Friedrich-Alexander-Universität Erlangen-Nürnberg

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

Master Thesis

Towards Automatic Audio Segmentation of Indian Carnatic Music

submitted by

Venkatesh Kulkarni

submitted

July 29, 2014

Supervisor / Advisor

Dr. Balaji Thoshkahna

Prof. Dr. Meinard Müller

Reviewers

Prof. Dr. Meinard Müller

AUDIO LABS

International Audio Laboratories Erlangen

A Joint Institution of the

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and

Fraunhofer Institute for Integrated Circuits IIS

Fraunhofer IIS

Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Erlangen, July 29, 2014

Venkatesh Kulkarni

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Balaji Thoshkahna, whose expertise, understanding and patience added considerably to my learning experience. I appreciate his vast knowledge and skill in many areas (e.g., signal processing, Carnatic music, ethics and interaction with participants).He provided me with direction, technical support and became more of a friend, than a supervisor.

A very special thanks goes out to my Prof. Dr. Meinard Müller, without whose motivation and encouragement, I would not have considered a graduate career in music signal analysis research. Prof. Dr. Meinard Müller is the one professor/teacher who truly made a difference in my life. He was always there to give his valuable and inspiring ideas during my thesis which motivated me to think like a researcher. It was though his, persistence, understanding and kindness that I completed my thesis. I doubt that, I will ever be able to convey my appreciation fully, but I owe him my eternal gratitude.

Special thanks to Nanzhu Jiang, who was always available for technical discussions on music signal processing. She helped me in annotating the Carnatic music database for my thesis.

I am indebted to many of my friends who support me during the writing of my thesis. Dr. Balaji Thoshkahna, Nanzhu Jiang, VedaShruti Kulkarni, Praveen Kulkarni, Shruti Shukradas and Pavan Kantharaju have all helped me in proof reading my thesis report. They have given me the detailed and helpful feedback. I am grateful for their help. In particular, some of them were quite busy with their own work, however they still took their time for correcting my errors and offered suggestions.

Last but not the least important, I owe more than thanks to my family members which includes my parents, elder brother and my sister, for their financial support and encouragement throughout my life. Without their support, it is impossible for me to finish my college and graduate education seamlessly.

Abstract

In this thesis, we introduce the segmentation task typically encountered in multimedia signal processing. In particular, we focus on music data. Music is a very complex and highly structured data. The structure in music arises from verses, bridges, refrain and homogeneity in musical aspects such as tempo, melody, dynamics or timbre. The extraction of musical information from an audio recording is a challenging task in the field of music information retrieval. The audio data can be segmented based on the musical aspects. As an example scenario, in this thesis, we consider the Indian Carnatic music as an interesting case study. We consider the Carnatic music recording, which consists of three contrasting parts. They are, *Alapana*, *Krithi* and *Tani-Avarthanam*. The three parts are different concerning to the musical properties. The *Alapana* and *Krithi* parts have a clear notion of melody whereas, the *Tani-avarthanam* has a vague notion of melody. Similarly, the *Alapana* part has no clear notion of tempo whereas, the other two parts have a clear notion of the tempo.

The main contribution of this thesis is to apply signal processing techniques to design several novel features aimed towards the automatic segmentation of the three contrasting parts. We use tempo and melodic properties to design novel features for the segmentation. To this end, we perform qualitative and quantitative analysis of the novel features for the segmentation task.

Contents

Chapter 1

Introduction

1.1 Music Background

Carnatic music is popular in the southern part of India [3], roughly confined to four states of, Andhra Pradesh, Tamil Nadu, Karnataka and Kerala. Many concerts are performed in large music festivals conducted in India and abroad. Many of the performances are recorded and uploaded on audio or video sharing websites such as, YouTube or Sangeethapriya[1], a non-commercial service specialized for the exchange of Indian classical music. However, the uploaded audio material is often poorly segmented and annotated. Despite the massive amounts of data and its cultural relevance, only few attempts have been made to develop automated methods for making this material better accessible for users.

A typical concert may last for about $2-3$ hours in which various pieces are performed. A concert typically has $7-8$ pieces. Each piece is performed for roughly about $10-15$ minutes except for the main piece, which may last for upto 60 minutes. A concert is performed by a small ensemble of musicians, consisting of a lead artist (a vocalist or a harmonic instrument), a melodic accompaniment and rhythmic accompaniments (one or more percussive instrumentalist). In addition, a *drone* instrument (like *Tanpura*) is also played throughout a concert to set up the harmonic base for the music.

The main piece of a concert consists of three contrasting parts, known as *Alapana*, *Krithi* and *Tani-Avarthanam* [6] as shown in the Figure 1.1. The first part, *Alapana* is a purely melodic improvisation of *Raga*[2] accompanied with the harmonic instruments. *Raga* is considered as one of the most important aspects of Carnatic music. In the *Alapana* part, exposition of *Raga* takes place with slow improvisation with no rhythm involved in it. In the second part, *Krithi*, a lyrical composition is performed by the lead artist in a *Raga* and *Tala*[3]. *Tala* is a rhythmic framework. Finally, in the concluding part, *Tani-Avarthanam*, the percussionist(s) show their virtuosic skills by further exploring the *Tala*. Further details on Carnatic music compositions and musical properties of different parts are discussed in the Chapter 2.

The motivation of this thesis is not to aim at perfect segmentation of the main piece but to

[1] `http://www.sangeethapriya.org`

[2] *Raga* is one of the melodic modes used in Indian classical music and is also known as *Ragam*

[3] *Tala* is a repeating rhythmic phrase rendered on a percussive instrument

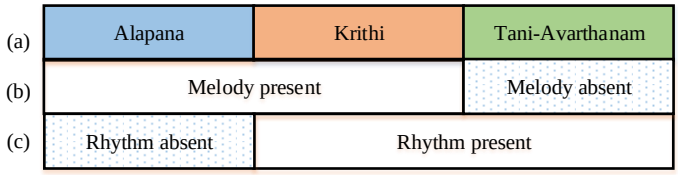

Figure 1.1: Typical structure of a Carnatic main piece. **(a)** Main piece constituting of three contrasting parts: *Alapana*, *Krithi* and *Tani-Avarthanam*. **(b)** Melody property distinguishes *Tani-Avarthanam* with respect to *Alapana* and *krithi*. **(c)** Tempo property distinguishes *Alapana* with respect to *Krithi* and *Tani-Avarthanam*.

analyze and investigate the extent to which the musical parts can be characterized by their musical properties.

As mentioned earlier, a drone instrument is played throughout a concert to provide the reference pitch for the performers. Neglecting the drone instrument, we can summarize the musical properties of the main piece as below,

1. *Alapana* is purely harmonic as the lead performer(vocalist or harmonic instrumentalist) improvises the melodic mode without the presence of percussion. Therefore, it has the existence of melody and the absence of tempo.

2. *Krithi* is a lyrical composition performed by the lead artist, with the harmonic and percussive instruments played in the background. Hence, this part has the presence of both melody and tempo.

3. *Tani-Avarthanam* is a purely percussive part performed by percussionist(s). Thus, this part has a clear notion of tempo and an absence of melody.

1.2 Main Contributions

From the above observations , we can infer that, it is possible to musically distinguish the different parts of the main piece based on melody and tempo information as shown in the Figure 1.1. The *Tani-Avarthanam* part can be musically distinguished from the *Alapana* and *Krithi* based on melodic information. The *Alapana* part can be musically distinguished with respect to *Krithi* and *Tani-Avarthanam* based on tempo information. The main technical contribution of this thesis is as follows.

Firstly, we design several tempo salience features based on the cyclic tempogram representation, which captures the absence or the presence of the tempo in the constituent parts. The cyclic tempogram representation is obtained from the existing tempogram toolbox [16]

Secondly, we design several chroma salience features based on the chroma representation, which captures the absence or the presence of the melody in the constituent parts. The chroma representation is obtained from the existing chroma toolbox [25].

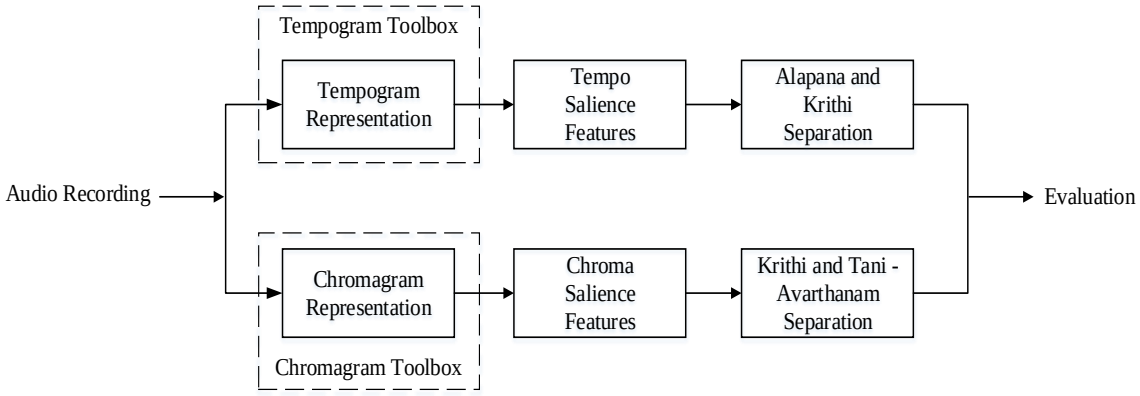

Figure 1.2: Model of thesis structure.

Finally, we provide some insights into the tempo and chroma salience features by a quantitative and qualitative analysis of their properties. For each of the features, we compute the average feature value and its standard deviation for each of the three Alapana, Krithi and Tani-Avarthanam parts separately. From this statistics, we can investigate how well the three different musical parts are characterized by the novel features. The chroma and tempo salience features can be further used in an automated algorithm for segmentation in conjunction with other features for retrieval tasks.

1.3 Thesis Organization

In the following, we give an outline of this thesis and briefly describe the contents of each chapter.

In Chapter 2, we introduce the building blocks of Carnatic music with detailed explanation of the different pieces performed in a concert. We also list a few differences between Indian classical music and Western classical music.

In Chapter 3, we discuss how the tempo cue can musically distinguish *Alapana* with respect to *Krithi* and *Tani-Avarthanam* parts of a main piece audio recording. We design several tempo salience features based on the cyclic tempogram representation, which captures the absence or the presence of the tempo in the constituent parts.

In Chapter 4, we discuss how the melody cue can musically distinguish *Alapana* and *Krithi* with respect to *Tani-Avarthanam* part of a main piece audio recording. We design several chroma salience features based on the chroma representation, which captures the absence or the presence of the melody in the constituent parts.

In Chapter 5, we discuss how well the three different musical parts are characterized based on quantitative and qualitative analysis of the chroma and tempo salience features. We also lists the strengths sand limitations of the features for analyzing and segmenting Carnatic music recordings.

At last in Chapter 6, we make conclusions about the main contributions of this thesis and discuss about the future work.

Chapter 2

Carnatic Music

The two main classical music forms of India are Hindustani music found in northern part of the country and Carnatic music popular in South India. Both the music forms claim to have *Vedic* origins and it is also believed that they diverged from a common musical root in the 13^{th} century [5].

The traditional Indian classical music system is based on a just tempered scale[1] [23]. It is possible to get a lot of insight into Indian music using equal tempered scale which essentially makes it easy and simple to use Western keyboards. There are seven basic *swaras*[2] in Indian classical music and they are *Shadja(Sa), Rishabha (Ri), Gandhara (Ga), Madhyama(Ma), Panchama (Pa), Dhaivata (Dha)* and *Nishadam (Ni)* [7]. The remaining five *swaras* are derived from the basic notes to make it to twelve notes per octave. The notation for the notes used in Indian classical music and Western music as shown in the Table 2.1.

The three main pillars of Indian classical music are, *Raga* (providing the melodic framework), the *Tala* (providing the rhythmic framework) and improvisation.

Raga is a set of *swaras*, *gamakas*[3] given to these *swaras* and the sequence in which they occur. *gamaka* is the variation of a pitch oscillating between the adjacent and the current note (micro tones). Each *Raga* has specific rules for using the type of *gamaka* that is applied to a specific note. A *Raga* is described by the sequence of notes in *arohana* and *avarohana*. *Arohana* is the sequence of notes used in a *Raga* in the ascending order with the pitch going upwards. *Avarohana* is a sequence of notes used in descent. For example, *Lavangi Raga* has four notes. The *arohana* and *avarohana* are given as (*S R1 M1 D1*) and (*S D1 M1 R1*). Every *Raga* is associated with a certain time of day or night, or season, resulting in particular moods or feelings being evoked [3].

Tala provides a rhythmic framework for the musicians. It roughly corresponds to metre in Western music. There are two main characteristics of the *Tala* which differentiate it from Western music [1]. In Western music each segment has the same number of beats (like, 4+4+4+4), whereas in Indian classical music each segment may have a different number of beats (like,

[1] Just tempered scale (just intonation) is a method of tuning intervals and notes based exclusively on rational numbers.

[2] *Swara* refers to a note in the octave.

[3] *Gamaka* refers to ornamentation that is used in the performance of Indian classical music. The unique character of each *Raga* is given by its *gamaka*, making their role essential rather than decorative in Indian music.

Poistions	Western Music System(notes)	Indian Music System		
		Swara	Notation 1	Notation 2
1	C	Shadja	Sa	S1
2	$C\sharp$	Shuddha Rishabha	Ri 1	R1
3	D	Chatushruti Rishabha	Ri 2 (Ga 1)	R2 (G1)
4	$D\sharp$	Sadharana Gandhara	Ri 3 (Ga 2)	R3 (G2)
5	E	Antara Gandhara	Ga 3	G3
6	F	Shuddha Madhyama	Ma 1	M1
7	$F\sharp$	Prati Madhyama	Ma 2	M2
8	G	Panchama	Pa	P1
9	$G\sharp$	Shuddha Dhaivata	Dha 1	D1
10	A	Chatushruti Dhaivata	Dha 2 (Ni 1)	D2 (N1)
11	$A\sharp$	Kaisiki Nishada	Dha 3 (Ni 2)	D3 (N2)
12	B	Kakali Nishada	Ni 3	N3

Table 2.1: Western music and Indian music system

2+3+2+3 pattern). The beat pattern can be presented in three ways, as a series of counts made by wave of the hand or tap of the hand on the lap or using both the hands in a manner of clap. For example, *Jhoomra Tala* has 14 beats, counted as 3+4+3+4.

There are several music compositions. Each composition is performed in a given Raga and *Tala* along with the lyrics. A artist decides which Raga and *Tala* to be used based on the mood or the audience choice. Here the artist has freedom to improvise *Raga* and *Tala* to impress the audience. There are different *Ragas* and *Talas* used in both of the main forms of Indian classical music. The origin and historical development of both forms of the music are discussed further.

Hindustani music was developed, refined in 13^{th} and 14^{th} centuries AD. It was not only influenced by ancient Hindu musical traditions and *Vedic* philosophy but also enriched by the Persian performance practices of the *Mughal* era. The practice of singing based on notes was popular from the *Vedic* times [4].

Carnatic music is based on historical developments in 15^{th} and 16^{th} centuries AD. Carnatic in Sanskrit means soothing to hear [23]. It is a form of music, emphasizing on creativity and improvisation. Most of the compositions are written to be sung and they are accompanied with some instruments performed in a singing style. Purandara Dasa is considered as the "Sangeetha Pitamaha" or the "grandfather of Carnatic music". He has been credited to elevate Carnatic music by systematizing the teaching methods and framing a series of lessons on various forms of Carnatic music. Most of the songs and poems performed today in the concerts were written and composed way back in 14th century. Composers such as *Tyagaraja*, *Muthuswami Dikshitar* , *Shyama Shastri*, *Annamacharya*, *Bhadrachala Ramadasu*, *Annamacharya*, *Kanakadasa*, etc wrote lyrical compositions in various *Ragas* in languages such as Sanskrit, Telugu, Kannada and Tamil. Most of the songs performed in concerts today come from one of the above mentioned composers.

Indian classical music differs from Western classical music in the aspects listed in the Table 2.2 [23]. In this thesis, we shall analyze Carnatic music compositions. There are different types of Carnatic music compositions and improvisational aspects of compositions which are discussed in the next Section 2.1.

	Indian Classical Music	Western Classical Music
1	Musicians have freedom for improvisation.	Musicians have lesser freedom for improvisation.
2	It is homophonic focusing on melody created using a set of notes.	It is polyphonic focusing on melody and texture created using multiple voices.
3	Complex beat cycles.	Simple beat cycles.
4	Use of micro tones.	Restricted to semi tones.
5	No use of Dissonance.	Use dissonance to add texture to the composition.

Table 2.2: Indian Classical Music versus Western Classical Music

2.1 Carnatic Music Compositions

Carnatic music can be classified in two basic formats. *Abyasa Gaanam* (literally singing for practice) - for the purpose of learning and practicing music. *Sabha Gaanam* (literally concert singing) - for performing in the concerts or public gathering. We mainly focus on the second music format.

There are various forms of Carnatic music compositions [5]. The following types of pieces popular in Carnatic music are listed below.

1. *Varnam* : A typical Carnatic concert begins with a *Varnam*. It is a lyrical composition which may last for about 5 minutes. A *Varnam* usually starts off slow and requires a doubling of tempo towards the end. This warms up the musicians and sets the mood and pace of the concert. There are two types of *Varnam* and they are *Tana Varnam* and *Pada Varnam*. *Tana Varnam* is performed in music concerts. *Pada Varnam* primarly intended for classical dance.

2. *Krithi*: *Krithi* is an important piece of a Carnatic music concert. It is usually based on a lyrical composition praising a personal deity or a patron king. Composers choose a certain *Raga*, *Tala*, style and can also improvise the piece to impress the audience. Besides the lead artist and melodic instruments, there are also accompanying percussion instruments that establish a clear rhythmic framework based on the *Tala*. Since the percussionist(s) provide rhythmic framework, the *Krithi* has a clear notion of tempo, which usually stays roughly constant for the entire piece. A *Krithi* may last for upto 30 minutes.

   Krithi is sub-divided into three parts and they are, *Pallavi*, *Anu-pallavi* and *Charanam*. *Pallavi* is same as refrain in the Western music. *Anu-pallavi* is the second verse and sometimes optional. *Charanam* is the longest and the final verse that concludes the piece. *Charanams* last line usually contains the signature of the composer.

3. *Ragam Talam Pallavi* : It is the middle piece of a typical Carnatic music concert. It consists of *Raga Alapana, Tanam* and *Pallavi*. It may last for about 15-20 minutes. *Raga Alapana* is the improvised piece in the concert followed by *Tanam*. *Raga* is improvised by using the words *Anantam Anandam* (bliss). It uses the syllables like *aa, nam, taa, tham, na, thom, tha* in a repetition form. Rhythmic pulse plays an important role in the *Tanam*

exposition. *Pallavi* is usually a chosen lyrical line that is explored in the *Raga* and *Tala* framework.

4. *Viruttam* : In this piece devotional songs are performed. This may include various *Slokas*[4], *Bhajans*[5] and other compositions performed in honor of the performer's teacher. It does not posses a set *Tala* but solely improvises one or more *Raga* in the same piece. Each verse of the piece is improvised with different *Raga* followed by a song. The song performed will have the same *Raga* as that of last verse of *Viruttam*. This piece may perform the same verse with different *Raga* in different concerts. It is mainly performed at the end of the concert. Apart from music concerts, it is also performed in traditional celebrations in the praise of Lord Muruga and Lord Ayyapa [1].

5. *Tillana* : It is a rhythmic piece performed almost at the end of the concert. It is a composed piece intended mainly for dance performances. It uses certain syllables denoting division of the *Tala* like, *Ta, Deem, Thom* and *Takadimi*.

6. *Mangalam* : This is the last piece of a Carnatic music concert. In this piece the artist performs a thankful prayer in honor of his/her teacher and concludes the musical event.

An artist (vocalist or instrumentalist) has the freedom to improvise based on a given *Raga* and *Tala* in any of the compositions. There are several improvisational aspects of a composition as listed below.

1. *Alapana* : The *Alapana* part is a slow improvisational exposition, which introduces a *Raga* and its underlying mood. This part may last up to 30 minutes, is performed without any percussion and involves only the lead artist (often a singer) and a main melodic instrument (often a violin) that follows, imitates, accompanies and interacts with the lead artist. In the *Alapana*, various tonal and melodic aspects of the *Raga* are explored, and typical phrases built from the *Raga* are presented to the audience. Being performed in a relaxed and free manner, the *Alapana* part has only a vague sense of tempo.

2. *Neraval*: In this part, artist takes the lines from the *Krithi* and sing the lines over and over each time. It may last for about $5-10$ minutes. The*Neraval* is accompanied with percussion instruments. The artist keeps the track of *Tala* to ensure that the lines of the composition occur in the correct position of the *Tala* where the line starts. This is the main aspect of Pallavi of the *Krithi* or the *Ragam*, *Talam* and *Pallavi* piece. The purpose of *Neraval* is to elaborate the selected lines to bring out the underlying music and lyrical beauty in it.

3. *Kalpana Swara*: In this part, musicians display various phrases of the *Raga* through the *Swara* syllables namely *Sa, Ri, Ga, Ma, Pa, Dha* and *Ni*. It may last for about $5-10$ minutes. The artists sings several *Swaras* and finish on the same line very time, ensuring the *Swara* exposition had ended on the exact position of the *Tala*. The choice of *Swaras* used must stick to the grammar of the *Raga*. For example, when singing *Kalpana Swaras* in the *Raga Hamsadhwani* which consist of *Sa, Ri, Ga, Pa* and *Ni* notes. The artist is usually restricted from using any other note than the notes of the*Raga Hamsadhwani*.

[4] *Slokas* is a category of verse line developed during *Vedic* times
[5] *Bhajans* is a type of Hindu devotional song

4. *Tani-Avarthanam* : This part is dedicated to the solo performances by the percussionists. If more than one percussion instrument is involved in the concert, each percussionist takes turns to exhibit his or her creativity in presenting interesting rhythmic patterns. The percussionists finally join together in a grand crescendo following the same rhythmic patterns. In the *Tani-Avarthanam*, there are no melodic instruments involved except for a drone. Sometimes a *morsing* (Jaw harp) may be present acting as a kind of melodic percussionist. While exploring the nuances of the underlying *Tala*, the rhythms presented by the percussionists are often of high speed involving complex and syncopated patterns. Usually, the *Tani-Avarthanam* part, which also may take up to 20 minutes, has a clear notion of tempo. However, the tempo may change several times, in particular between the various solo sections.

2.2 Carnatic Music Concert Structure

A Carnatic concert typically features a lead artist (often a vocalist), who is supported by a lead melodic instrument (usually a violin). One or more percussion instruments (like the *Mridangam*, *Ghatam*, or *kanjira*) may accompany the artists to provide both rhythmic and timbral variety. A drone (the *Tanpura*) is often used to support the melodic instruments or the singer. A list of harmonic and percussive instruments used in Carnatic music are shown in Figure 2.1 and Figure 2.2.

A typical concert has around 7 to 8 pieces performed over a duration of 2 to 3 hours. There is usually 1 *Varnam* piece, 2 or 3 small *Krithi* pieces, 1 main piece, 1 or 2 elaborate *Krithi* pieces (one of which may incorporate *Viruttam*), 1 *Thillana* and 1 *Mangalam* piece. The smaller pieces may last for about $5 - 20$ minutes while the main piece may last upto 60 minutes. The lead performer chooses the set of pieces to perform depending on the occasion and the taste of the audience.

The main piece may include *Alapana*, *Krithi* and *Tani-Avarthanam* or either *Alapana* or *Tani-Avarthanam* along with *Krithi* or it can be only *Krithi*, as it is the considered the heart of the main piece. The lead artist renders the main piece of the concert in his personal style. The artist(vocalist or instrumentalist) displays improvisation of *Raga* and creativity in *Alapana and Krithi* parts and then gives an opportunity for the percussionists to show their creativity in the *Tani-Avarthanam* part.

In this thesis, we shall consider the main piece having all the three parts, *Alapana*, *Krithi* and *Tani-Avarthanam*. The state of the art and the research in the filed of automatic segmentation techniques for Carnatic music are discussed in the next Section.

2.3 Segmentation Task

Given an audio recording of the main piece consisting of *Alapana, Krithi* and *Tani-Avarthanam*, we perform segmentation using various musical cues such as rhythm, tempo, tonal content, harmony, melody, instruments, timbre and so on. Specific cues are selected that can differentiate the segments with respect to each other. Main contribution of this thesis is to make use of two

Figure 2.1: Harmonic instruments used in Carnatic music (for more details, refer Appendix A).

musical cues. The first cue is the tempo which captures the information on the existence or absence of local tempo changes. The second cue is the chroma which captures the information of existence or absence of melody and tonal information in constituent parts of the main piece audio recording.

Till date only little work has been done on automatic segmentation of Carnatic pieces of the music. Padi Sarala and Hema Murthy proposed an idea to segment Carnatic music recordings into individual items for archival purposes using applauses between the different parts [29].

Figure 2.2: Percussion instruments used in Carnatic music (for more details, refer Appendix A).

Further Hema Murthy et al. extended their work to segment the pieces based on identifying inter and intra applauses in the music [28]. Inter applauses were used to locate the end of each part where as intra applauses were identified which helped to merged the parts belonging to the same item. Apart from segmenting Carnatic music pieces, many techniques are proposed on *Raga* identification. Note identification based on frequency spectrum [27]. Identification of Raga was possible using midi representation [26] and Hidden Markov Model [9].

In the next Chapter 3, we discuss how the tempo cue can musically distinguish *Alapana* with respect to *Krithi* and *Tani-Avarthanam* parts of a main piece audio recording. We design several tempo salience features based on the cyclic tempogram representation, which captures the absence or the presence of the tempo in the constituent parts.

Chapter 3

Tempo Salience Features

For any audio recording, the extraction of tempo and beat information is a challenging task, in particular for music with soft onsets[1] and local tempo variations. Carnatic music has very complex predefined rhythmic beat structure (*Tala*). The artist has a freedom of speeding up or slowing down of the tempo as a piece progresses, especially in the *Krithi* and *Tani-Avarthanam* parts of the main piece. Sometimes there is also a vague notion or non-existence of beat in the *Alapana* part. We exploit this musical property to segment the *Alapana* from the *Krithi* and *Tani-Avarthanam* in which there exists a clear notion of beat and local tempo changes (see Figure 3.1). Hence, we design few tempo salience features from the time-tempo representation to musically distinguish the constituent parts of a Carnatic concert main piece audio recording.

The local tempo variations and beat information is captured in two steps. Firstly, note onset from the music signal are extracted by exploiting the fact that the note onset typically occur due to the sudden changes in the signal's energy or spectrum. Based on this property, the note onset detection novelty curves are derived. Secondly, the novelty curves are analyzed for local periodic patterns using tempogram from which the local tempo variations are estimated.

Tempogram is a time-tempo representation that encodes the local tempo of a music signal over time [16]. This can be obtained from comb-filter, Fourier or autocorrelation methods. In the Fourier-based tempogram [14, 15], the novelty curve is compared with sinusoidal kernels each representing a specific tempo. It reveals the local similarity of the novelty curve and is suitable for analyzing tempo on tatum and tactus level [20]. In the autocorrelation-based tempogram [12], novelty curve is compared with time-lagged windowed sections of itself. It reveals the novelty similarity and is suitable for analyzing tempo on tactus and measure level. In the comb filter-based tempogram, the tempo tracker is modeled as stochastic dynamical system and is estimated by Kalman filtering [12].

Inspired by the concept of chroma features. We use the concept of cyclic tempogram, where the idea is to form tempo equivalence classes by identifying tempi that differ by a power of two [21]. It is a mid level representation which robustly tracks the beat and local tempo changes [22].

In order to derive the novel features (see Figure 3.2) based on tempo property, we follow three steps. Firstly, we obtain the note onset detection novelty curve using spectral based

[1] Onset is the time position where a note is played or finding start times of perceptually relevant to acoustic events in music signal.

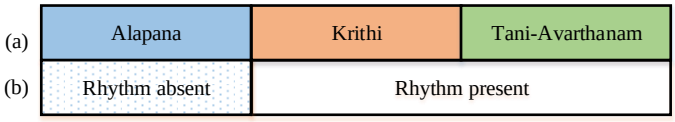

Figure 3.1: Typical structure of a Carnatic main piece.**(a)** Main piece constituting of three contrasting parts: *Alapana*, *Krithi* and *Tani-Avarthanam*.**(b)** Tempo property distinguishes *Alapana* with respect to *Krithi* and *Tani-Avarthanam*.

method [11, 32]. Secondly, we obtain the autocorrelation based tempogram using the novelty curve. Finally, cyclic tempogram is used to track local tempo changes in each part of the main piece audio recording.

The main contribution of this Chapter is to design the tempo salience features based on the cyclic tempogram representation, which captures the absence or the presence of the tempo in the constituent parts of the main piece audio recording. The cyclic tempogram representation is obtained from the existing tempogram toolbox [16]. This Chapter is organized as follows. We introduce the tempogram toolbox algorithms and parameter settings in the first two Sections. In the Section 3.3, we design the tempo salience features based on cyclic tempogram representation to musically distinguish the constituent parts of the main piece .

3.1 Note Onset Detection

A note onset is a time position where the note is played. The note onset detected can be obtained in 2 steps. Firstly, transform the signal to a suitable feature representation. Secondly, derive a novelty function based on some kind of derivative operative to detect the note onsets. There are different methods to compute the novelty function such as energy based or spectral based novelty functions [16]. Energy based method is good for percussive instruments having hard onsets but not for harmonic string instruments having weak onsets. To increase the robustness of the onset detection we use spectral based method which is more refined and used for onset detection [11, 32]. The steps involved to compute the novelty function are discussed in the Section 3.1.1.

3.1.1 Spectral Based Novelty

Onset detection becomes much harder when we have polyphonic music. The low intensity and high intensity events may occur at the same time. The low intensity events may be masked by the high intensity events. It is very hard to detect all onsets when we have multiple instruments played at the same time having different energy fluctuations in the sustain phase. It is difficult to obtain all the onsets using energy based approach. Characteristics of note onset events may differ based on different categories of instruments. Percussive instruments have impulse like note onsets with sudden increase of energy across all the frequencies of the spectrum. For harmonic instruments most of the energy is concentrated in lower frequency bands for harmonic instruments. Transients are often well detected in the higher frequency bands.

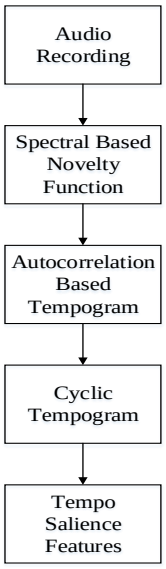

Figure 3.2: Tempo salience feature extraction.

Hence motivated by this observation, we use the concept of spectral based novelty detection to first convert the signal to time-frequency representation (spectrogram) and then capture the spectral changes throughout the frequency range [16]. The steps involved in computing the onset detection novelty function are mentioned below.

1. *Spectrogram*: Given a discrete time domain signal x of an audio recording. Convert the time domain signal into time-frequency representation to detect the spectral changes over the entire frequency range. Aspects concerning pitch, harmony or timbre are captured by spectrogram. Let X be the discrete STFT (short-time Fourier transform) of the discrete time signal x with parameters including sampling rate $F_s = 1/T$, the discrete window w with window length N and hop size H. The discrete STFT signal mathematically represented as $X(n, k) \in \mathbb{C}$. It denotes the k^{th} Fourier coefficient for frequency index $k \in [0 : K]$ and time frame $n \in \mathbb{N}$, where $K = \frac{N}{2}$ is the frequency index corresponding to Nyquist frequency.

2. *Logarithmic compression*: In order to enhance the spectral coefficients, we apply logarithmic compression to the spectral coefficients which mimics the processing within the auditory systems. The advantage of such a compression is to enhance weak high frequency spectrum, low intensity values and balance out the dynamic range of the signal [20]. We apply a logarithm to the magnitude spectrogram $|X|$ of the signal yielding to,

$$Y := \log(1 + \mathcal{C} \cdot |X|), \tag{3.1}$$

with suitable constant $\mathcal{C} > 1$ which acts as compression factor for spectral coefficients. Different values of $\mathcal{C}$ are chosen. For example, for $\mathcal{C} = 1$, the low frequency component is visible but may not track the weak vertical line corresponding to the beat positions in the spectrogram. For $\mathcal{C} = 1000$, even the weak vertical lines are prominent and can track the weak onsets. Large value of $\mathcal{C}$ corresponding to larger compression may end up in

amplifying the non relevant noise like components, hence an optimum value of $\mathcal{C}$ is to be selected.

3. *Differentiation*: To capture the changes in the spectral content, we compute the discrete temporal derivative of the log compressed spectrum [11, 32]. . Increase in the intensity is obtained by considering only the positive differences and by discarding the negatives ones which, results in spectral based novelty function, $\triangle_{Spectral} : \mathbb{Z} \to \mathbb{R}$ for $n \in \mathbb{Z}$.

$$\triangle_{Spectral}(n) := \sum_{k=0}^{K} |Y(n+1,k) - Y(n,k)|_{\geq 0}, \tag{3.2}$$

4. *Normalization*: One can further enhance the properties of novelty function by applying post-processing techniques to suppress the small fluctuations [11, 32]. . Enhanced novelty function $\bar{\triangle}_{Spectral}$ can be obtained in two steps. Firstly, subtract $\bar{\triangle}_{Spectral}$ with the local mean $\mu(n)$ as in equation (3.3). Finally, we perform half-wave rectification to obtain only the positive part that enhances the peak structure and reduce the spurious note onset peaks. This results in a function, which is known as note onset detection novelty function

$$\mu(n) := \frac{1}{(2M+1)} \sum_{k=-M}^{M} \triangle_{Spectral}(n+m), \tag{3.3}$$

$$\bar{\triangle}_{Spectral}(n) := | \triangle_{Spectral}(n) - \ \mu(n)|_{\geq 0}, \tag{3.4}$$

The resulting novelty function serve as a basis for deriving the time-tempo representation, which is explained in the Section 3.2.

3.2 Tempogram Representation

Similar to the idea of a spectrogram, a tempogram is a time-tempo representation, which indicates for each time instance the local relevance of a specific tempo for a given audio recording. Mathematically, a tempogram can be modeled as a function,

$$\Gamma(t,\tau) : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}, \tag{3.5}$$

where, time $t \in \mathbb{R}$ measured in seconds and a tempo $\tau \in \mathbb{R}_{>0}$ measured in beats per minute (BPM). A large value $\Gamma(t,\tau)$ indicates that the music signal has at time $t \in \mathbb{R}$ a dominating tempo $\tau \in \mathbb{R}_{>0}$. $\Gamma(t,\tau)$ indicates to what extent the signal contains the locally periodic pulse of a given tempo τ in a neighborhood of time instance t [16]. For example, let us suppose that the music signal has a dominant tempo of 200 $\mathbb{BPM}$ around the time position $t = 20sec$ then, the resulting tempogram Γ has a large value at $t = 20$ sec and $\tau = 200$ $\mathbb{BPM}$.

Tempogram representation of a music recording can be derived in two steps. Firstly, based on the two assumptions which were discussed earlier, to convert the given music signal into note

onset novelty function. In the second step, we analyze the locally periodic behaviour of the novelty function $\bar{\triangle}_{Spectral}$.

