{
    "adi_kalai2": { // Info about lower speed tala.
        "type": "tala",
        "structure": [4,2,2],
        "beats_per_count": 2,
        "pulses_per_count": 4,
        "tempo_bpm": 70
    },

    "adi_kalai1": { // Info about higher speed tala.
        "type": "tala",
        "structure": [4,2,2],
        "beats_per_count": 1,
        "pulses_per_count": 4,
        "tempo_bpm": 70
    },

    "contents": ["pallavi", "anupallavi", "muktayisvaram",
                 "ending_before_caranam", "caranam", "cittasvaram"],

    "pallavi": {
        "tala": "adi_kalai2",
        "contents": ["line1", "line2"],
        "line1": {
```

```
            "lyrics": [["ka:2", "ru:2", "Nim:8", "pa:4"], ["i:4", "di:4", "man:5", "ci:3"]],
            "presc": [
                ["pa:2", "ma1:2", "ga3:2", "ga3", "ma1", "ri2:2", "ga3", "ri2", "sa:4"],
                ["ni2-", "sa", "ri2", "ga3", "ri2", "sa", "ni-", "sa", "da2-", "pa-", "ma1-", "da2-:2", "ni2-", "sa", "ri2"]
                    ]
        },
        "line2": {
            "lyrics": [["ta:7", "ru:7", "Na:5", "mu:3"], ["sa:7", "mi:3"]],
            "presc": [
                ["pa", "ma1", "ga3", "ma1", "ri2", "ga3", "ri2", "sa:3"],
                ["ri2", "ga3", "ma1", "pa", "ma1"],
                ["da2:2", "ni2:2", "sa+", "pa:3"],
                ["ri2", "ga3", "ma1"],
                ["da2", "pa", "pa", "ma1", "ga3", "ma1", "ri2"]
                    ]
        }
    },

    // ...
}
```

Appendix I

Gamaka selection logic

Listings I.1 and I.2 describe scoring functions that determine how compatible two adjacent gamakās are. These functions determine the weights of the connections of a DAG representing the options available for a phrase, with the selection from the possibilities done using optimal path selection. Compound scores that express relative priorities of the preferences involved are expressed using sum and product operations. Note that there is a difference in the signatures of the two functions. Where the `pasrScore` function directly calculates the score, the `dpasrScorer` produces a function that calculates the score. This difference is incidental and should be overlooked as irrelevant to the model.

Listing I.1: Score calculation for PASR based gamakā selection

```
function pasrScore(note1, gamaka1P, note2, gamaka2P, params) {

    var gamaka1 = gamaka1P.pasr;
    var gamaka2 = gamaka2P.pasr;
    var score = 1.0;
    var gclass1 = pasrEqClass(gamaka1);
    var gclass2 = pasrEqClass(gamaka2);
    var gclass1_s = gclass1.filter(function (p) { return p[2] === 'S'; });
    var gclass2_s = gclass2.filter(function (p) { return p[2] === 'S'; });

    // Junction focal pitch rules.
    //
    // 0. When the two notes are plain, they must match
    // the respective note pitches. Otherwise we penalize them.
    if (gclass1.length === 1) {
        if (gclass1[0][0] !== note1.pitch_value) {
```

```
            return 0;
        }
    }

    if (gclass2.length === 1) {
        if (gclass2[0][0] !== note2.pitch_value) {
            return 0;
        }
    }

    // 1. If the "s count" of the two pasr equivalence classes
    // are different, penalize accordingly. This biases the
    // rendering towards a steady pacing. Also bias slightly
    // against differing movement rates.
    score += Math.pow(2, - 0.25 * Math.abs(gclass1_s.length / note1.duration -
        gclass2_s.length / note2.duration));
    score += Math.pow(2, - 0.1 * Math.abs(gclass1.length / note1.duration -
        gclass2.length / note2.duration));

    // 2. It is undesirable for a gamaka that precedes a pitch
    // inflection to start with transients.
    if (isInflection(note2) && gclass1[0][1] === 'T') {
        score *= 0.6;
    }

    // 3. Gamaka1 ending with an intermediate or transient with
    // gamaka2 being a plain note is undesirable, but this is not
    // a critical rule. Just bias against it a little bit.
    if (gclass1[gclass1.length - 1][1] !== 'S' && gclass2.length === 1) {
        score *= 0.85;
    }

