NOVEL AUDIO FEATURES FOR CAPTURING TEMPO SALIENCE IN MUSIC RECORDINGS

Balaji Thoshkahna[2], *Meinard Müller*[1], *Venkatesh Kulkarni*[1,2], *Nanzhu Jiang*[1]

[1]International Audio Laboratories Erlangen
[2]Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany
balaji.thoshkahna@iis.fraunhofer.de, meinard.mueller@audiolabs-erlangen.de

ABSTRACT

In music compositions, certain parts may be played in an improvisational style with a rather vague notion of tempo, while other parts are characterized by having a clearly perceivable tempo. Based on this observation, we introduce in this paper some novel audio features for capturing tempo-related information. Rather than measuring the specific tempo of a local section of a given recording, our objective is to capture the existence or absence of a notion of tempo, a kind of tempo salience. By a quantitative analysis within an Indian music scenario, we demonstrate that our audio features capture the aspect of tempo salience well, while being independent of continuous fluctuations and local changes in tempo.

Index Terms— Audio, Music, Tempo, Salience, Analysis, Segmentation, Classification

1. INTRODUCTION

Tempo and beat are fundamental aspects of music, and the automated extraction of such information from audio recordings constitutes one of the central and well-studied research areas in music signal processing [1, 2, 3, 4, 5, 6, 7, 8, 9]. When assuming a steady beat and a single global tempo, many automated methods yield accurate tempo estimates and beat tracking results [10, 11]. However, the task becomes much more difficult when one deals with music with weak note onsets and local tempo changes [6]. A discussion of difficult examples for beat tracking can also be found in [12, 11]. Instead of extracting tempo and beat information explicitly, various spectrogram-like representations have been proposed for visualizing tempo-related information over time. Such mid-level representations include tempograms [13, 14, 15], rhythmograms [16], or beat spectrograms [4, 17]. Cyclic versions of time-tempo representations, which possess a high degree of robustness to pulse level switches, have been introduced in [15, 17].

For certain types of music, however, there is hardly a notion of tempo or beat. For example, this is often the case for music that is played in an improvisational style. Even within a single composition, there may be parts with a rather vague notion of tempo, while other parts are characterized by having a clearly perceivable tempo and rhythm. As an example, let us consider the song "In the year 2525" by Zager and Evans, which has the musical structure $IV_1V_2V_3V_4V_5V_6V_7BV_8O$, see Figure 1a. The song starts with a slow contemplative intro, which is represented by the I-part. The eight verse sections of the song, which are represented by the V-parts, have a clear rhythm with a well-defined tempo. Between the seventh verse V_7 and eighth verse V_8, the improvisational style of the beginning is resumed in the bridge part B—some kind of melancholic retrospect.

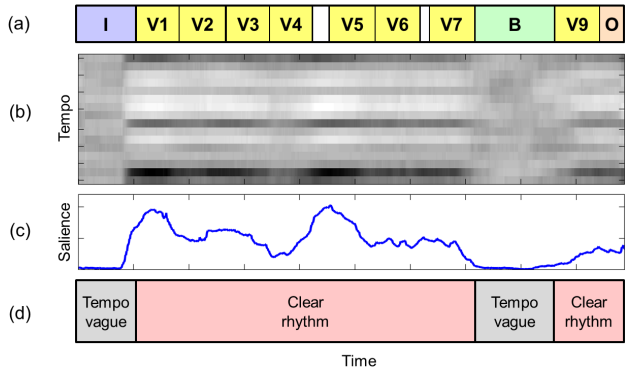

Fig. 1. **(a)** Musical structure of the song "In the year 2525" by Zager and Evans. **(b)** Tempogram representation of the recording. **(c)** Tempo salience feature. **(d)** Manual annotation of parts with a clear rhythm and parts with a vague tempo.

In this paper, we deal with an aspect of tempo which we refer to as *tempo salience*. Rather than measuring the concrete tempo of a local section of a given recording, our objective is to capture the existence or absence of a notion of tempo. For our computations we start with a mid-level representation known as cyclic tempogram [15]. As our main technical contribution, we then derive several novel one-dimensional audio features that locally measure the tempo salience, see Figure 1c for an example. Note that it is not our objective to determine the tempo itself. Instead the salience features should only express the degree to which there may be any sense of a perceivable tempo or not—irrespective of possible abrupt tempo changes or continuous tempo fluctuations.