The novelty curves are analyzed for the locally periodic patterns for various periods $T > 0$ in a neighborhood of a given time instance. The period T ($T = 1/\omega$, ω in Hz) and the tempo τ (in BPM) are related by,

$$\tau = 60 \cdot \omega. \tag{3.6}$$

If we consider a musical recording that reveals significant tempo changes, the detection of locally periodic patterns becomes a challenging task. Furthermore, there are various pulse levels that contribute to the human perception of tempo such as the *tatum*, *tactus*, and *measure* levels [20]. Tempo on *tactus* level matches to the foot tapping rate, *measure* to fast music and tatum refers to the fastest music.

Due to the ambiguity concerning the pulse levels, the tempogram Γ takes into account the existence of different pulse levels. Higher pulse level corresponding to integral multiples of τ, 2τ, 3τ... of a given tempo τ (referred as *harmonics* of τ) and integer fractions τ, $\tau/2$, $\tau/3$... of a given tempo τ (referred as *subharmonics* of τ).

As mentioned before, we can derive two different types of tempograms, one emphasizing on tempo *harmonics* (using Fourier method, see Figure 3.3 b) and the other on tempo *sub-harmonics* (using autocorrelation method, see Figure 3.3 d). In our implementation, we use autocorrelation based tempogram, which is discussed in Section 3.2.1.

3.2.1 Autocorrelation Tempogram

In this Section, we discuss the autocorrelation based approach for computing a tempogram [12]. Autocorrelation is a mathematical tool for measuring the degree of similarity between a given time series and a delayed version of itself over successive time intervals. It can also be interpreted as calculating the correlation between two different time series, except that the same time series is used twice, once in its original form and once lagged one or more time instants.

Let us consider a discrete-time real valued signal $x \in l^2(\mathbb{Z})$ having finite energy then the autocorrelation $R_{xx} : \mathbb{Z} \to \mathbb{C}$ of signal x is defined as,

$$R_{xx}(l) = \sum_{m \in M} x(m)x(m-l), \tag{3.7}$$

where $l \in \mathbb{Z}$ is the time lag parameter. R_{xx} is well defined in space $l^2(\mathbb{Z})$, is maximal for $l = 0$ and symmetric in l

To analyze the given novelty function ($\bar{\triangle}_{Spectral}$, see Figure 3.3 a), we now apply autocorrelation in a local fashion with a time parameter n. Let the window function be $w : \mathbb{Z} \to \mathbb{R}$ of finite length at $n = 0$, then the windowed version of $\bar{\triangle}_{Spectral_{(w,n)}}$ is given by,

$$\bar{\triangle}_{Spectral_{(w,n)}}(m) := \bar{\triangle}_{Spectral}(m)w(m-n), \tag{3.8}$$

where $m \in \mathbb{Z}$. If we assume, that the window function w lies between the interval $[-L : L]$ and $L \in \mathbb{N}$, then the unbiased local autocorrelation $\Re(n, l) : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ is given by,

$$\Re(n, l) := \frac{\sum_{m \in M} \bar{\triangle}_{Spectral}(m) w(m-n) \bar{\triangle}_{Spectral}(m-l) w(m-n-l)}{2L+1-l}. \tag{3.9}$$

To obtain the time-lag representation from time-tempo representation, we need to convert the lag parameter into tempo parameter. Let the time frame be r seconds and the time-lag be l seconds. The shift of each time frame corresponds to a rate $1/\ (l \cdot r)$ Hz, then the time-tempo can be obtained from time-lag by,

$$\tau = \frac{60}{l \cdot r} \text{BPM}. \tag{3.10}$$

If we assume that, there is a high correlation of the windowed section of novelty function with a shift a l lags where $k \in \mathbb{N}$ then, l corresponds to tempo τ and the lags $k.l$ corresponds to sub-harmonics τ/k.

3.2.2 Cyclic Tempogram

As an analogy, the different tempo levels like measure, tactum and tactus may be compared to the existence of harmonics in the pitch context. Inspired by the concept of chroma features, we introduce the concept of cyclic tempogram which reduces the effect of harmonics, where the idea is to form tempo equivalence classes by identifying tempi that differ by a power of two [21].

To be more precise, if we assume two tempi say τ_1 and τ_2, they are said to be octave equivalent, if they are related by $\tau_1 = 2^{\mu}\tau_2$ for some $\mu \in \mathbb{Z}$.

Given a tempogram representation $\Gamma : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$, cyclic tempogram is defined as

$$\Im(t, [\tau]) := \sum_{\alpha \in [\tau]} \Gamma(t, \alpha). \tag{3.11}$$

Note that the tempo equivalence classes topologically corresponds to a circle. Fixing a reference tempo ρ (e. g., $\rho = 60BPM$), the cyclic tempogram can be represented by a mapping $\Im_\rho : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ defined by,

$$\Im_\rho(t, s) := \Im(t, [s \cdot \rho]), \tag{3.12}$$

for $t \in \mathbb{R}$ and $s \in \mathbb{R}_{>0}$. Note that $\Im_\rho(t, s) = \Im_\rho(t, 2^k s)$ for $k \in \mathbb{Z}$ and $\Im_\rho$ is completely determined by its relative tempo values $s \in [1, 2)$. Figure 3.3c shows an example for cyclic tempogram using autocorrelation based method.

So far we assumed that, the time and tempo parameters are continuous. In practice, one computes a cyclic tempogram only for a finite number of time points t and a finite number of relative tempo parameters s. In the following, let N be the number of time points (corresponding to frames) and M the number of considered scaling parameters (logarithmically spaced on the

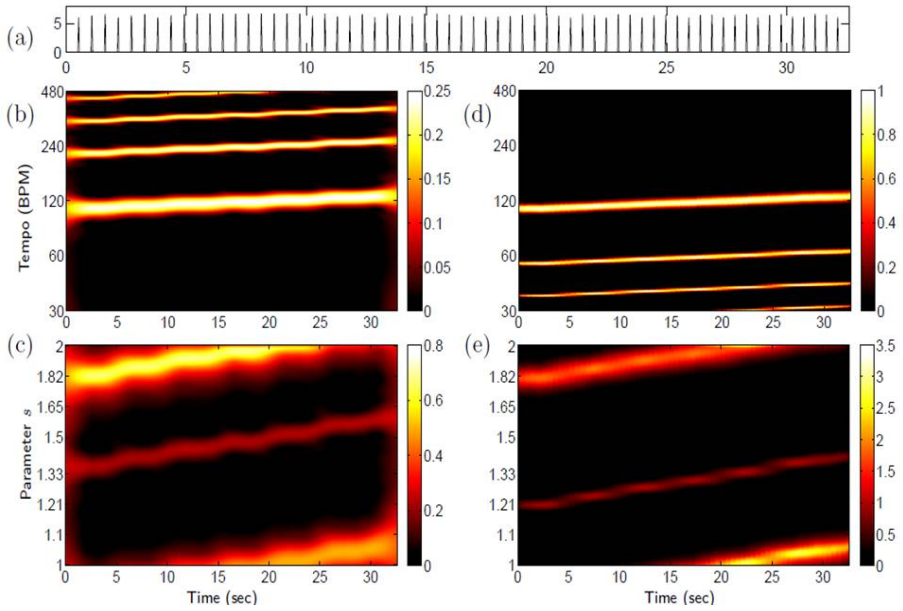

Figure 3.3: Tempo features for a click track with increasing click frequency: **(a)** Novelty curve, **(b)** Fourier-based tempogram(showing harmonics), **(c)** Cyclic Fourier-based tempogram **(d)** Autocorrelation-based tempogram(showing sub-harmonics), **(e)** Cyclic autocorrelation-based tempogram. Reproduced from [17].

tempo axis). By abuse of notation, let $\Im_\rho(n, m)$ denote the values of the cyclic tempogram for discrete time parameters $n \in [0 : N-1]$ and relative tempo parameters $m \in [0 : M-1]$, refer [16] for more details.

There are different ways for computing tempograms and its cyclic versions. In the following, we use a cyclic tempogram computed from an autocorrelation tempogram as described in [14, 17]. In order to obtain the autocorrelation based cyclic tempogram representation, we use the existing tempogram toolbox in this thesis with the following parameter settings [16] . A higher-order smoothed differentiator [8] of filter length 0.3 seconds is used. The spectrum is processed in a band wise fashion using five bands, which are logarithmically spaced and non-overlapping (with logarithmic compression factor $\mathcal{C} = 1000$). Each band is roughly one octave wide. The lowest band covers the frequencies from 0 Hz to 500 Hz, the highest band from 4000 Hz to 11025 Hz. The five novelty curves which are summed up to obtain the resulting novelty function. From the novelty function, the autocorrelation based cyclic tempogram is obtained. There are three main parameter setting for autocorrelation based cyclic tempogram, which specify the length L (measured in seconds) of the analysis window used in the local autocorrelation, a hop size parameter that determines the final feature rate F_s (measured in Hertz) of the tempogram, and the number M of relative tempo parameters that determines the dimension of the feature vectors. In our setting, using $L = 16$ sec, $F_s = 5$ Hz, and $M = 15$ turned out to be a reasonable setting for the experiments, which are further discussed in chapter 5.

3.3 Salience Features

This Section contains one of the main technical contribution of the thesis. In this Section, we design the tempo salience features based on tempogram representation. The tempogram representation is obtained from the existing tempogram toolbox [16] as described in the Section 3.1 and in the Section 3.2.

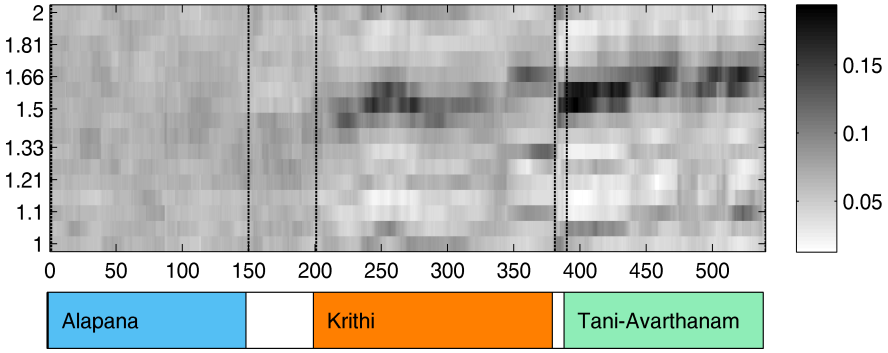

Figure 3.4: Discrete version of a normalized cyclic tempogram representation based on an autocorrelation tempogram using the parameters $L = 16$ sec, $F_s = 5$ Hz, and $M = 15$.

Recall from the Section 1.1 that, the *Krithi* and *Tani-Avarthanam* parts have a strong notion of tempo, as opposed to Alapana. As mentioned earlier that, *Alapana* is an improvisational part where only exposition of *Raga* takes place, with no rhythm involved in it. This observation can be better explained with a cyclic tempogram representation of a Carnatic concert main piece audio recording as shown in Figure 3.4. In *Alapana* part, the tempo looks rather diffused with no large coefficients that would indicate the clear notion of tempo. On the other hand, *Krithi* and *Tani-Avarthanam* parts have stronger coefficients that would indicate the presence of specific dominating tempo entry. As we can also notice that the dominating tempo entries may vary over time in *Krithi* and *Tani-Avarthanam*, which reflects the fact that the tempo is changing.

Our objective is to not find the specific tempo in each part but to differentiate the parts with the absence or presence of notion of tempo. In the the following we refer this property as *tempo salience*. We now describe several kinds of salience features derived from a tempogram representation as mentioned below.

3.3.1 Entropy Feature

A first idea is to apply the concept of *entropy*, which is usually used to express the uncertainty when predicting the value of a random variable. For a probability vector $p = (p_0, \ldots, p_{M-1})^T \in \mathbb{R}^M$, the (normalized) entropy is defined by

$$\mathcal{H}(p) = -\big(\textstyle\sum_{m=0}^{M-1} p_m \log_2(p_m)\big) / \log_2(M), \tag{3.13}$$

which assumes a maximal value of one if the vector p corresponds to a uniform distribution and a minimal values of zero if the vector p corresponds to a dirac distribution.

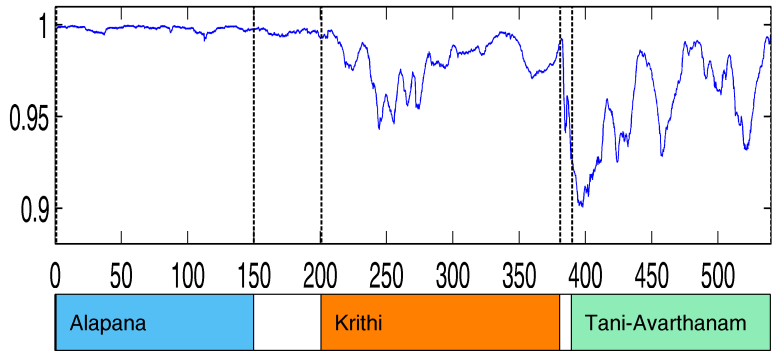

Figure 3.5: Entropy feature $\mathcal{H}(X)$ obtained from the discrete version of a normalized autocorrelation based cyclic tempogram representation (see, Figure 3.4) of a Carnatic concert main piece audio recording.

In our scenario, we normalize the columns of the cyclic tempogram $\Im_\rho \in \mathbb{R}^{N\times M}$ with regard to the Manhattan norm to obtain a matrix $X \in [0,1]^{N\times M}$. Then, each column $X[n] \in \mathbb{R}^M$, $n \in [0:N-1]$, of X can be interpreted as a probability vector. Applying the entropy to each column, we obtain the sequence

$$\mathcal{H}(X) := (\mathcal{H}(X[0]), \ldots, \mathcal{H}(X[N-1])) \tag{3.14}$$

of numbers $\mathcal{H}(X[n]) \in [0,1]$, see Figure 3.5 for an example. To obtain a measure of salience (rather than one of uncertainty), we consider $1 - \mathcal{H}(X[n])$. Further smoothing this sequence by applying an averaging filter of some length $\lambda \in \mathbb{N}$ yields our first feature that we refer to as $f_\lambda^{\mathcal{H}}$. As demonstrated by Figure 3.6, this feature has the desired property of being close to zero in the *Alapana* part and much larger in the other parts, see Chapter 5 for a more detailed investigation.

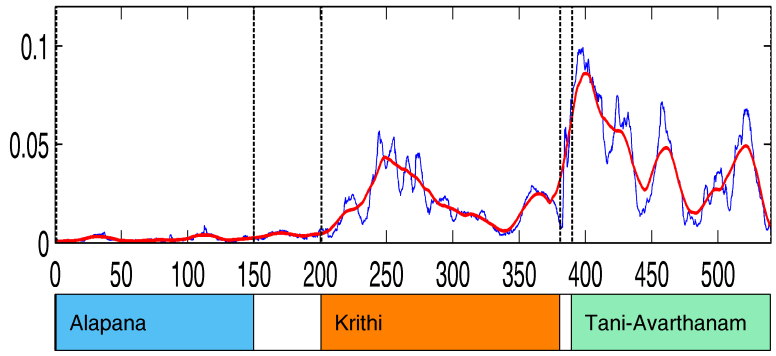

Figure 3.6: Feature $f_\lambda^{\mathcal{H}}$ with $\lambda = 1$ (blue) and $\lambda = 100$ corresponding to 20 sec (red). It is obtained from the discrete version of a normalized autocorrelation based cyclic tempogram representation (see, Figure 3.4) of a Carnatic concert main piece audio recording.

3.3.2 Max Median Feature

As an alternative to the entropy, one may also look at the difference of the maximum value and the median value of a probability vector. This yields a number

$$\mathcal{M}(p) := \max\{p_0, \ldots, p_{M-1}\} - \text{median}\{p_0, \ldots, p_{M-1}\} \tag{3.15}$$

in the interval $[0, 1]$, which assumes the value 0 in the case that p is a uniform distribution and the value 1 if p is dirac distribution. Applying $\mathcal{M}$ to each column of X and smoothing the resulting sequence with an averaging filter of length $\lambda \in \mathbb{N}$ yields our second feature we refer to $f_\lambda^{\mathcal{M}}$. As illustrated by Figure 3.7, this feature behaves similarly to $f_\lambda^{\mathcal{H}}$.

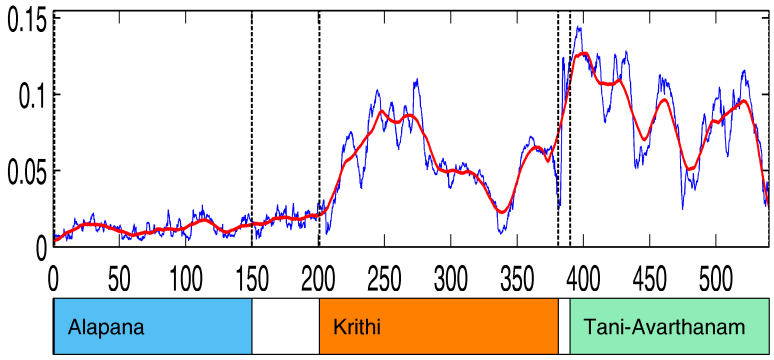

Figure 3.7: Feature $f_\lambda^{\mathcal{M}}$ with $\lambda = 1$ (blue) and $\lambda = 100$ (red). It is obtained from the discrete version of a normalized autocorrelation based cyclic tempogram representation (see, Figure 3.4) of a Carnatic concert main piece audio recording.

3.3.3 Tempo Density Feature

Next, we introduce a conceptually different salience feature, which measures a kind of density of abrupt and significant tempo changes. To this end, we first compute the maximizing tempo index for each column of X:

$$m^{\max}(n) := \text{argmax}_{m \in [0:M-1]}(X(n, m)). \tag{3.16}$$

Then the idea is to look at differences of the resulting sequence of tempo indices over subsequent time frames. However, when computing these differences, one needs to take into account that we are dealing with *cyclic* tempogram features. Therefore, we define a cyclic distance by setting

$$d^{\text{cyc}}(m_1, m_2) := \min\left\{|m_1 - m_2|, M - |m_1 - m_2|\right\} \tag{3.17}$$

for $m_1, m_2 \in [0 : M - 1]$. With this definition at hand, we then define

$$\mathcal{I}(n) := d^{\text{cyc}}\big(m^{\max}(n), m^{\max}(n-1)\big) \tag{3.18}$$

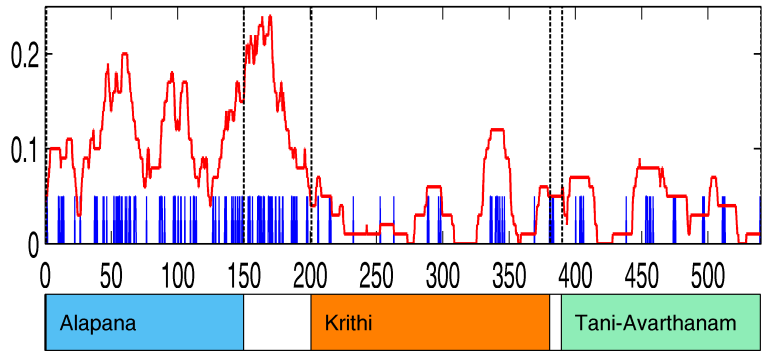

Figure 3.8: Feature $f^{\mathcal{I}}_{\tau,\lambda}$ with $\tau = 0$ and $\lambda = 1$ (blue, size of binary values reduced for visibility reasons) and $\lambda = 100$ (red). It is obtained from the discrete version of a normalized autocorrelation based cyclic tempogram representation (see, Figure 3.4) of a Carnatic concert main piece audio recording.

for $n \in [1 : N-1]$. Intuitively, any value $\mathcal{I}(n) > 0$ expresses that there has been a tempo change at time n. Now, smooth tempo changes and small local tempo fluctuations may result in a value $\mathcal{I}(n) = 1$ as illustrated by the tempogram in the *Krithi* part of Figure 3.6. Therefore, being interested in measuring abrupt tempo changes rather than small deviations, we introduce a tolerance parameter $\tau \in \mathbb{N}_0$ and define the feature $f^{\mathcal{I}}_{\tau}$ by setting

$$f^{\mathcal{I}}_{\tau}(n) := \begin{cases} 0 & \text{if } \mathcal{I}(n) \leq \tau, \\ 1 & \text{if } \mathcal{I}(n) > \tau. \end{cases} \tag{3.19}$$

As before, applying an averaging filter of length $\lambda \in \mathbb{N}$ yields a feature we refer to as $f^{\mathcal{I}}_{\tau,\lambda}$.

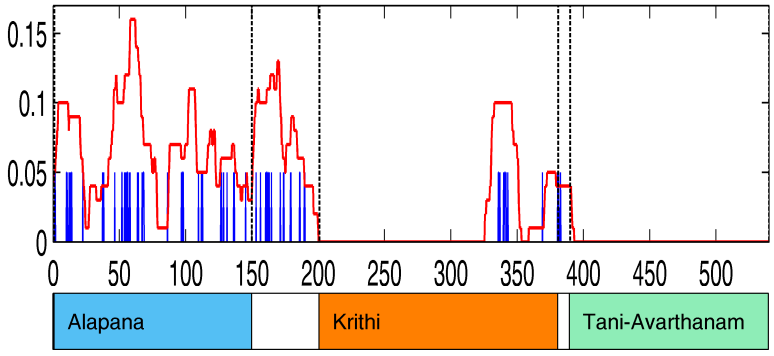

Figure 3.9: Feature $f^{\mathcal{I}}_{\tau,\lambda}$ with $\tau = 1$. It is obtained from the discrete version of a normalized autocorrelation based cyclic tempogram representation (see, Figure 3.4) of a Carnatic concert main piece audio recording.

These definitions are illustrated by Figure 3.8, which shows the binary feature $f^{\mathcal{I}}_{\tau}$ for $\tau = 0$ and its averaged version $f^{\mathcal{I}}_{\tau,\lambda}$ using λ corresponding to 20 sec. One important observation is that this density feature tends to assume large values in sections with a diffuse tempo (such as in the *Alapana* part). In such noise-like sections, the maximizing index randomly jumps from frame

to frame, which results in many non-positive values of $\mathcal{I}$. Using a tolerance parameter $\tau = 1$ results in the features shown in Figure 3.9. In this case, smooth tempo changes as occurring in the *Krithi* and *Tani-Avarthanam* parts do not contribute to the density feature.

In this Chapter, we derived the tempo salience features which has a strong relevance of tempo in *Krithi* and *Tani-Avarthanam* but low value in *Alapana*. In the next Chapter, we use the melody property to musically distinguish the constituent parts of a Carnatic concert main piece audio recording. We design few novel features which has a strong relevance of melody in *Alapana* and *Krithi* but low value in *Tani-Avarthanam*.

Chapter 4

Chroma Salience Features

A music signal has several musical aspects such as tempo, beat, played notes, melody, harmony, timbre of different instruments, dynamics of the sound, etc. For a given music processing task, only few of them may be relevant. In the last Chapter, we derived few novel features to distinguish the *Alapana* from the *Krithi* and the *Tani-Avarthanam* based on the tempo cue. Similarly, in this chapter, we use the melody property to design the novel features to musically differentiate the *Tani-Avarthanam* from the *Alapana* and the *Krithi* parts of the main piece audio recording. In doing so, we accomplish the task of segmenting the *Alapana*, *Krithi* and *Tani-Avarthanam* into individual parts.

As mentioned earlier in the Section 1.1 that, a drone instrument is played throughout the main piece to provide a pitch reference to the artists. By neglecting the effect of the drone, we can make a few observations based on the melody property as follows. The *Tani-Avarthanam* is a purely percussive and has an absence of melody. The *Alapana* and the *Krithi* parts have clear notion of melody as shown in Figure 4.1b. Hence, we use the *chroma*[1] cue, which captures the information of the existence or absence of melody and tonal information of the constituent parts.

This Chapter is organized as follows. We present pitch[2] features obtained from the existing chroma toolbox, which serve as a basis for the other features in our experiments (see Section 4.1). We discuss the tuning of the drone instruments used in the Carnatic concerts and also state the art to remove the drone from the pitch feature representation of the main piece audio recording. After the drone removal, we now obtain enhanced chroma features (see Section 4.2). The enhanced chroma features are computed based on the singing octave range of the singers in the Carnatic music. Finally, we design the novel features based on the chroma features (see Section 4.4).

The main contribution of this Chapter is to design a few chroma novel features based on pitch and chroma feature representations. The pitch and chroma feature representations are obtained by using the existing chroma toolbox [25].

[1] *Chroma* is a set of all the pitches belonging to the same pitch class which are perceived as having a similar "quality" or "color".

[2] Pitch is a property of a sound that correlates to its perceived frequency [24].

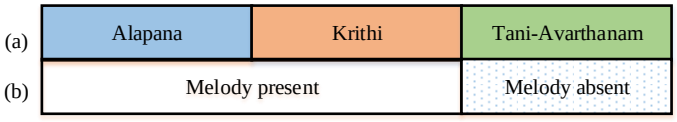

Figure 4.1: Typical structure of a Carnatic main piece. **(a)** Main piece constituting of three contrasting parts: *Alapana*, *Krithi* and *Tani-Avarthanam*. **(b)** Melody property distinguishes the *Tani-Avarthanam* part with respect to the *Alapana* and the *Krithi* parts.

4.1 Pitch Features

In order to obtain the chroma features, we first decompose a given audio signal into 88 frequency bands, which corresponds to 88 musical notes. The 88 frequency bands have center frequencies corresponding to the pitches $A0$ to $C8$ [25]. The pitches $A0$ to $C8$ are the MIDI[3] pitches corresponding to $p = 21$ to $p = 108$, where p is the MIDI note number. These musical notes are of equal-tempered scale. Every note is associated with a certain frequency range with a fixed center frequency. For example, the note $A4$ corresponds to MIDI note number $p = 69$ has the center frequency as 440 Hz.

Let p be the MIDI note number, $p \in [0 : 127]$ with f_p as its center frequency, then p and f_p are related by,

$$f_p = 2^{\frac{p-69}{12}} \cdot 440, \tag{4.1}$$

where, MIDI note number $p = 69(A4)$ with center frequency 440 Hz is taken as a reference for computing the center frequency f_p for the corresponding MIDI note number p. For example, $p = 81(A5)$ has a center frequency $f_p = 880$ Hz. From this, we can infer that, the pitch of a note that is one octave higher than a reference note is twice that of the reference note [19]. As increase in the MIDI note number p leads to an increase in f_p in logarithmic fashion.

The decomposition of a given audio signal into pitch sub-bands can be achieved by using multirate filter banks, which consist of an array of band-pass filters. As the pitch gets higher, the bandwidth of the corresponding filter gets wider (see Figure 4.2). For better spectral resolution either reduce the sampling rate or increase the temporal window length. Hence, we use different sampling rate for different pitches. Higher the pitch, lower will be the sampling rate and vice-versa.

Let X be the discrete STFT of the discrete time signal x with parameters including sampling rate $F_s = 1/T$, the discrete window w with window length N and hop size H. The discrete STFT is denoted $X(n, k) \in \mathbb{C}$. It denotes the k^{th} Fourier coefficient for frequency the index $k \in [0 : K]$ and time frame $n \in \mathbb{N}$, where $K = \frac{N}{2}$ is the frequency index corresponding to Nyquist frequency.

The frequency corresponding to spectral coefficient $X(n, k)$ is given by,

$$f_{coeff}(k) := \frac{k}{N} \cdot \frac{1}{T}. \tag{4.2}$$

[3]MIDI stands for Musical Instrument Digital Interface, is essentially a communications protocol for computers and electronic musical instruments

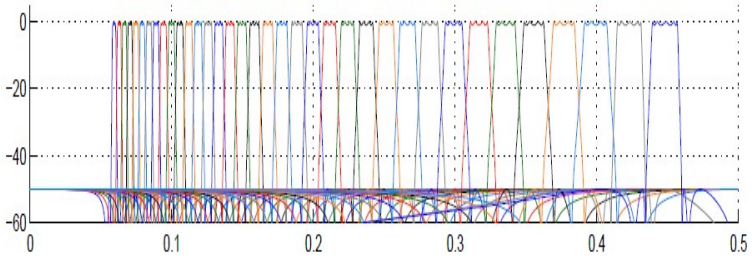

Figure 4.2: A sample array of filters with their respective magnitude responses in dB. (reproduced from [25]).

Let f_p be the MIDI note number of a pitch $p \in [0:127]$ (refer, equation (4.1)) and S(p) be the set of frequency indexes assigned to a MIDI note number p. The set S(p) is given by,

$$S(p) := \{k : f_p(p-0.5) \leq f_{coeff}(k) < f_p(p+0.5)\}. \tag{4.3}$$

For each sub-band, we now compute short time mean square power $\mathcal{P}(n,p)$ for a MIDI note number p and is given by,

$$\mathcal{P}(n,p) := \sum_{k \in S(p)} |X(n,k)|^2. \tag{4.4}$$

The resulting $\mathcal{P}(n,p)$ is referred to as pitch features for a given frame n. The pitch features measure the short time mean square power of the signal within each sub-band.

Further, global tuning of an audio recording is taken into account by suitably shifting the center frequencies of the sub-band filters of the multirate filter bank. This is done in two steps. Firstly, we compute average spectrogram vector. Secondly, we derive an estimate for the tuning deviation by simulating the filter banks shifts using weighted binning techniques [25].

In order to obtain MIDI pitch representation, we use the chroma toolbox with the parameter settings as follows. We employ a constant-Q multirate filter bank of 88 sub-bands with a sampling rate of 22050 Hz for high pitches, 4410 Hz for medium pitches and 882 Hz for low pitches (see [13, 18, 24] for further details). Short time mean square power for each sub-band is computed by using a fixed window length of 200 milliseconds with overlap of 50% (leads to feature rate = 10 features per second). For tuning of an audio recording, we consider pre-computed six multirate filter banks corresponding to a shift of $\sigma \in \{0, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}\}$ semi-tones respectively. According to the estimated tuning deviation, we choose the most suitable filter bank.

In the next Section, we discuss how to obtain the standard chroma features from the pitch feature representation.

4.2 Chroma Features

As we know that, the human perception of a pitch is periodic in the sense that two pitches are perceived similar in color, if they differ by an octave [13, 24, 10]. The pitch has two components, namely chroma and tone height (octave number) [30]. From the pitch representation, we can compute chroma representation by simply adding up the corresponding values that belong to the same chroma.

Let us consider the musical notes of equal-tempered scale. The 12 dimension chroma C is defined as a set of pitch classes $\mathsf{C} \triangleq \{\mathcal{C}, \mathcal{C}^\sharp, \mathcal{D} \ldots, \mathcal{B}\}$, each pitch class corresponding to a chroma. For example, chroma $\mathcal{C}$ is computed by adding up values corresponding to the musical pitches $\mathcal{C} \triangleq \{C_1, C_2, \ldots, C_8\}$ (MIDI pitches $p = 24, 36, \ldots, 108$).

In the Section 4.3, we next apply the concept of pitch and chroma features on a Carnatic music main piece audio recording for the analysis of the constituent parts.

4.3 Enhanced Chroma Representation

To compute the chroma features from the MIDI pitch representation of a Carnatic music main piece audio recording, we follow two steps. In the first step, we remove the drone from the MIDI pitch representation. Secondly, we consider the ideal singing octave range to compute the chroma features. The resulting chroma representation is referred to as enhanced chroma representation.

As discussed earlier that, a drone instrument is played through out a concert to provide a pitch reference for the performers to stay in tune. The *Tani-Avarthanam* part is not purely percussive, due to the presence of the melody from the drone instrument. By removing the effect of the drone from the main piece audio recording, we can musically distinguish the *Tani-Avarthnam* part with respect to the *Alapana* and the *Krithi* parts based on chroma representation.

Let us now discuss the various types and their characteristics of the drone instruments used in the Carnatic music. Drone instrument is a long plucked string instrument whose function is to continuously sound one or more notes providing the harmonic base for the performers. The drone became a definite component of the Carnatic music in the late 17^{th} century [2]. The most commonly used drone instrument in the Carnatic music are *Tanpura*, *Tambura* or *Sruti petti* (see Figure 2.1). A *drone* instruments can have 1-7 strings. A drone instrument with single string is known as the primary drone, which is always tuned to the note *Sa* often at $C^\sharp$. The note *Sa* is tuned to the male vocalist pitch or to the female singer (usually a fifth higher). For two string drone instrument, the first string is the primary drone and second string is known as the secondary drone. If the primary drone is tuned to *Sa* referring to $C^\sharp$, then the second string is tuned to seven notes higher *Pa* corresponding to $G^\sharp$ in the Western music scale. These tonic notes may vary according to the preference of the singer, as there is no absolute or fixed pitch-reference in the Indian classical music system. The most commonly used drone instrument has $3-4$ strings. The three string drone instrument is tuned to (*Sa* - *Pa* - *Sa*), where first two strings belong to a pitch octave and the last note *Sa* refers to the next immediate octave note , i. e., ($C_4^\sharp$ – $G_4^\sharp$ – $C_5^\sharp$) on Western music scale. Similarly, four string drone instrument is tuned to (*Sa* - *Sa* - *Pa* - *Sa*), in which the last three strings is same as that of the three string instrument, except its first note referring to an immediate lower octave note *Sa*, i. e., ($C_3^\sharp$ – $C_4^\sharp$ – $G_4^\sharp$ – $C_5^\sharp$).

To compute the chroma features, we consider the ideal singing octave range of the singers in the Carnatic music. An ideal Carnatic music voice has three octave range of singing. They are Mandhara sthaayi (lower octave), Madhya sthaayi (middle or main octave) and Tara sthaayi (higher octave) [2, 31]. The lower octave is usually the third octave, middle octave is the fourth octave and higher octave corresponds to the fifth octave of the MIDI pitch representation. In Carnatic music, the singer normally starts at a frequency higher than 240 Hz and refers to the starting frequency as the *Sa* note. In addition, a typical Carnatic music song is performed in two octaves. The two octaves refer to the second half of the lower octave, the full of the middle octave and the first half of the higher octave [31].

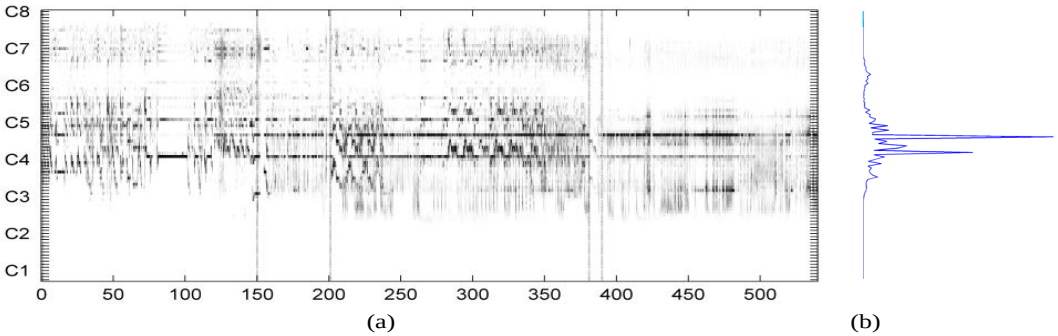

Figure 4.3: MIDI representation of the Carnatic music main piece audio recording. **(a)** MIDI pitch representation ($\mathcal{P}(n,p)$). **(b)** pitch energy across the frames ($D_e(p)$).

