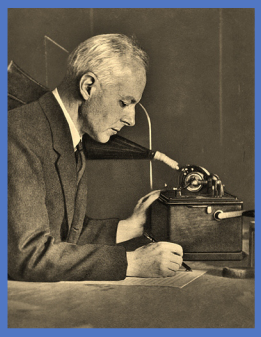
Proceedings of the Third International Workshop on Folk Music Analysis



Publishers:
Meertens Institute
Department of Information and Computing Sciences,
Utrecht University

Editors:
P. van Kranenburg
C. Anagnostopoulou
A. Volk

Proceedings of the Third International Workshop on Folk Music Analysis (FMA2013)

6 and 7 June, 2013
Amsterdam, Netherlands

Editors:
Peter van Kranenburg
Christina Anagnostopoulou
Anja Volk

Amsterdam: Meertens Institute
Utrecht: Department of Information and Computing Sciences, Utrecht University

Title Proceedings of the Third International Workshop on Folk Music Analysis (FMA2013)

Editors P. van Kranenburg, C. Anagnostopoulou, A. Volk

Publishers Meertens Institute; Department of Information and Computing Sciences, Utrecht University

ISBN 978-90-70389-78-9

Program Committee

Chairs

- Christina Anagnostopoulou (University of Athens)
- Anja Volk (Utrecht University, Netherlands)

Members

- Aggelos Pikrakis (University of Piraeus)
- Andre Holzapfel (Universitat Pompeu Fabra)
- Aline Honingh (University of Amsterdam)
- Ashley Burgoyne (University of Amsterdam)
- Bas de Haas (Utrecht University)
- Damien Sagrillo (University of Luxembourg)
- Darrell Conklin (Universidad del País Vasco UPV/EHU)
- Emilios Cambouropoulos (Aristotle University of Thessaloniki)
- Ewa Dahlig (The Institute of Art of the Polish Academy of Sciences)
- Joren Six (University College Ghent, Belgium)
- Matija Marolt (University of Ljubljana)
- Olivier Lartillot (University of Geneva)
- Olmo Cornelis (University College Ghent, Belgium)
- Paco Gomez (Technical University of Madrid)
- Polina Proutskova (Goldsmiths, London)

Organizing Committee

- Peter van Kranenburg (Meertens Institute, Amsterdam)
- Berit Janssen (Meertens Institute, Amsterdam)
- Anja Volk (Utrecht University)
- Frans Wiering (Utrecht University)
- Dániel P. Biró (University of Victoria)

Preface

The present volume contains the proceedings of the Third International Workshop on Folk Music Analysis. As the third in a series, this workshop offers an excellent opportunity to present and discuss ongoing research in the area of computational ethnomusicology. There are two important motivations. Firstly, recent advances in computer science, artificial intelligence, etc. have great potential to be employed for (ethno)musicology. This implies an empirical approach to music studies. The current research in this area is only in its beginnings. Therefore, much attention should be paid to explore these methods and their relation to the research traditions of musicology. Secondly, most of the current research in music information retrieval is exclusively aimed at western music. With this workshop we want to stimulate a broader focus that also includes non-western musics.

Computational study of music is inherently interdisciplinary. Musicologists, computer scientists, engineers and programmers need to collaborate. Therefore, we are exited that this workshop will bring together researchers from different backgrounds.

We are grateful to everybody who made this event possible, including The Meertens Institute, for hosting the workshop and for practical support (in particular Hetty Garcia and Marianne van Zuijlen); The members of the program committee; The members of discussion panel; The e-Humanities Group of the Royal Netherlands Academy of Arts and Sciences for sponsoring the keynote talk; the Study Group on Digital Musicology of the International Musicological Society; and of course the authors and participants.

Amsterdam, June 2013

The organizers

Contents

Full papers

Extended Abstracts

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

TEMPO AND PROSODY IN TURKISH TAKSIM IMPROVISATION

André Holzapfel
Boğaziçi University, Istanbul, Turkey
{xyzapfel}@gmail.com

ABSTRACT

Instrumental improvisation in Turkish makam music, the *taksim*, is considered to be free-rhythm, that is its rhythm develops without the underlying template of a meter or continuous organized pulsation. In this paper, we want to examine how in this setting, rhythmic idioms are formed and maintained throughout a performance. For this, we will apply a simple signal processing approach. We show differences that can be observed between performers, and raise the question if a tempo could be evoked by certain regularities in the occurring rhythmic elaborations.

1. INTRODUCTION

In Makam music of Turkey, we can distinguish between metered pieces and free-rhythm improvisation. In our paper, we focus on the latter in the form of instrumental improvisation, which is called *taksim* in Turkish art music. While rhythm in metered pieces of Turkish music was analyzed previously by Holzapfel & Bozkurt (2012), a detailed study of rhythm in Turkish improvisation still remains to be approached. Until now studies on taksim concentrated on aspects of melodic development (Stubbs, 1994), and scale aspects (Bozkurt, 2008). A study on rhythm is timely because improvisation in Turkish music is widely considered as free-rhythm (Clayton, 2009), which means that its surface rhythm is not related to an organized and continuous pulsation. Instead, it has been mentioned that taksim is characterized by pulsations in non-metrical flowing rhythm (Feldman, 1993). To the best of my knowledge it has not been investigated how such a pulsation is formed; *i.e.* how it appears throughout a performance, and if there is some degree of continuity of such pulsation as it was observed by Widdess (1994) for a specific Hindustani *alap* performance.

In the presented work we apply a simple signal processing framework in order to investigate the occurrence of pulsation in taksim. We restrict ourselves to taksimler (plural of taksim) played on the instrument *tanbur*, which is a plugged string instrument. This restriction is imposed in order to avoid any variance in style possibly encountered on different instruments, and because the sound of the tanbur has the advantage, from a signal processing point of view, that the strokes of the pick can be detected relatively easily. This enables us to study some basic rhythmic properties of taksim using a fairly simple signal processing approach.

We compiled a dataset of 52 tanbur taksimler played by five renowned masters, and observe how pulsation develops over the individual taksim. Interesting differences are pointed out that seem to be related to personal style, or to the style predominating the recording period. The creation of a tempo in taksim is discussed, and relations to speech utterances are pointed out.

2. PROCESSING APPROACH: DESCRIPTION, MOTIVATION AND EXAMPLES

First, we need to emphasize signal transients which are positioned at the time instances where the player hits a string. For this, we convert our original audio signal to an onset function by examining positive changes in its spectral magnitude (Holzapfel et al., 2010). Then autocorrelations of this onset function are computed in small shifting windows of 3s length and a hop size from one window to the next of 0.5s, similar to Holzapfel & Stylianou (2011). The obtained autocorrelation vectors are stringed together in a two-dimensional representation, referred to as pulsation matrix hereafter. This matrix has the time of the initial recording on its x-axis, and the lags of the autocorrelations (in seconds) on the y-axis.