    // 4. The focal pitch that is penultimate to a phrase end point
    // is biased against being a plain note.
    if (note2.next === null && gclass1.length === 1) {
        score *= 0.8;
    }

    // 5. Give preference to repeated gamakas. It is enough if the
    // gamaka is repeated as a prefix. Repetition is detected by
    // the amount of overlap between the two gamakas.
    var strify = function (f) { return f.join(''); };
    score += Math.pow(2, 0.5 * overlap(gclass1.map(strify), gclass2.map(strify)));
```

```

    // 6. Bias against joining focal pitches being both sustained ones.
    if (gclass1[gclass1.length - 1][0] === gclass2[0][0]) {
        if (gclass1[gclass1.length - 1][1] === 'S' && gclass2[gclass2.length -
            1][1] === 'S') {
            score *= 0.6;
        }
    }

    // 7. If joining focal pitches differ, then bias slightly against both being
    // sustained ones. Note the difference in the bias levels.
    if (gclass1[gclass1.length - 1][0] !== gclass2[0][0]) {
        if (gclass1[gclass1.length - 1][1] === 'S' && gclass2[gclass2.length -
            1][1] === 'S') {
            score *= 0.9;
        }
    }

    // 8. Prefer pitch sequence matches.
    var pitchOnly = function (f) { return f[0]; };
    score += Math.pow(2, 0.5 * overlap(gclass1.map(pitchOnly), gclass2.map(
        pitchOnly)));

    // 9. Prefer pitch sequence matches of the sustain subsets.
    if (gclass1_s.length < gclass1.length || gclass2_s.length < gclass2.length) {
        score += Math.pow(2, 0.5 * overlap(gclass1_s.map(strify), gclass2_s.map(
            strify)));
    }

    // 10. If the movement directions of the end of previous gamaka and the start
    // of the next one don't match, bias against it.
    if (gclass1.length > 1 && gclass2.length > 1) {
        var dir1 = gclass1[gclass1.length - 1][0] - gclass1[gclass1.length -2][0];
        var dir2 = gclass2[1][0] - gclass2[0][0];
        if (dir1 * dir2 < 0) {
            score *= 0.75;
        }
    }

    // 11. Bias towards plain sa and pa.
    if (note1.pitch_class === 'sa' || note1.pitch_class === 'pa') {
        score += Math.pow(2, 0.1 * (gclass1.length - 1));
    }
```

```
    // 12. If second note is an ending note, is either sa or pa,
    // and the previous one has movement in it, bias the second
    // against any movement on the sa or pa.
    if (note2.context[2].pitch === '-' && (note2.pitch_class === 'sa' || note2.
        pitch_class === 'pa')) {
        if (gclass1.length > 1) {
            score += Math.pow(2, - 0.2 * (gclass2.length - 1));
        }
    }

    // 13. Add a score that controls the number of movements per pulse.
    var limit = 240 / params.tempo_bpm;
    var movementsPerDur = (gclass1.length + gclass2.length) / (note1.duration +
        note2.duration);
    score *= Math.pow(2, - (Math.max(limit, movementsPerDur) - limit));

    return score;
}
```

Listing I.2: Score calculation for DPASR based gamakā selection

```
function dpasrScorer(phrase, speed, options) {

    return function (note1, note2, i) {
        return (stageCompat(note1, note2, i)
                * kampita(note1, note2, i)
                * rateLimit(note1, note2, i)
                * rateMatch(note1, note2, i));
    };

    function sqdiff(x1,x2) {
        return (x1[0] - x2[0]) * (x1[0] - x2[0]);
    }


    // 1. Compatibility of stage components of the two gamakas being considered.
    function stageCompat(note1, note2, i) {
        var s1 = note1.soasr.stage;
        var s2 = note2.soasr.stage;
        var duri = noteDur(phrase[i], options.tala, options.speed);
        var durip1 = noteDur(phrase[i+1], options.tala, options.speed);
        s1 = s1.length < 2 ? [s1[0],s1[0]] : [s1[0],s1[1]];
        s2 = s2.length < 2 ? [s2[0],s2[0]] : [s2[0],s2[1]];
        var dist = sqdiff(s1[1],s2[0]);
        var speedBias = note1.soasr.stage.length * note2.soasr.stage.length / (
            duri * durip1);
        var dp1 = Math.min(Math.abs(phrase[i].pitch_value - DB.pasr_pitch_value(s1
            [0])),
                           Math.abs(phrase[i].pitch_value - DB.pasr_pitch_value(s1
                               [1])));
        var dp2 = Math.min(Math.abs(phrase[i+1].pitch_value - DB.pasr_pitch_value(
            s2[0])),
                           Math.abs(phrase[i+1].pitch_value - DB.pasr_pitch_value(
                               s2[1])));
        var naturalBias = Math.exp( - (dp1 + dp2) ) / Math.LN2;
        return dist * speedBias * naturalBias;
    }