Let us consider the MIDI pitch representation of a Carnatic music main piece audio recording as shown in the Figure 4.3a. From the figure, we can make three observations. Firstly, there are three predominant pitches (three horizontal lines between the octaves four and five of the MIDI pitch representation) throughout the music piece. This is mainly because of the presence of three string drone instrument. The three horizontal lines corresponds to (*Sa* - *Pa* - *Sa*) notes, i. e., ($C_4^\sharp$ – $G_4^\sharp$ – $C_5^\sharp$) in the Western music scale. Secondly, we notice that the singing range of the singer is between three octaves, i.e., third to fifth MIDI octaves. In other words, most of the pitch energy is predominant between third to fifth MIDI octaves. Finally, we also observe that, pitches (*Sa* - *Pa*) are tuned to the ($C_4^\sharp$ – $G_4^\sharp$) corresponds to the main or the middle octave. The pitch energy is more predominant in the main octave due to the presence of the drone instrument as shown in the Figure 4.3b.

From the MIDI pitch representation of a Carnatic music main piece audio recording, we now remove the drone and later compute the chroma features based on the singing octave range of the Carnatic music as follows.

Let $\mathcal{P}(n,p)$ (see Figure 4.3a) be the pitch features (short time mean square power), where p is the MIDI note number $p \in [0:127]$ and time frame $n \in \mathbb{N}$. The pitch energy $D_e(p)$ (see Figure 4.3b) for all the time frames is given by,

$$D_e(p) := \sum_{\forall n} \mathcal{P}(n,p). \tag{4.5}$$

As defined earlier, Let $\mathbf{C} \triangleq \{C_1, C_2, \ldots, C_8\}$ be a set of octaves of the MIDI pitch representation,

then the pitch energy $\overline{E}(C_i)$ for MIDI octave C_i be given by,

$$\overline{E}(C_i) := \sum_{p \in C_i} D_e(p) \tag{4.6}$$

where, i is the octave number, $i \in [1:8]$ and p is the MIDI note number of the octave C_i. As mentioned earlier, the middle octave has the maximum chroma energy. Let the middle (main) octave with maximum energy $C_{\max}$ is given by,

$$C_{\max} := \text{argmax}_{i \in [1:8]} \, \overline{E}(C_i). \tag{4.7}$$

As mentioned earlier, the four string drone instrument is tuned to (*Sa - Sa - Pa - Sa*) notes i. e., ($C_4^\sharp$ – $G_4^\sharp$ – $C_5^\sharp$ – $C_5^\sharp$) in the Western music scale. The second and third notes are present in the middle octave which differ by seven semi-tones. The lower note *Sa* (first note) and the higher note *Sa* (fourth note) differs by an octave with respect to the second note of the drone instrument. It may so happen that, the *Pa* note is more predominant than *Sa* note of a main octave. To avoid such confusion in finding the *Sa* note of the drone instrument, we add the pitch energies of the notes differing by seven semitones of the middle octave. Let $\tilde{\mathcal{D}}_e(p)$ be the resulting pitch energies of the middle octave and is given by,

$$\tilde{\mathcal{D}}_e(p) := D_e(p) + D_e(p+7), \tag{4.8}$$

where, $p \in C_{\max}(\tilde{S})$ middle octave and the set $\tilde{S} = \{1, 2, 3, 4, 5\}$ corresponding to the first five notes of the octave $C_{\max}$. The predominant *Sa* note present in the main octave given by,

$$D_{\max} := \text{argmax}_{p \in C_{\max}} \, \tilde{\mathcal{D}}_e(p). \tag{4.9}$$

If $\hat{p} = \{p_{\max-12},\ p_{\max},\ p_{\max+7},\ p_{\max+12}\}$ is a set of MIDI note numbers tuned to a drone instrument. The resulting pitch energy $\tilde{\mathcal{P}}(n, p)$ without the drone is given by,

$$\tilde{\mathcal{P}}(n, p) := \mathcal{P}(n, p) \,.\, \kappa \,, \quad where \quad \begin{cases} \kappa = 0 & \text{for} \quad p \in \hat{p} \\ \kappa = 1 & \text{for} \quad p \notin \hat{p} \end{cases} \tag{4.10}$$

As mentioned earlier that, an ideal Carnatic music voice has three octave range of singing. Hence, we consider the relevant three octaves ($C_{\max-1},\ C_{\max},\ C_{\max+1}$) having dominant pitch energy from the MIDI pitch representation to compute the chroma features. The chroma feature is computed by simple addition of the corresponding values that belong to the same chroma. We compute the chroma features for four different combinations of MIDI octaves. The combination having maximum discriminability of the constituent parts is selected to design the novel features. The four combinations are, a) Lower and middle octave; b) Middle octave; c) Middle and upper octave; and d) Lower to upper octave.

The Figure 4.4 shows chroma features for all the four different combinations of the MIDI octaves. We select the middle and upper octave because majority of the time the singer usually sings in the middle and the upper octave and thus, this combination has a maximum discriminability of the constituent parts (see Figure 4.4d) as compared to the other octave combinations as

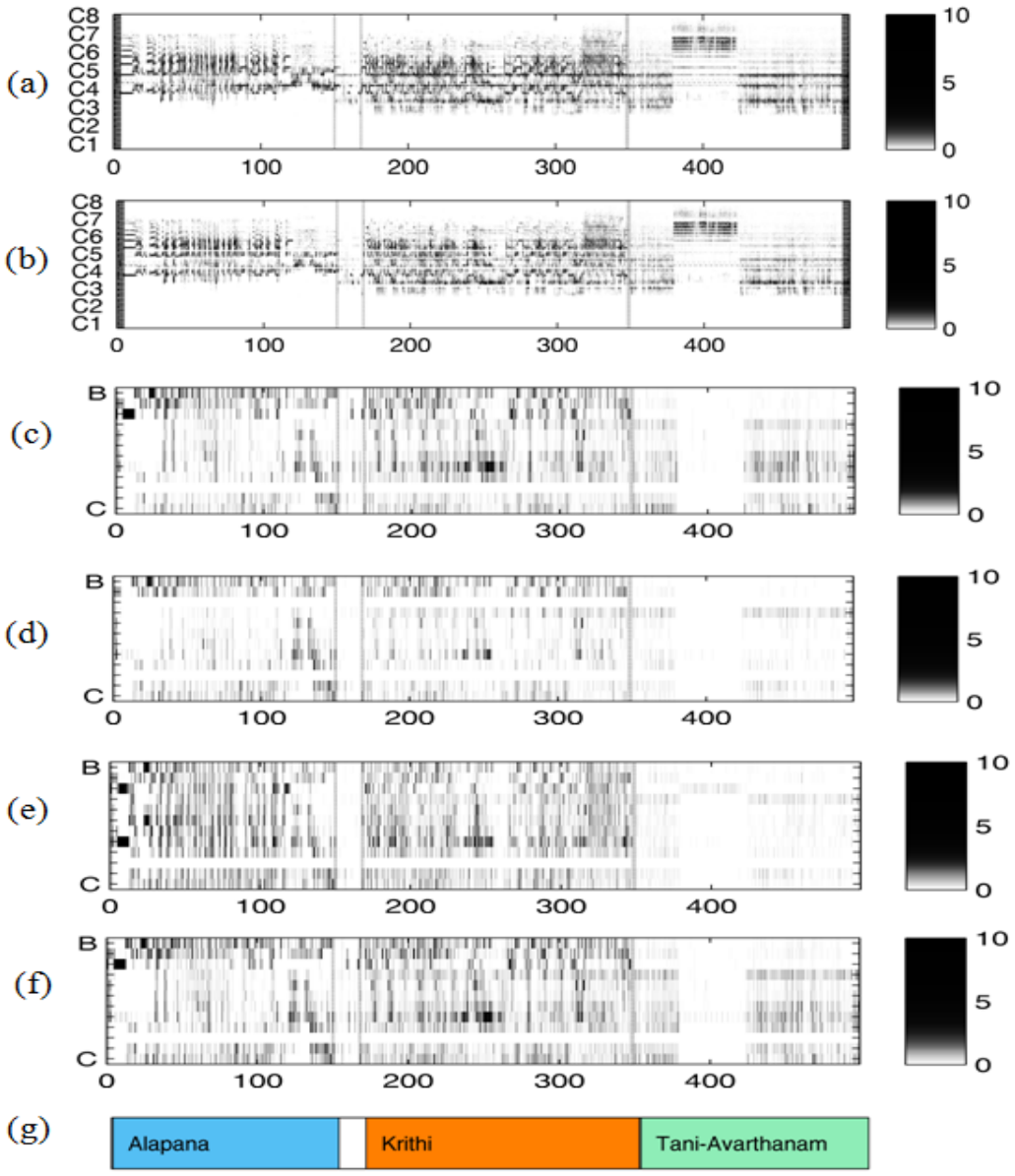

Figure 4.4: The MIDI pitch and chroma representation of the Carnatic music main piece audio recording. **(a)** MIDI pitch representation (drone present). **(b)** MIDI pitch representation (drone removed). **(c)** Chroma representation (lower to middle octave). **(d)** Chroma representation (middle octave). **(e)** Chroma representation (middle and upper octave). **(f)** Chroma representation (lower to upper octave). **(g)** Manual segmentation of the recording. The white areas indicate transition regions (often pauses, sometimes used for tuning the instruments) between the respective parts.

mentioned above. We neglect the lower octave because the singer rarely sings in the lower octave and the bass noise present in the lower octave pops up in the chroma feature (see Figure 4.4c). Hence, we only consider the middle octave ($C_{\max}$) and the upper octave ($C_{\max+1}$) for computing

the enhanced chroma features.

Let the pitch class $\Theta = \{C_{\max}, C_{\max+1}\}$ where, $C_{\max}$ and $C_{\max+1}$ corresponds to the middle octave and upper octave. The resulting chroma feature is represented by a twelve dimension vector $\mathbf{v}_n = [v_n(1), v_n(2), v_n(3), \ldots, v_n(12)]^T$, where n is the frame index. For example, The twelve dimension vector $\mathbf{v}_n$ corresponds to C, $v_n(1)$ corresponds to chroma $\mathcal{C}$, $v_n(2)$ to $\mathcal{C}^\sharp$, $v_n(3)$ to $\mathcal{D}$ and so on. From the resulting enhanced chroma features $\mathbf{v}_n$, we now design the novel features as described in the next Section.

4.4 Salience Features

This Section contains the main technical contribution of the thesis. In this Section, we design the chroma salience features based on enhanced chroma representation (see Section 4.3). The enhanced chroma representation is derived from the pitch feature representation, which in turn is computed from the existing chroma toolbox [25] as described in the Section 4.1 and in the Section 4.2.

Recall from 1.1 that, a drone instrument is played throughout a concert to provide a reference pitch for the performers. Neglecting the drone instrument, the *Tani-Avarthanam* part of a Carnatic music main piece tends to have no strong notion of chroma, as opposed to the *Alapana* and the *Krithi* parts. This observation is reflected well by the enhanced chroma representation of a Carnatic music main piece audio recording as shown in the Figure 4.4e.

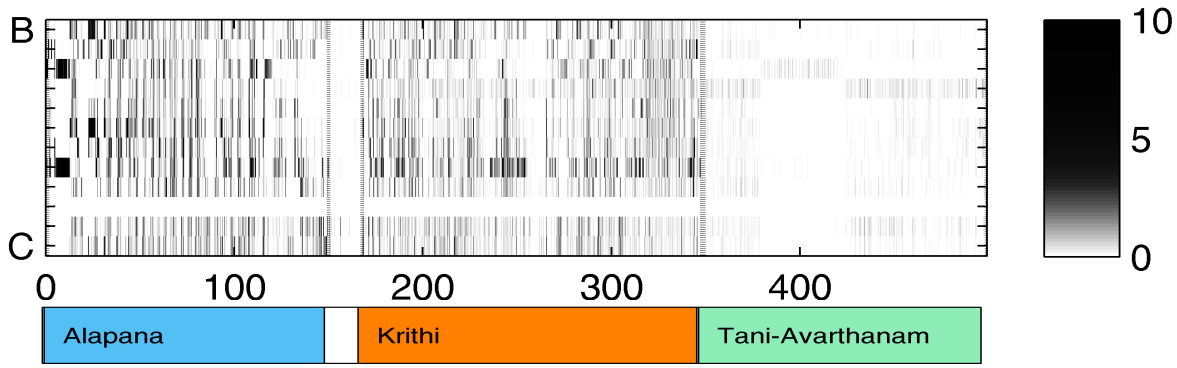

Figure 4.5: The enhanced chroma representation of a Carnatic music main piece audio recording (considering only middle and upper octaves of MIDI representation).

In the *Tani-Avarthanam* part, the enhanced chroma representation looks rather diffuse having no larger coefficients that would indicate the presence of a specific chroma. In contrast, most of the chroma vectors that belong to the *Alapana* and the *Krithi* parts possess a dominant entry. Furthermore, one can notice that the chroma class of the dominating entry may vary over time. Which reflects the fact that, the chroma is changing especially in the *Alapana* and the *Krithi* parts.

It is our objective to capture the property of having a dominating chroma regardless of the specific value of the chroma or a possible change in chroma. In the following, we refer to this property as *chroma salience*. We now describe several kinds of salience features derived from the

enhanced chroma representation.

4.4.1 Max Chroma Feature

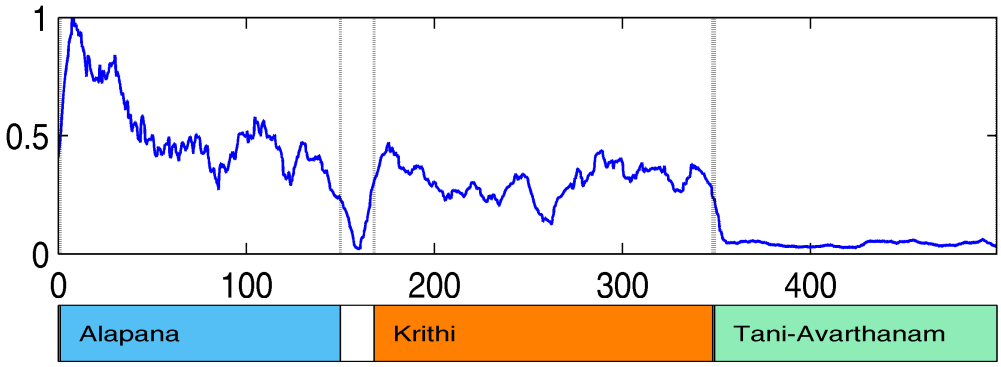

Figure 4.6: Max chroma feature $f_{\lambda}^{\mathcal{M}_C}$ with $\lambda = 15$ sec obtained from the enhanced chroma representation (see Figure 4.5) of a Carnatic music main piece audio recording (considering only upper and lower octaves of MIDI representation).

The Figure 4.5 shows the enhanced chroma representation of a Carnatic music main piece audio recording. From the figure, we can notice that, the chroma is more predominant in the *Alapana* and *Krithi* parts as compared to the *Tani-Avarthanam*. Motivated by this observation, a first idea is to apply the concept of the maximum chroma for a given. The maximum chroma $\mathcal{M}_{\mathcal{C}}(n)$ for a frame index $n \in \mathbb{N}$ is given by,

$$\mathcal{M}_{\mathcal{C}}(n) := max_{\forall n}\mathbf{v}_n. \tag{4.11}$$

Furthermore, as the chroma energy is very low in the *Tani-Avarthanam* part as compared to the dominant chroma entries in the *Alapana* and *Krithi* parts. We normalize the feature $\mathcal{M}_{\mathcal{C}}(n) \in [0:1]$ with its most dominant chroma entry. To this end, smoothing this sequence by applying an averaging filter of some length $\lambda \in \mathbb{N}$ yields our first feature that we refer to as $f_{\lambda}^{\mathcal{M}_C}$. As demonstrated by Figure 4.6, this feature has the desired property of being close to zero in the *Tani-Avarthanam* part and much larger in the other parts, see Section 5.2 for a more detailed investigation.

4.4.2 Sum Chroma Feature

As an alternative to the maximum chroma feature, one may also compute entire chroma energy per frame resulting in the sum chroma feature. If the twelve dimensional chroma vector is given by $\mathbf{v}_n = [v_n(1), v_n(2), v_n(3), \ldots, v_n(12)]^T$, then the resulting sum chroma feature is computed by adding all the twelve chromas. The chroma feature $\mathcal{S}_{\mathcal{C}}(n)$ for a frame index $n \in \mathbb{N}$ is given by,

$$\mathcal{S}_{\mathcal{C}}(n) := \sum_{i=1}^{12} v_n(i), \tag{4.12}$$

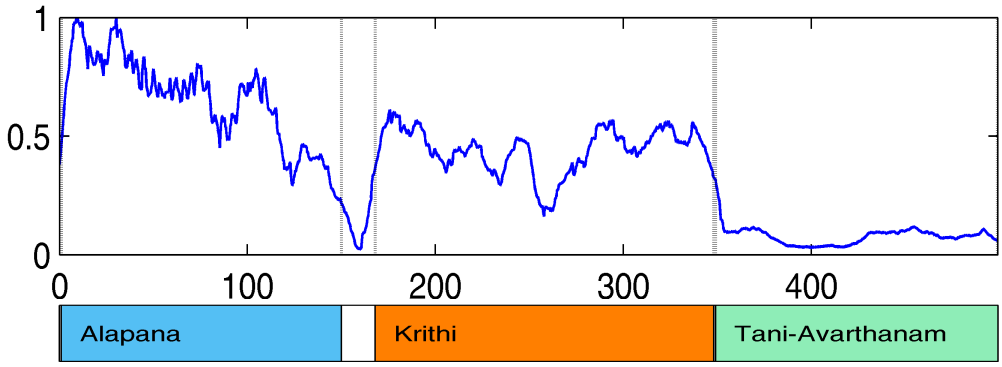

Figure 4.7: Sum chroma feature $f_{\lambda}^{\mathcal{S}_C}$ with $\lambda = 15$ sec obtained from the enhanced chroma representation (see Figure 4.5) of a Carnatic music main piece audio recording (considering only middle and upper octaves of MIDI representation).

where, $v_n(i)$ is the i^{th} chroma for the n^{th} frame index.

Furthermore, the feature $\mathcal{S}_{\mathcal{C}}(n)$ behaves similar as the feature $\mathcal{M}_{\mathcal{C}}(n)$. Hence, we normalize the feature $\mathcal{S}_{\mathcal{C}}(n) \in [0:1]$ with its most dominant chroma entry. To this end, smoothing this sequence by applying an averaging filter of some length $\lambda \in \mathbb{N}$ yields our second feature that we refer to as $f_{\lambda}^{\mathcal{S}_C}$. As illustrated by Figure 4.7, this feature behaves similarly to $f_{\lambda}^{\mathcal{M}_C}$.

4.4.3 Relative Chroma Strength Feature

The Figure 4.5 shows the enhanced chroma representation of a Carnatic music main piece audio recording. From the figure, we can make 2 observations. Firstly, there is a strong notion of melody in the *Krithi* and the *Alapana* parts and has predominant chroma energy. Secondly, a vague notion of melody in *Tani- Avarthanam* part and has very low chroma energy.

Hence, we can infer that, the relative strength of chroma is predominant in the *Alapana* and the *Krithi* parts and has low relative chroma strength in the *Tani-Avarthanam* part of a main piece.

From the above observation, we now introduce a conceptually different salience feature, i.e., relative chroma strength. Let the most predominant chroma $\tilde{\mathcal{M}}_{\text{E}}$ for the entire piece be given by,

$$\tilde{\mathcal{M}}_{\text{E}} = maximum(\mathcal{M}_{\mathcal{C}}(n)), \tag{4.13}$$

where, $\mathcal{M}_{\mathcal{C}}(n)$ is the maximum chroma for the n^{th} time frame. Let vector $\hbar_n$ represent all the chroma entries greater than $(\tilde{\mathcal{M}}_{\text{E}} \cdot \tau)$ and is given by,

$$\hbar_n(i) = \begin{cases} 0 & if \quad v_n(i) \geq (\tilde{\mathcal{M}}_{\text{E}} \cdot \tau) \\ 1 & if \quad v_n(i) < (\tilde{\mathcal{M}}_{\text{E}} \cdot \tau) \end{cases}, \tag{4.14}$$

where, τ is a threshold and n is the frame index. We now define relative chroma strength $\mathcal{R}_{\mathcal{C}}(n)$

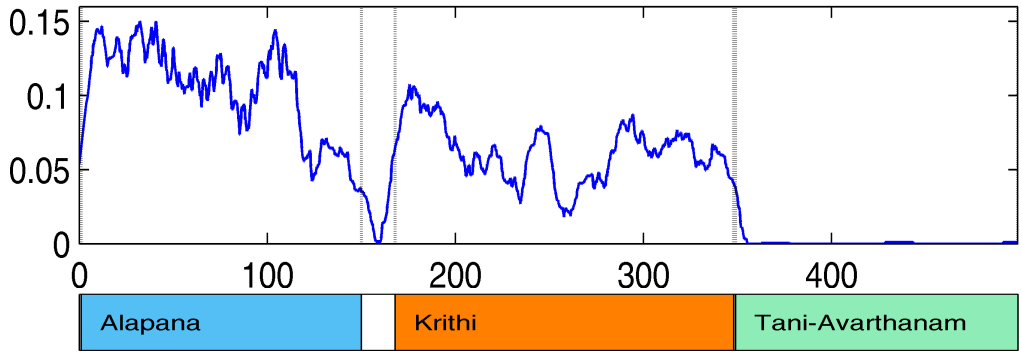

Figure 4.8: Relative chroma strength feature $f_\lambda^{\mathcal{R}_\mathcal{C}}$ with $\lambda = 15$ sec and $\tau = 0.05$ obtained from the enhanced chroma representation (see Figure 4.5) of a Carnatic music main piece audio recording (considering only middle and upper octaves of MIDI representation).

for n time frames and is given by,

$$\mathcal{R}_\mathcal{C}(n) := \frac{1}{12}\sum_{i=1}^{12} \hbar_n(i). \tag{4.15}$$

Furthermore, for any time frame n, the maximum value of $\hbar_n(i)$ can have all the chroma entries, i.e., $\hbar_n(i) = 12$. Hence, we normalized the feature $\mathcal{R}_\mathcal{C}(n)$ by the total number of notes of an octave. To this end, smoothing this sequence by applying an averaging filter of some length $\lambda \in \mathbb{N}$ yields our third feature that we refer to as $f_\lambda^{\mathcal{R}_\mathcal{C}}$. As demonstrated by Figure 4.8, this feature has the desired property of relative chroma strength being nearly close to zero in the *Tani-Avarthanam* part and much larger in the other parts, see Section 5.2 for a more detailed investigation.

In this Chapter, we designed the chroma salience features based on the melody property to musically distinguish the *Tani-Avarthanam* part from the *Alapana* and *Krithi* parts. Similarly, in the Chapter 3, we derived the tempo salience features based on the tempo property to musically distinguish *Alapana* part from the *krithi* and *Tani-Avarthanam* parts. Hence, we can use the tempo and chroma salience features to aim towards the segmentation task.

We now investigate in the next Chapter, how well the three different musical parts are characterized based on quantitative and qualitative analysis of the chroma and tempo salience features.

Chapter 5

Evaluation

In this Chapter, we provide some insights into the tempo and chroma salience features by a quantitative and qualitative analysis of their properties. We shall describe examples of audio where the features behave anomalously. In our experiments, we used music recordings of good audio quality from the Sangeethapriya website[1]. In total, our dataset consists of 15 main pieces from various Carnatic concerts with an overall duration of more than 15 hours. We manually annotated the music recordings (for more details, refer Appendix B.2) and determined the segment boundaries for the *Alapana*, *Krithi* and *Tani-Avarthanam* parts. This chapter is organized as follows. Firstly, evaluation of tempo salience features is discussed in the Section 5.1. Secondly, we discuss the evaluation of chroma salience features in the Section 5.2.

5.1 Tempo Salience Features

Based on the annotations, we computed some statistics to investigate how well the three different musical parts are characterized by our tempo salience features. To this end, we first computed for each audio file cyclic tempogram representations[2]. Based on these representations, we computed the salience features as introduced in Section 3.3.

Then, for each of the features, we computed the average feature value and its standard deviation for each of the three *Alapana*, *Krithi* and *Tani-Avarthanam* parts separately. These values, in turn, were averaged over the 15 different pieces. As a result, we obtained for each salience feature and each part a mean $\overline{\mu}$ and a standard deviation $\overline{\sigma}$. These results are shown in Table 5.1. Note that, rather than the absolute values, the relative relation between the values across the three different parts are of interest.

First, let us have a look at the statistics for the features $f_\lambda^{\mathcal{H}}$ and $f_\lambda^{\mathcal{M}}$. As can be seen from Table 5.1, the mean statistics of $f_\lambda^{\mathcal{H}}$ assume a value of 0.0032 for the *Alapana* part, which is roughly ten times smaller than the value 0.0285 for the *Krithi* part and the value 0.0246 for the *Tani-Avarthanam* part. Also, the standard deviation $\overline{\sigma}$ for $f_\lambda^{\mathcal{H}}$ shows a similar trend: it assumes

[1] `http://www.sangeethapriya.org`

[2] For the computation we used the MATLAB implementations supplied by the *Tempogram Toolbox*, see [16] and `www.mpi-inf.mpg.de/resources/MIR/tempogramtoolbox`.

Feature	$\overline{\mu}$			$\overline{\sigma}$		
	A	*K*	*T*	*A*	*K*	*T*
$f_\lambda^{\mathcal{H}}$	0.0032	0.0285	0.0246	0.0016	0.0181	0.0154
$f_\lambda^{\mathcal{M}}$	0.0169	0.0685	0.0585	0.0054	0.0228	0.0237
$f_{0,\lambda}^{\mathcal{I}}$	0.1045	0.0059	0.0115	0.0489	0.0154	0.0181
$f_{1,\lambda}^{\mathcal{I}}$	0.0705	0.0059	0.0115	0.0436	0.0154	0.0181

Table 5.1: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of various salience features shown for the three different parts *Alapana* (*A*), *Krithi* (*K*), and *Tani-Avarthanam* (*T*). For the full table of all the 15 pieces of dataset, refer Appendix C.

the value 0.0016 for the *Alapana* part, which is much lower than the value 0.0181 for the *Krithi* and the value 0.0154 for the *Tani-Avarthanam* part. Recall from Section 3.3 that the feature $f_\lambda^{\mathcal{H}}$ measures the column-wise entropy of a normalized tempogram. Therefore, a low value of $f_\lambda^{\mathcal{H}}$ indicates a flat distribution (no clear notion of a tempo), whereas a high value indicates a dirac-like distribution (the presence of a dominating tempo value). The average values of $f_\lambda^{\mathcal{H}}$ in the three parts exactly reflect the musical property that there is no sense of tempo in the *Alapana* part, whereas there is a clearly perceivable tempo (either constant or changing) in the other two parts. For the feature $f_\lambda^{\mathcal{M}}$, one can observe similar trends as for $f_\lambda^{\mathcal{H}}$. Both features are suitable for discriminating the *Alapana* part from the other two parts.

Next, we examine the behavior of the features $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$. As shown by Table 5.1, these features also assume quite different values in the *Alapana* part compared to the other two parts. However, this time the features assume comparatively high values in the *Alapana* part. For example, the mean of $f_{0,\lambda}^{\mathcal{I}}$ is 0.1045 for the *Alapana* part, which is much higher than the mean value 0.0059 for *Krithi* and 0.0115 for *Tani-Avarthanam* part. The relative differences between the parts behave almost similar of the feature $f_{1,\lambda}^{\mathcal{I}}$. Recall from Section 3.3 that the features $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$ measure some kind of density for tempo changes by considering differences of maximizing bin indices between subsequent frames. Since the tempogram in the *Alapana* part is rather diffuse, the maximizing entries are unstable leading to more or less random jumps when considering subsequent frames. This results in large values of $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$. In contrast, there usually exists a dominating tempo in the *Krithi* and *Tani-Avarthanam* part for most of the frames, which results in a more or less constant sequence when considering maximizing bin indices in the columns of the tempogram. Small tempo fluctuations may lead to bin differences of plus or minus one, which are filtered out when considering the feature $f_{1,\lambda}^{\mathcal{I}}$. As a result, only occasional index jumps due to abrupt and significant tempo changes are captured by this feature. Since such tempo changes are rare in the *Krithi* and *Tani-Avarthanam* part, the overall mean values are small compared to the *Alapana* part. Interestingly, the mean and standard deviations of Table 5.1 also indicate that abrupt tempo changes seem to occur more often in the final *Tani-Avarthanam* part compared to the *Krithi* part. This observation is also reflected by our listening inspections and the representative examples as shown in Figure 5.1.

As a supplement to our quantitative evaluation, we further illustrate the behavior of our salience features by showing representative examples computed from three different recordings[3] in

[3]For the sake of a better visual understanding, the figure shows the various feature representations only for representative subsections of the four parts (also including the transition regions between subsequent parts) instead

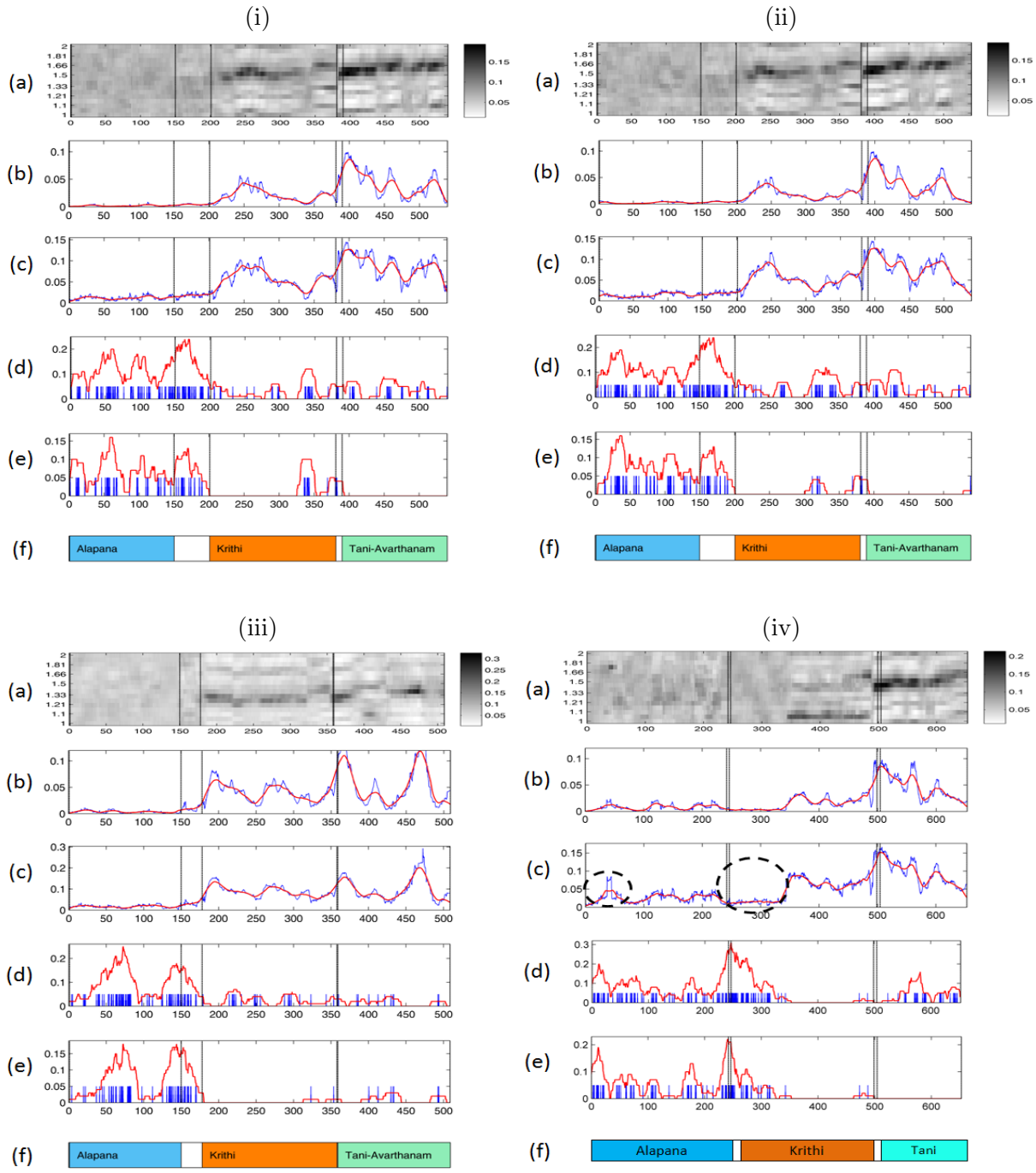

Figure 5.1: **(a)** Discrete version of a normalized cyclic tempogram representation based on an autocorrelation tempogram using the parameters $L = 16$ *sec*, $F_s = 5$ Hz, and $M = 15$. **(b)** Feature $f_\lambda^{\mathcal{H}}$ with $\lambda = 1$ (blue) and $\lambda = 100$ corresponding to 20 *sec* (red).**(c)** Feature $f_\lambda^{\mathcal{M}}$ with $\lambda = 1$ (blue) and $\lambda = 100$ (red). **(d)** Feature $f_{\tau,\lambda}^{\mathcal{I}}$ with $\tau = 0$ and $\lambda = 1$ (blue, size of binary values reduced for visibility reasons) and $\lambda = 100$ (red). **(e)** $f_{\tau,\lambda}^{\mathcal{I}}$ with $\tau = 1$. **(f)** Manual segmentation of the recording. The white areas indicate transition regions (often pauses, sometimes used for tuning the instruments) between the respective parts.

Figure 5.1. We start with the inspection of the *Alapana* parts of the examples. In Figure 5.1(i), the behavior of the features exactly corresponds to our previous discussion. The features $f_\lambda^{\mathcal{H}}$ and $f_\lambda^{\mathcal{M}}$ consistently assume small values in the curve, while the features $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$ present large values. A similar behavior of the features in the *Alapana* part can be observed in the other three examples Figure 5.1.

However, in Figure 5.1(iv), one can notice an outlier in the feature representation of $f_\lambda^{\mathcal{M}}$ (indicated by a dotted circle). By listening inspection, we noticed that in this section of the *Alapana* part, the audience started to rhythmically clap along the music, thus introducing some clear notion of tempo. Such clapping is not unusual in Carnatic music concerts.

Turning to the *Krithi* part, the features $f_\lambda^{\mathcal{H}}$ and $f_\lambda^{\mathcal{M}}$ assume consistently large values (compared to the ones in the *Alapana* part) for the first three examples of Figure 5.1. Also, as expected, the features $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$ assume small values thus reinforcing our observation that there is usually a clear notion of tempo in the *Krithi* part with some occasional tempo changes. In the third example shown in Figure 5.1(iv), one can observe a deviation of this tendency in the first section of the *Krithi* part (indicated by a dashed circle). In this part, it turned out that the lead artist has started the *Krithi* part without being accompanied by percussion. Even though the lyrical composition has already started, the artist still continues in an *Alapana*-like style. After a while, the percussion finally sets in, which leads to the expected feature values in the *Krithi* part. Even though such deviations in the *Krithi* part do not happen often in concerts, they give an idea of the wide range of challenges that Carnatic music poses to any segmentation algorithm.