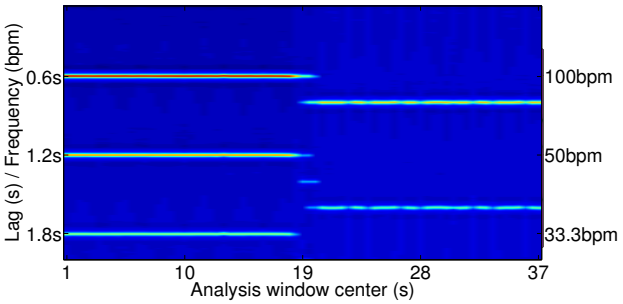
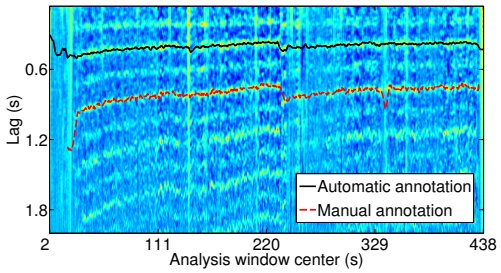


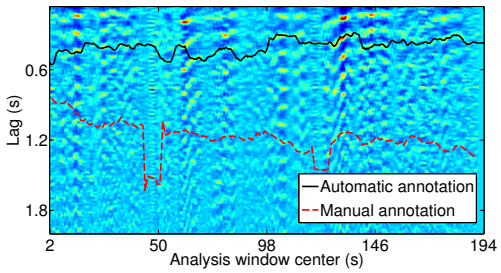
Figure 1: Example for a pulsation matrix derived from an artificial signal, containing a series of noise bursts

We clarify this process using a simple artificial example. We generate a signal, which contains a series of equidistant impulsive sounds (here: noise bursts, could be also *e.g.* hand claps). In the first half of the signal, each sound is 0.6s from its neighbors apart, while in the second half this period is increased to 0.75s. The onset function derived from this signal has peaks only at the onsets of the impulsive sounds and is zero, or at least very small, otherwise. An autocorrelation of a 3s excerpt from the first half of this signal will have peaks at the period of 0.6s, and at its multiples. This can be seen in Figure 1 from the bright colors located at these lags (0.6s,1.2s,1.8s, higher multiples not shown in the Figure). These lags l, in seconds, can be interpreted as specific tempo locations s in beats per minute (bpm), by the simple conversion $s = 60/l$. For

example, the series of sounds in the first half of our example is related to the tempo of $100bpm$, meaning that we have a regular sequence of 100 impulses per minute. However, in the middle of our example we change this period, which causes the shown pulsation matrix to have its peaks related to this second series of pulses, which has a period of $0.75s$ or a tempo of $80bpm$. We can see in this simple example that if a pulsation is maintained stable over a period of time, we will observe a relatively stable comb-like structure over several columns of the matrix, which are related to the tempo of the pulsation in this signal. If, as usually in music interpreted by humans without the use of metronomes, the tempo changes gradually, we will observe bright parallel lines that do not remain at a constant position as in our example but change their place gradually. On the other hand, if we have a signal that has no pulsation at all, we will end up with a matrix having an almost uniform color.



(a) Pulsation matrix of metered piece



(b) Pulsation matrix of taksim

Figure 2: Two examples of pulsation matrices for recordings of Turkish makam music

In Figures 2a and 2b, we depict such pulsation matrices for a metered piece of Turkish music, and for a taksim, respectively. The lines on top of the pulsation matrices have been obtained by running a beat tracking algorithm (Davies & Plumbley, 2007) (bold black line), and by manually tapping to the piece of music (dotted red line). Both tapping and the beat tracker provide us with a series of time values for the position of the pulses. The value on the y-axis of these lines represents the time-interval between these pulses. For the metered piece in Figure 2a, equidistant horizontal lines are characteristic for the pulsation matrix, with mutual distances of about 0.2s. We can observe that the black and the red lines of the annotations are exactly on top of one of the ridges formed by the horizontal lines. This clarifies that the piece has indeed strong, continuous and relatively stable periodicities in its surface rhythm. For the taksim, on the other hand, Figure 2b shows parallel line structures which imply the existence of pulsation in the piece. Here, however, they are less stable, which means that they change rapidly, and they are interrupted with sequences that lack pulsation completely (*e.g.* at about 50s). Neither beat tracker nor manual annotation follow the pulsation indicated by the ridges in this matrix consistently. This example indicates that, while pulsation occurs in taksim, this does not lead to a clearly trackable pulse throughout a performance.

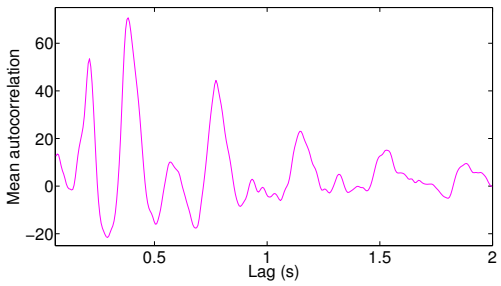
These two examples seem to be representative for the "behavior" of beat tracking algorithms; In our recent work (Srinivasamurthy et al., 2013) we observed that two different beat tracking algorithms often estimate either the true tempo or a tempo related to ground truth annotation with a factor of 2 on a collection of 63 Turkish makam music recordings. This confirms that for metered pieces of Turkish makam music, tempo obtained from algorithms and human performance tend to be strongly related. For pieces with no or highly ambiguous meter, our work on the mutual agreement of beat tracking algorithms documents that algorithmic output on such signals changes randomly between approaches, which is reflected in the arbitrary relation between algorithmic and human tempo annotation in Figure 2b.

Regarding human behavior it is less apparent how the two examples generalize to other metered pieces or taksims, and the beat or tempo humans would generally perceive in such pieces. We are currently conducting a series of experiment to evaluate for the sensorimotor synchronization (Repp, 2005) of Turkish musicians to metered pieces. We can observe that musicians tend to differentiate between clapping to music in a "technical" way by aligning their strokes to the underlying *usul* (*i.e.* rhythmic mode), or by freely accompanying the surface rhythm with their claps. While the claps can show a wide variety of behavior, the technical way of clapping is less limited in its variation because musicians are aware of the alignment between rhythm and rhymic mode (usul). For taksim, such a behavioral study is even more complex, if not impossible. There is no doubt among practicing musicians that a taksim has no meter. When asked about the rhythmic elaboration of taksim, they usually state that they not consciously maintain a tempo. On the other hand, some of them do not want to exclude that at least in some examples a continuous pulse might exist.