The remainder of this paper is structured as follows. In Section 2, we further motivate our salience concept by considering a scenario from Indian Carnatic music. We review in Section 3 the concept of tempogram representations and then describe in Section 4 the technical details for deriving our novel salience features from these representations. Finally, in Section 5, we report on a some quantitative evaluations that indicate the potential of our salience features.

2. MOTIVATING APPLICATION SCENARIO

As a motivating scenario for our features, we consider a music genre that goes beyond Western music. Carnatic music plays an important role in the culture of South India, where huge music festivals with hundreds of concerts are held. Many of the large-scale compositions performed at such occasions consist of several contrasting parts [18]. A typical example for structural parts of a Canatic music composi-

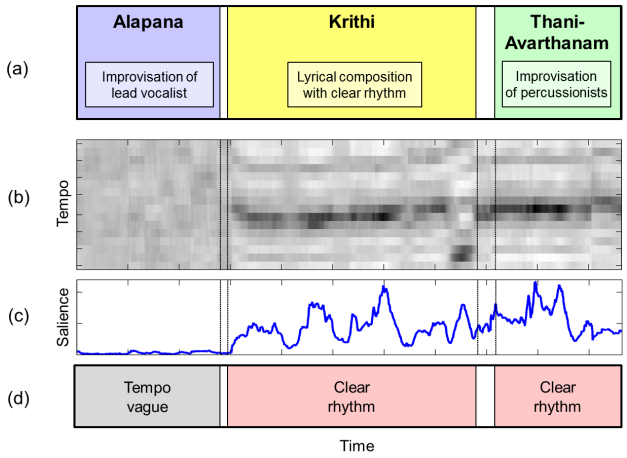

Fig. 2. **(a)** Typical structure of a Carnatic piece of music. **(b)** Tempogram representation of an audio recording. **(c)** Tempo salience feature. **(d)** Manual annotation of parts with a clear rhythm and parts with a vague tempo.

tion is shown in Figure 2. In an opening improvisational part, called *Alapana*, the melodic mode (referred to as *Raga*) that underlies the composition is introduced. The lead artist explores the melodic and tonal material in an improvisational style with no clear rhythm involved. In a second part, called *Krithi*, a lyrical composition is presented. Unlike the first part, the lead artist is accompanied by percussive instruments that play some predefined rhythmic patterns determined by the *Tala* (the rhythmic framework underlying the composition). Finally, in a closing part, called *Thani-Avarthanam*, the percussionists take over and present their virtuosic skills. While exploring the nuances of the underlying *Tala*, the rhythms presented by the percussionists are often of high speed involving complex and syncopated patterns with many tempo changes.

One important observation is that in improvisational *Alapana* part there is a very vague notion of rhythm and tempo. In contrast, the *Krithi* and *Thani-Avarthanam* parts are of rhythmic nature with a much clearer notion of tempo. While the tempo in the *Krithi* part stays roughly constant, there may be sudden tempo changes in the *Thani-Avarthanam* part, in particular between the various solo sections. One main motivation for designing tempo salience features is to find low-dimensional representations that discriminate the first part, the *Alapana*, from the other two parts, see Figure 2c. Opposed to previously studied features, we want to design features that only capture the salience while being more or less invariant to local tempo fluctuations.

We want to mention that only little work has been done so far on automatically segmenting Carnatic music recordings, even though there is an explosion of available audio material thanks to video sharing websites such as YouTube or Sangeethapriya[1].

To segment such audio material, Sarala et al. [19] proposed to use applause detection as a form of locating important segments in a concert. Sarala and Murthy [20] extended this work to segment a concert into separate pieces using the applause between two pieces as a cue. Both contributions are based on the assumption that music recordings have been performed in front of a live audience. Ranjani and Sreenivas [21] used various features followed by a hierarchical classification method to label the pieces in a concert into *Alapana*, *Thanam*, *Krithi*, *Viruttham*, and *Thillana*. In this context, we hope that our tempo salience features are a valuable extension to existing audio features (MFFCs, spectral features, chroma features etc.) as often used in automated segmentation and classification procedures.