A similar behavior of the features can be seen in the *Tani-Avarthanam* part of the four examples. As an interesting tendency, one can observe that the features $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$ vary more in the *Tani-Avarthanam* part compared to the *Krithi* part, as also verified by the standard deviations $\overline{\sigma}$ shown in Table 5.1. This nicely reflects the fact that in this final part the percussionists often change the tempo abruptly to present a new solo.

Let us now discuss a quantitative statistics and a qualitative analysis of the chroma salience features in the Section 5.2.

5.2 Chroma Salience Features

Based on the annotated dataset, we first computed for each audio file chroma representations[4]. Based on these representations, we computed the salience features as introduced in Section 4.4. Then, for each of the features, we computed the average feature value and its standard deviation for each of the three *Alapana*, *Krithi* and *Tani-Avarthanam* parts separately. These values, in turn, were averaged over the 15 different pieces. As a result, we obtained for each salience feature and each part a mean $\overline{\mu}$ and a standard deviation $\overline{\sigma}$. These results are shown in Table 5.2. As before, the relative values across the three parts are of interest rather than their absolute values.

First, let us have a look at the statistics for the features $f_\lambda^{\mathcal{M}_\mathcal{C}}$ and $f_\lambda^{\mathcal{S}_\mathcal{C}}$. It can be seen from Table 5.2, the mean statistics of $f_\lambda^{\mathcal{M}_\mathcal{C}}$ assume a value of 0.0393 for the *Tani-Avarthanam* part,

of showing the representations for the entire pieces.

[4]For the computation we used the MATLAB implementations supplied by the *Chroma Toolbox*, see [25] and `www.mpi-inf.mpg.de/resources/MIR/chromatoolbox/`.

Feature	$\overline{\mu}$			$\overline{\sigma}$		
	A	*K*	*T*	*A*	*K*	*T*
$f_\lambda^{\mathcal{M}c}$	0.3976	0.3376	0.0393	0.1954	0.1672	0.0269
$f_\lambda^{\mathcal{S}c}$	0.4209	0.3912	0.0691	0.2026	0.1889	0.0445
$f_\lambda^{\mathcal{R}c}$	0.0492	0.0466	0.0013	0.0288	0.0308	0.0028

Table 5.2: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of various salience features shown for the three different parts *Alapana* (*A*), *Krithi* (*K*), and *Tani-Avarthanam* (*T*). For the full table for all the 15 pieces of dataset, refer Appendix C.

which is roughly ten times smaller than the value 0.3376 for the *Krithi* part and the value 0.3976 for the *Alapana* part. Also, the standard deviation $\overline{\sigma}$ for $f_\lambda^{\mathcal{M}c}$ shows a similar trend: it assumes the value 0.0269 for the *Tani-Avarthanam* part, which is much lower than the value 0.1672 for the *Krithi* and the value 0.1954 for the *Alapana* part. Recall from Section 4.4.1 that the feature $f_\lambda^{\mathcal{M}c}$ measures the maximum chroma of a chroma feature representation. Therefore, a low value of $f_\lambda^{\mathcal{M}c}$ implies no clear notion of a chroma, whereas a high value indicates the presence of a dominating chroma value. The average values of $f_\lambda^{\mathcal{M}c}$ in the three parts exactly reflect the musical property that there is no obvious melody in the *Tani-Avarthanam* part, whereas there is a clearly perceivable melody in the other two parts. For the feature $f_\lambda^{\mathcal{S}c}$ (refer, Section 4.4.2), one can observe similar trends as for $f_\lambda^{\mathcal{M}c}$. Both features are suitable for differentiating the *Tani-Avarthanam* from the other two parts.

Next, we examine the behavior of the feature $f_\lambda^{\mathcal{R}c}$. As shown in the Table 5.2, this feature also assume quite different values in the *Tani-Avarthanam* part compared to the other two parts. However, this time the feature assumes comparatively very low values in the *Tani-Avarthanam* part. For example, the mean of $f_\lambda^{\mathcal{R}c}$ is 0.0013 for the *Tani-Avarthanam* part, which is very less as compared to the mean value 0.0466 for *Krithi* and 0.0492 for *Alapana* part.

The relative differences between the parts become even larger for the average values of the feature $f_\lambda^{\mathcal{R}c}$. Recall from Section 4.4.3 that the feature $f_\lambda^{\mathcal{R}c}$ measures some kind of relative chroma changes. Since, the melody in the *Tani-Avarthanam* part is rather diffused as compared to that of the other two parts. This results in large values of $f_\lambda^{\mathcal{R}c}$ in the remaining two parts of the main piece audio recordings. In contrast, there usually exists a dominating chroma in the *Krithi* and *Alapana* parts for most of the frames, which results in a more or less constant sequences of the chroma representation. Since such chroma changes are rare in the *Tani-Avarthanam* part (because of tuned percussive instruments), the overall mean values are small as compared to that of the *Krithi* and the *Alapana* parts. Interestingly, the mean and standard deviations of Table 5.2 also indicate that abrupt chroma changes seem to occur more often in the first two parts compared to the *Tani-avarthanam* part. This observation is also reflected by our listening inspections and the representative examples as shown in Figure 5.2.

As a supplement to our quantitative evaluation, we further illustrate the behavior of our salience features by showing representative examples computed from four different recordings[5] in Figure 5.2. We start with the inspection of the *Tani-Avarthanam* parts of the examples. In Figure 5.2(i),

[5]For the sake of a better visual understanding, the figure shows the various feature representations only for representative subsections of the three parts (also including the transition regions between subsequent parts) instead of showing the representations for the entire pieces.

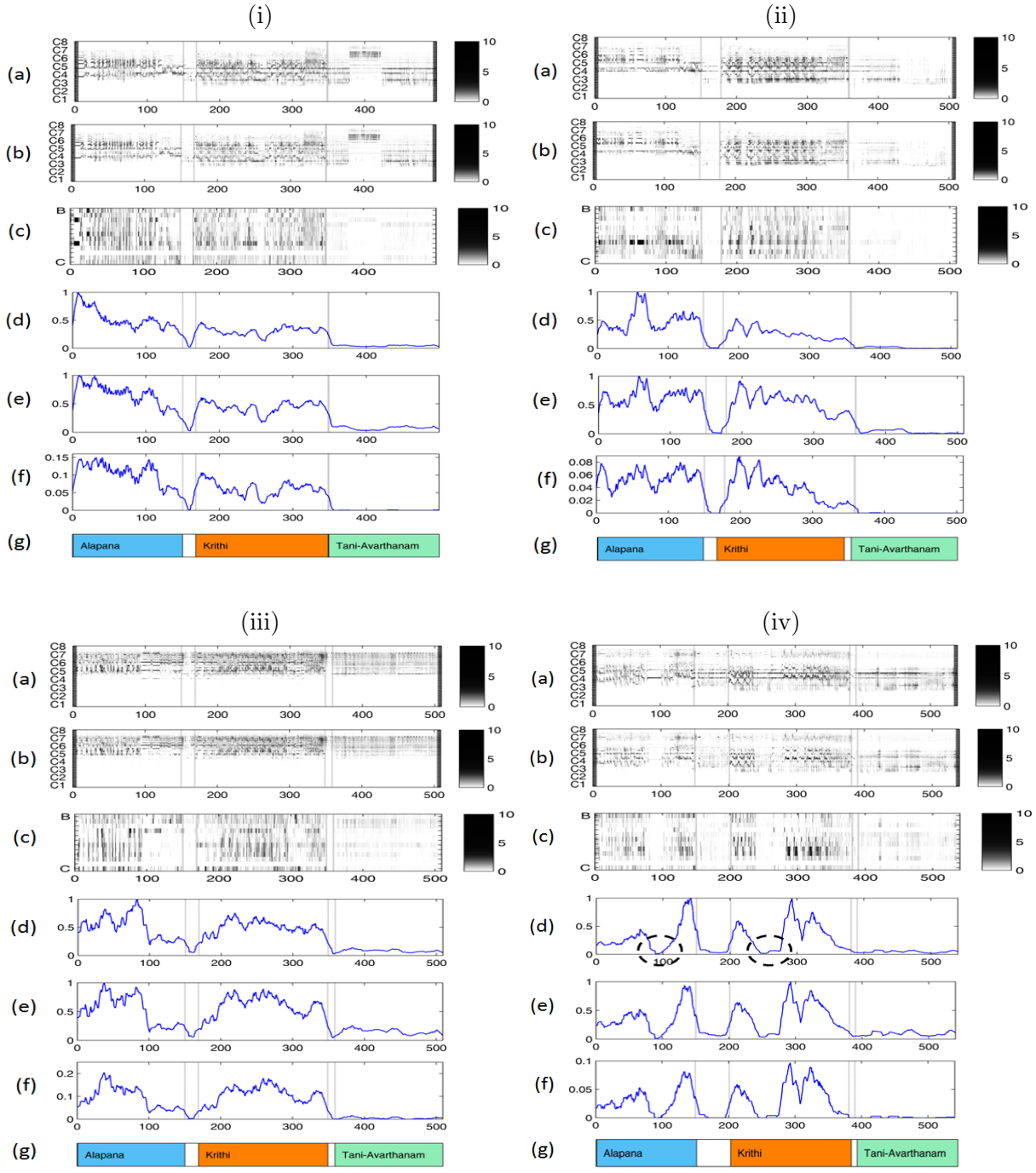

Figure 5.2: Representation of a Carnatic music recordings and the resulting feature representations with parameter settings, short time mean square power for each sub-band is computed by using a fixed window length of 200 milliseconds with overlap of 50% (leads to feature rate = 10 features per second) and $F_s = 22050\ Hz$. **(a)** midi pitch representation (drone present) **(b)** midi pitch representation (drone removed). **(c)** chroma representation (considering middle and upper midi octave). **(d)** Maximum chroma feature $f_\lambda^{\mathcal{M}_C}$ (with $\lambda = 15$ sec). **(e)** Sum chroma feature $f_\lambda^{\mathcal{S}_C}$ (with $\lambda = 15$ sec). **(f)** Relative chroma strength feature $f_\lambda^{\mathcal{R}_C}$ (with $\lambda = 15$ sec). **(g)** Manual segmentation of the recording. The white areas indicate transition regions (often pauses, sometimes used for tuning the instruments) between the respective parts.

the behavior of the features exactly corresponds to our previous discussion. The features $f_\lambda^{\mathcal{M}_\mathcal{C}}$ and $f_\lambda^{\mathcal{S}_\mathcal{C}}$ consistently assume small values in the curve, while the feature $f_\lambda^{\mathcal{R}_\mathcal{C}}$ presents the lowest values as compared to the other two parts.

A similar behavior of the features in the *Tani-Avarthanam* part can be observed in the other three examples of Figure 5.2. However, in Figure 5.2(iv), one can notice an outlier in the feature representation of $f_\lambda^{\mathcal{S}_\mathcal{C}}$ (indicated by a dotted circle). By listening inspection, we noticed that in this section of the *Alapana* and *Krithi* parts, the singer stopped singing for sometime resulting in low chroma energy as compared to its remaining parts, thus losing a notion of melody.

Turning to the *Tani-Avarthanam* part, the features $f_\lambda^{\mathcal{M}_\mathcal{C}}$ and $f_\lambda^{\mathcal{S}_\mathcal{C}}$ assume consistently small values (compared to the ones in the *Alapana* and *Krithi* parts) for all the examples of Figure 5.2. Also, as expected, the feature $f_\lambda^{\mathcal{R}_\mathcal{C}}$ assume lowest values thus reinforcing our observation that there is usually a vague notion of chroma in the *Tani-Avarthnam* part with some occasional chroma changes. The occasional changes are due to the harmonically tuned percussive instruments performed in the *Tani-Avarthanam* part. A similar behavior of the features can be seen in the *Tani-Avarthanam* part of the four examples. As an interesting tendency, one can observe that the features $f_\lambda^{\mathcal{R}_\mathcal{C}}$ has a very low relative chroma strength in the *Tani-Avarthanam* part compared to the *Krithi* and the *Alapana* parts, as also verified by the standard deviations $\overline{\sigma}$ shown in Table 5.2. This nicely reflects the fact that in this final part the there is a vague notion of melody.

5.3 Summary

Based on the evaluation of tempo and chroma salience features by a quantitative and qualitative analysis, one can now investigate how well the three musical parts are characterized as follows.

1. The tempo salience features $f_\lambda^{\mathcal{H}}$ and $f_\lambda^{\mathcal{M}}$ has a low μ and σ statistics for the *Alapana* part as compared to the *Krithi* and *Tani-Avarthanam* parts. Similarly, the tempo salience features $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$ has a high μ and σ statistics for the *Alapana* part as compared to the *Krithi* and *Tani-Avarthanam* parts. Hence, we can infer that, tempo salience feature musically distinguish *Alapana* part with respect to the *Krithi* and the *Tani-Avarthanam* parts of the main piece audio recording (see Figure 5.3b).

2. The chroma salience features $f_\lambda^{\mathcal{M}_\mathcal{C}}$, $f_\lambda^{\mathcal{S}_\mathcal{C}}$ and $f_\lambda^{\mathcal{R}_\mathcal{C}}$ has a high μ and σ statistics for the *Alapana* and *Krithi* parts as compared to the *Tani-Avarthanam* part. Hence, we can infer that, chroma salience feature musically distinguish *Alapana* and *Krithi* parts with respect to the *Tani-Avarthanam* part of the main piece audio recordings (see Figure 5.3c).

3. The tempo and chroma salience features can musically distinguish all the three parts of a Carnatic music main piece audio recordings into *Alapana*, *Krithi* and *Tani-Avarthanam* (see Figure 5.3d).

Furthermore, the designed tempo and chroma features aim towards the automatic segmentation of the main piece audio recording into *Alapana*, *Krithi* and *Tani-Avarthanam*. For example, we considered an audio recording, which depicts the automatic segmentation of all the three parts using $f_\lambda^{\mathcal{M}}$ and $f_\lambda^{\mathcal{R}_\mathcal{C}}$ salience features as shown in the Figure 5.4. A simple threshold technique is applied for both the features $f_\lambda^{\mathcal{M}}$ and $f_\lambda^{\mathcal{R}_\mathcal{C}}$. The threshold is obtained by considering the average

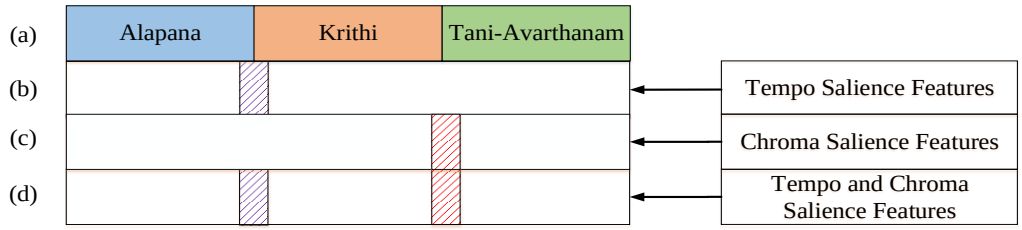

Figure 5.3: Typical structure of a Carnatic main piece. **(a)** Main piece constituting of three contrasting parts: *Alapana*, *Krithi* and *Tani-Avarthanam*. **(b)** Tempo salience features musically distinguishes *Alapana* with respect to *Krithi* and *Tani-Avarthanam*. **(c)** Chroma salience features musically distinguishes *Alapana* and *Krithi* with respect to *Tani-Avarthanam*. **(d)** Chroma and Tempo salience features together musically distinguishes all the three parts into *Alapana*, *Krithi* and *Tani-Avarthanam*.

μ and σ of the lowest chroma energy of the contrasting part of the 15 main piece audio recordings (i.e., *Alapana* part for tempo feature $f_\lambda^{\mathcal{M}}$ and *Tani-Avarthanam* for chroma feature $f_\lambda^{\mathcal{R}_C}$). From Table 5.1, the statistics of the *Alapana* for the tempo feature $f_\lambda^{\mathcal{M}}$ is $\mu = 0.0169$ and $\sigma = 0.0054$, then the resulting threshold is 0.0223 (i.e., $\mu + \sigma$) as shown in Figure 5.4a. Similarly, from Table 5.2, the statistics of the *Tani-Avarthanam* part for the tempo feature $f_\lambda^{\mathcal{M}}$ is $\mu = 0.0013$ and $\sigma = 0.0028$, then the resulting threshold is 0.0041 as shown in Figure 5.4c. By applying the tempo and chroma thresholds, we can automatically segment the three contrasting parts as shown in Figure 5.4b,d.

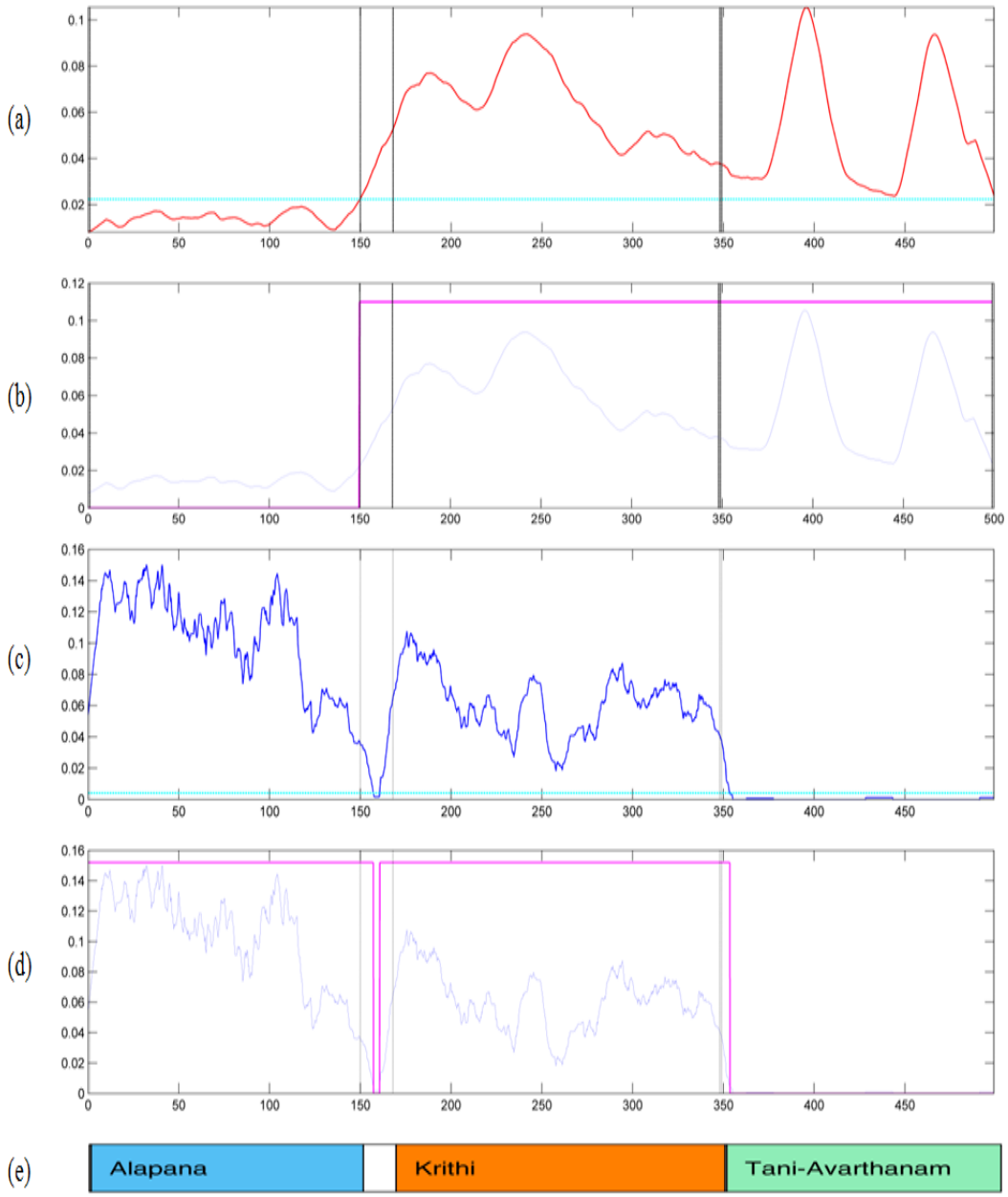

Figure 5.4: $Wavfile$: $Raga\_02\_excerpt\_s\_152.wav$. The chroma and tempo salience features together musically distinguish and segment all the three parts into *Alapana*, *Krithi* and *Tani-Avarthanam*. **(a)** The tempo salience feature $f_\lambda^{\mathcal{M}}$ with $\lambda = 100$ (red) and $tempo\_threshold = 0.0223$ (cyan). **(b)** After segmentation (by using a simple global $tempo\_threshold = 0.0223$), the tempo salience feature $f_\lambda^{\mathcal{M}}$ musically segment *Alapana* with respect to *Krithi* and *Tani-Avarthanam*. **(c)** The chroma salience feature $f_\lambda^{\mathcal{R}c}$ with $\lambda = 15$ (blue), $\tau = 0.05$ and and $chroma\_threshold = 0.0223$ (cyan). **(d)** After segmentation (by using ta simple global $chroma\_threshold = 0.0223$) task, the chroma salience feature $f_\lambda^{\mathcal{R}c}$ musically segment *Alpana* and *Krithi* with respect to *Tani-Avarthanam*. **(e)** Manual segmentation of the recording. The white areas indicate transition regions (often pauses, sometimes used for tuning the instruments) between the respective parts.

Chapter 6

Conclusions

In this thesis, we considered a segmentation problem of Carnatic music audio recordings, where the different musical parts are characterized by the presence or absence of certain musical properties. A music signal may contain several musical aspects characterized by timbre, tempo, beat, played notes, melody, harmony, timbre of different instruments, dynamics of the sound, etc. For a given music processing task, only few of them may be relevant. As main technical contribution, we described several novel features that capture tempo and melody related information.

By means of the Carnatic music scenario, we demonstrated that these features reflect well whether there is a notion of a clear tempo or melody in the constituent parts of a Carnatic music main piece audio recording. Based on tempo property alone, one can clearly distinguish the *Alapana* part from the other two parts. If we consider melody property, one can also musically differentiate the *Tani-Avarthanam* part from the other two parts. In the study presented in this thesis, we focused on the aspect of tempo and melody salience that should also be useful for analyzing and understanding other types of music.

Our main contributions can be summarized as follows. We introduced the concert format in Carnatic music, explained in detail the significance and structure of the main piece and its musical characteristics. We further exploited the musical characteristics of each part of the main piece to design tempo and chroma salience features which define a notion of tempo and melody in any part of a music signal. These features can be used for any type of music as descriptors of a tempo salience and melody salience. Furthermore, we also showed the consistency of the features in analyzing the segments, corroborating with our experimental evidence. Further research into this area could lead to interesting analytical approaches to other forms of music and also new features which offer insights into musical structure. We intend to explore this technique for analyzing other structures in Carnatic music such as *RTP* and *Thillana* parts.

The chroma and tempo salience features can be further used in an automated algorithm for segmentation as shown in Figure 5.4 in conjunction with other features for retrieval tasks. The tempo salience and the melody salience can also be used as descriptors for Carnatic music.

Appendix A

Carnatic Music Instruments

The instruments used in the Carnatic music are broadly divided into two groups. They are as follows.

1. Harmonic or melodic instruments : The Melodic instruments may be further sub-divided into three groups. They are as follows.
   (a) String instruments : The string instruments are categorized by the way they are played, plucked or bowed. The string instruments used in Carnatic music are as follows.
      i. Tambura : is a long-necked plucked string instrument(3-6 strings) found. It is a drone instrument, which is an important part of the Carnatic music concerts. It is used to support and sustain the melody of instruments or a singer, and played in a continuous loop.
      ii. Tanpura : is also a long-necked plucked string instrument(3-6 strings) found in various forms in Indian music. It is similar to *Tambura*. It also act as a drone instrument, which is played throughout the Carnatic concerts to provide a pitch reference to the artist.
      iii. Veena : is a 7-string instrument which was invented in ancient India. It is made out of the dried wood of the jack fruit tree. The instrument is played through breath retention (in other words, you neither inhale nor exhale while playing).
      iv. Violin : is a string instrument generally used for high pitched sounds. The violinist produces the sound by having a bow over the strings. This was mainly used with gut, nylon, synthetic or steel strings. It is also used in Western music.
   (b) Wind instruments : The wind instruments used in Carnatic music as as follows.
      i. Indian Flute : is one of the oldest wind instruments used in the Carnatic music. The most commonly used flute in Indian classical music is made up of bamboo. A typical Indian *Flute* is about fourteen inches in length and 0.75 of an inch in diameter.
      ii. Nagaswaram : is a musical instrument, mostly used in the Hindu weddings and in the South Indian temples. This was played with pair accompanied with a pair of drums called *Thavil* or *Ottu*. It is the most popular classical musical instrument

and also used in the concerts. Its body is made of hard wood, and its flaring bell is made of wood or metal.

iii. Ottu : is a drone instrument. It resembles the *Nagaswaram* in shape and construction but is slightly longer. The player holds the reed at the upper end of the instrument in his mouth and blows into it to produce a single note which provides the drone for the Carnatic music.

(c) Bellowed instruments : The bellowed instruments used in Carnatic music as as follows.

i. Harmonium : is used to provide a drone sound. It has a keyboard of over two and one-half octaves and works on a system of bellows. This instruments is very popular in the North India. In South India, it is used more commonly in concerts and Bhajans (devotional) songs.

ii. Sruti-peti : is similar to *Harmonium* and it is used to provide a drone sound. It is a small box with few buttons used for adjusting the drone sounds.

2. Percussion or rhythm instruments : The Percussive instruments used in Carnatic music as as follows.

(a) Mridangam : is a percussive instrument from ancient Indian origin. It is like a drum shape with membranes on both of its ends. One of the membrane is smaller and the other one on larger membrane. Like other drums, it is also used as an accompaniment for instruments, vocals and dance performances.

(b) Ghatam : is pot like structure. The player uses his fingers, thumbs, palms and heels of the hands to play on the outer surface of the pot. A low pitch bass sound is obtained by hitting the mouth of the pot with a hand. Different tones generated while hitting on different regions of the pot. It accompanies with *Mridangam* instrument played in the concerts.

(c) Kanjira : is used as a supporting instrument for Mridangam. It is made from the wood of jack fruit tree. It is $7-9$ inch in width and $2-4$ inches in depth.

(d) Moorsing : is mainly used in the Carnatic concerts and for Indian folk music. It is a tuned percussive instrument. It is mostly tuned to higher octave as compared to the remaining percussive instruments. It is made of a metal ring which looks similar to a shape of a horse shoe.

(e) Thavil : The Thavil is the main percussion instrument for the *Nagaswaram*. It is a barrel-shaped drum which is hollowed out of a solid block of wood. The one head of the Thavil is made from the skin of Buffalo and the other head is made from goat skin.

(f) Jalra : is made of metal and connected with a copper cord which passes through holes in their center. They produce a rhythmic sound when struck together. The sound's pitch varies according to their size, weight and the material of their construction. By varying the point of contact, one can adjust its timbre.

Appendix B

Annotation of Carnatic Music Database

In our experiments, we have used the music recordings of good audio quality from the Sangeethapriya website[1]. In total, our dataset consists of 15 main pieces from various Carnatic concerts with an overall duration of more than 10 hours. In the chapter, we discuss about the naming convention of the concert files (see Appendix B.1) and how to manually annotate the music recordings with the help of Sonic Visualiser[2]. software (see Appendix $B.2$) to find the segment boundaries for the *Alapana*, *Krithi* and *Tani-Avarthanam* parts.

B.1 Database Naming Convention

As mentioned in Chapter 2, Every piece in a Carnatic concert is performed with a predefined *Raga* and *Tala*. It may also so happen that, main piece may include *Alapana* , *Krithi* and *Tani-Avarthanam* or either *Alapana* or *Tani-Avarthanam* along with *Krithi* or it can be only *Krithi*, as it is the heart of the main piece. In this thesis, we shall consider the main piece having all the three parts, *Alapana*, *Krithi* and *Tani-Avarthanam*. Considering all the above criteria for the naming convention of the main piece. We have come up with the following naming convention which has information regarding *Raga*, *Krithi*, *Tala*, *Performer*, *year* and *month* of the performance as shown below.

$$raga\_krithi\_tala\_performer\_year\_month.wav$$

For example, *madyamaavati_paalinchukaamaakSi_aadi_tnseshagopalan*_2003_01.*wav* is name given to a Carnatic music main piece audio recording with madyamaavati as the name of the *Raga*, paalinchukaamaakSi as *Krithi*, aadi as *Tala*, T.N.Seshagopalan as *performer*, 2003 as *year* and 01 as the *month* of the performance. A metadata file is also provided with the database for every main piece of the concert which almost has all information like, *file formats*, *Raga*, *Tala* lyrical composition of *Krithi* part and soon as shown below.

[1] `http://www.sangeethapriya.org`
[2] `http://http://www.sonicvisualiser.org/`

```
######################Filename#############################################

raaga_kriti_tala_Performer_year_xx.wav(format)
madyamaavati_paalinchukaamaakSi_aadi_tnseshagopalan_2003_01.wav
madyamaavati_paalinchukaamaakSi_aadi_tnseshagopalan_2003_01.mp3

########################################################################
   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

         Raga     : madyamaavati
         Krithi   : paalinchu kaamaakSi
         TaaLam   : aadi
         Singer   : t n seshagopalan
         Composer : shyaamaa shaastree
         Concert  : madurai
         Year     : 2003

   %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
*************************Raga*********************************************

paalinchu kaamaakSi
raagam: madyamaavati
22 kharaharapriya janya
Aa: S R2 M1 P N2 S
Av: S N2 P M1 R2 S
*************************Tala*********************************************

Taalam: aadi
Composer: Shyaamaa Shaastree
*************************Lyrics*******************************************

pallavi:
pAlincu kAmAkSi pAvani pApa shamani

anupallavi:
cAla bahu vidhamugA ninu sadA vEDukonaDina endEla iLAgu sEvu veta
harincavE vEgamE nanu

swara saahitya:
kanaka giri sadana lalita ninu bhajana santatamu jEya nijanuDana vinumu
nikhila bhuvana janavini ipuDu mA duritamu dIrcci varAlicci

caraNam 1:
svAntambulOna ninnE dalacE sujanula kellanE vELa santOSamu
losagEvani nIvu manOratha phaladAyinivani kAntamagu pEru ponditivi
kAruNya mUrti vaijagamu kApADina talli gadA nEnu nIdu biTTanu lAlinci

caraNam 2:
I mUrti inta tEjOmayamai iTuvale kIrti visphUrti viTalAnaya guNa
mUrti trilOkamulO jUcindaina galadA Emi toli nOmu nOcitinO nI pAda
padma darshanamu vEmAru labhinci krtArtuDainati nA manaviyAlinci

caraNam 3:
rAjAdhirAja rAjanmakuTI taTamaNi rAj abhAjAla nija sannidhi
dEvi samasta janula kella varadA rAjamukhi shyAmakrSNanuta
kAnci purIshvari vikaca rAjIva daLAkAi jagat sAkSiyau prasanna parAshakti

swara:
ni sa ri pa ma ri sa ni sa ri ma ri sa ni sa ri sA, ri sa ni pA,
pa ma pa ni sa rI
pa ma pa sa ri sa pa ma pa ni pa sa ni ri sa ma rI , sa ni pa
ri sa , ni pa ma , rI sa

http://www.karnatik.com/c2442.shtml

************************************************************************
```

The naming convention of the the main piece audio recordings are very long. In order to have simple names, we perform mapping of naming conventions of the dataset as shown in Figure B.1.

Dataset Mapping

aabheri_nagumomuganaleni_aadi_tnseshagopalan_2002_01.wav	⇒	*Raga_01.wav*
bhairavi_thanayunibrova_aadi_tnseshagopalan_1993_01.wav	⇒	*Raga_02.wav*
brindaavanasaaranga_kamalaaptakula_deshaadi_tnseshagopalan_1980_01.wav	⇒	*Raga_03.wav*
kaambhoji_ohrangasaayi_aadi_tnseshagopalan_2002_01.wav	⇒	*Raga_04.wav*
kaambhoji_ohrangasaayi_aadi_tnseshagopalan_xxxx_01.wav	⇒	*Raga_05.wav*
kalyaani_thallininunera_mishracaapu_tnseshagopalan_1983_01.wav	⇒	*Raga_06.wav*
kalyani_etaavunara_aadi_tnseshagopalan_1988_01.wav	⇒	*Raga_07.wav*
kalyani_kamalambambhajareremaanasa_aadi_sanjaisubramaniam_2007_01.wav	⇒	*Raga_08.wav*
madyamaavati_paalinchukaamaakSi_aadi_tnseshagopalan_2003_01.wav	⇒	*Raga_09.wav*
mukhaari_elaavathaaram_aadi_tmkrishna_2005_01.wav	⇒	*Raga_10.wav*
shankaraabharanam_endhukupeddhala_deshaadi_tnseshagopalan_1979_01.wav	⇒	*Raga_11.wav*
shankaraabharanam_sridakshinaamoortte_mishrajhampa_tmkrishna_2006_01.wav	⇒	*Raga_12.wav*
shankarabharanam_edhutanilichithe_aadi_tnseshagopalan_1988_01.wav	⇒	*Raga_13.wav*
todi_kaddhanuvaariki_aadi_tnseshagopalan_1989_01.wav	⇒	*Raga_14.wav*
todi_shrikrishnambhajamanasa_aadi_tnseshagopalan_1985_01.wav	⇒	*Raga_15.wav*

Table B.1: Mapping of dataset naming conventions

B.2 Annotation

For annotating a Carnaitc music main piece audio recording, we manually find the boundaries of the *Alapana*, *Krithi* and *Tani-Avarthanam* parts with the help of the *Sonic Visualiser* tool. We listen to the audio recordings in the Sonic Visualiser and manually annotate the boundaries of all the three parts (for example, see Table B.2).

Parts	$Start_{time}$(in sec)	End_{time}(in sec)
Alapana	$t_{alapana\_start}$	$t_{alapana\_end}$
Krithi	t_{krithi_start}	t_{krithi_end}
Tani-Avarthanam	t_{tani_start}	t_{tani_end}

Table B.2: Annotation of the main piece of the concert

B.3 Excerpt Database

A typical Carnatic concert main piece may last for about 60 minutes. Testing on such a huge audio recording is time consuming. Hence, we obtain excerpt files from the audio recordings. The excerpt files contain small segments of each part (say, 2 minutes) along with the transition region between the parts of the main piece.