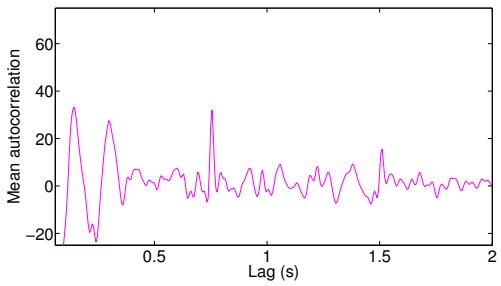
We believe that an access different from a sensorimotor synchronization experiment has to be found to shed light on the elaboration of rhythm in taksim. The first reason to assume that is the observation that musicians already in a free-form tapping experiment to metered music showed little enthusiasm for tapping rhythmic patterns in an experimental setup. Furthermore, language impedes an explicit access needed in such an experiment, as a term like "pulse" or "beat" is hard to translate to Turkish. Its meaning would be either interpreted as not musical, or as related to a rhythmic mode. However, as the taksim obviously do not have a rhythmic mode, a direct access using language to form a suitable question for an experiment seems impossible.

Finally, sensorimotor synchronization tasks were usually conducted using highly simplified sounds (Repp, 2005). It can be expected that the complexity of taksim sounds in terms of rhythm and other aspects represent another reason to hesitate in conducting such a study.

For these reasons, we want to apply our simple signal processing approach in order to obtain some insights into the rhythmic structure of taksim, as this might help us to form a more precise hypothesis about rhythmic elaboration in taksim. We observe that some periods seem to be prominent throughout a performance, which is exemplified in Figures 3a and 3b, which depict the mean over time of the pulsation matrices in Figure 2a and Figure 2b, respectively. Clear peaks related the tempo exist for the metered piece (Figure 3a) at the lag related to the tempo (at a period of 0.77s) and at multiples and $1/2$ and $1/4$ if $0.77s$. The taksim shows some clear maxima as well (Figure 3b), however they are not spread over a wide range as for the metered piece. The sharp peak at 0.76s is caused by the periodic noise from the record of the original historic recording. The two peaks below 0.5s are caused by the rhythmic properties of the performance, which might be indicative for a tempo impression in this form. While it is apparent that the pulsation in our example frequently changes its period (the pulsation matrix in Figure 2b is not characterized by parallel, continuous lines over time), we would like to take this observation as a starting point for a stylistic comparison between players and for developing a hypothesis about tempo that is evoked at least in some taksim performances.



(a) Pulsation profile of peşrev



(b) Pulsation profile of taksim

Figure 3: Pulsation profiles, which are obtained by computing the mean over time of a pulsation matrix.

We will first describe the collection of taksim performances which we use in our experiments in Section 3. In Section 4 we will determine the pulsation profiles for all the performances in the collection and use them to obtain a first orientation among the possibly existing different rhythmic idioms in the recordings. Based on these findings, we will focus on comparing two specific players in the collection, and address the question if the differences in their pulsation profiles are indeed in some way related to different styles in rhythmic elaboration. In Section 5, we will give some perspective on how pulsation matrices can be used to evaluate for the existence of a tempo in the sense of a continuous pulsation in some taksim. We will discuss how our representations motivate for searching relations to signals of human speech. Finally, Section 6 concludes the paper.

3. MUSIC COLLECTION

Our music collection contains 52 recordings of taksim by five renowned masters of tanbur in Turkey. The players and the numbers of recordings from each player are given in Table 1. These players cover a range of a century of recordings, with Tanburi Cemil Bey marking the beginning of recording history for Turkish music in the beginning of the century. His recordings became influential for generations of players since then, which is why we hope to be able to shed some light on the rhythmic aspects of his playing, and how it possibly differed from other players.

Table 1: Tanbur players and numbers of pieces in the collection

Name	Number of pieces
Ercümend Batanay	11
Mesut Cemil Bey	8
Murat Aydemir	5
Necdet Yašar	15
Tanburi Cemil Bey	13

4. FATHER AND SON

As a first step we computed all the pulsation profiles for the 52 taksim recordings. In order to compare the profiles we chose the cosine distance, which converts the angle between two vectors to a distance measure in the range between 0 and 1. As we detailed previously (Holzapfel & Stylianou, 2011), this measure is adequate for rhythmic descriptors that contain information about a range of periodicities present in a music signal.

In order to obtain a first orientation, we computed all the mutual distances between the pulsation profiles. We then ordered the distances according to their size, and determined for each taksim which other recording is most similar. In Table 2 shows the results of that experiment, which can be interpreted as a k-nearest-neighbor (kNN) classification with k=1.

It is not the goal to derive some means to classify a recording of a taksim to a specific player, and therefor we will rather try to interpret the meaning of the numbers shown in Table 2. The highest accuracy in kNN classification is related to the taksimler played by Tanburi Cemil

Table 2: Nearest neighbor classification (kNN), with k=1

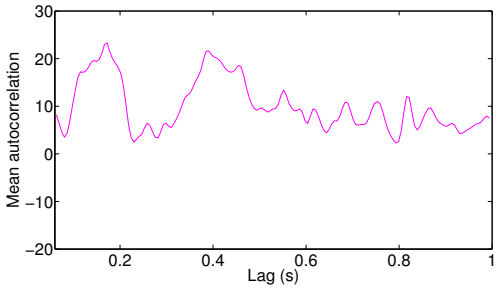
Player	Batanay	Mesut C.	Aydemir	Yaşar	Tanburi C.
Acc.	72.7%	37.5%	80.0%	20.0%	84.6%

Bey. They seem to be related to pulsation profiles with a very consistent shape, and therefor they should be characterized by pulsations that are concentrated at specific values. As we will see in the following, this is related to a quite characteristic way, in which rhythm is elaborated in his improvisations. While a similar conclusion can be drawn for Ercümend Batanay and Murat Aydemir, the situation seems to be different for Mesut Cemil and Necdet Yaşar. The latter two players seem to be characterized by a wider variety of pulsation profiles, what however does not yet enable us to say anything specific about their rhythmic idioms.

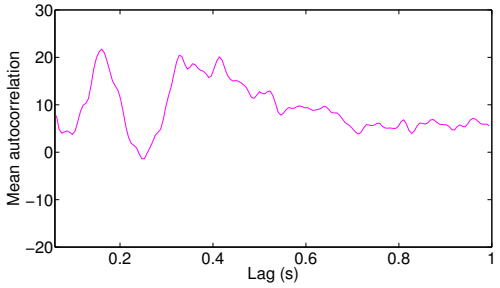
At this point, we would like to take a turn from the overview over the music collection towards a more focused comparison. This more focused comparison will shed light on the reasons for the differences in the pulsation profiles. We chose to compare two players, namely Tanburi Cemil bey and Mesut Cemil. The former became a legend with his recordings in the beginning of the last century, in the last phase of the Ottoman empire. The latter was his son, and contributed significantly to many changes in style in Turkish music with the beginning of the Turkish republic. Therefor, as shortly pointed out by Feldman (1993), their musical styles in taksim differed in terms of the applied rhythmic idioms. This might be a cause for the differences we observe in their pulsation measurements. A comparison of two examples taken from Mesut Cemil's taksimler (Figure 4) with two taksimler by Tanburi Cemil Bey (Figure 5) reveals some differences. For this, we focus on the lags smaller than 1s, as according to Figure 2b for non-metered pieces these short period pulsations seem to be important. Tanbur Cemil Bey's taksimler seem to contain strong pulsations concentrated at 0.15s and 0.3s, indicated by the maxima at these values in Figure 5. For Mesut Cemil's taksimler, the peaks are clearly less concentrated which indicates a larger variation of the pulsations in his taksimler. Especially in the Hisarbuselik taksim, the leftmost peak (at 0.16s) is not accompanied by a second, harmonically related and clear peak. These aspects indicate a difference in the rhythmic content in the related pieces.