[1] A non-commercial service specialized for the exchange of Indian Classical Music, see `http://www.sangeethapriya.org`

3. CYCLIC TEMPOGRAM FEATURES

Our salience features are built on a mid-level representation referred to as *cyclic tempogram*. In this section, following [15], we review this concept while introducing some notation. Similar to the idea of a spectrogram, a *tempogram* $\mathcal{T} : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ is a time-tempo representation of a given music signal. A large value $\mathcal{T}(t, \tau)$ indicates that the music signal has at time $t \in \mathbb{R}$ (measured in seconds) a dominating tempo $\tau \in \mathbb{R}_{>0}$ (measured in beats per minute or BPM). To increase the robustness while drastically reduce the dimensionality, one can convert a tempogram into a cyclic mid-level representation, where tempi differing by a power of two are identified. More precisely, two tempi τ_1 and τ_2 are said to be *octave equivalent*, if they are related by $\tau_1 = 2^k \tau_2$ for some $k \in \mathbb{Z}$. For a given tempo parameter τ, the resulting tempo equivalence class is denoted by $[\tau]$. The *cyclic tempogram* $\mathcal{C}$ induced by $\mathcal{T}$ is defined by

$$\mathcal{C}(t, [\tau]) := \sum_{\alpha \in [\tau]} \mathcal{T}(t, \alpha). \quad (1)$$

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

$$\mathcal{C}_\rho(t, s) := \mathcal{C}(t, [s \cdot \rho]), \quad (2)$$

for $t \in \mathbb{R}$ and $s \in \mathbb{R}_{>0}$. Note that $\mathcal{C}_\rho(t, s) = \mathcal{C}_\rho(t, 2^k s)$ for $k \in \mathbb{Z}$ and $\mathcal{C}_\rho$ is completely determined by its relative tempo values $s \in [1, 2)$. An example for a cyclic tempogram is shown in Figure 3a.

So far we have 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 tempo axis). By abuse of notation, let $\mathcal{C}_\rho(n, m)$ denote the values of the cyclic tempogram for discrete time parameters $n \in [0 : N - 1]$ and relative tempo parameters $m \in [0 : M - 1]$, see [15] for 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 [15, 7]. There are three main parameters, 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 in the experiments, which are further discussed in Section 5.

4. SALIENCE FEATURES

In this section, we introduce our novel salience features based on the tempogram features. Recall from Section 2 that the improvisational *Alapana* part of a Carnatic composition tends to have no strong notion of tempo, opposed to the *Krithi* and *Thani-Avarthanam* parts. This observation is reflected well by the cyclic tempogram of the example shown in Figure 3a. In the *Alapana* part, the tempogram looks rather diffuse having no single large coefficient that would indicate