The excerpt files created from a main piece audio recording is as shown in the Figure B.1. The Figure B.1(a) represents the main piece constituting parts, the*Alapana*, the *Krithi* and the

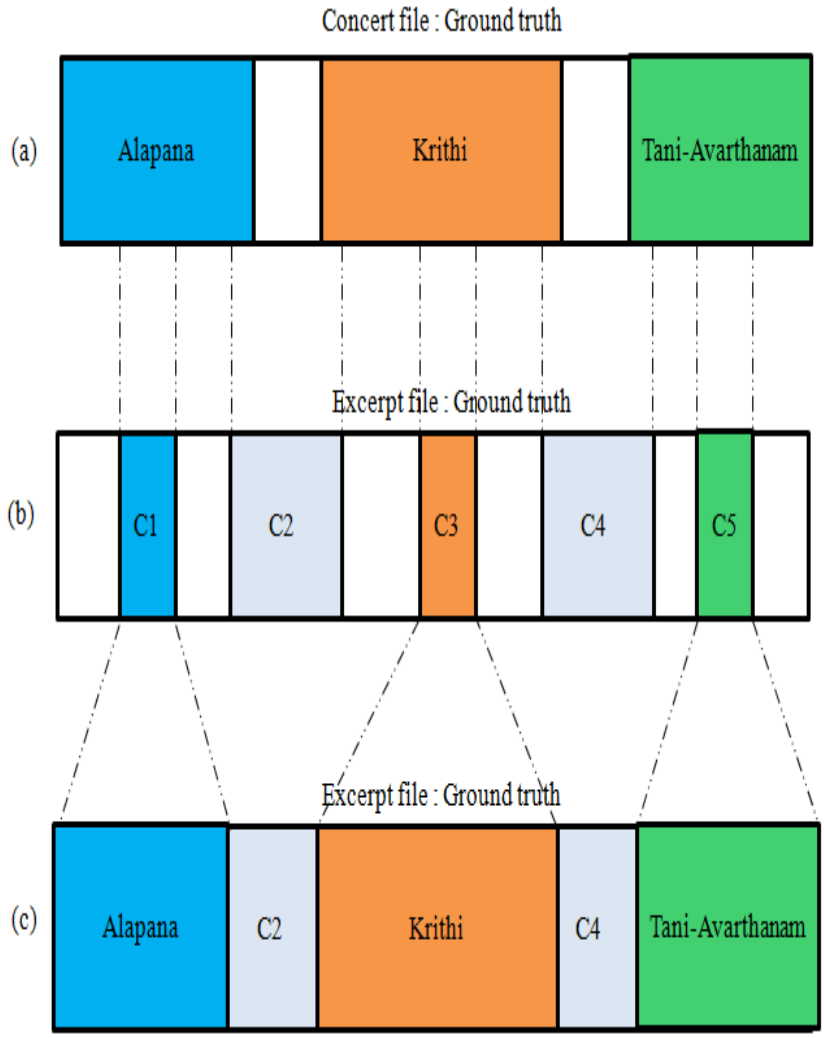

Figure B.1: For example, *bhairavi_thanayunibrova_aadi_tnseshagopalan_*1993_01.*wav*, a main piece of the concert file. **(a)** Shows the ground truth of main piece of the concert of length roughly around 60-90 minutes. **(b)** Taking chunks [C1, C3, C5] of length 2 minutes from each parts of main piece along with the transition parts [C2, C4]. **(c)** Concatenate [C1 ... C5] to get an excerpt file, which is used for testing the salience features.

Tani-Avarthanam along with the transition regions. We now take chunks of 2 minutes of each of the parts as shown in Figure B.1 (b) as C1, C3 and C5 whereas, C2 and C4 are also taken to preserve the information about the transition between the parts. Now by concatenating these chunks [C1 ... C5], results in excerpt file as shown in Figure B.1 (c)).

By taking different chunks from each main audio recording, 10 such excerpt audio files are obtained from each main piece audio recording. As we have 15 such huge main piece audio recordings and 10 excerpt audio files per audio recording, resulting in (15×10) 150 excerpt audio files. Hence, we use 150 total excerpt files in our dataset.

Appendix C

Dataset Overview

This chapter is organized as follows. Firstly, Based on the manually annotation dataset, we compute statistics to investigate how well the three different musical parts are characterized by our tempo and chroma salience features. Secondly, we show an overview of the results of tempo and chroma salience features on the Carnatic music excerpt dataset. The tempo salience features are obtained from the concept of tempogram which, helps to musically differentiate the*Alapana* with respect to the *Krithi* and the *Tani-Avarthanam* parts based on the tempo cue. Similarly, the salience features can be derived from the concept of Chroma features which, helps to musically differentiate the *Tani-Avarthanam* with the *Krithi* and the *Alapana* parts based on melody cue.

Feature	$f_\lambda^{\mathcal{M}_\mathcal{C}}$					
	$\overline{\mu}$			$\overline{\sigma}$		
	A	K	T	A	K	T
Raga_01.wav	0.3424	0.3418	0.0502	0.2391	0.1992	0.0185
Raga_02.wav	0.6049	0.4756	0.0653	0.1507	0.0891	0.0334
Raga_03.wav	0.1822	0.4044	0.0512	0.1158	0.2267	0.0299
Raga_04.wav	0.1235	0.2719	0.0261	0.0667	0.2303	0.0158
Raga_05.wav	0.4706	0.3288	0.0311	0.2229	0.1722	0.0114
Raga_06.wav	0.6292	0.5607	0.1419	0.2025	0.1479	0.0757
Raga_07.wav	0.2147	0.0947	0.0189	0.2227	0.0487	0.0227
Raga_08.wav	0.4188	0.0734	0.0182	0.2126	0.0318	0.0122
Raga_09.wav	0.3614	0.2593	0.0202	0.2515	0.2276	0.0101
Raga_10.wav	0.4730	0.3023	0.0091	0.2173	0.1559	0.0066
Raga_11.wav	0.5100	0.4936	0.0791	0.2032	0.1409	0.0225
Raga_12.wav	0.2436	0.5881	0.0107	0.1913	0.1674	0.0091
Raga_13.wav	0.5017	0.3537	0.0120	0.1519	0.1983	0.0146
Raga_14.wav	0.4568	0.2756	0.0320	0.1581	0.1855	0.0260
Raga_15.wav	0.4317	0.2403	0.0237	0.2303	0.1209	0.0079
Average	0.3976	0.3376	0.0393	0.1954	0.1672	0.0269

Table C.1: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of maximum chroma salience feature ($f_\lambda^{\mathcal{M}_\mathcal{C}}$) shown for the three different parts *Alapana* (A), *Krithi* (K), and *Tani-Avarthanam* (T).

Feature	$f_\lambda^{\mathcal{S}_\mathcal{C}}$					
	$\overline{\mu}$			$\overline{\sigma}$		
	A	K	T	A	K	T
Raga_01.wav	0.4001	0.4522	0.1076	0.2381	0.2126	0.0427
Raga_02.wav	0.6142	0.5021	0.0934	0.1983	0.0925	0.0396
Raga_03.wav	0.1608	0.3491	0.0718	0.1069	0.2156	0.0429
Raga_04.wav	0.1084	0.2820	0.0437	0.0613	0.2518	0.0277
Raga_05.wav	0.4447	0.3321	0.0441	0.2225	0.1625	0.0172
Raga_06.wav	0.5753	0.5703	0.2378	0.2010	0.1612	0.1211
Raga_07.wav	0.3128	0.2128	0.0445	0.2023	0.1291	0.0476
Raga_08.wav	0.5257	0.1321	0.0493	0.2196	0.0513	0.0326
Raga_09.wav	0.3806	0.2339	0.0252	0.2773	0.2083	0.0139
Raga_10.wav	0.4938	0.3824	0.0179	0.2098	0.2037	0.0149
Raga_11.wav	0.5339	0.5902	0.1559	0.2370	0.1786	0.0453
Raga_12.wav	0.2795	0.6441	0.0199	0.2201	0.1906	0.0153
Raga_13.wav	0.5389	0.5065	0.0189	0.1150	0.2490	0.0178
Raga_14.wav	0.5087	0.3581	0.0585	0.1856	0.2316	0.0471
Raga_15.wav	0.4355	0.3204	0.0475	0.2284	0.1773	0.0135
Average	0.4209	0.3912	0.0691	0.2026	0.1889	0.0445

Table C.2: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of sum chroma salience feature ($f_\lambda^{\mathcal{S}_\mathcal{C}}$) shown for the three different parts *Alapana* (A), *Krithi* (K), and *Tani-Avarthanam* (T).

Feature	$f_\lambda^{\mathcal{RC}}$					
	$\overline{\mu}$			$\overline{\sigma}$		
	A	K	T	A	K	T
Raga_01.wav	0.0358	0.0388	0.0008	0.0214	0.0243	0.0014
Raga_02.wav	0.0734	0.0517	0.0007	0.0279	0.0152	0.0030
Raga_03.wav	0.0206	0.0426	0.0012	0.0159	0.0300	0.0018
Raga_04.wav	0.0204	0.0646	0.0029	0.0150	0.0617	0.0034
Raga_05.wav	0.1055	0.0680	0.0003	0.0527	0.0416	0.0005
Raga_06.wav	0.0809	0.0797	0.0099	0.0324	0.0253	0.0088
Raga_07.wav	0.0131	0.0039	0.0003	0.0114	0.0044	0.0007
Raga_08.wav	0.0125	0.0003	0	0.0107	0.0011	0
Raga_09.wav	0.0512	0.0244	0.0000	0.0411	0.0301	0.0001
Raga_10.wav	0.0233	0.0112	0	0.0135	0.0079	0
Raga_11.wav	0.0838	0.0895	0.0025	0.0397	0.0326	0.0024
Raga_12.wav	0.0352	0.0990	0.0002	0.0348	0.0322	0.0009
Raga_13.wav	0.0630	0.0526	0.0002	0.0176	0.0361	0.0012
Raga_14.wav	0.0714	0.0452	0.0007	0.0280	0.0380	0.0024
Raga_15.wav	0.0477	0.0277	0.0000	0.0305	0.0216	0.0003
Average	0.0492	0.0466	0.0013	0.0288	0.0308	0.0028

Table C.3: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of relative chroma strength salience feature ($f_\lambda^{\mathcal{RC}}$) shown for the three different parts *Alapana* (A), *Krithi* (K), and *Tani-Avarthanam* (T).

Feature	$f_\lambda^{\mathcal{H}}$					
	$\overline{\mu}$			$\overline{\sigma}$		
	A	K	T	A	K	T
Raga_01.wav	0.0023	0.0223	0.0235	0.0011	0.0162	0.0135
Raga_02.wav	0.0031	0.0237	0.0168	0.0013	0.0100	0.0105
Raga_03.wav	0.0034	0.0220	0.0330	0.0028	0.0069	0.0094
Raga_04.wav	0.0059	0.0694	0.0115	0.0025	0.0562	0.0056
Raga_05.wav	0.0030	0.0145	0.0149	0.0019	0.0064	0.0088
Raga_06.wav	0.0021	0.0265	0.0264	0.0009	0.0072	0.0141
Raga_07.wav	0.0042	0.0304	0.0152	0.0022	0.0166	0.0047
Raga_08.wav	0.0038	0.0385	0.0419	0.0017	0.0129	0.0362
Raga_09.wav	0.0039	0.0293	0.0342	0.0013	0.0133	0.0204
Raga_10.wav	0.0020	0.0287	0.0186	0.0007	0.0124	0.0124
Raga_11.wav	0.0035	0.0156	0.0258	0.0012	0.0074	0.0090
Raga_12.wav	0.0027	0.0140	0.0176	0.0013	0.0096	0.0052
Raga_13.wav	0.0038	0.0288	0.0401	0.0013	0.0125	0.0259
Raga_14.wav	0.0026	0.0315	0.0238	0.0016	0.0094	0.0102
Raga_15.wav	0.0024	0.0314	0.0255	0.0010	0.0101	0.0099
Average	0.0032	0.0285	0.0246	0.0016	0.0181	0.0154

Table C.4: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of tempo entropy salience feature ($f_\lambda^{\mathcal{H}}$) shown for the three different parts *Alapana* (A), *Krithi* (K), and *Tani-Avarthanam* (T).

Feature	$f_{\lambda}^{\mathcal{M}}$					
	$\overline{\mu}$			$\overline{\sigma}$		
	A	K	T	A	K	T
Raga_01.wav	0.0132	0.0580	0.0618	0.0039	0.0273	0.0223
Raga_02.wav	0.0161	0.0660	0.0477	0.0037	0.0146	0.0171
Raga_03.wav	0.0187	0.0564	0.0708	0.0102	0.0112	0.0141
Raga_04.wav	0.0229	0.1212	0.0425	0.0067	0.0595	0.0105
Raga_05.wav	0.0170	0.0461	0.0446	0.0062	0.0110	0.0185
Raga_06.wav	0.0134	0.0679	0.0606	0.0025	0.0075	0.0253
Raga_07.wav	0.0180	0.0718	0.0417	0.0061	0.0276	0.0128
Raga_08.wav	0.0187	0.0856	0.0777	0.0055	0.0178	0.0477
Raga_09.wav	0.0202	0.0750	0.0758	0.0047	0.0213	0.0299
Raga_10.wav	0.0130	0.0713	0.0497	0.0040	0.0155	0.0225
Raga_11.wav	0.0176	0.0508	0.0501	0.0041	0.0126	0.0124
Raga_12.wav	0.0156	0.0442	0.0532	0.0044	0.0150	0.0100
Raga_13.wav	0.0177	0.0710	0.0813	0.0047	0.0204	0.0410
Raga_14.wav	0.0161	0.0732	0.0579	0.0049	0.0149	0.0193
Raga_15.wav	0.0157	0.0688	0.0625	0.0046	0.0146	0.0140
Average	0.0169	0.0685	0.0585	0.0054	0.0228	0.0237

Table C.5: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of tempo maximum median salience feature ($f_{\lambda}^{\mathcal{M}}$) shown for the three different parts *Alapana* (A), *Krithi* (K), and *Tani-Avarthanam* (T).

Feature	$f_{0,\lambda}^{\mathcal{I}}$					
	$\overline{\mu}$			$\overline{\sigma}$		
	A	K	T	A	K	T
Raga_01.wav	0.1296	0.0136	0.0032	0.0446	0.0287	0.0078
Raga_02.wav	0.1205	0.0008	0.0145	0.0385	0.0033	0.0208
Raga_03.wav	0.1015	0.0042	0.0093	0.0493	0.0095	0.0142
Raga_04.wav	0.0956	0.0017	0.0108	0.0453	0.0060	0.0143
Raga_05.wav	0.0916	0.0049	0.0326	0.0609	0.0115	0.0308
Raga_06.wav	0.1152	0.0051	0.0080	0.0520	0.0115	0.0155
Raga_07.wav	0.1043	0.0216	0.0120	0.0420	0.0337	0.0131
Raga_08.wav	0.0814	0.0037	0.0151	0.0427	0.0084	0.0181
Raga_09.wav	0.0873	0.0058	0.0056	0.0488	0.0168	0.0170
Raga_10.wav	0.1224	0.0026	0.0210	0.0616	0.0050	0.0324
Raga_11.wav	0.1224	0.0031	0.0210	0.0421	0.0068	0.0238
Raga_12.wav	0.0925	0.0026	0.0091	0.0454	0.0072	0.0129
Raga_13.wav	0.1055	0.0032	0.0081	0.0562	0.0086	0.0141
Raga_14.wav	0.0951	0.0085	0.0025	0.0470	0.0231	0.0065
Raga_15.wav	0.1022	0.0071	0.0004	0.0501	0.0102	0.0023
Average	0.1045	0.0059	0.0115	0.0489	0.0154	0.0181

Table C.6: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of tempo stability salience feature ($f_{0,\lambda}^{\mathcal{I}}$) shown for the three different parts *Alapana* (A), *Krithi* (K), and *Tani-Avarthanam* (T).

Feature	$f^{\mathcal{I}}_{1,\lambda}$					
	$\overline{\mu}$			$\overline{\sigma}$		
	A	K	T	A	K	T
Raga_01.wav	0.0900	0.0136	0.0032	0.0426	0.0287	0.0078
Raga_02.wav	0.0833	0.0008	0.0145	0.0434	0.0033	0.0208
Raga_03.wav	0.0738	0.0042	0.0093	0.0462	0.0095	0.0142
Raga_04.wav	0.0643	0.0017	0.0108	0.0363	0.0060	0.0143
Raga_05.wav	0.0629	0.0049	0.0326	0.0546	0.0115	0.0308
Raga_06.wav	0.0799	0.0051	0.0080	0.0440	0.0115	0.0155
Raga_07.wav	0.0593	0.0216	0.0120	0.0358	0.0337	0.0131
Raga_08.wav	0.0609	0.0037	0.0151	0.0392	0.0084	0.0181
Raga_09.wav	0.0664	0.0058	0.0056	0.0413	0.0168	0.0170
Raga_10.wav	0.0757	0.0026	0.0210	0.0507	0.0050	0.0324
Raga_11.wav	0.0720	0.0031	0.0210	0.0429	0.0068	0.0238
Raga_12.wav	0.0514	0.0026	0.0091	0.0404	0.0072	0.0129
Raga_13.wav	0.0701	0.0032	0.0081	0.0480	0.0086	0.0141
Raga_14.wav	0.0713	0.0085	0.0025	0.0418	0.0231	0.0065
Raga_15.wav	0.0767	0.0071	0.0004	0.0422	0.0102	0.0023
Average	0.0705	0.0059	0.0115	0.0436	0.0154	0.0181

Table C.7: Mean $\overline{\mu}$ and standard deviation $\overline{\sigma}$ of tempo stability salience feature ($f^{\mathcal{I}}_{1,\lambda}$) shown for the three different parts *Alapana* (A), *Krithi* (K), and *Tani-Avarthanam* (T).

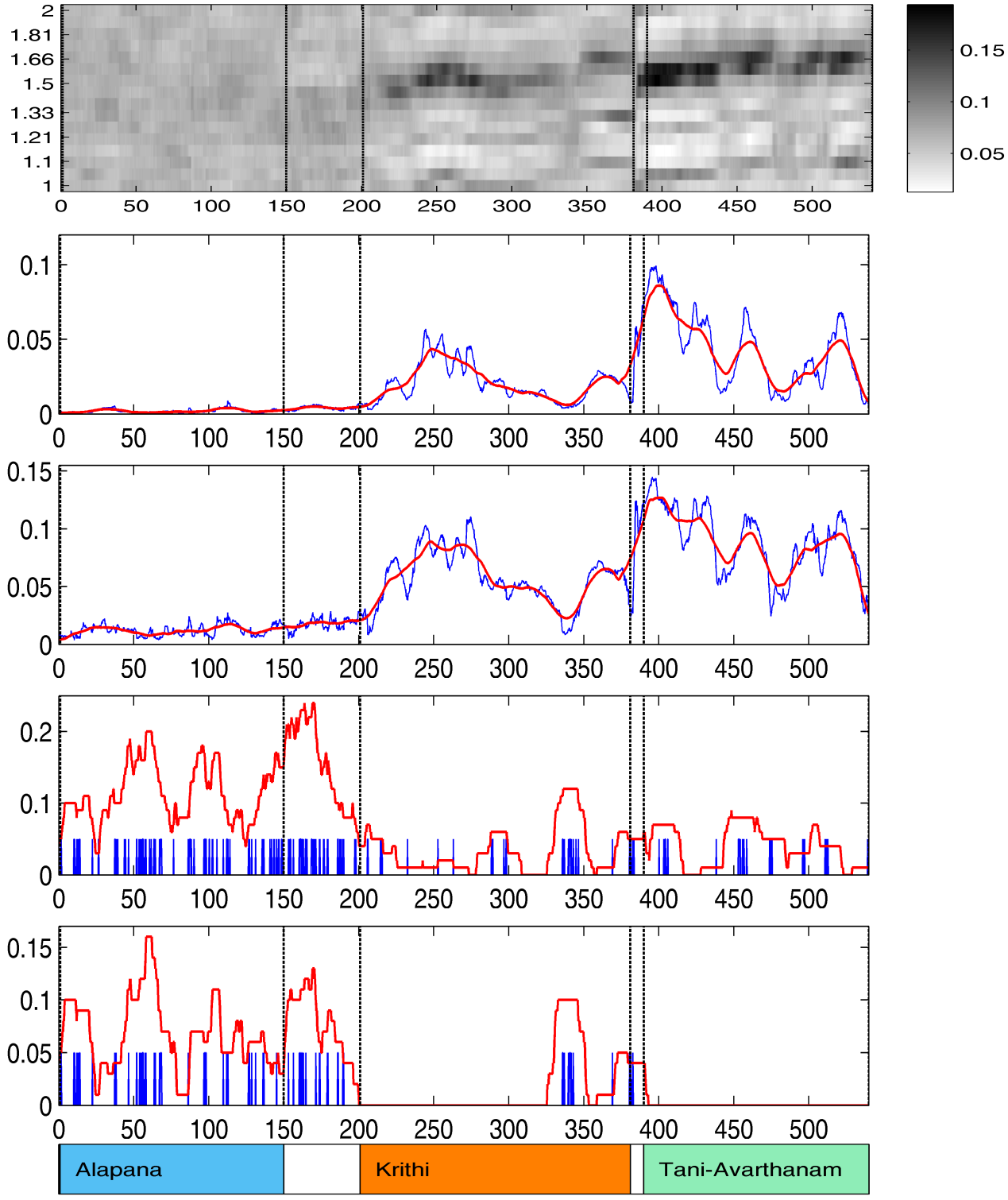

Figure C.1: *Wavfile* : *Raga_01_excerpt_s_152.wav*: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

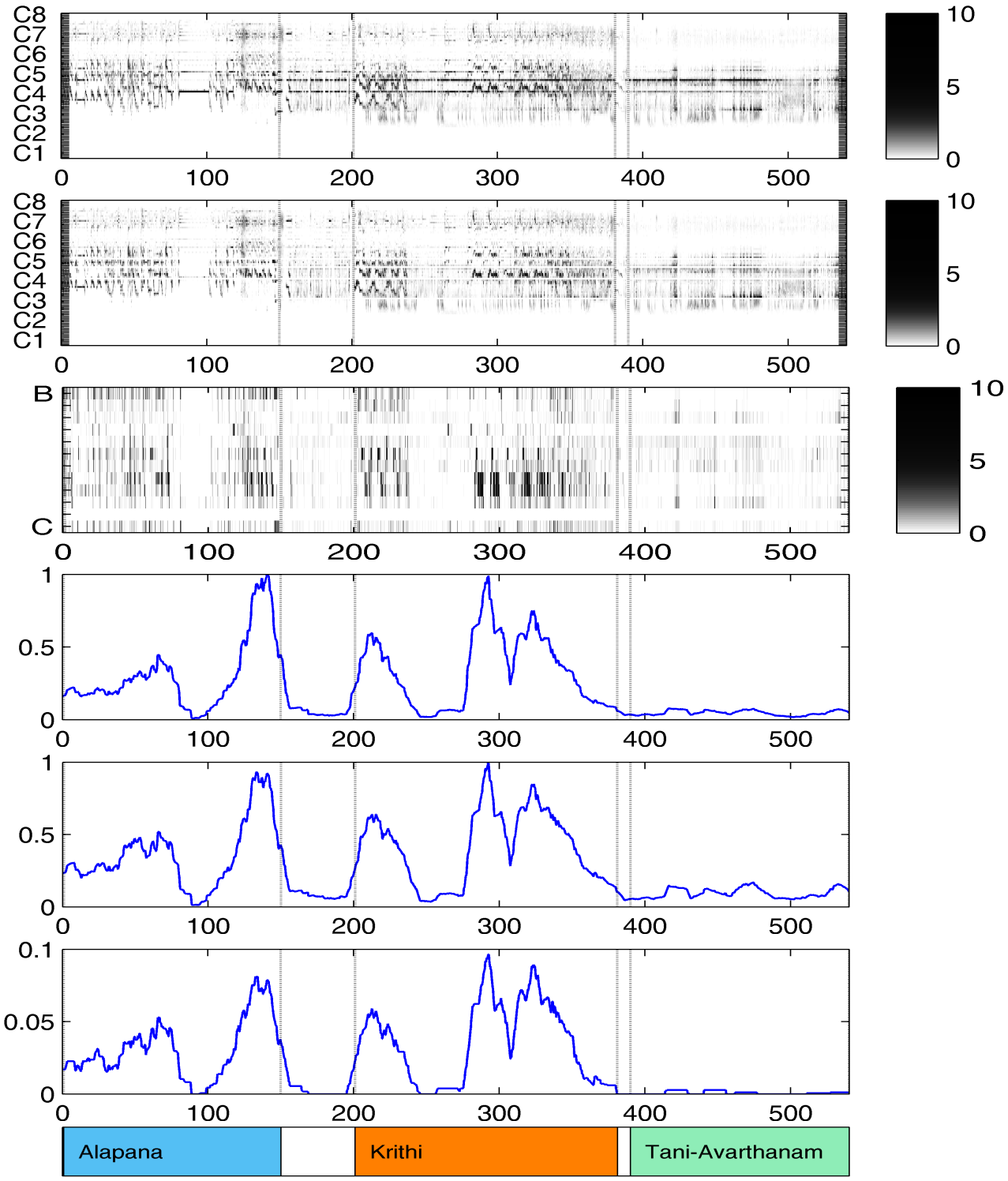

Figure C.2: $Wavfile$: $Raga\_01\_excerpt\_s\_152.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_\lambda^{\mathcal{M}_c}$, $f_\lambda^{\mathcal{S}_c}$ and $f_\lambda^{\mathcal{R}_c}$. The same parameter setting as in Figure 5.2 are used.

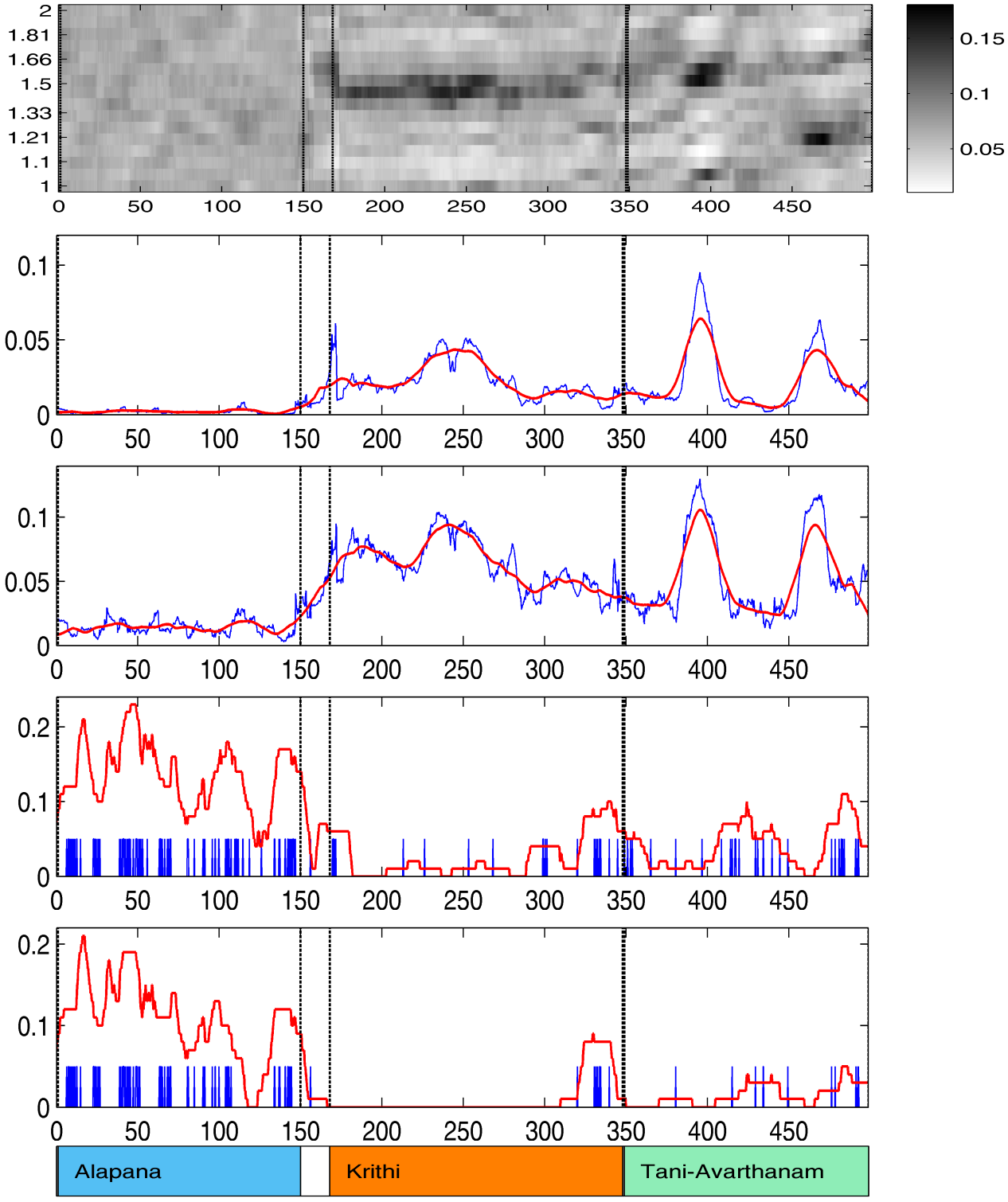

Figure C.3: $Wavfile$: $Raga\_02\_excerpt\_s\_152.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

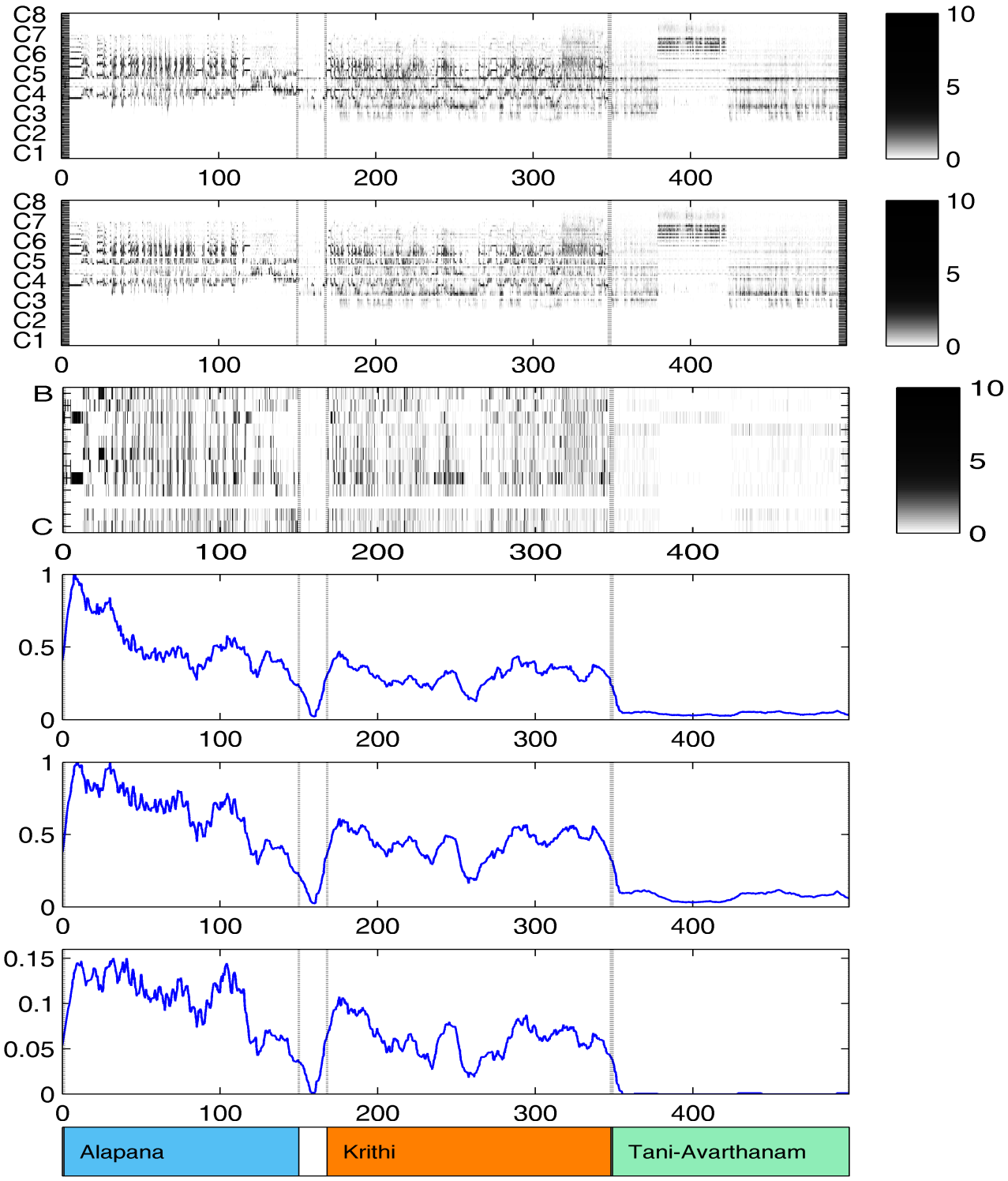

Figure C.4: $Wavfile$: $Raga\_02\_excerpt\_s\_152.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_\lambda^{\mathcal{M}_c}$, $f_\lambda^{\mathcal{S}_c}$ and $f_\lambda^{\mathcal{R}_c}$. The same parameter setting as in Figure 5.2 are used.