The pulsation profiles cannot tell us anything how pulsation develops throughout a piece. While the clear peaks for Tanburi Cemil bey imply strong pulsation, the lack of them for Mesut Cemil does not necessarily imply the absence of pulsation, but might as well indicate a high variation of pulsation tempi throughout a piece. In order to understand more about the temporal development, we need to look at the pulsation matrices of the pieces in question.

They are depicted in Figures 6 and 7 and reveal clear differences in rhythmic elaboration between the two taksimler by Mesut Cemil and the taksimler by Tanburi Cemil.

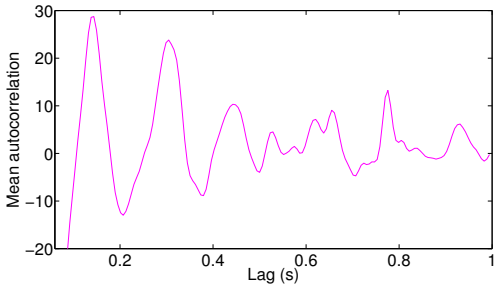


(a) Müstear taksim

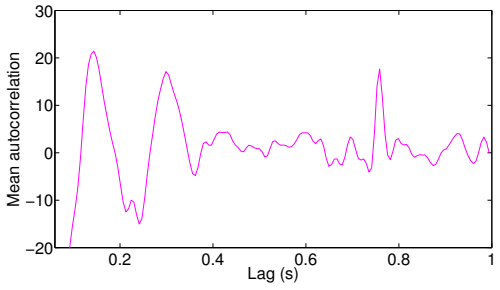


(b) Hisarbuselik taksim

Figure 4: Pulsation profiles for two taksim of Mesut Cemil Bey.



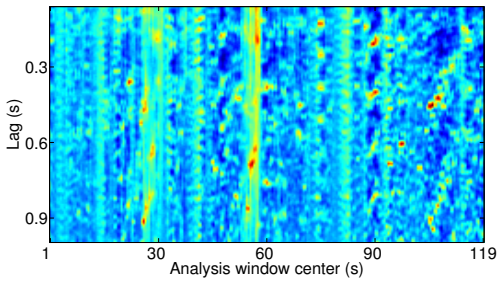
(a) Rast taksim



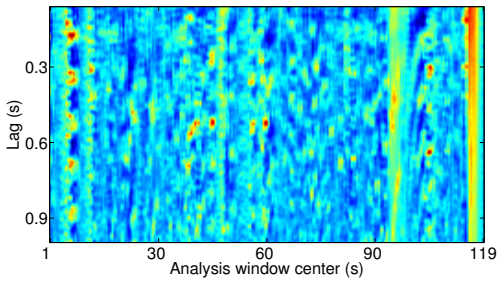
(b) Suzidil taksim

Figure 5: Pulsation profiles for two taksim of Tanburi Cemil Bey.

In the taksimler by Tanburi Cemil pulsations are maintained over large durations, especially in the example of the Rast taksim. This conclusion can be drawn by observing the bright horizontal line patterns in Figures 7a and 7b. In both figures, the second lines (from the top) of these patterns are graphically emphasized by overlaying them with white polygons. We can observe, that *e.g.* in the Rast taksim a continuous pulsation is established at about 20s, which is then increased in tempo until 70s, and then slowly fades out. The other depicted taksim by Tanburi Cemil does not have such a clear continuous development,

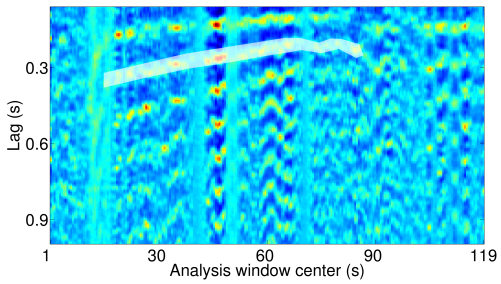


(a) Müstear-taksim by Mesut Cemil

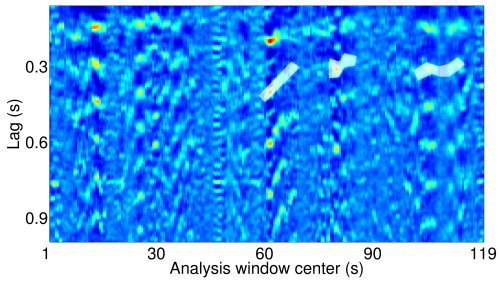


(b) Tahirbuselik-taksim by Mesut Cemil

Figure 6: Pulsation matrices related to the taksimler depicted in Figure 4. The first two minutes are depicted for better comparability.



(a) Rast-taksim by Tanburi Cemil Bey



(b) Suzidil-taksim by Tanburi Cemil Bey

Figure 7: Pulsation matrices related to the taksimler depicted in Figure 5. The first two minutes are depicted for better comparability.

but still *e.g.* from 100s-115s a continuous area is marked. The establishment of such a continuous pulsation seems to occur rarely for Mesut Cemil, which is exemplified by the lack of such line patterns in the depicted two pulsation matrices for the Müstear and the Tahirbuselik taksim. Only after 10s of the beginning of the Tahirbuselik taksim, we can observe such a short pattern in Figure 6b.

5. PROSODY OF TAKSIM

As we discussed in the previous section, some players, such as Tanburi Cemil bey, seem to emphasize pulsations of specific frequencies in their playing, which seems to make them differ regarding their style from other players. We were able to observe that this emphasis is expressed by a continuous pulsation of up to 50s for Tanburi Cemil bey, while Mesut Cemil seems not to elaborate rhythm in such a continuous way. There might be two reasons for such differences, the first simply being differences in individual playing style, and the second, a difference that is caused by the changed stylistic preferences of the society at different historic periods. The second hypothesis is attractive, because Mesut Cemil is widely known to have broken with many concepts of the court music tradition of the former Ottoman empire. He contributed to defining the new national identity of Turkish music by introducing chorus singing, and by banning styles such as the vocal improvisation *gazel* that were considered not to fit to an orientation towards Europe. However, while our results might indicate such a direction, other recordings from the final period of the Ottoman empire would have to be examined.