    // 2. Adjustment in the case of kampitas based on the tempo and the number of
        oscillations.
    function kampita(note1, note2, i) {
        var k1 = endingKampita(note1);
```

```
        var k2 = startingKampita(note2);
        if (k1.length > 0 && k2.length > 0) {
            return 1 / (1 + kampitaSpeedBoost(note1, note2, i) * Math.max(0, k1.
                length + k2.length - 3));
        } else {
            return 1;
        }
    }

    // Utility function.
    function endingKampita(note) {
        var k = [];
        var i;
        var d = note.soasr.dance.shape_partial;
        for (i = d.length - 1; i > 0; --i) {
            if (Math.abs(d[i] - 1 + d[i-1]) < 0.05) {
                return k.push(d[i]);
            } else {
                break;
            }
        }
        k.reverse();
        return k;
    }

    // Utility function.
    function startingKampita(note) {
        var k = [];
        var i;
        var d = note.soasr.dance.shape_partial;
        for (i = 1; i < d.length; ++i) {
            if (Math.abs(d[i] - 1 + d[i-1]) < 0.05) {
                return k.push(d[i-1]);
            } else {
                break;
            }
        }
        return k;
    }


    // Utility function
    function kapitaSpeedBoost(note1, note2, i) {
```

```
        var tala = options.tala;
        var speed = options.speed;
        if (tala.beats_per_count / speed < 1.01) {
            return 2;
        } else {
            return 1;
        }
    }

    // 3. Score gamaka choices based on whether the rate of movements
    // in them are excessive or acceptable.
    function rateLimit(note1, note2, i) {
        var dur1 = noteDur(phrase[i], options.tala, options.speed);
        var dur2 = noteDur(phrase[i+1], options.tala, options.speed);
        var limitOscil = options.rateLimit * (dur1 + dur2);
        var foundOscil = note1.soasr.dance.shape_partial.length + note2.soasr.dance.shape_partial.length;
        return Math.exp(-0.5*Math.min(0,limitOscil - foundOscil));
    }