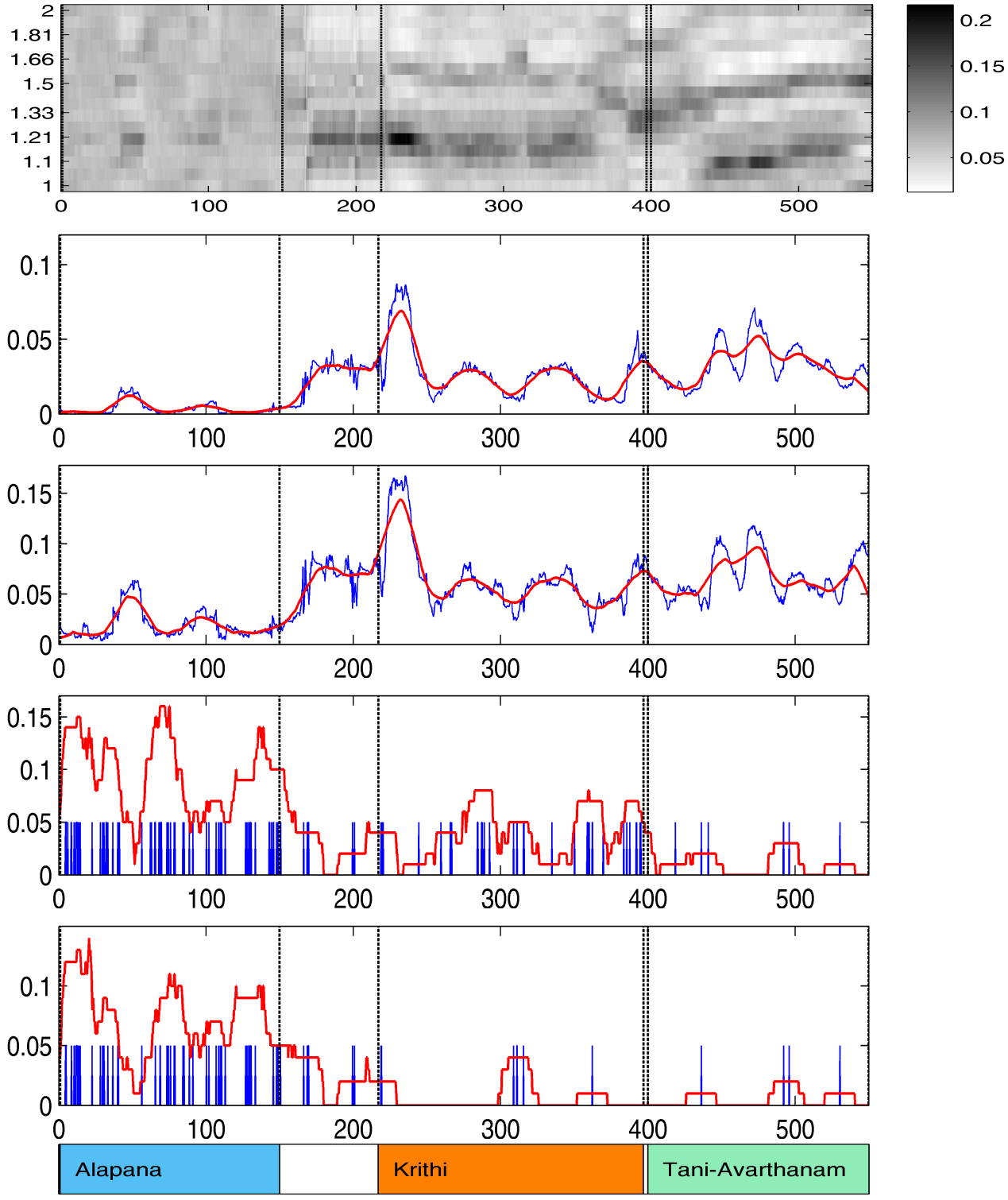

Figure C.5: $Wavfile$: $Raga\_03\_excerpt\_s\_248.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

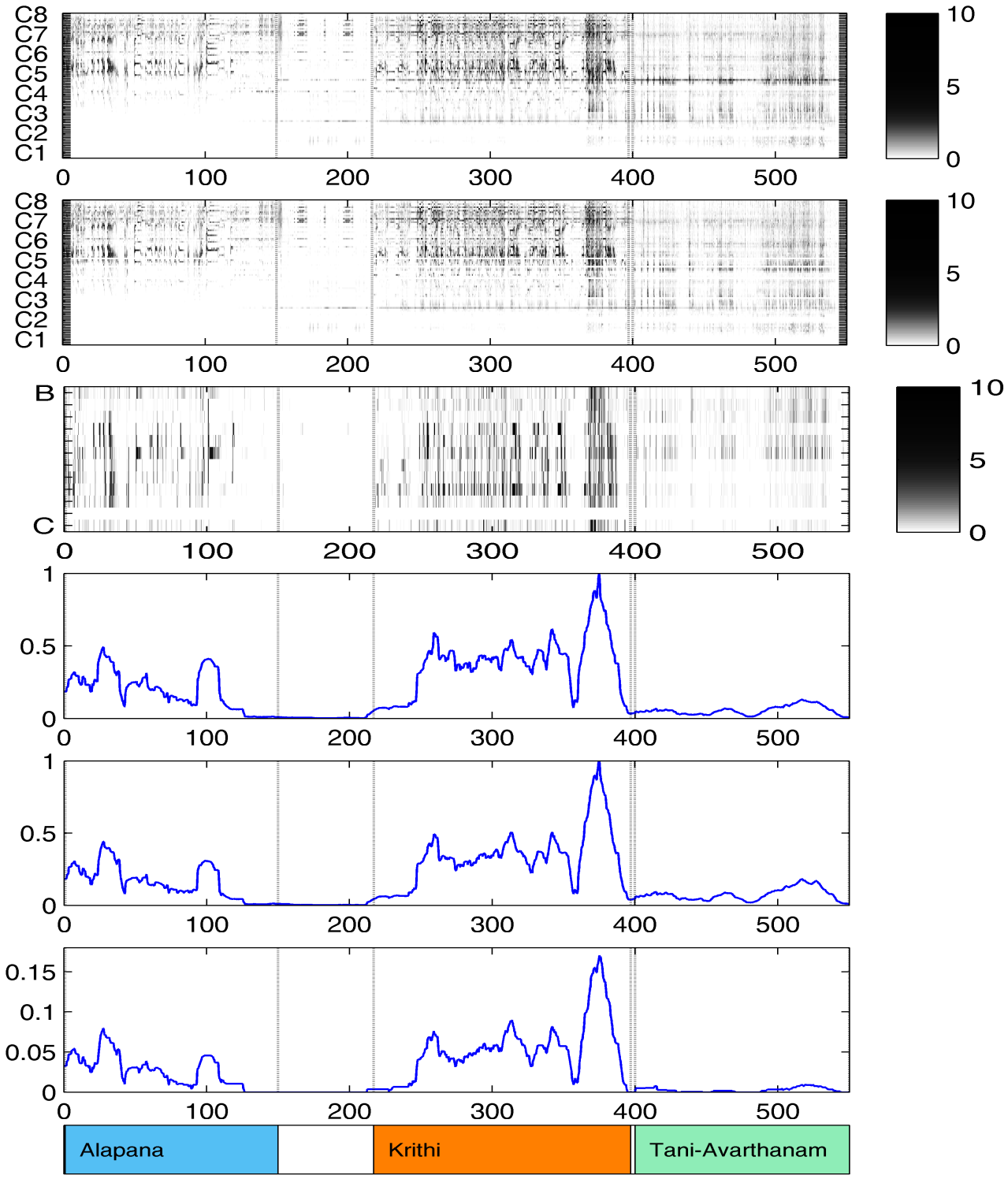

Figure C.6: $Wavfile$: *Raga_03_excerpt_s_248.wav*: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}c}$, $f_{\lambda}^{\mathcal{S}c}$ and $f_{\lambda}^{\mathcal{R}c}$. The same parameter setting as in Figure 5.2 are used.

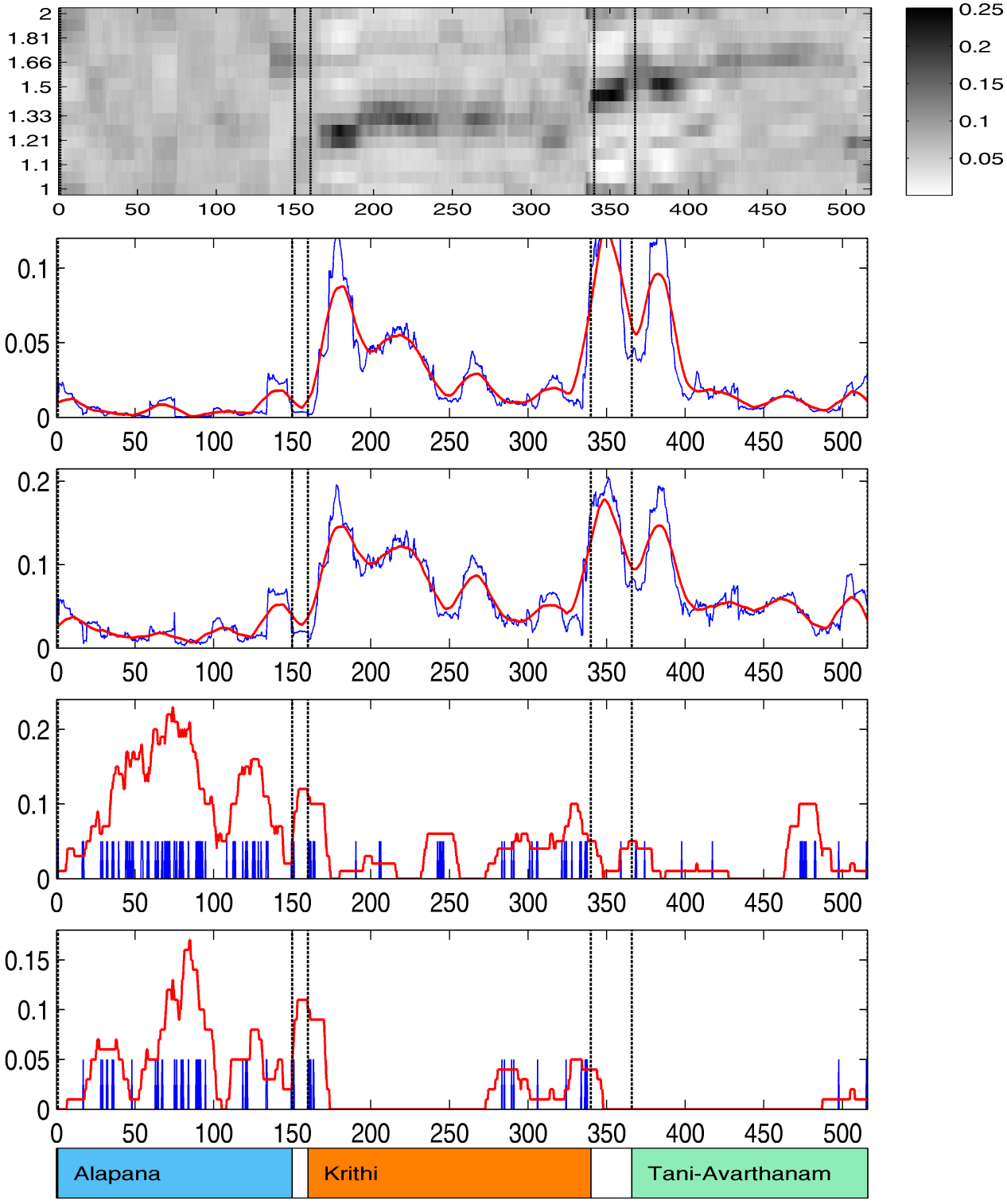

Figure C.7: *Wavfile* : *Raga_04_excerpt_s_260.wav*: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

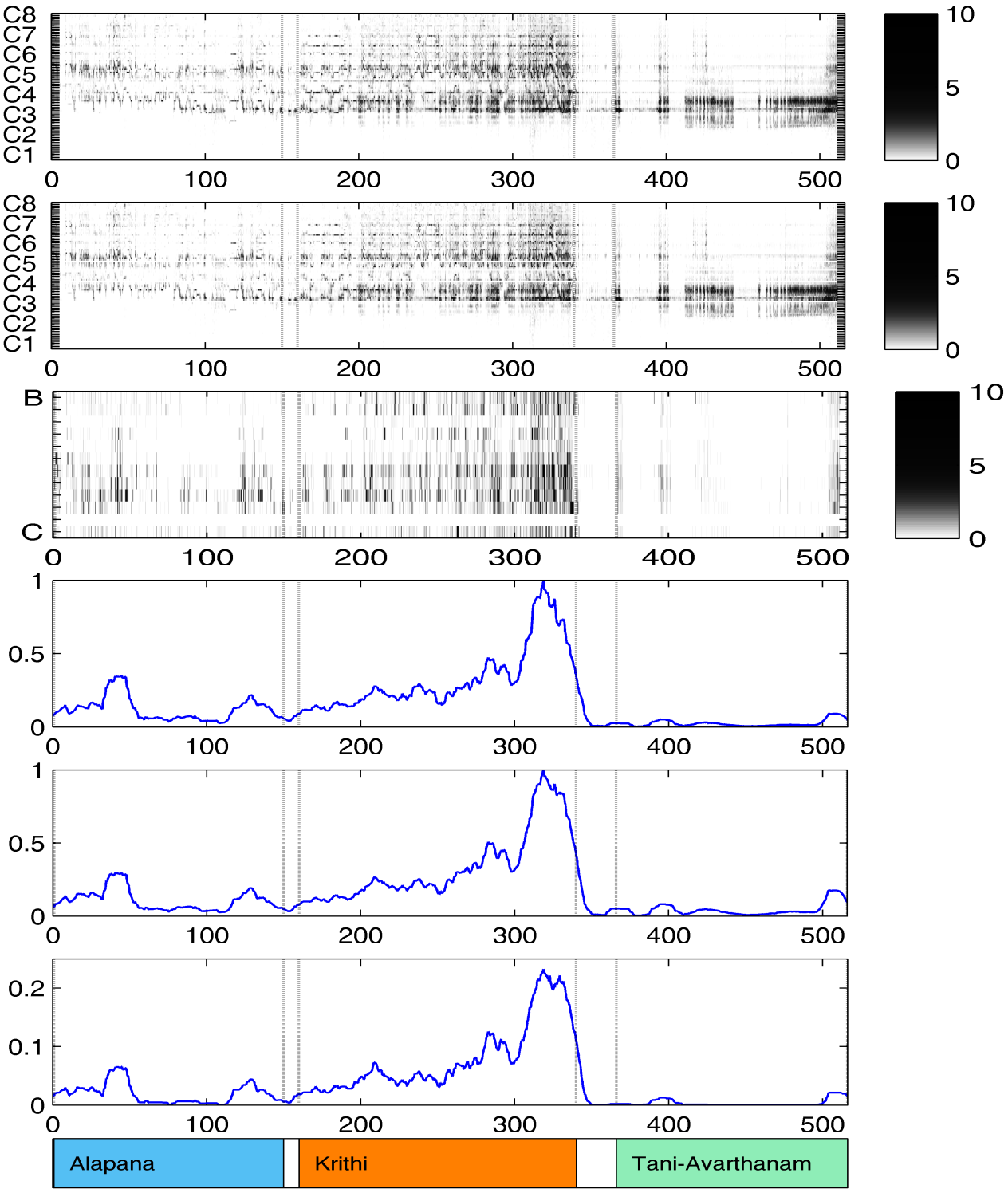

Figure C.8: $Wavfile$: $Raga\_04\_excerpt\_s\_260.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_\lambda^{\mathcal{M}_C}$, $f_\lambda^{\mathcal{S}_C}$ and $f_\lambda^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

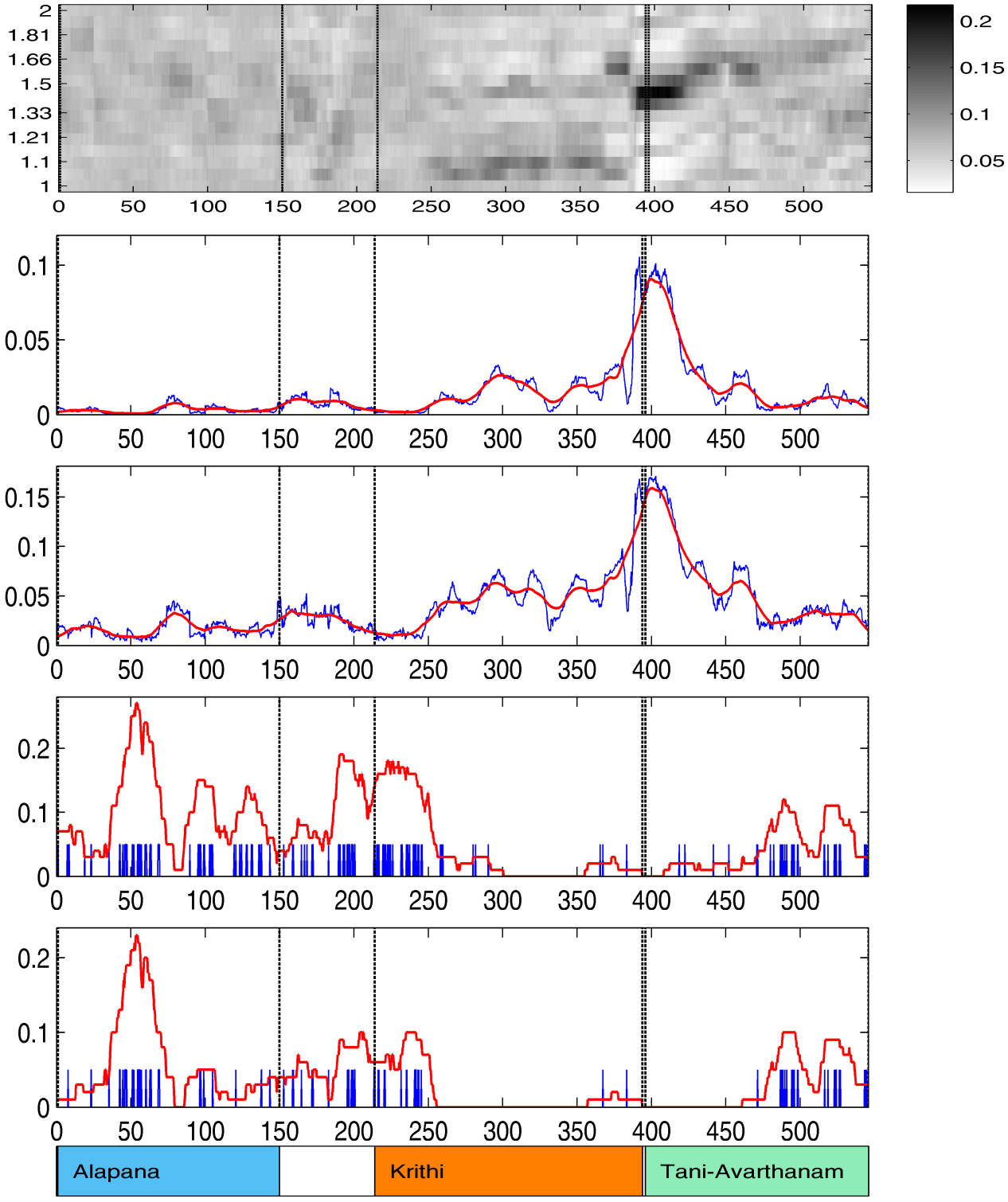

Figure C.9: $Wavfile$: $Raga\_05\_excerpt\_s\_212.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

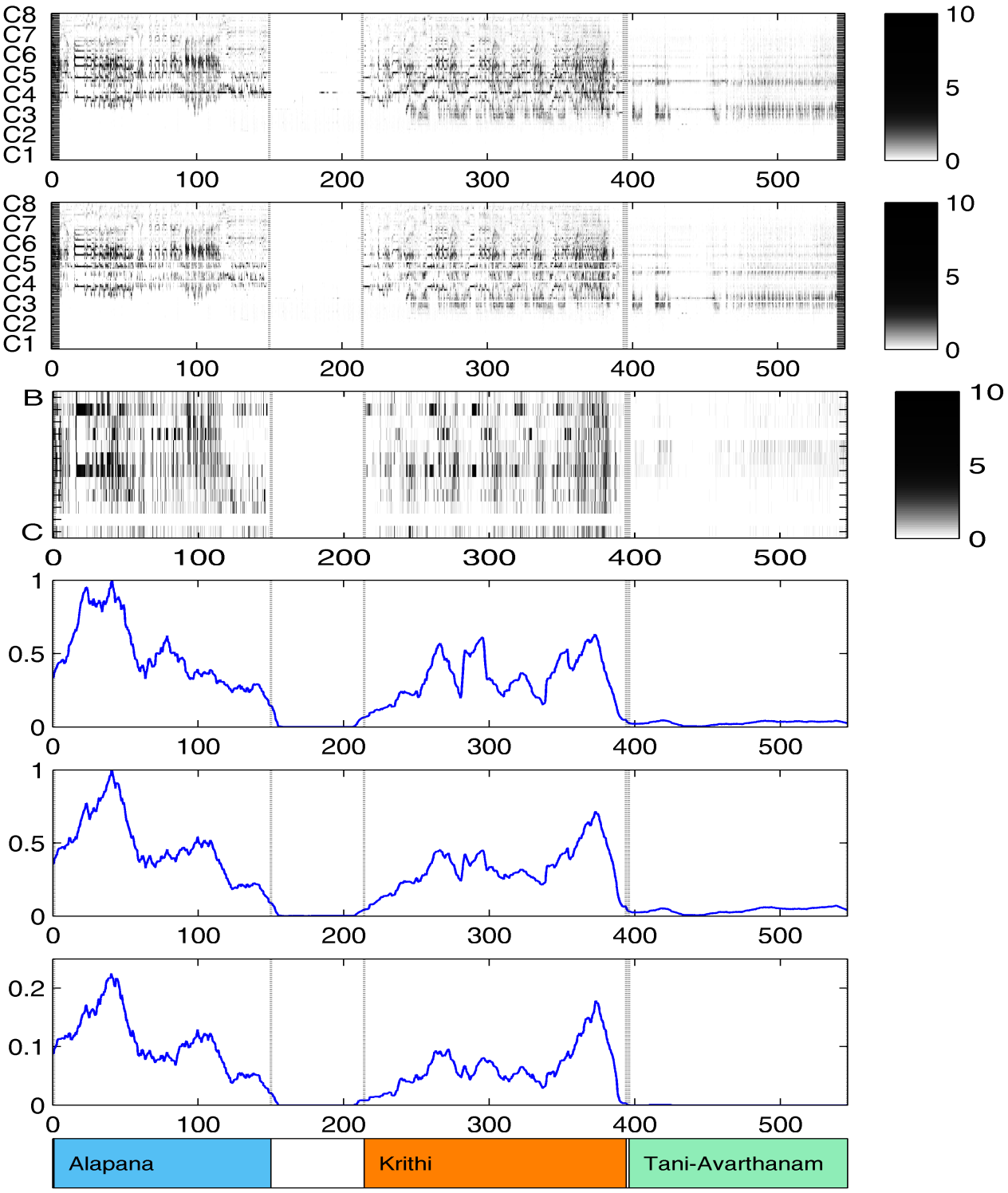

Figure C.10: $Wavfile$: $Raga\_05\_excerpt\_s\_212.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}_C}$, $f_{\lambda}^{\mathcal{S}_C}$ and $f_{\lambda}^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

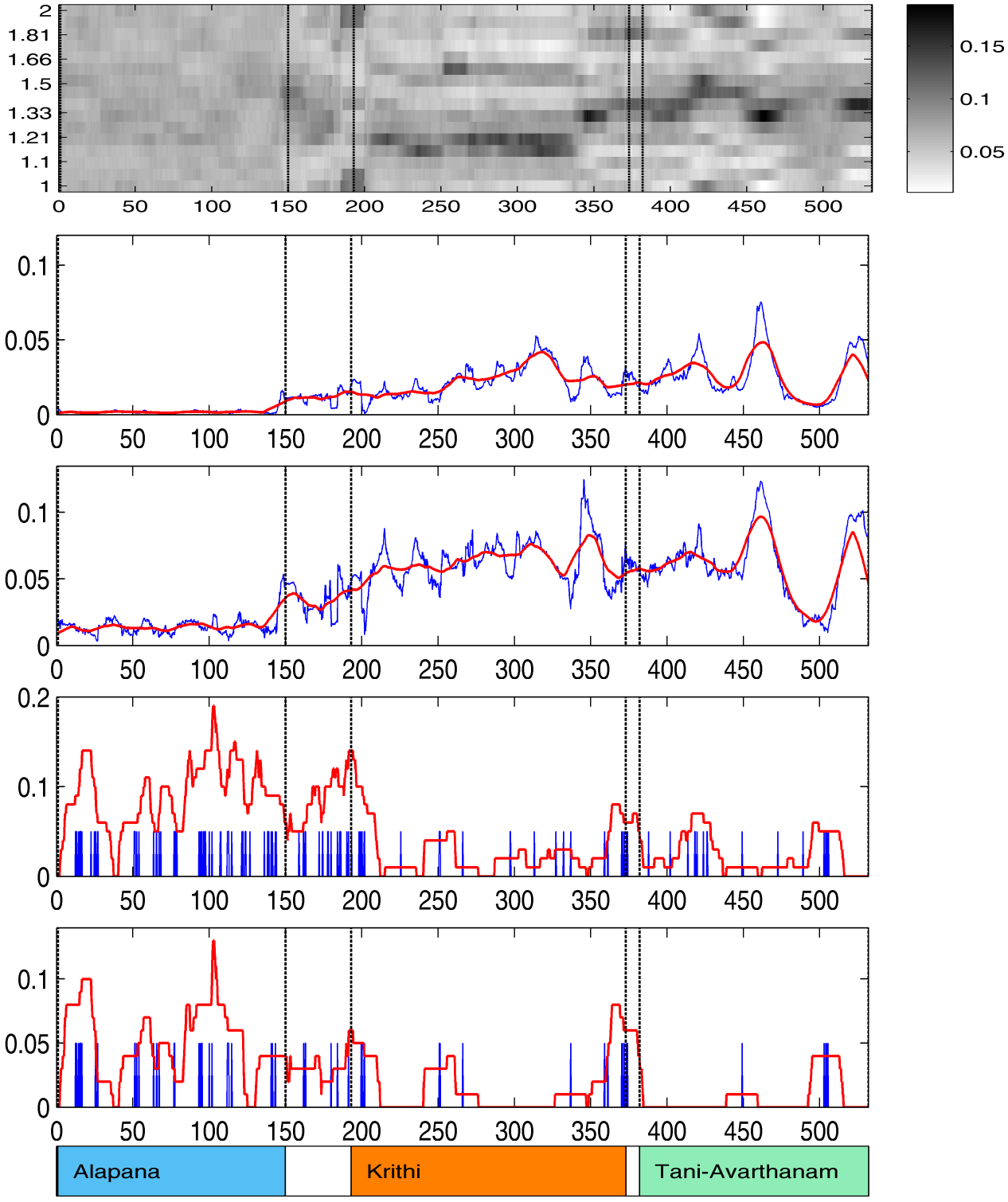

Figure C.11: $Wavfile$: $Raga\_06\_excerpt\_s\_164.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

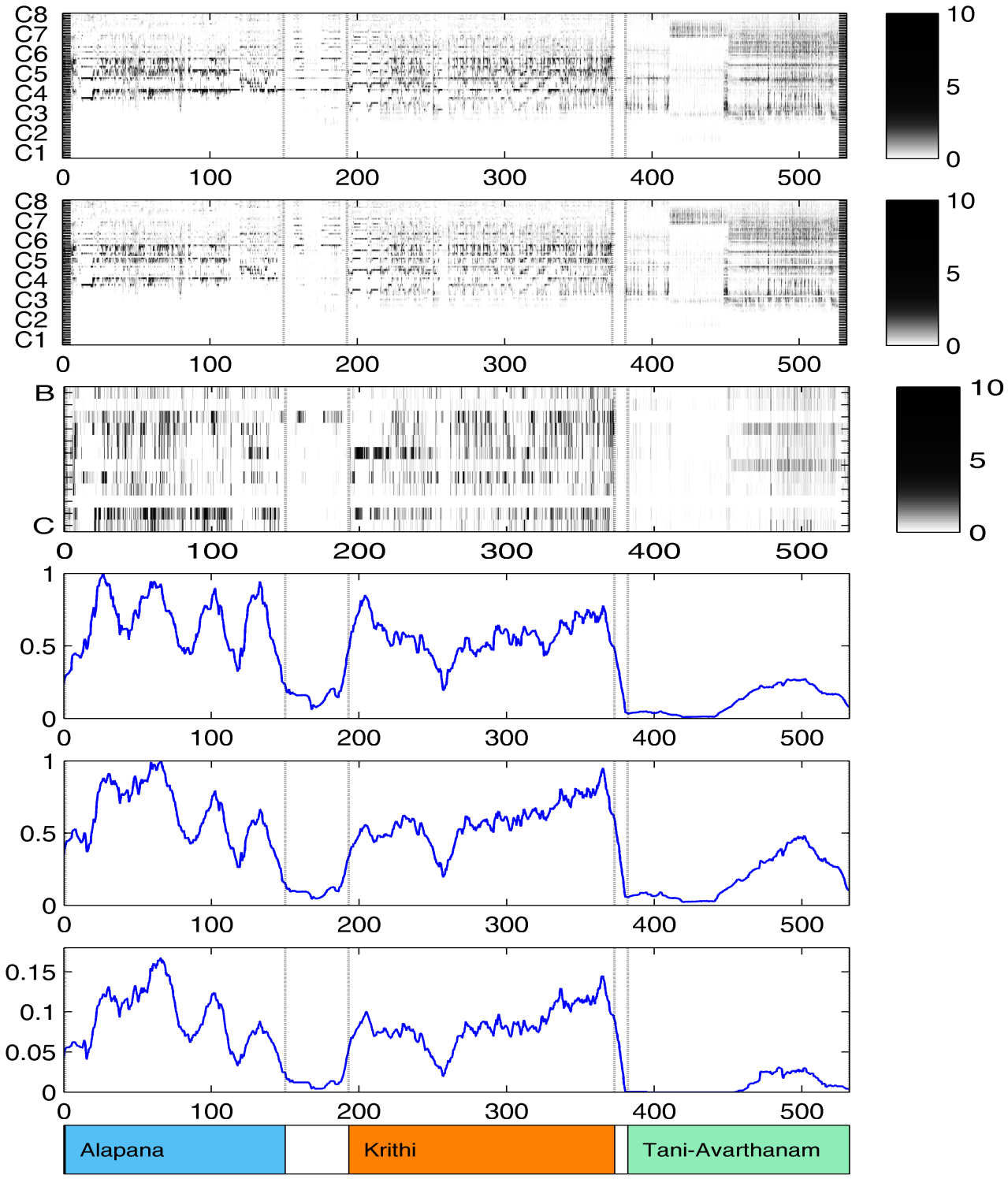

Figure C.12: $Wavfile$: *Raga_06_excerpt_s_164.wav*: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}_C}$, $f_{\lambda}^{\mathcal{S}_C}$ and $f_{\lambda}^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

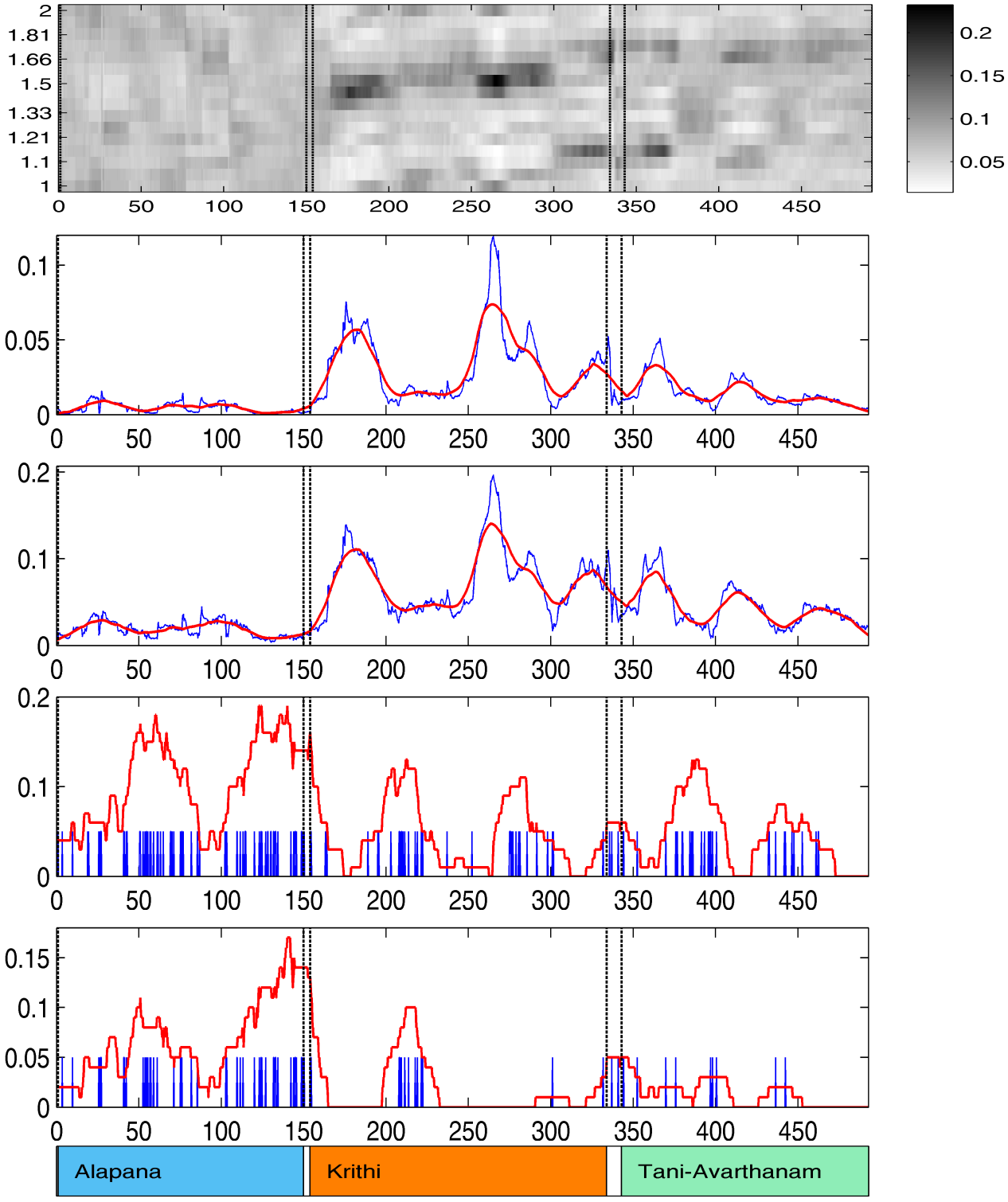

Figure C.13: $Wavfile$: $Raga\_07\_excerpt\_s\_248.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

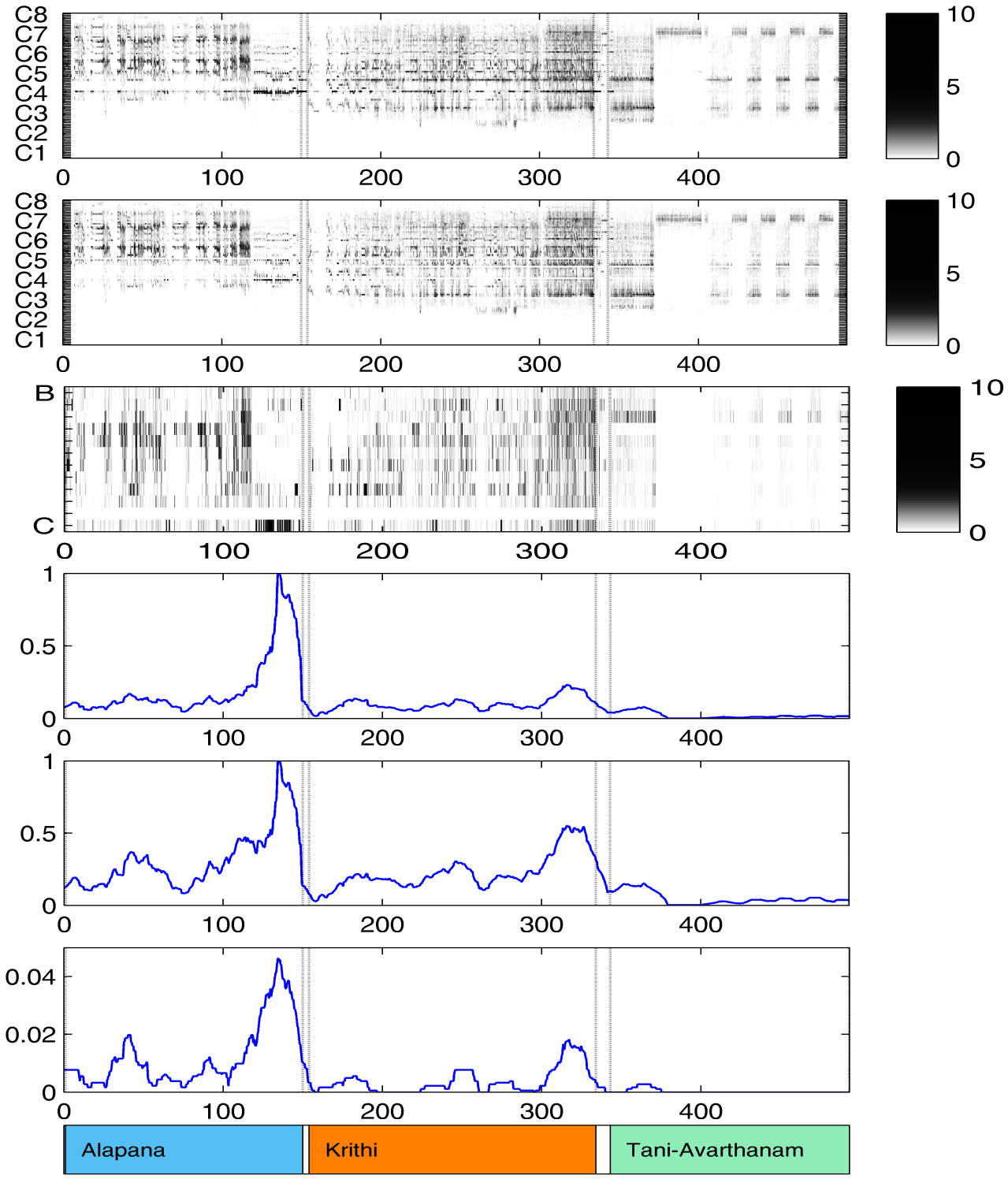

Figure C.14: $Wavfile$: *Raga_07_excerpt_s_248.wav*: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}_c}$, $f_{\lambda}^{\mathcal{S}_c}$ and $f_{\lambda}^{\mathcal{R}_c}$. The same parameter setting as in Figure 5.2 are used.