The peaks in the pulsation profiles, and their temporal continuity for some taksim motivates to ask if these phenomena evoke the impression of a tempo in the listener. We could ask if listeners can perceive a tempo development in a taksim that follows the shape of the patterns we observe in the pulsation matrices. It is difficult, however, to quantify the agreement of a listener with the measurements. Therefor, we might establish a tempo curve for a taksim, which follows *e.g.* the white shaded area in the Rast taksim by Tanburi Cemil bey depicted in Figure 7a. Then, a stimulus in form of a click sequence can be generated that follows this tempo curve, and the resulting click sequence can be superimposed to the sound of the taksim, to ask listeners regarding the relation between the click sound and the music. This way we could for the first time establish some rules how a tempo is established in taksim.

Obviously, a taksim is not based upon a musical meter. This is apparent for various reasons; First, musicians are absolutely clear in the differentiation of forms that follow a rhythmic mode (*usul*), and forms that do not have any *usul*, such as the taksim. Furthermore, in literature taksim was always referred to as free-rhythm. When intending to understand in more detail how rhythm in taksim is shaped, investigating relations to rhythm in speech might be helpful instead. Relations between musical expression and speech were frequently used in music seminars of Turkish makam music, which I attended. For instance, teachers might motivate their students to play a short phrase, or their names, on a musical instrument, expressing the sound of the name with the instrument. For that reason, it appears as an interesting question if the pulsation in taksim is in some way related to syllable and word rates in Turkish language. It is interesting to observe that the poetry of the Ottoman was mainly following quantitative meter, hence being based on schemes of syllable durations. This poetry, in the form of gazel, had surely an influence on artists like Tanburi Cemil,

while in the times of the Turkish republic a stronger emphasis was given on folk poetry with its qualitative meter.

The discussed relations between poetry and taksim, as well as the potential perception of tempo in taksim can only be examined after a careful annotation of timing in related recordings. As a next step, we intend to manually annotate for some taksimler the time instances, at which the player hits the string. This will enable us to obtain more detailed insights into the rhythmic elaboration of the pieces. Furthermore, it appears meaningful to attempt the same for recitations of poetry or some free speech samples in Turkish language, to be able to eventually compare the occurring timing patterns.

6. CONCLUSION

By applying a simple signal processing approach, we were able to observe differences in the ways two renowned master players of Turkish makam music shape(d) rhythm in their free-rhythm improvisations. Differences are related to the continuity in which a pulsation is encountered over time. These differences might be related to personal style, or to style preferences of different historical periods. The following steps that will help to illustrate these aspects in more detail will lie in conducting some interviews with listeners, and by detailed manual annotations of onset instances in some of the taksim. We will have to address the question, if there are some general styles present in the prosody of taksim, and if the two examined masters might be representative for such styles.

7. ACKNOWLEDGEMENTS

This work is supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583);

8. REFERENCES

Bozkurt, B. (2008). An automatic pitch analysis method for turkish maqam music. *Journal of New Music Research*, *37*(1), 1–13.

Clayton, M. (2009). Free rhythm: Ethnomusicology and the study of music without metre. *Bulletin of the School of Oriental and African Studies*, *59*(2), 323–332.

Davies, M. E. P. & Plumbley, M. D. (March 2007). Context-dependent beat tracking of musical audio. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(3), 1009–1020.

Feldman, W. (1993). Ottoman sources on the development of the taksim. *Yearbook for Traditional Music*, *25*, 1–28.

Holzapfel, A. & Bozkurt, B. (2012). Metrical strength and contradiction in turkish makam music. In *2nd CompMusic Workshop*, Istanbul, Turkey.

Holzapfel, A. & Stylianou, Y. (2011). Scale transform in rhythmic similarity of music. *IEEE Trans. on Speech and Audio Processing*, *19*(1), 176–185.

Holzapfel, A., Stylianou, Y., Gedik, A. C., & Bozkurt, B. (2010). Three dimensions of pitched instrument onset detection. *IEEE Transactions on Audio, Speech and Language Processing*, *18*(6), 1517–1527.

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review*, *12*(6), 969–992.

Srinivasamurthy, A., Holzapfel, A., & Serra, X. (2013). In search of automatic rhythm analysis methods for turkish and indian art music. *Journal of New Music Research*, *submitted*.

Stubbs, F. W. (1994). *The art and science of taksim: an empirical analysis of traditional improvisation from 20th century Istanbul*. PhD thesis.

Widdess, R. (1994). Involving the performers in transcription and analysis: A collaborative approach to *Dhrupad*. *Ethnomusicology*, *38*(1), 59–79.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

QUANTIFYING TIMBRAL VARIATIONS IN TRADITIONAL IRISH FLUTE PLAYING

Islah Ali-MacLachlan[1], Münevver Köküer[1,2], Peter Jančovič[2], Ian Williams[1], Cham Athwal[1]
[1] School of Digital Media Technology, Birmingham City University, UK
[2] School of Electronic, Electrical & Computer Engineering, University of Birmingham, UK
{islah.ali-maclachlan, munevver.kokuer, ian.williams, cham.athwal}
@bcu.ac.uk, p.jancovic@bham.ac.uk

ABSTRACT

In this paper, we investigate inter-flute and inter-player timbral variations in traditional Irish flute playing. A scale of individual notes played by three players by using six dimensionally and materially different flutes are digitally recorded and sound timbre is analysed by their spectral harmonic content. The long-term average spectrum (LTAS) and the short-term magnitude values at harmonic peaks are used to characterise the spectral harmonic content. The latter avoids the effect of any fluctuations in the fundamental frequency of played notes. The analysis quantifies the amount of variations across the flute players and flute models at each harmonic and explores the consistency of the timbral profile of each player. Experimental results demonstrate that there are in overall larger differences across the flute players than across the flute models from different manufacturers.

1. INTRODUCTION

The flute is a popular instrument within traditional Irish music, having a long history and enduring popularity (Breathnach, 1971; Duggan, 2009). Most analysis of flute timbre so far has focused on the cylindrical metal flute, developed in 1847 by Theobald Boehm, as it is regarded as the standard in classical and jazz (Coltman, 1971; Widholm et al., 2001).

Traditional Irish musicians predominantly play a wooden concert flute of the type used commonly by classical players before Boehm's metal flute. Whilst the wooden concert flute can have up to eight keys to make it fully chromatic, many of the flutes played by traditional musicians are unkeyed and play diatonically in the key of D, having a bottom note of D_4. The popular corpus of traditional tunes relies heavily on melodies collected by O'Neill (1998), published in 1907, and this includes mostly melodies that can be played on a model without keys. Notes outside the key of D can be produced by partially covering toneholes or using alternative fingerings.