    // 4. Prefer two adjacent gamakas to have similar movement rates.
    function rateMatch(note1, note2, i) {
        var dur1 = noteDur(phrase[i], options.tala, options.speed);
        var dur2 = noteDur(phrase[i+1], options.tala, options.speed);
        var rate1 = note1.soasr.dance.shape_partial.length / dur1;
        var rate2 = note2.soasr.dance.shape_partial.length / dur2;
        return Math.exp(4 * Math.abs(rate1 - rate2));
    }
}
```

Glossary

ālāpana A form of improvisation in Carnatic and Hindustani music (where it is referred to as "*ālāp*") wherein one or more rāgās are elaborated on without a metrical time structure. An *ālāpana* often precedes the rendition of a composition in a typical Carnatic music performance and can be brief or elaborate. 33, 71, 74, 78, 104, 120, 142

anupallavi Usually the line that follows a Carnatic composition's *pallavi*. 36, 63, 143

ārōhaṇa Ascent up the scale of a rāgā. Note that ascent may not be strictly directional and may involve *vakra* or zig zagging movements. 81, 83, 119, 145

avarōhaṇa Descent down the scale of a rāgā. Note that descent may not be strictly directional and may involve *vakra* or zig zagging movements. 81, 83, 90, 119, 145

BNF Backus-Naur Form 129

caraṇam The terminal lyrical lines of a composition. The word literally means "feet". In *varṇam* type of compositions, the *caraṇam* appears repeated and interwoven with solfa lines called *ciṭṭasvaram*. 36, 63, 141

ciṭṭasvaram Lines of solfa compositions mostly found in *varṇam*s interlaced with the terminal thematic lines of the composition called *caraṇam*. 36, 63, 141

DAG Directed Acyclic Graph 30, 42, 87, 102, 104, 134

descriptive notation The term "descriptive notation", introduced by ethnomusicologist Charles Seeger [Seeger, 1958], stands for a notation made of a specific performance, as opposed to prescriptive notation that serves as instructions for performers. In the context of Carnatic music, it also refers to a way of notating the internal movements within a phrase using discrete *svara*s, a pedagogical approach introduced in [Viswanathan, 1977]. 2, 6, 13, 14, 25, 34, 38, 45, 72, 90

DPASR Dual-PASR form 28, 30, 32, 39, 40, 44, 52, 66, 68–70, 79, 83–86, 89, 90, 97, 98, 100, 102, 104, 105, 127, 138

EMI Experiments in Musical Intelligence 19, 103

gamakā A broad term used to refer to movements that connect two or more tonal positions. Gamakās are a distinct characteristic of both the South Indian and North Indian classical music traditions. Although the term encompasses both discontinuous and continuous connective movements, it is largely used to refer to continuous connective movements. Ontological literature describes various types of gamakās such as *kampita*, *jāru*, *nokku*, *orikai*, *odukkal*, *sphuritam* and *pratyāhatam*. i, 2, 8–11, 24, 26–34, 41, 42, 55, 62, 66, 68, 75, 76, 98, 101–104, 113, 142–144

jāru A type of gamakā. It is a long sliding movement between two tonal positions, applicable when the tonal positions are far enough apart for a slide to be perceived as such. On the *vīṇā*, a *jāru* is performed by sliding between two frets without pulling on the string. 23, 88, 89, 142

JND Just Noticeable Difference 23, 37

JSON JavaScript Object Notation 131

kampita A type of gamakā. Repeated oscillatory movements between two tonal positions are called "kampita gamakās". i, 67, 142

krti A compositional form with greater emphasis on lyrics compared to *varṇam*s. Renditions of *krti*s are expected to be faithful to the lyrics, therefore musicians attempt to interpret its poetry in their choice of gamakās, dynamics, variations and repetitions. *Krti*s are always performed in a single speed, unlike *varṇam*s, although the composition itself may feature multiple speeds in different parts. The word "k¯" may also be used in the case where the work is devotional in nature. 33, 103, 104

laya Refers to a broad notion of a "sense of time". The term encompasses metric time structures – i.e. the tāḷā – as well as the sense of time necessary for free-time improvisatory forms such as *ālāpana* and *tānam*. It is also used to refer to expressive timing relevant to the performance of lyrical compositions. 72

MIDI Musical Instrument Digital Interface 4, 21

MIDI MIDI (abbrev. Musical Instrument Digital Interface) is a protocol for communicating musical events and other data to synthesizers and notation tools. The "Standard MIDI file" is a computer file format for the storage and transmission of MIDI data. 143

MIR Music Information Retrieval 3, 4

muktāyisvaram A solfa section of a composition that may follow its *anupallavi* and which marks the end of the first part of the composition. 36, 63

NFP Normal Focal Pitch 62, 69

nokku A type of gamakā. It is a movement that stresses a tonal position by a quick continuous movement from a relatively higher tonal position. 142

odukkal A type of gamakā. It is a movement to arrive at a tone from a lower tone with a brief overshoot before landing on the target tone. i, 67, 142

orikai A type of gamakā. It is a momentary flick at the end of a principal tone to a higher tone. i, 14, 67, 142

pallavi The opening line of a composition in Carnatic music. The term is used to refer to both the opening line of lyrics as well as the melody to which it is sung. During a performance of a composition, the performer will usually return to the *pallavi* as a theme at several points. 36, 63, 141