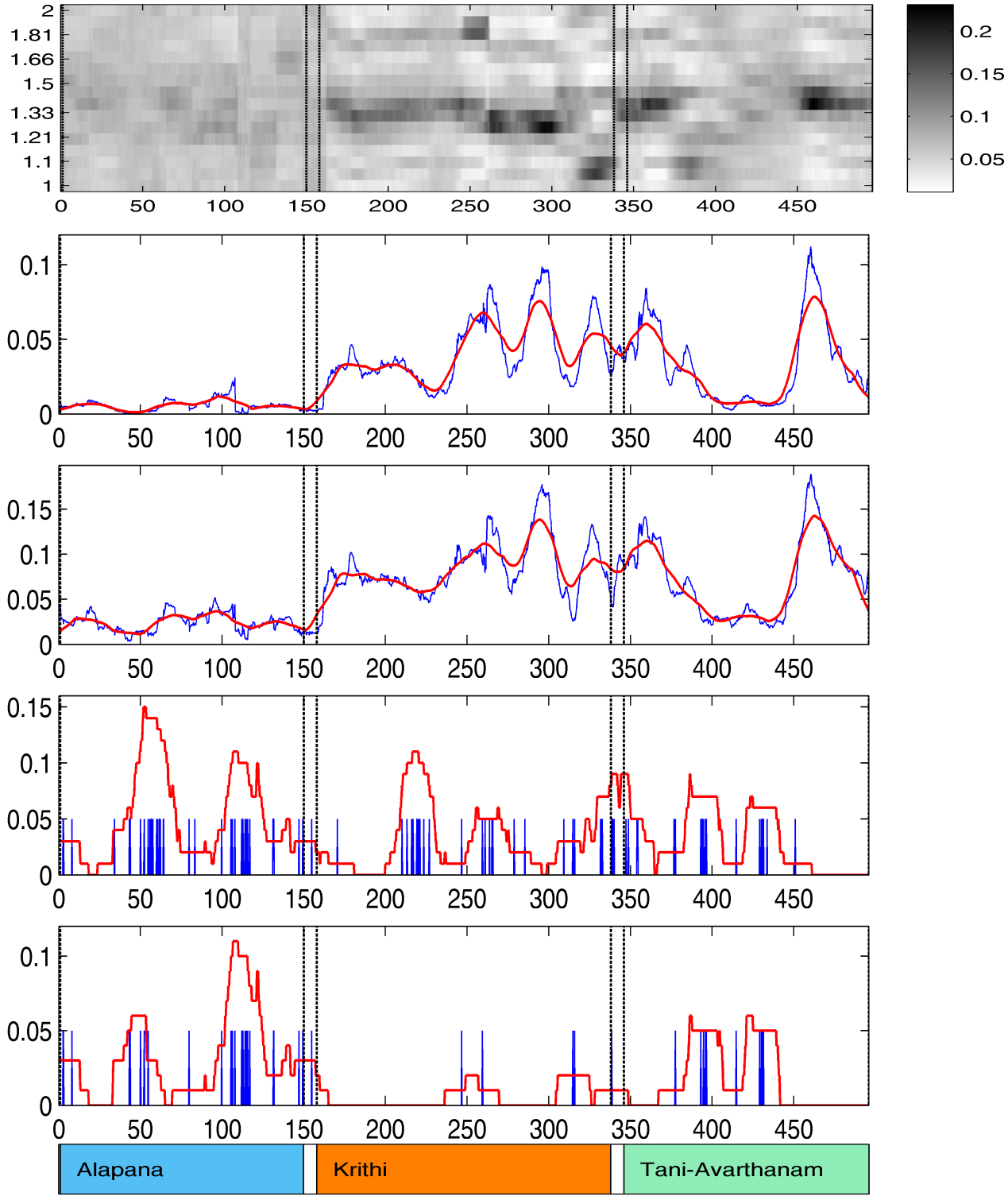

Figure C.15: $Wavfile$: $Raga\_08\_excerpt\_s\_260.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

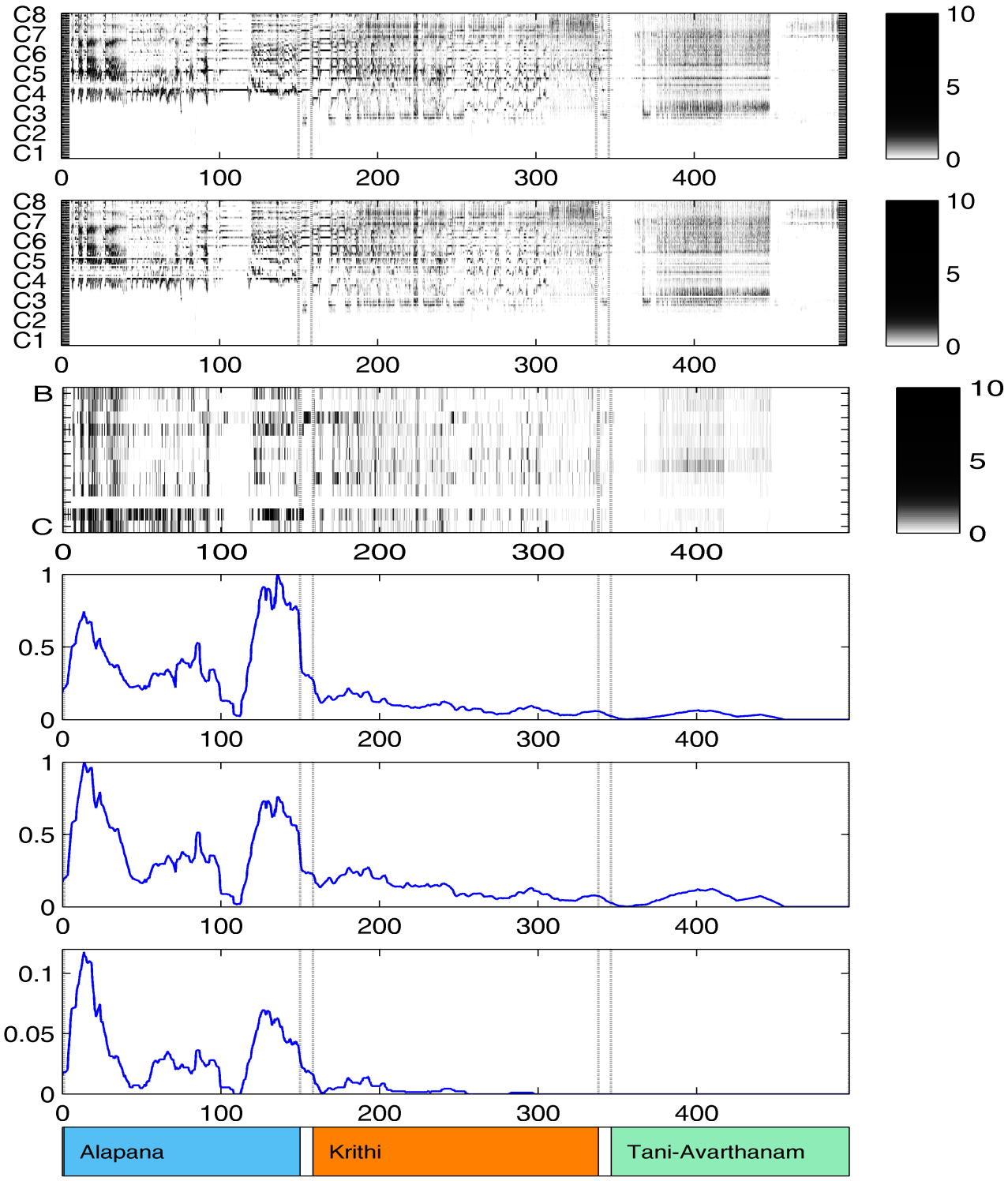

Figure C.16: $Wavfile$: $Raga\_08\_excerpt\_s\_260.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_\lambda^{\mathcal{M}_c}$, $f_\lambda^{\mathcal{S}_c}$ and $f_\lambda^{\mathcal{R}_c}$. The same parameter setting as in Figure 5.2 are used.

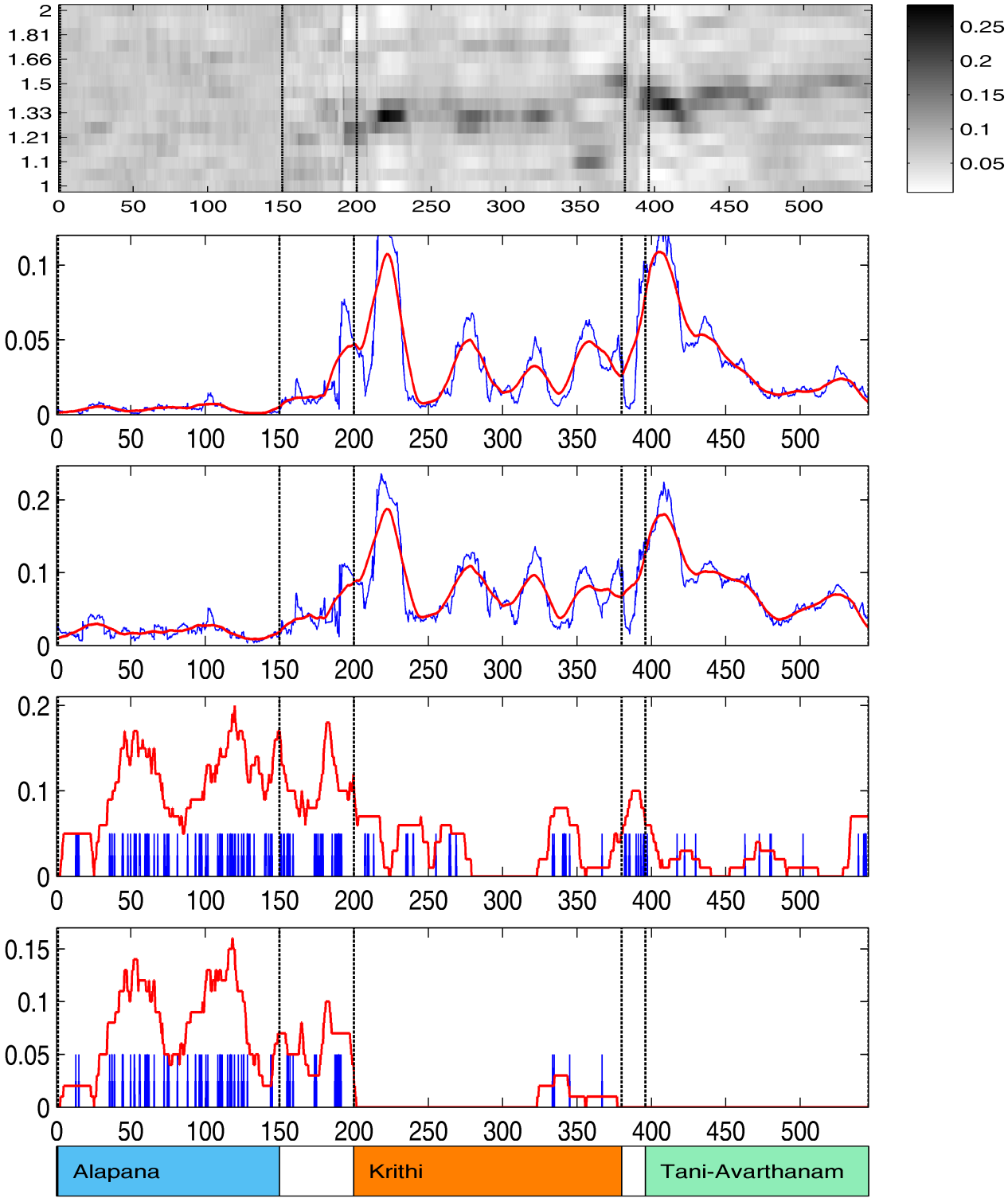

Figure C.17: $Wavfile$: $Raga\_09\_excerpt\_s\_224.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

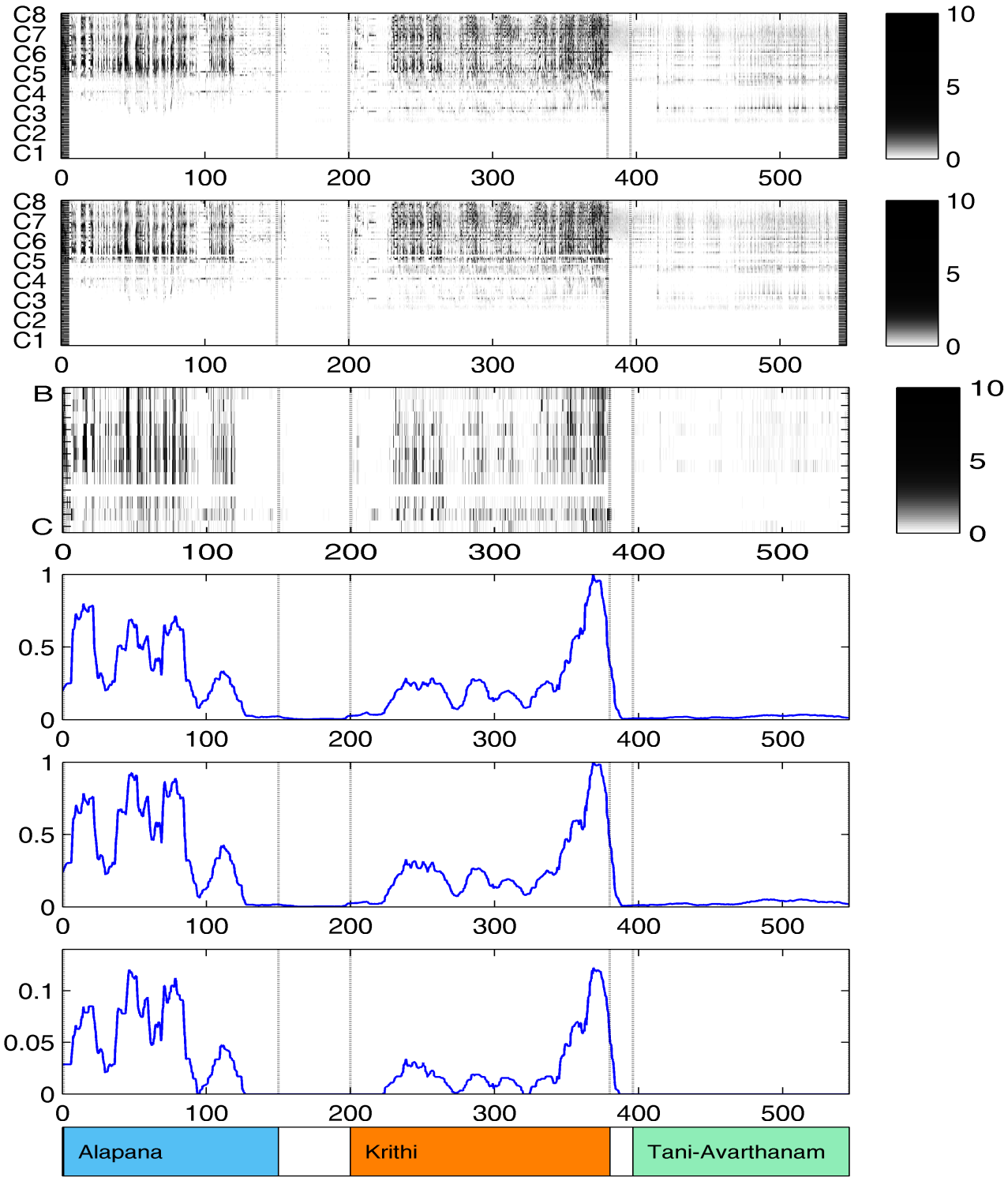

Figure C.18: $Wavfile$: $Raga\_09\_excerpt\_s\_224.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}_C}$, $f_{\lambda}^{\mathcal{S}_C}$ and $f_{\lambda}^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

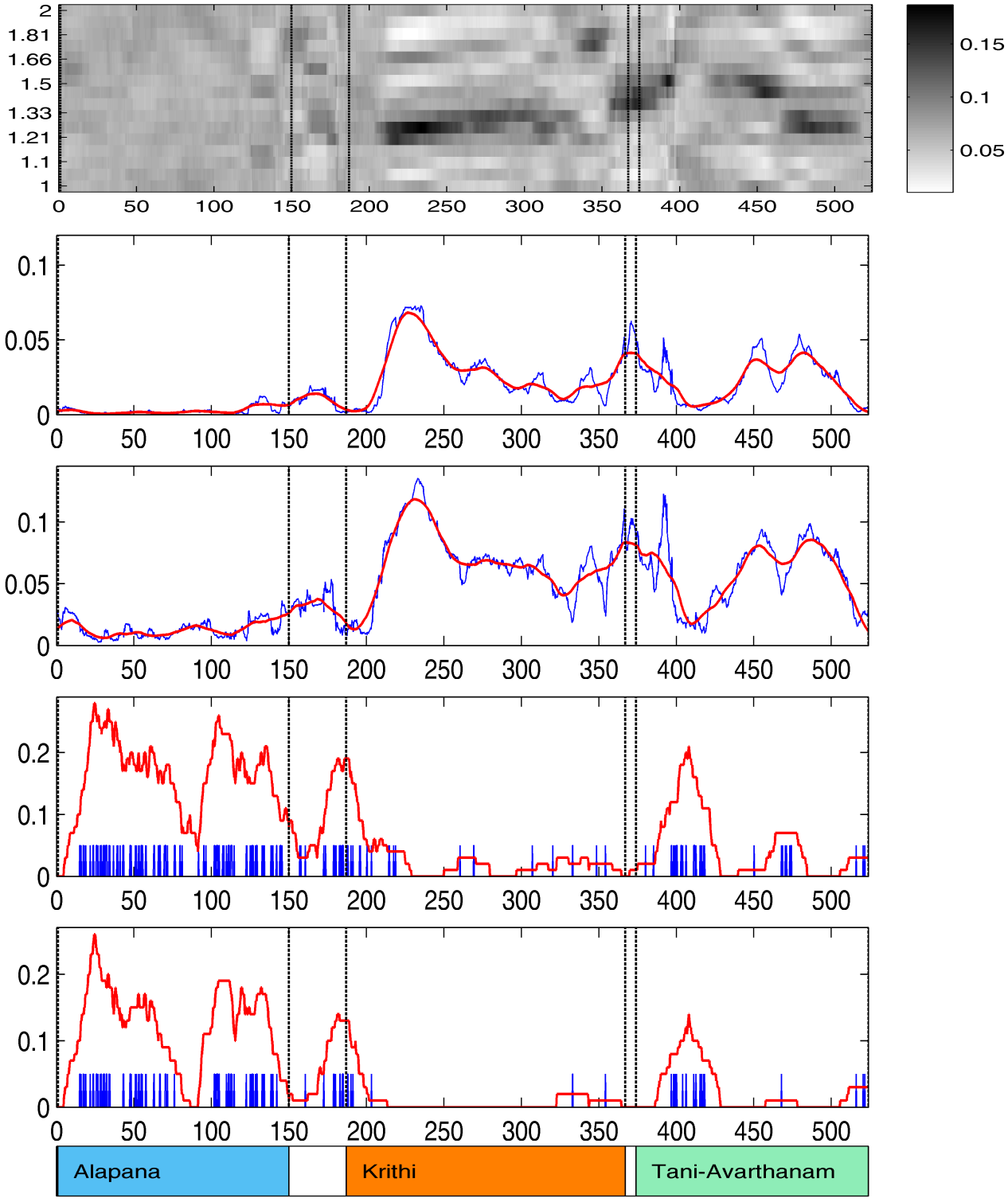

Figure C.19: $Wavfile$: $Raga\_10\_excerpt\_s\_176.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

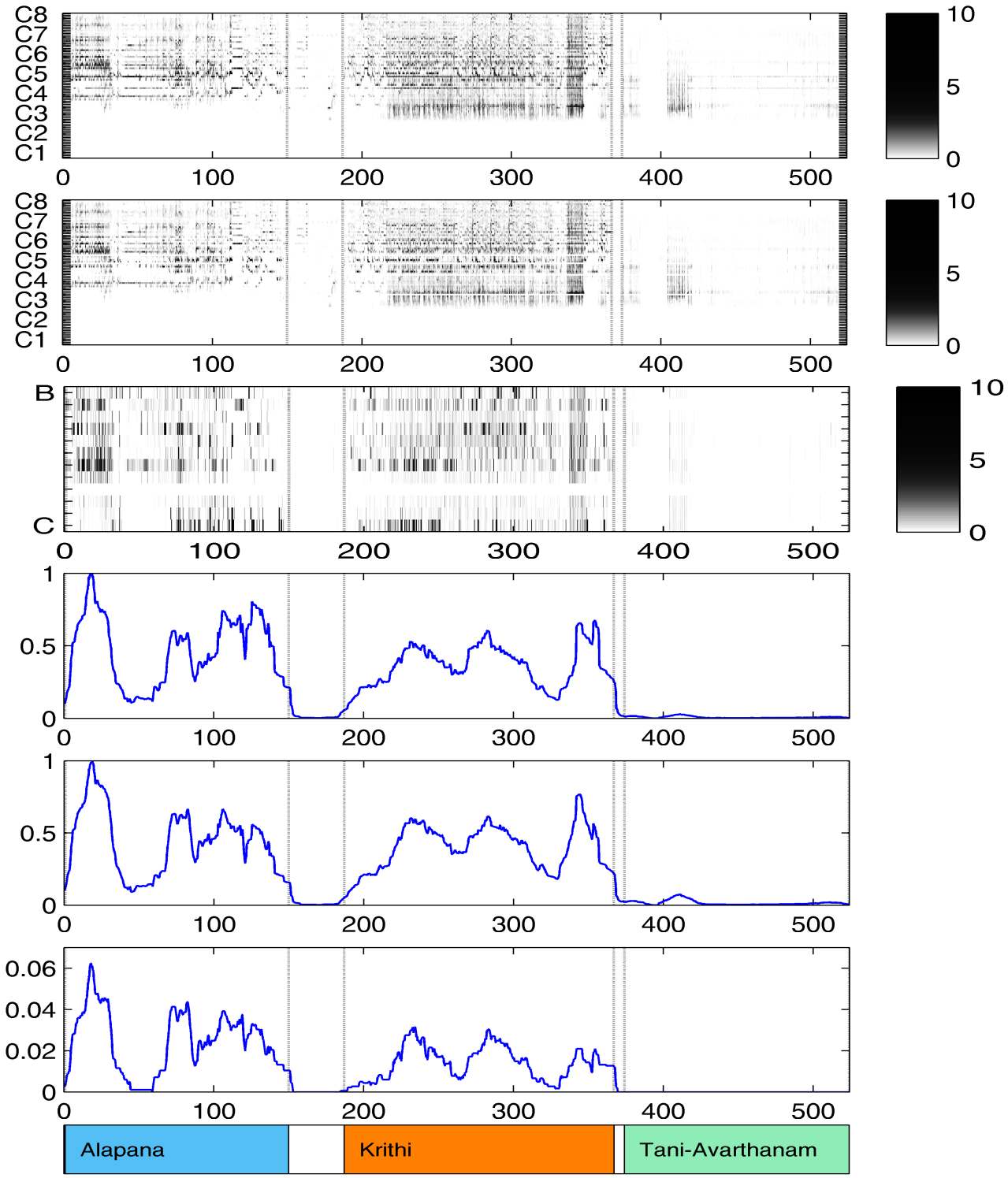

Figure C.20: $Wavfile$: $Raga\_10\_excerpt\_s\_176.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_\lambda^{\mathcal{M}_C}$, $f_\lambda^{\mathcal{S}_C}$ and $f_\lambda^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

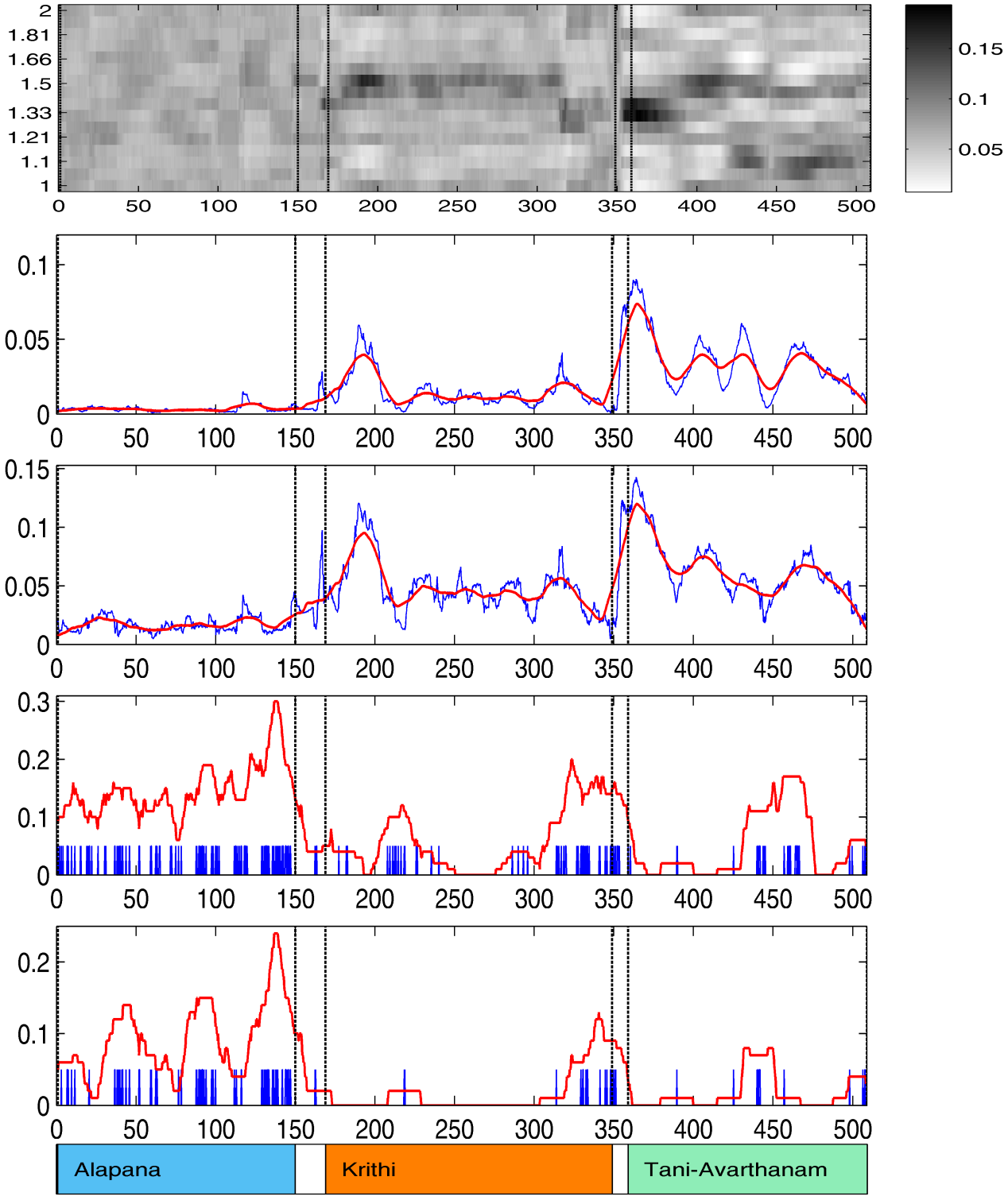

Figure C.21: $Wavfile$: $Raga\_11\_excerpt\_s\_200.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

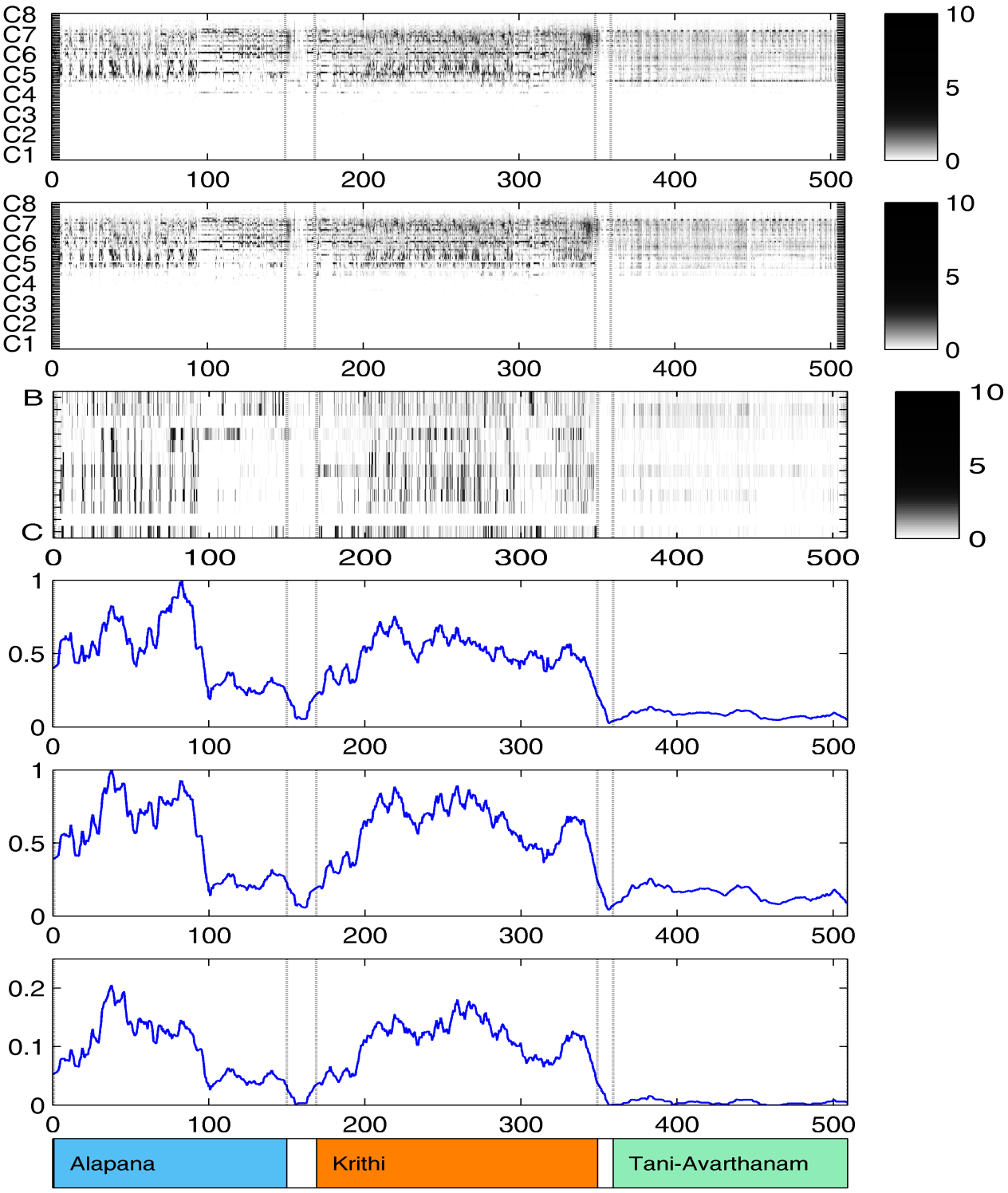

Figure C.22: $Wavfile$: $Raga\_11\_excerpt\_s\_200.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}_C}$, $f_{\lambda}^{\mathcal{S}_C}$ and $f_{\lambda}^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

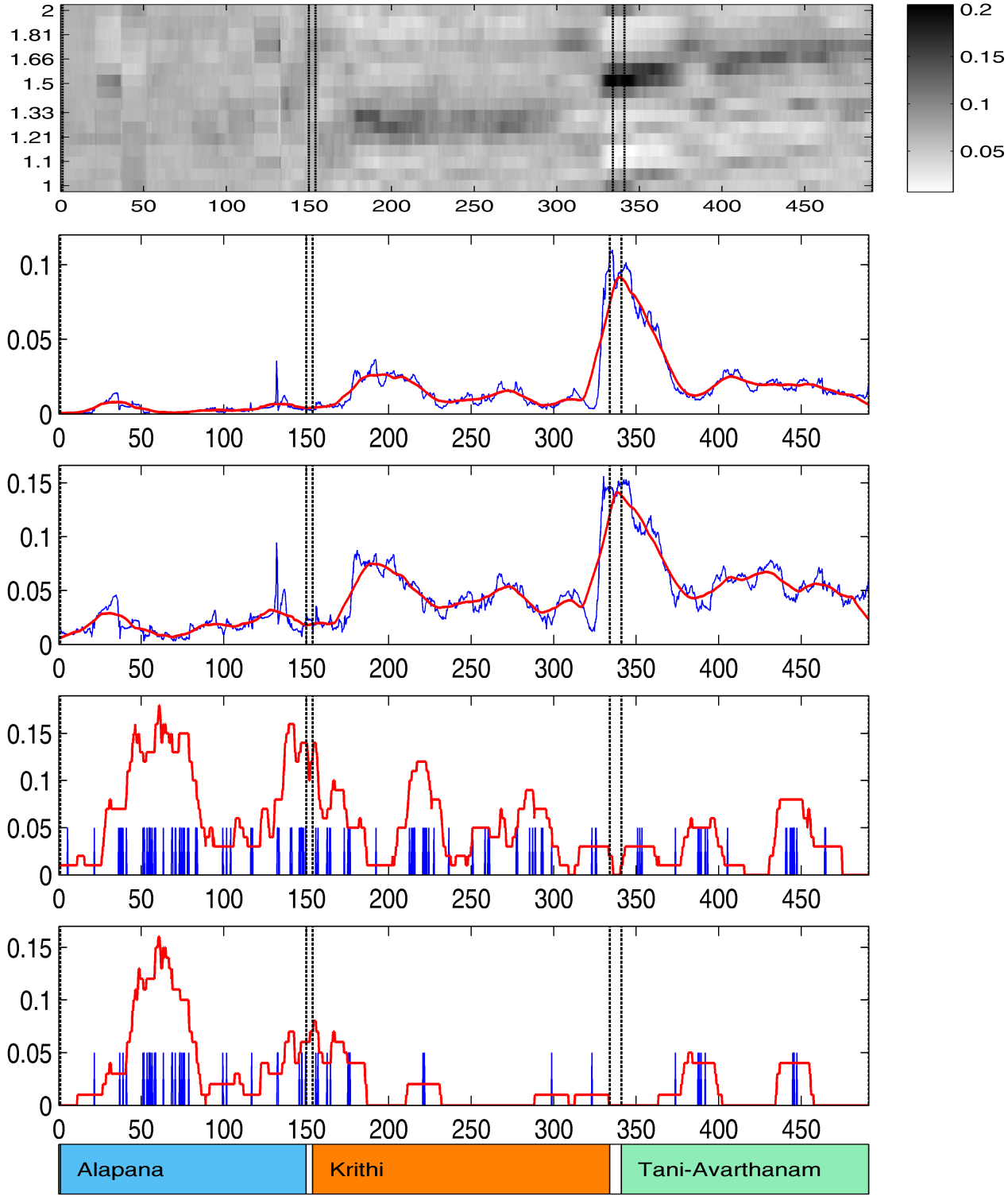

Figure C.23: $Wavfile$: $Raga\_12\_excerpt\_s\_152.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