Many current models are based on designs that were originally developed by Rudall & Rose and Boosey (Pratten), both London flutemakers in the mid 1800s (Larsen, 2003; Hamilton, 1990). In comparison to the standard design of metal "Boehm System" flutes, the wooden or concert flute preferred by traditional encompasses a range of designs that deliver the same notes. There are differences in materials, bore profiles and lengths, and variances in use of keys, tonehole and embouchure hole dimensions and position.

The role of the instrument in the production of a flute players' overall timbre has been studied on a number of occasions in a classical context. The work of Backus (1964) and Backus & Hundley (1966) initially developed the argument that wall material and thickness make little difference to the overall timbre of the flute by testing a range of artificially blown instruments. Coltman (1971) tested three unkeyed flutes made from silver, copper and wood and found that listeners, whether musically trained or not, were not able to distinguish between flutes made of the different materials or with different wall thicknesses. This work was followed by Widholm et al. (2001) who tested seven flutes manufactured by Muramatsu that were identical apart from their material. Widholm also found that the material used to manufacture a flute has a negligible effect on the overall timbre. The study did however find that individual players produce an almost consistent timbre across a range of flutes.

The aim of this study is to explore the timbral variations between diffrent players and a range of typical flute designs in traditional Irish flute playing, to ascertain which has more influence on the overall sound. We made recordings of a range of individual notes played by three different players with six different flute models. The analysis in this paper is performed using the sustained part of note G_4 recordings. Due to using only the sustained parts of individual note recordings, the sound timbre is determined by the spectral harmonic content. We perform the analysis by employing the long-term average spectrum (LTAS) and the short-term magnitude values at harmonic peaks. The LTAS is a classical method in speech and audio processing. It provides representative information of sound timbre and has been employed in a number of studies on analysis of speech (White, 2001; Leino, 2009; Sergeant & Welch, 2009) and music (Boersma & Kovacic, 2006). As fluctuations in the fundamental frequency of a played sound affect the conventional LTAS, and slight fluctuations were observed in our recordings, we also perform the analysis using the short-term magnitudes of the harmonic peaks. The harmonic peaks are localised semi-automatically. Experimental results show that the differences in individual players' timbres are larger than the timbral differences between flute models from different manufacturers. The timbral profiles of individual Irish flute players are distinctive even when playing various flute models.

2. DATA COLLECTION

For this analysis a set of flutes were selected to reflect the varying properties possible in wooden concert flutes. This was aimed at being reflective of the full representation of concert flutes available and cover several alternatives in manufacturing material and style. The flutes used, with their specifications, are presented in Table 1.

Recording controls were maintained throughout the experimentation, these included the use of a semi anechoic chamber, and a controlled microphone position, 20 cm away from and approximately one third of the distance between the head and the foot. The position was chosen to maintain tonal balance and minimize direct wind noise from the embouchure hole. Recordings were performed using the DPA 4090 microphone – this has a flat frequency response between 20Hz and 20kHz (+/-2dB) (Robjohns, 2006) and is therefore suitable for recordings of this type. The audio signal was sampled at 44100 Hz.

Three players were selected for the experiment. The players chosen varied from a beginner level, with under a year experience, to a player with over fifteen years of experience. All players were asked to maintain tonal consistency for all the recordings across each flute and to play a scale of individual notes from D_4 to B_5. Recordings of single notes were edited to remove the attack and decay sections as these areas are harmonically less stable (Keeler, 1972). As noted by Bamberger (2004), playing the lower notes on the flute require attention to the pressure as well as the size of the lip opening. Playing an unfamiliar flute also adds to this difficulty. Thus, the note G_4 was selected for analysis in this paper as it is further up the scale and not as susceptible to timbral differences based on these difficulties.

3. ACOUSTIC ANALYSIS

First, the volume differeces between the recordings are normalised out to ensure that each recording has the same average energy.

The sampled audio signal of each recording is then processed using the short-term Fourier analysis. The signal $x(n)$ is segmented into short overlapping analysis signal frames, with the length of the frame N set to 2048 samples (corresponding to approx. 46 ms) and the shift between adjacent frames L set to 512 samples. Each signal frame is multiplied by the Hamming window function. The windowed frames are then zero padded to have R samples, set to 4096 in our case, and the Fourier transform is aplied to provide the short-term Fourier spectrum $X(l,k)$ as given below

$$X(l,k) = \sum_{n=0}^{N-1} x(n+lL)w(n)e^{-j\frac{2\pi kn}{R}} \qquad (1)$$

where k and l is the frequency bin and frame-time index, respectively, and $w(n)$ denotes the analysis window function. Taking the absolute value of the $X(l,k)$ gives the short-term magnitude spectrum. The range of magnitude values is usually compressed by applying $20log_{10}$, resulting in decibel scale. The collection of the short-time magnitude spectrum over time is also referred to as spectrogram. An example of the time domain signal (with the attack and decay sections being removed) and its corresponding spectrogram is depicted in Figure 1 for player A playing G_4 note on flute 'Wallis'. It can be seen that the magnitude values at frequencies above approximately 5 kHz are small. Since similar trend was also observed for all other recordings, only the frequencies up to 5 kHz were used in our analysis.

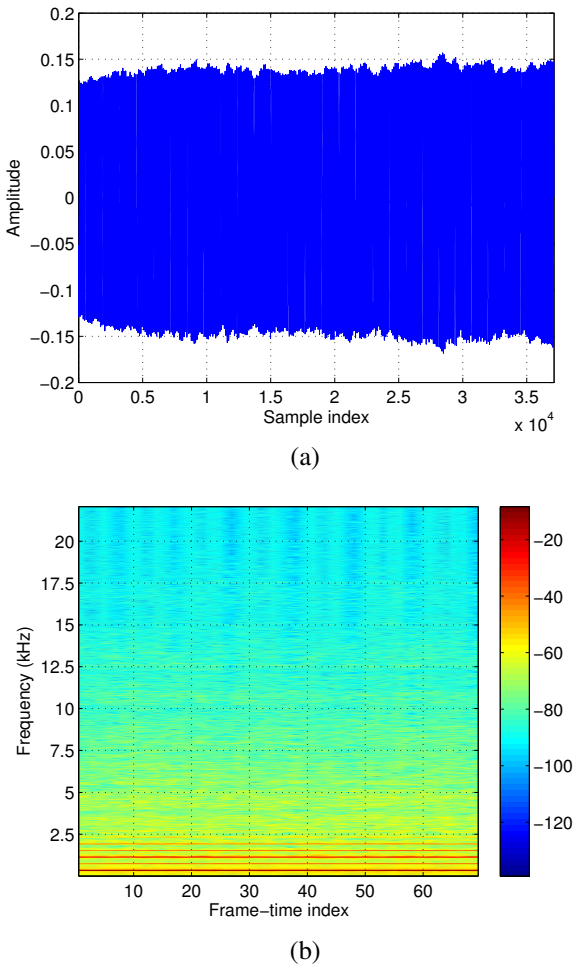


Figure 1: Flute 'Wallis' playing note G_4: waveform (a) and the corresponding spectrogram (b).