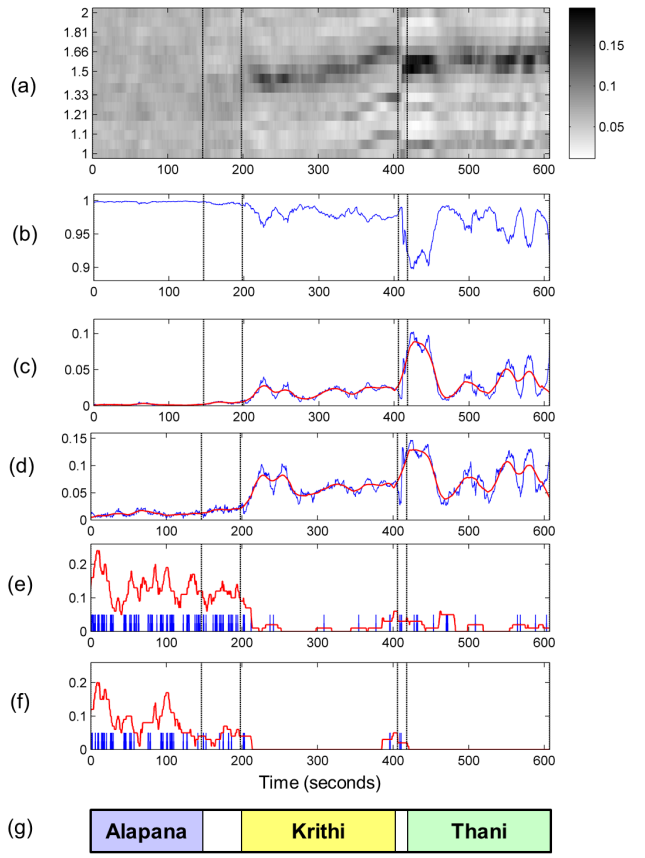

Fig. 3. **(a)** Discrete version of a normalized cyclic tempogram representation based on an autocorrelation tempogram using the parameters $L = 16$ sec, $F_s = 5$ Hz, and $M = 15$. **(b)** Entropy feature $\mathcal{H}(X)$. **(c)** Feature $f_\lambda^{\mathcal{H}}$ with $\lambda = 1$ (blue) and $\lambda = 100$ corresponding to 20 sec (red). **(d)** Feature $f_\lambda^{\mathcal{M}}$ with $\lambda = 1$ (blue) and $\lambda = 100$ (red). **(e)** 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). **(f)** $f_{\tau,\lambda}^{\mathcal{I}}$ with $\tau = 1$. **(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 presence of a specific tempo. In contrast, most of the tempo vectors that belong to the *Krithi* and *Thani-Avarthanam* parts possess a dominating entry. Furthermore one can notice that the tempo class of the dominating entry may vary over time, which reflects the fact that the tempo is changing.

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

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 $p = (p_0, \ldots, p_{M-1})^{\mathrm{T}} \in \mathbb{R}^M$, the (normalized) entropy is defined by

$$\mathcal{H}(p) = -\big(\textstyle\sum_{m=0}^{M-1} p_m \log_2(p_m)\big) / \log_2(M), \tag{3}$$

which assumes a maximal value of one if the vector p corresponds to a uniform distribution and a minimal value of zero if the vector p corresponds to a dirac distribution. In our scenario, we normalize the columns of the cyclic tempogram $\mathcal{C}_\rho \in \mathbb{R}^{N\times M}$ 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

$$\mathcal{H}(X) := (\mathcal{H}(X[0]), \ldots, \mathcal{H}(X[N-1])) \tag{4}$$

of numbers $\mathcal{H}(X[n]) \in [0,1]$, see Figure 3b for an example. As a measure of salience (rather than one of uncertainty), we consider $1 - \mathcal{H}(X)$. 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 3c, this feature has the desired property of being close to zero in the *Alapana* part and much larger in the other parts, see Section 5 for a more detailed investigation.

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

$$\mathcal{M}(p) := \max\{p_0, \ldots, p_{M-1}\} - \mathrm{median}\{p_0, \ldots, p_{M-1}\} \tag{5}$$

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 gives us an indication of tempo distribution. By 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 3d, this feature behaves similarly to $f_\lambda^{\mathcal{H}}$.

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:

$$m^{\max}(n) := \mathrm{argmax}_{m\in[0:M-1]}(X(n,m)). \tag{6}$$

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

$$d^{\mathrm{cyc}}(m_1, m_2) := \min\big\{|m_1 - m_2|, M - |m_1 - m_2|\big\} \tag{7}$$

for $m_1, m_2 \in [0 : M-1]$. Based on this definition, we define

$$\mathcal{I}(n) := d^{\mathrm{cyc}}\big(m^{\max}(n), m^{\max}(n-1)\big) \tag{8}$$

for $n \in [1 : N-1]$. Intuitively, any value $\mathcal{I}(n) > 0$ expresses that there has been a tempo change at time n. Smooth tempo changes and small local tempo fluctuations (see, e. g., the *Krithi* part in the tempogram of Figure 3a) may result in a value $\mathcal{I}(n) = 1$. 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 $f_\tau^{\mathcal{I}}$ by setting

$$f_\tau^{\mathcal{I}}(n) := \begin{cases} 0 & \text{if } \mathcal{I}(n) \le \tau, \\ 1 & \text{if } \mathcal{I}(n) > \tau. \end{cases} \tag{9}$$

As before, applying an averaging filter of length $\lambda \in \mathbb{N}$ yields a feature we refer to as $f_{\tau,\lambda}^{\mathcal{I}}$. These definitions are illustrated by Figure 3e, which shows the binary feature $f_\tau^{\mathcal{I}}$ for $\tau = 0$ and its averaged version $f_{\tau,\lambda}^{\mathcal{I}}$ using λ corresponding to 20 sec. Note that, for illustration purposes, we scaled down the features by a factor of 20 (blue bars in Figure 3e,f). 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 3f. In this case, smooth tempo changes as occurring in the *Krithi* and *Thani-Avarthanam* parts do not contribute to the density feature.