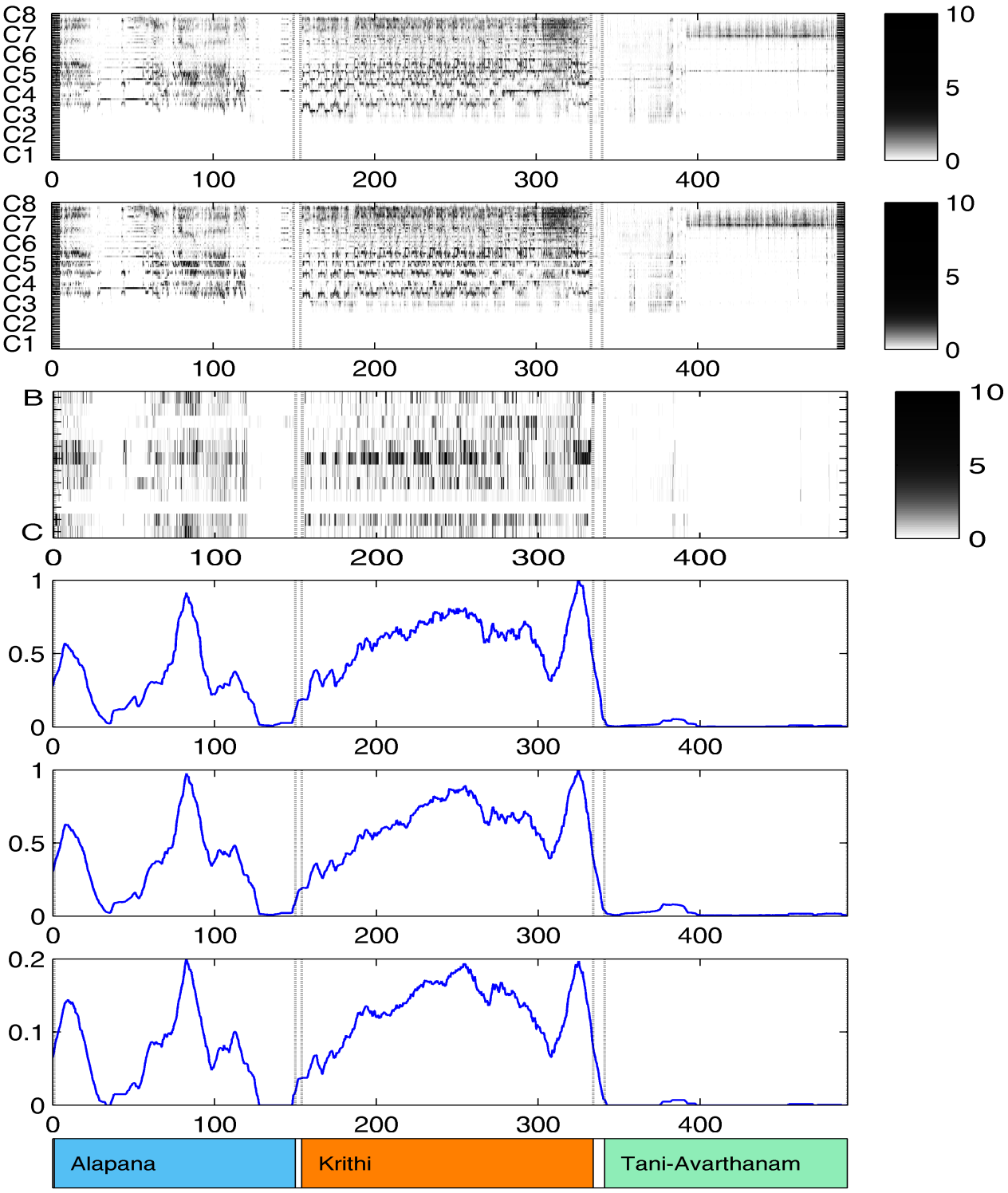

Figure C.24: *Wavfile* : *Raga_12_excerpt_s_152.wav*: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_\lambda^{\mathcal{M}_c}$, $f_\lambda^{\mathcal{S}_c}$ and $f_\lambda^{\mathcal{R}_c}$. The same parameter setting as in Figure 5.2 are used.

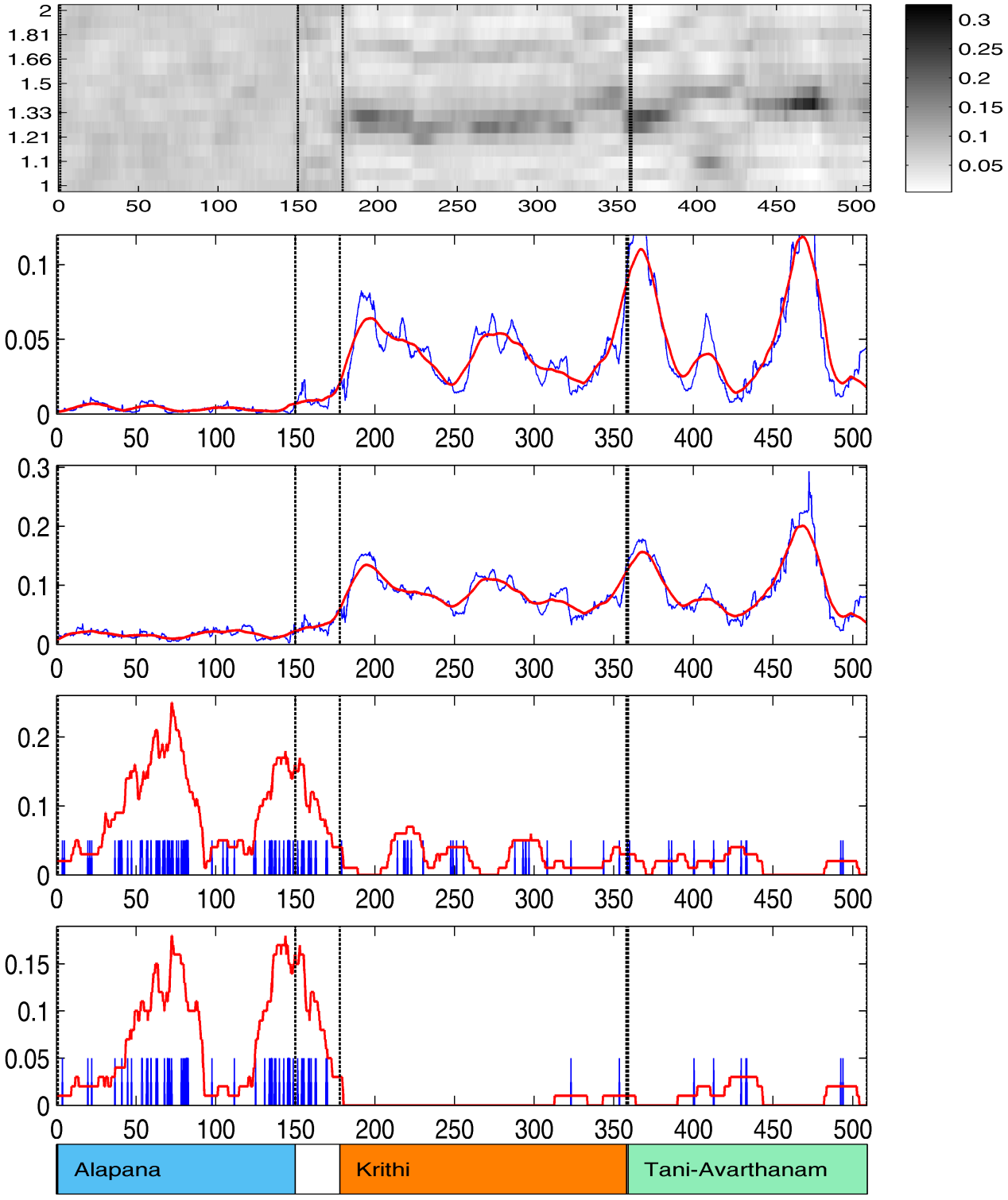

Figure C.25: $Wavfile$: $Raga\_13\_excerpt\_s\_152.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

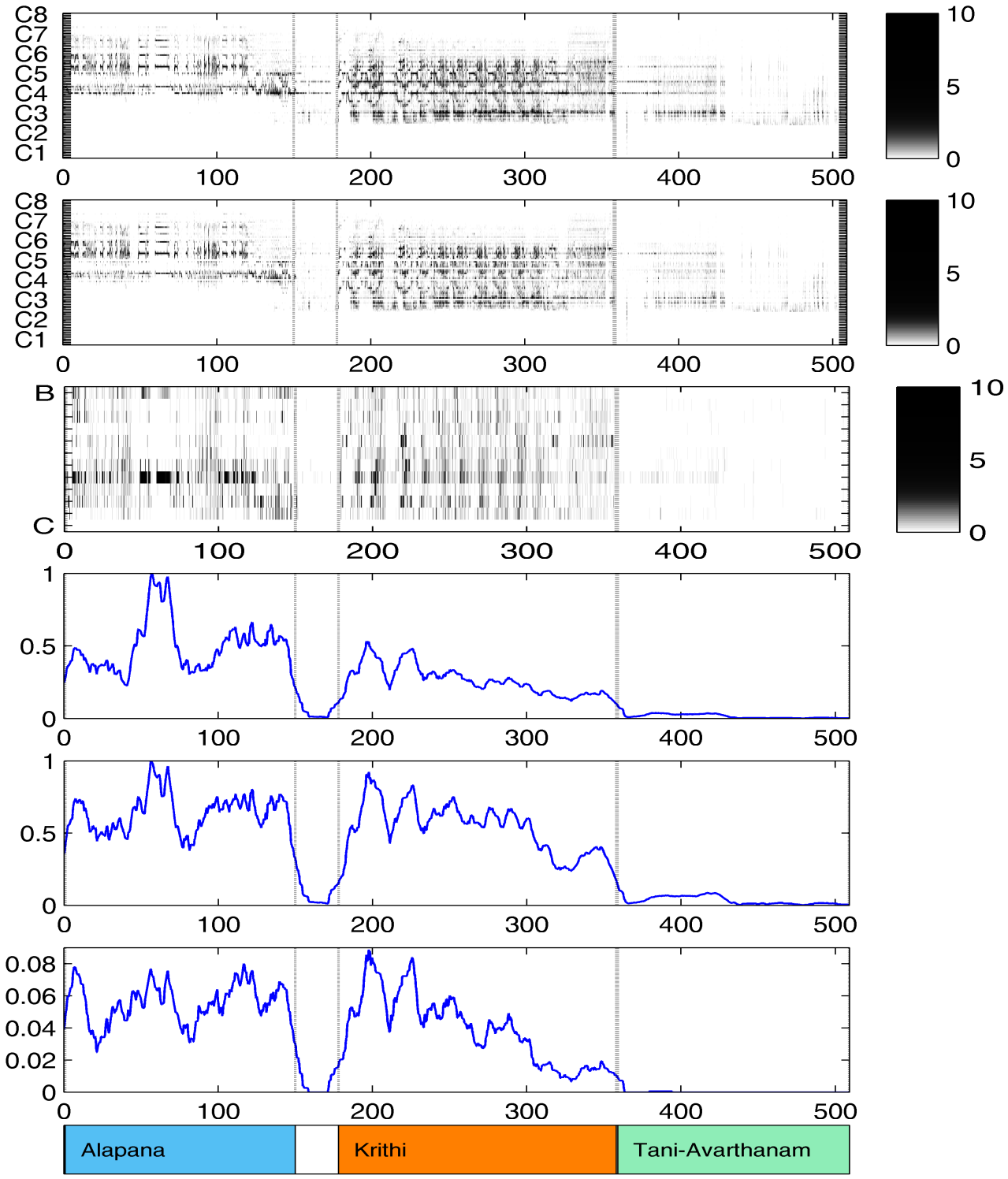

Figure C.26: $Wavfile$: $Raga\_13\_excerpt\_s\_152.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_\lambda^{\mathcal{M}c}$, $f_\lambda^{\mathcal{S}c}$ and $f_\lambda^{\mathcal{R}c}$. The same parameter setting as in Figure 5.2 are used.

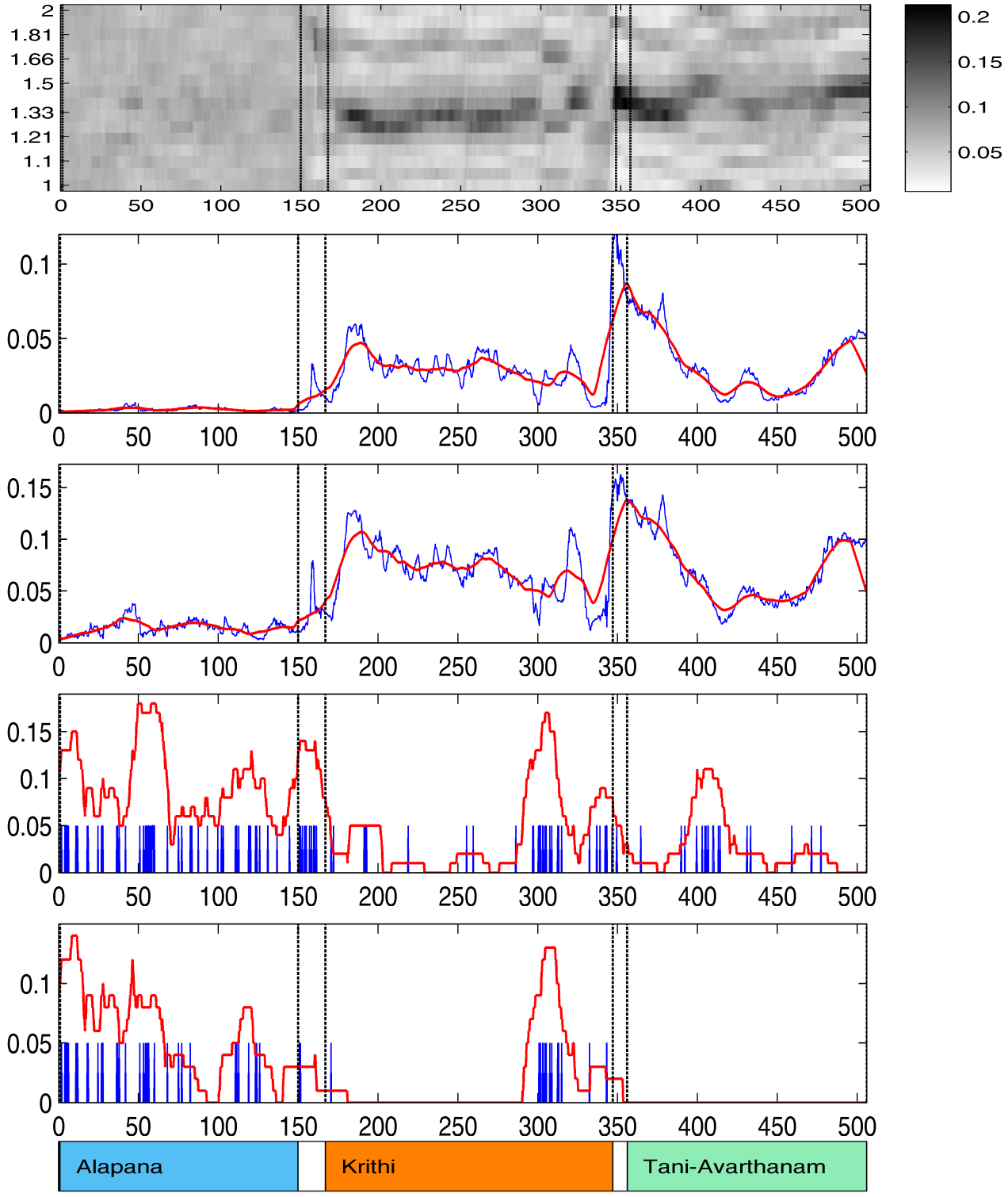

Figure C.27: $Wavfile$: $Raga\_14\_excerpt\_s\_224.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

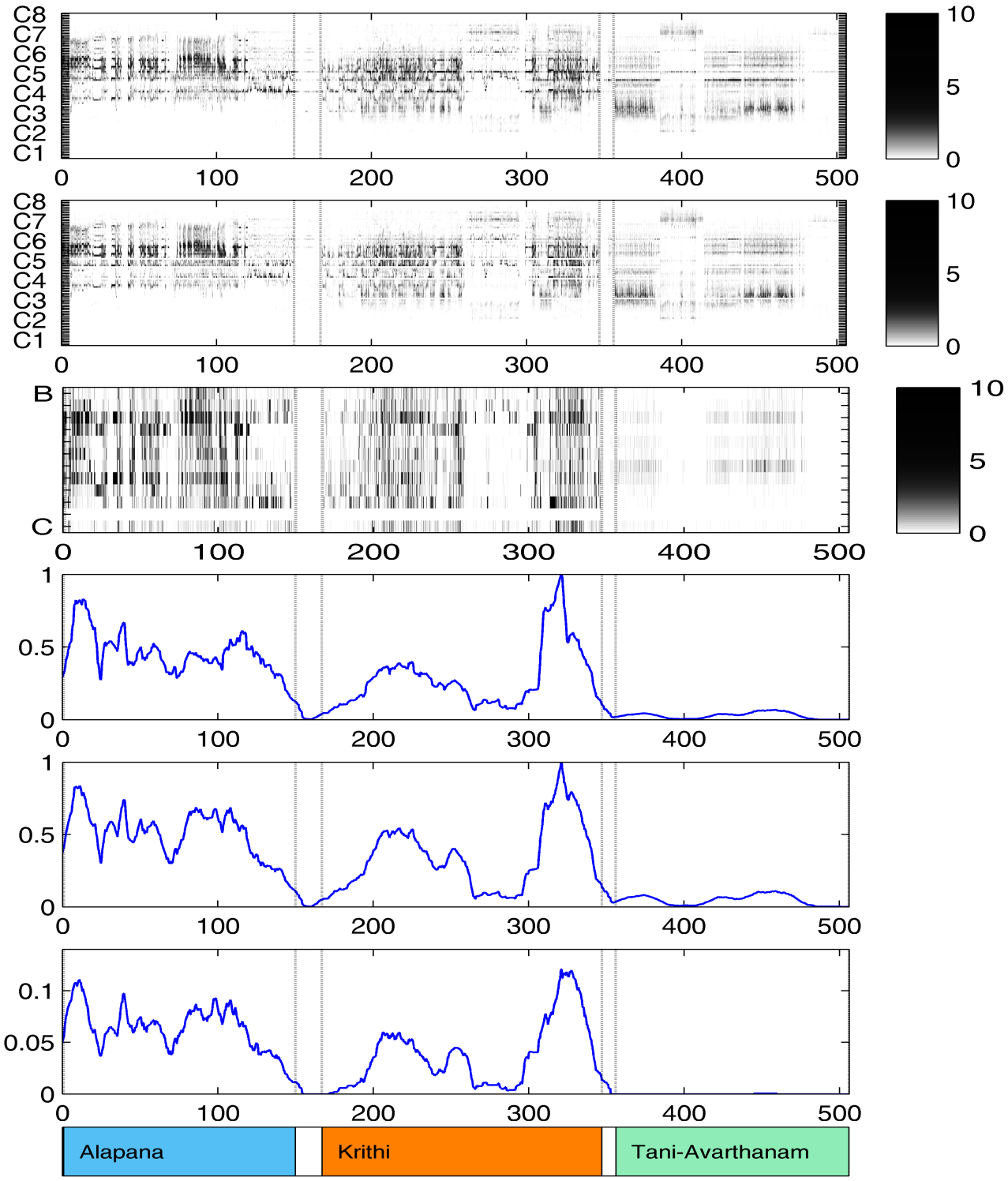

Figure C.28: $Wavfile$: $Raga\_14\_excerpt\_s\_224.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}_C}$, $f_{\lambda}^{\mathcal{S}_C}$ and $f_{\lambda}^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

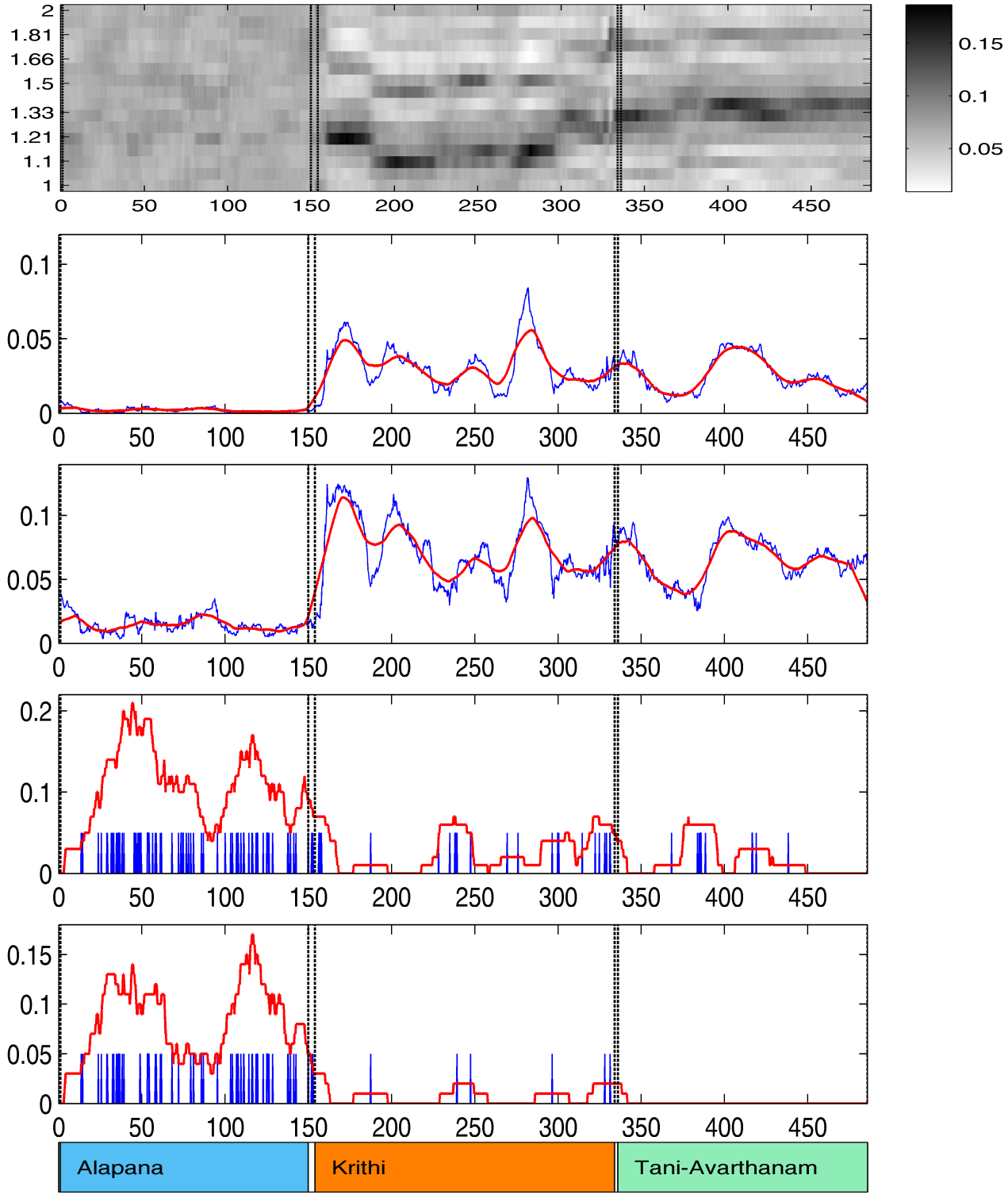

Figure C.29: $Wavfile$: $Raga\_15\_excerpt\_s\_224.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the normalized cyclic tempogram representation as well as the salience features $f_{\lambda}^{\mathcal{H}}$, $f_{\lambda}^{\mathcal{M}}$, $f_{0,\lambda}^{\mathcal{I}}$, and $f_{1,\lambda}^{\mathcal{I}}$. The same parameter setting as in Figure 5.1 are used.

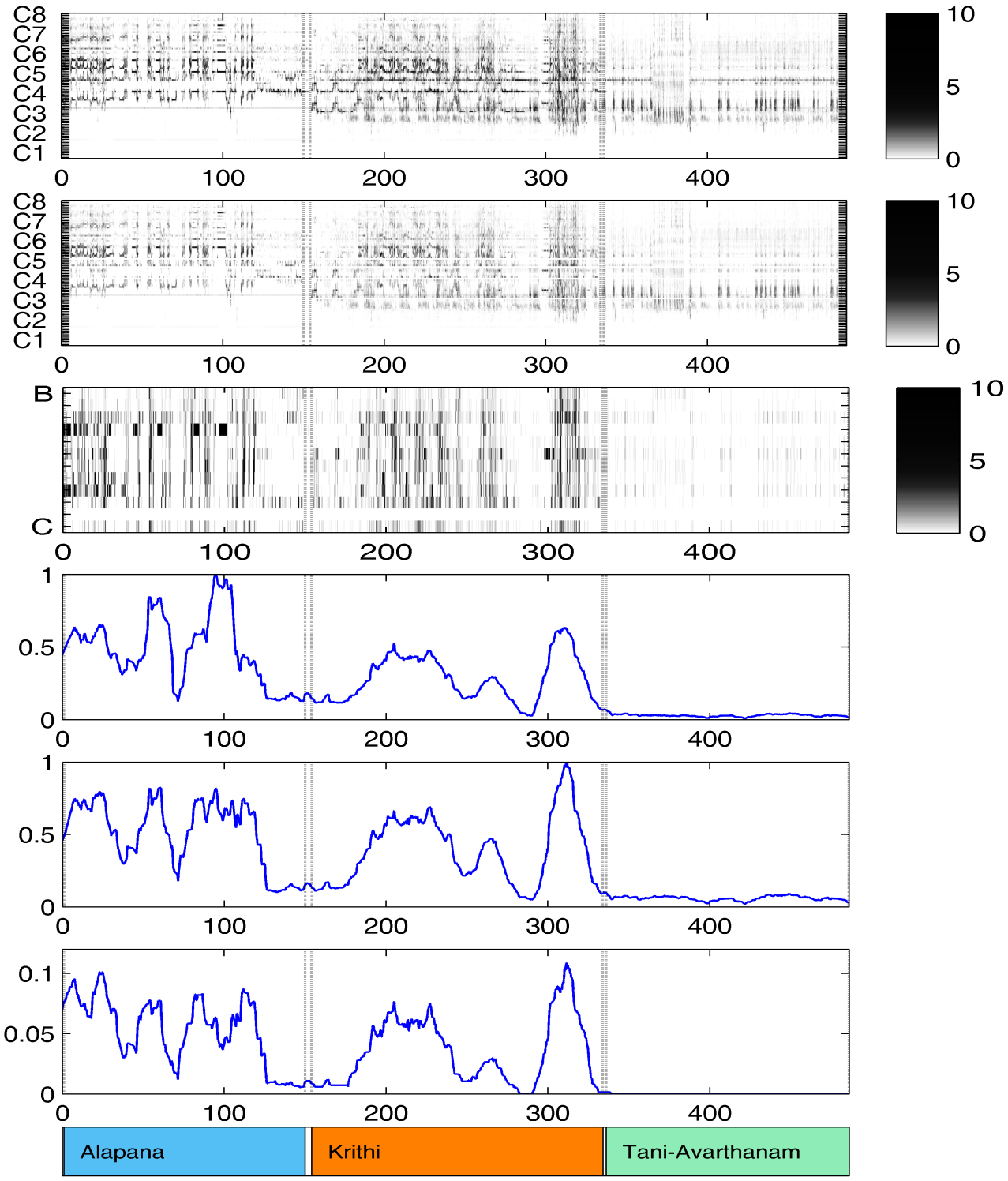

Figure C.30: $Wavfile$: $Raga\_15\_excerpt\_s\_224.wav$: Representation of a Carnatic music recordings and the resulting feature representations. The figure shows the midi pitch (with and without drone), chroma representation as well as the salience features $f_{\lambda}^{\mathcal{M}_C}$, $f_{\lambda}^{\mathcal{S}_C}$ and $f_{\lambda}^{\mathcal{R}_C}$. The same parameter setting as in Figure 5.2 are used.

Bibliography

[1] A carnatic music primer. http://www.ae.iitm.ac.in/~sriram/karpri.html.

[2] Drone in indian classical music. http://chandrakantha.com/articles/indian_music/drone.html.

[3] History of carnatic music. http://carnatica.net/origin.htm.

[4] The history of hindustani classical music. http://en.wikipedia.org/wiki/Hindustani_classical_music.

[5] Introduction to carnatic music. http://en.wikipedia.org/wiki/Carnatic_music.

[6] Sangeetanubhava concert structure. http://carnatica.net/sangeet/concertpresentation.htm.

[7] Swaras in indian classical music. http://en.wikipedia.org/wiki/Swara.

[8] Miguel Alonso, Bertrand David, and Gaël Richard. Tempo and beat estimation of musical signals. In *In Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2007.

[9] A.S.Krishna, P.V.Rajkumar, K.P.Saishankar, and M.John. Identification of carnatic raagas using hidden markov models. In *Proceedings of IEEE Applied Machine Intelligence and Informatics (SAMI).*, pages 107 – 110, Smolenice,Slovakia, Jan 27-29, 2011.

[10] Mark A. Bartsch and Gregory H. Wakefield. Audio thumbnailing of popular music using chroma-based representations. In *IEEE Transactions on Multimedia*, page 7(1):96104, February 2005.

[11] Juan P. Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. A tutorial on onset detection in music signals. In *IEEE Trans. Speech and Audio Processing*, pages vol. 13, no. 5, pp.10351047, 2005.

[12] Daniel P. W. Ellis. Beat tracking by dynamic programming. In *J. New Music Research, Special Issue on Beat and Tempo Extraction*, pages 51 – 60, March 2007.

[13] Emilia Gômez. *Tonal Description of Music Audio Signals.* PhD thesis, 2006.

[14] P. Grosche and Meinard Müller. Time variable tempo detection and beat marking. In *In Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, 2005.

[15] P. Grosche and Meinard Müller. Computing predominant local periodicity information in music recordings. In *in Proc. IEEE WASPAA*, New Paltz, New York, USA, 2009.

[16] P. Grosche and Meinard Müller. Tempogram toolbox: Matlab implementations for tempo and pulse analysis of music recordings. In *In Late-Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, FL, USA, 2011.

[17] P. Grosche, Meinard Müller, and F. Kurth. Cyclic tempogram a mid-level tempo representation for music signals. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5522 – 5525, Dallas, Texas, USA, Mar. 2010.

[18] Peter M. Grosche. *Signal Processing Methods for Beat Tracking, Music Segmentation, and Audio Retrieval*. PhD thesis, 2012.

[19] Nanzhu Jiang. An analysis of automatic chord recognition procedures for music recordings. Master's thesis, , 2011.

[20] Anssi P. Klapuri, Antti J. Eronen, and Jaakko Astola. Analysis of the meter of acoustic musical signals. In *IEEE Trans. Audio, Speech, and Language Processing, vol. 14, no. 1,*, pages 342–355, Victoria, Canada, 2006.

[21] Frank Kurth and Thorsten Gehrmann and Meinard Müller. The cyclic beat spectrum: Tempo-related audio features for time-scale invariant audio identification. In *in Proc. ISMIR*, pages 35 – 40, Victoria, Canada, 2006.

[22] Frank Kurth, Thorsten Gehrmann, and Meinard Müller. The cyclic beat spectrum: Tempo-related audio features for time-scale invariant audio identification. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, pages 35–40, Victoria, Canada, October 2006.

[23] Ramesh Mahadevan. A ggentle introduction to south indian classical music (1-4). `http://www.shivkumar.org/music/basics/ramesh/gentle-intro-ramesh-mahadevan-I.pdf`.

[24] Meinard Müller. *Information Retrieval for Music and Motion*. Springer Verlag, 2007.

[25] Meinard Müller and Sebastian Ewert. Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*, pages 215–220, Miami, USA, 2011.

[26] R.Sudha, A.Kathirvel, and R.M.D.Sundaram. System of tool for identifying ragas using midi. In *Proceedings of IEEE Computer and Electrical Engineering, ICCEE.*, pages 644 – 647, Dubai, Dec 28-30, 2009.

[27] R.Venugopalan and T.R.Prashanth. Note identification in carnatic music from frequency spectrum. In *Proceedings of IEEE Communications and Signal Processing (ICCSP),*, pages 87 – 91, Calicut, Feb 10-12, 2011.

[28] Padi Sarala and Hema A.Murthy. Inter and intra item segmentation of continuous audio recordings of carnatic music for archival. In *Proceedings of ISMIR International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kobe, Japan, Sep 2013.

[29] Padi Sarala, Vignesh Ishwar, Ashwin Bellur, and Hema A.Murthy. Applause identification and its relevance to archival of carnatic music. In *Proceedings of Second CompMusic Workshop*, Istanbul, Turkey, July 12-13, 2012.

[30] Roger N. Shepard. Circularity in judgments of relative pitch. In *Journal of the Acoustic Society of America*, page 36(12), 1964.

[31] Rajeswari Sridhar, Karthiga S, and Geetha T V. Fundamental frequency estimation of carnatic music songs based on the principle of mutation. In *IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 7*, July 2010.

[32] Ruohua Zhou, Marco Mattavelli, and Giorgio Zoia. Music onset detection based on resonator time frequency image. In *IEEE Trans. Audio, Speech, and Language Processing*, pages vol. 16, no. 8, pp. 16851695, Dallas, Texas, USA, 2008.