In order to obtain information about average spectral properties of the instruments and players, each recording was subjected to a long-term average spectrum (LTAS) analysis. LTAS analysis can provide information on sound timbre and is particularly useful when persistent spectral features are under investigation (White, 2001). LTAS is computed by averaging the short-term Fourier magnitude spectra over time, resulting in a single feature vector representing each of the recording.

In addition to using the LTAS, we also conducted the analysis using only the information on the short-term magnitudes of the first few harmonics. In this paper, the harmonics were identified semi-automatically based on the

Flute ID	Manufacturer	Keys	Material	Details
1	deKeyser	0	African Blackwood	C foot, Pratten
2	Dixon	0	Polymer	D foot, Rudall & Rose
3	McMahon	0	African Blackwood	C foot, Rudall & Rose
4	Sweetheart	0	Maple	One piece body, D foot
5	Vignoles	0	Blackwood Body, Polymer Head	C foot, Pratten
6	Wallis	8	African Blackwood	C foot

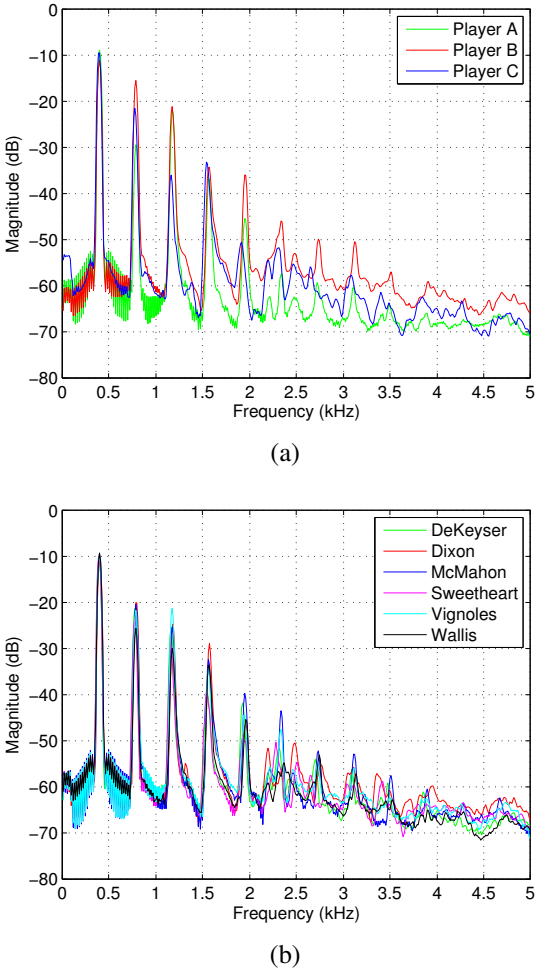
Table 1: Types of flutes used in the study.

knowledge of the note played, i.e., the harmonics were located by finding peaks around the multiples of the note frequency. In an automated system, the harmonics could be identified by employing methods for detection of sinusoidal components we presented in (Jančovič & Köküer, 2007, 2011b). These methods do not require the information about the fundamental frequency and have been employed for processing of speech and audio signals, e.g., (Jančovič & Köküer, 2009, 2011a).

4. EXPERIMENTAL RESULTS

4.1 Analysis employing the LTAS

Here we use the long-term average spectrum (LTAS) to analyse the inter-flute and inter-player timbral differences. This is performed by calculating the mean LTAS of each player, obtained by averaging over the flutes, and of each flute, obtained by averaging over the players. The resulting mean LTAS are depicted for each player in Figure 2(a) and for each flute in Figure 2(b). It can be seen that the mean LTAS of individual players vary mostly in the 2^{nd}, 3^{rd} and 5^{th} harmonic, while the mean LTAS of individual flutes vary mostly in the 3^{rd}, 4^{th}, and 5^{th} harmonic. In overall, the mean LTAS of individual players differ more than those of individual flutes. Closer look at Figure 2(a) reveals that player A is weaker on the 2nd harmonic, while player B has a strong 2^{nd} and 5^{th} harmonic and player C is weaker on the 3^{rd} and 5^{th} harmonic. Having observed the differences between the players in the mean LTAS averaged over all flutes in Figure 2(a), we examine in Figure 3 the LTAS of all the six flutes for each player A, B and C. It can be seen that there are common traits in the timbre of each player across all of their recordings with different flutes, for instance, player B has strong the 2^{nd} harmonic consistently over all the flutes.



(a)

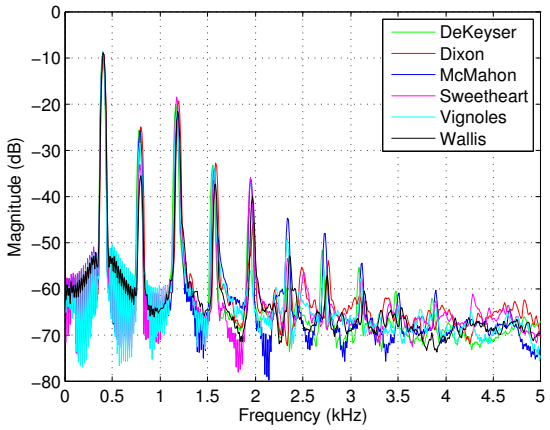
(b)

Figure 2: The LTAS of each player averaged over six flutes (a) and of each flute averaged over three players (b).

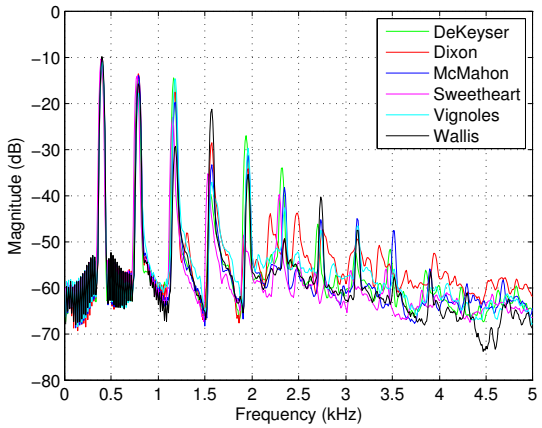
4.2 Analysis employing the short-term harmonic peaks magnitudes

While the conventional LTAS as used above can provides useful information, it is susceptible to variations in the fundamental frequency. We have noticed that in our recordings there were some small fluctuations in the fundamental frequency of the played note, resulting in the harmonics (especially the higher ones) of each short-term spectrum not being accurately aligned. This effect of misalignment of the harmonics could be avoided by using a pitch-corrected LTAS, similarly as in (Boersma & Kovacic, 2006), or alternatively by localising the harmonic peaks in the short-term specrum and using their magnitudes only. Our analyses in this section are performed using the latter, i.e., magnitudes of the harmonics peaks.