5. EVALUATION

In this section, we report on a quantitative evaluation of our tempo salience features within the Carnatic music scenario. In our experiments, we used music recordings of good audio quality from the Sangeethapriya website[2]. 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 and determined the segment boundaries for the *Alapana*, *Krithi* and *Thani-Avarthanam* parts. For reproducibility, the available links to the audio material, which has been published by Sangeethapriya under a creative common license, as well as the annotations can be found on a website[3].

Based on this annotation, 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[4]. Based on these representations, we computed the salience features as introduced in Section 4. Then, for each of the features, we computed the average feature value and its variance for each of the three 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 Figure 4. Note that, rather than the absolute values, the relative relation between the values across the three different parts are of interest. [5]

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 Figure 4, the mean statistics of $f_\lambda^{\mathcal{H}}$ assume a value of 0.0031 for the *Alapana* part, which is roughly ten times smaller than the value 0.0314 for the *Krithi* part and the value 0.0238 for the *Thani-Avarthanam* part. Also, the standard deviation $\overline{\sigma}$ for $f_\lambda^{\mathcal{H}}$ shows a similar trend: it assumes the value 0.0021 for the *Alapana* part, which is much lower than the value 0.0168 for the *Krithi* and the value 0.0184 for the *Thani-Avarthanam* part. Recall from Section 4 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 Figure 4, 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 0.1061 for the *Alapana* part, which is much higher than the mean value 0.0245 for *Krithi* and 0.0348 for *Thani-Avarthanam* part. The relative differences between the parts become even larger for the mean values of the feature $f_{1,\lambda}^{\mathcal{I}}$. Recall from Section 4 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 *Thani-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 *Thani-Avarthanam* part, the overall mean values are small compared to the *Alapana* part. Interestingly, the mean and standard deviations of Figure 4 also indicate that abrupt tempo changes seem to occur more often in the final *Thani-Avarthanam* part compared to the *Krithi* part.

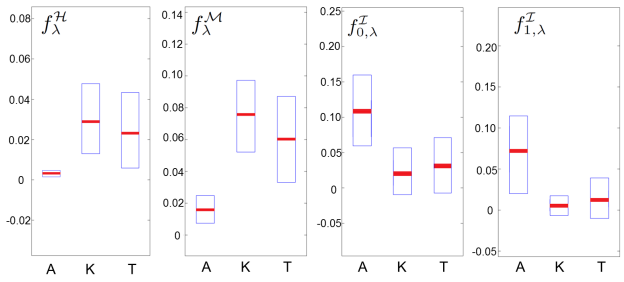

Fig. 4. Mean $\overline{\mu}$ (red line) and standard deviation $\overline{\sigma}$ (blue box) of various salience features shown for the three different parts *Alapana* (A), *Krithi* (K), and *Thani-Avarthanam* (T).

6. CONCLUSIONS

In this paper, we have introduced several novel audio features that capture a musical property we referred to as tempo salience. By means of the Carnatic music scenario, we demonstrated that these features reflect well whether there is a notion of a clear tempo or not. Besides their discriminative power, our salience features also have the benefit of having a low dimensionality and of possessing a direct musical interpretation. Therefore, we expect that such features are a valuable extension to existing audio features that correlate to other types of information such as instrumentation, timbre, or harmony. In future work, we plan to apply our salience feature for various audio analysis, classification and segmentation tasks including Indian music and beyond.

Acknowledgments: This work has been supported by the German Research Foundation (DFG MU 2686/5-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and the Fraunhofer-Institut für Integrierte Schaltungen (IIS).

[2] http://www.sangeethapriya.org

[3] www.audiolabs-erlangen.de/resources/MIR/2015_ICASSP_CarnaticMusic

[4] For the computation we used the MATLAB implementations supplied by the *Tempogram Toolbox*, see [22].