pāṇi Refers to a musical/stylistic lineage, often associated with a particular maestro or identified by the place where it gained prominence and developed. For example, the performer whose performance is used as reference for this work belongs to the "kāraikkudi pāṇi". The roughly equivalent term in the Hindustani tradition would be "Gharāna". 33, 71, 143

PASR Pitch Attack Sustain Release form 28, 30, 32, 39, 44–46, 48, 50, 52, 62, 66, 68, 70, 79, 83–86, 89, 97, 98, 100, 104, 105, 114, 127, 131, 134, 142

pāṭhāntra The repertoire taught in a particular school or "*pāṇi*" of Carnatic music. 71

pratyāhatam A type of gamakā similar to *sphuritam*, with the difference that the technique is applied during scale descent. The difference between *sphuritam* and *pratyāhatam* is only with respect to the fingering technique on the *vīṇā*. The sound of the two are in practice indistinguishable when sung. 98, 142, 144

prescriptive notation According to the original definition of the term "prescriptive notation" by ethnomusicologist Charles Seeger [Seeger, 1958], it is notation that is intended for interpretation by a performer and can assume what is considered to be common knowledge among practitioners of the genre it is intended for. The term, in the context of Carnatic music, is used to refer to the common forms of published sparse music notation using discrete *svara*s to outline the forms of phrases. i, 1, 2, 5, 6, 8–10, 16, 18, 23–25, 38, 44, 48, 54, 63–65, 72, 99, 113, 141

rāgā A loosely defined term that is used to refer to the set of melodic constraints that apply to a given compositional or improvisatory context. These constraints typically include scalar ascent and descent patterns. Many rāgās are also characterized by specific gamakās and phrases. 8, 9, 26–28, 32–34, 62, 63, 65, 71, 72, 74, 85, 89, 90, 92, 103, 104, 113, 119, 141, 144, 145

rāga lakśaṇā Literature that describes rāgās by giving their ascent/descent scales, frequency and usage of the pitch classes that feature in them and characteristic phrases and gamakās. 9, 10, 13, 27, 83, 119

SFP Sustained Focal Pitch 62

sphuritam A type of gamakā technique on the *vīṇā*. It consists of a discontinuous movement between two nearby tonal positions which stresses the higher tonal position. Used most commonly when repeated svaras such as "m m" occur in sequence, the second svara is stressed by approaching it from the immediately lower semitone, for example "m (Gm)". 98, 142, 144

svara A term used in Carnatic music to refer to a pitch class simultaneously as a tonal position and its solfa name. When peforming svara sections of compositions, singers articulate the syllable names of the svaras. Svaras are also sung in the type of improvisation known as "*svara kalpana*". 26, 27, 36, 65, 80, 90, 100, 101, 113, 114, 120, 126, 130, 131, 142, 144, 145

svara kalpana A melodic improvisatory form in Carnatic music where the *svara*s constituting the melody are themselves sung (in a vocal performance). This form of improvisation is done within a tāḷā structure. The term may also be found expressed as "kalpana svara". 74, 78, 144

svarasthāna Refers to one of the twelve tonal positions that constitute an octave. In this document, these are notated using the mnemonic letters "*SrRgGmMPdDnN*". 27, 114

synthesis bias A bias expressed by a musician or connoisseur for or against computer synthesized sound, particularly when used for genres they are intimate with. 74

tāḷā Cyclic temporal beat structure used for compositions and many forms of structured improvisation in Carnatic music. The beats constituting a tāḷa cycle are grouped into units that are indicated by hand gestures such as claps, finger-counts and waves. For example, the "ādi tāḷa" has a 4-2-2 beat structure that is gesturally indicated as clap-1-2-3-clap-wave-clap-wave. 24, 26, 27, 35, 40, 74, 104, 113, 142, 145

tānam A form of melodic improvisation in Carnatic music involving semi-rhythmic pulsating patterns constrained by a rāgā. Singing of *tānam* involves the use of syllables such as "nam", "nom", "tha" and "thom" interspersed with vocalizations of "ā" of "m". 33, 142

TFP Transient Focal Pitch 62, 69

vakra When either the "ascent" (*ārōhaṇa*) or "descent" (*avarōhaṇa*) progression of a rāgā are not strictly directional and have "crooked" or "zig zagging" movements, the progressions are referred to as "vakra". The term is used to classify rāgās based on their scalar constraints. 34, 65, 141

varṇam Elaboration compositional forms that feature sparse lyrics rendered with many variations as well as solfa sections. They are usually performed in multiple speeds related by simple integer ratios with at least two speeds featuring in every performance. *Varṇam*s also serve as early pedagogical material to introduce students to the characteristics of major rāgās. i, 27–29, 32–36, 40, 42, 103, 104, 131, 141, 142

vīṇā A traditional fretted instrument used in Indian classical music belonging to the lute family. The South Indian *vīṇā* has two resonating gourds connected by a fret board on which four main and three side strings are strung. The distinct tonal characteristic of a *vīṇā* derives from the use of a curved bridge plate to anchor the strings to the main resonator. The frets of the *vīṇā* are mounted on the fret board using a black wax mixture which is scalloped between frets to allow easy pitch bending by pulling a string along a fret. For detailed information about the instrument, history and lineages of practitioners see [Subramanian, 1985a]. 28, 30, 32–35, 40–42, 76, 97–99, 104, 130, 142, 144

Index