[5] Additionally, to show that the discrimination between the rhythmic Krithi/Tani parts and the non-rhythmic Alapana part is statistically significant, we performed a one-way ANOVA (Analysis of Variance) test [23]. We obtained p-values of 0.0034 and 0.0098 for the features $f_\lambda^{\mathcal{H}}$ and $f_\lambda^{\mathcal{M}}$, respectively. Furthermore, we obtained very low p-values below 0.00001 for the features $f_{0,\lambda}^{\mathcal{I}}$ and $f_{1,\lambda}^{\mathcal{I}}$.

7. REFERENCES

[1] Miguel Alonso, Bertrand David, and Gaël Richard, "Tempo and beat estimation of musical signals," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004.

[2] Simon Dixon, "Automatic extraction of tempo and beat from expressive performances," *Journal of New Music Research*, vol. 30, pp. 39–58, 2001.

[3] Daniel P.W. Ellis, "Beat tracking by dynamic programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.

[4] Jonathan Foote and Shingo Uchihashi, "The beat spectrum: A new approach to rhythm analysis," in *Proceedings of the International Conference on Multimedia and Expo (ICME)*, Los Alamitos, CA, USA, 2001.

[5] Masataka Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *Journal of New Music Research*, vol. 30, no. 2, pp. 159–171, 2001.

[6] Peter Grosche and Meinard Müller, "Extracting predominant local pulse information from music recordings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1688–1701, 2011.

[7] Geoffroy Peeters, "Time variable tempo detection and beat marking," in *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, 2005.

[8] Eric D. Scheirer, "Tempo and beat analysis of acoustical musical signals," *Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 588–601, 1998.

[9] George Tzanetakis and Graham Percival, "An effective, simple tempo estimation method based on self-similarity and regularity," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 241–245.

[10] Martin F. McKinney, Dirk Moelants, Matthew E.P. Davies, and Anssi P. Klapuri, "Evaluation of audio beat tracking and music tempo extraction algorithms," *Journal of New Music Research*, vol. 36, no. 1, pp. 1–16, 2007.

[11] André Holzapfel, Matthew E.P. Davies, Jose R. Zapata, J. Oliveira, and Fabien Gouyon, "On the automatic identification of difficult examples for beat tracking: towards building new evaluation datasets," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, 2012, pp. 89–92.

[12] Peter Grosche, Meinard Müller, and Craig Stuart Sapp, "What makes beat tracking difficult? A case study on Chopin Mazurkas," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands, 2010, pp. 649–654.

[13] Ali Taylan Cemgil, Bert Kappen, Peter Desain, and Henkjan Honing, "On tempo tracking: Tempogram representation and kalman filtering," *Journal of New Music Research*, vol. 28, no. 4, pp. 259–273, 2001.

[14] Peter Grosche and Meinard Müller, "A mid-level representation for capturing dominant tempo and pulse information in music recordings," in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, 2009, pp. 189–194.

[15] Peter Grosche, Meinard Müller, and Frank Kurth, "Cyclic tempogram – a mid-level tempo representation for music signals," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, USA, 2010, pp. 5522 – 5525.

[16] Kristoffer Jensen, Jieping Xu, and Martin Zachariasen, "Rhythm-based segmentation of popular chinese music," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, UK, 2005.

[17] 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)*, Victoria, Canada, 2006, pp. 35–40.

[18] Lakshmi Subramanian, "New mansions for music: Performance, pedagogy and criticism," in *Berghahn Books*, ISBN 978-81-87358-34-3, 2008.

[19] Padi Sarala, Vignesh Ishwar, Ashwin Bellur, and Hema A Murthy, "Applause identification and its relevance to archival of carnatic music," in *Proceedings of the 2nd CompMusic Workshop*, Istanbul, Turkey, 2012.

[20] Padi Sarala and Hema A.Murthy, "Inter and intra item segmentation of continuous audio recordings of carnatic music for archival," in *Proceedings of the 14th International Conference on Music Information Retrieval (ISMIR)*, Curitiba, Brazil, 2013.

[21] H. G. Ranjani and T. V. Sreenivas., "Hierarchical classification of carnatic music forms," in *Proceedings of the 14th International Conference on Music Information Retrieval (ISMIR)*, Curitiba, Brazil, 2013.

[22] Peter Grosche and Meinard Müller, "Tempogram Toolbox: MATLAB implementations for tempo and pulse analysis of music recordings," in *Late-Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, FL, USA, 2011.

[23] David Moore and George McCabe, "Introduction to the practice of statistics," 2003.