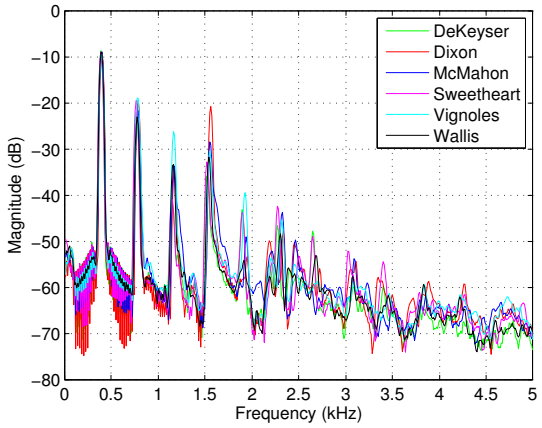
First, we quantify further the timbral variances between the players and flutes. For each recording, we calculate the mean magnitude value of each localised harmonic peak over all the signal frames. Then the standard deviation of these mean magnitudes of harmonics is calculated for each player across the flutes, presented in Table 2, and for each flute across the players, presented in Table 3. It can be seen that in most cases there are smaller variations across the flutes than across players. Figure 4 shows the average



(a)



(b)



(c)

Figure 3: The LTAS of six flutes for player A (a), B (b) and C (c).

standard deviations for each player and each flute, in both cases averaged over the first five harmonics. The overall average standard deviation across the flutes is 3.7 dB while across the players is 6.1 dB.

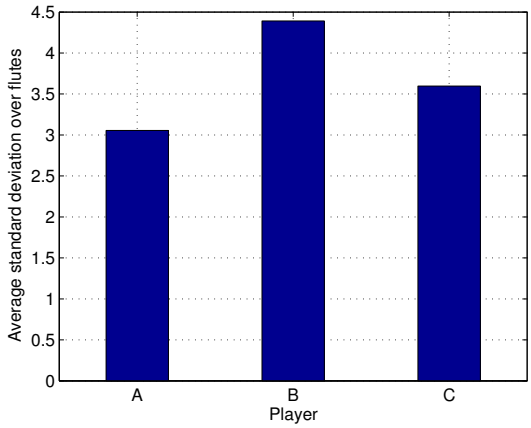
Next, we perform evaluations using the short-term harmonic magnitudes of all the signal frames. In addition to the effect of players and instruments, this will also reveal the effect of frame-to-frame variations. Figure 5 depicts 2-dimensional scatter plots of the magnitudes of one

Standard deviation across flutes for	Index of the harmonic				
	1	2	3	4	5
Player A	0.2	4.4	1.7	1.9	7.0
Player B	0.9	1.6	5.7	6.4	7.4
Player C	0.2	1.8	5.4	5.1	5.5
Average	0.4	2.6	4.3	4.5	6.6

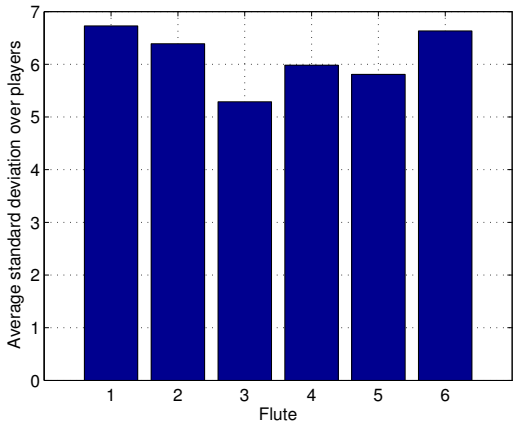
Table 2: The standard deviations (in dB) across flutes for the first five harmonics.

Standard deviation across players for flute	Index of the harmonic				
	1	2	3	4	5
deKeyser (1)	2.0	5.5	14.0	2.7	9.4
Dixon (2)	1.5	5.7	9.4	6.0	9.3
McMahon (3)	1.1	6.1	7.2	3.1	9.0
Sweetheart (4)	0.8	9.8	11.3	2.0	5.9
Vignoles (5)	1.2	6.0	5.9	3.2	12.7
Wallis (6)	0.6	10.0	5.7	8.2	8.7
Average	1.2	7.2	8.9	4.2	9.2

Table 3: The standard deviations (in dB) across players for the first five harmonics.



(a)



(b)

Figure 4: The standard deviations (in dB) of the harmonic peaks magnitudes for each player (a) and flute (b), averaged over the first five harmonic peaks.

harmonic peak against other harmonic peak for all signal frames of all recordings. The individual players are indicated by different shape and colour markers. It can be seen that the individual players can be well separated using a combination of the first five harmonic magnitudes. The figures show that the Player A, denoted by green square mark in Figure 5, is contained in few sub-clusters, which we found to correspond to different flutes. This can be observed in Figure 6 depicting the same scatter plots of the 2^{nd} against the 3^{rd} and the 2^{nd} against the 4^{th} harmonics with the markers indicating now the individual flutes. The variations for Player A within each flute are the smallest across the players, which indicates that the Player A plays each given instrument in most consistent way, i.e., most stable timbre. This is in contrast to Player C, who shows large variations for most of the flutes, especially, Sweetheart and deKeyser. Figure 6 also shows that the inter-player differences are larger than the inter-flute differences, confirming the results of variance analysis presented earlier.

5. CONCLUSION

This paper presented an analysis of the timbral variations in traditional Irish flute playing. The analysis were performed using isolated recordings of note G_4, played by three different players with six different flutes. The attack and decay sections of the recordings were not used in the analysis due to being harmonically less stable. The analysis was performed by employing the long-term average spectra (LTAS) and also employing the short-term magnitudes of the harmonic peaks. The latter avoids the effect of any fluctuations in the fundamental frequency of the played notes. Experimental results demonstrated that the LTAS of individual players differ more than those of individual flutes. The analysis of variations of the harmonic magnitudes showed that in most cases there are smaller variations across the flutes than across the players, with the overall average standard deviation being 3.7 dB across the flutes and 6.1 dB across the players. The short-term analysis of harmonic magnitudes of all signal frames revealed how consistent the timbral profile of each player was when using a particular flute.

Our near future work will focus on timbral analysis of other individual note recordings as well as further statistical analysis of the short-term harmonic magnitudes, for instance, an employment of the principal component analysis to determine the main eigenvectors representing the timbral variations due to players and due to instruments. Our overall aim is to research novel methods for analysis of musical styles in continuous song recordings.

6. REFERENCES

Backus, J. (1964). Effect of wall material on the steady-state tone quality of woodwind instruments. *The Journal of the Acoustical Society of America*, *36*(10), 1881–1887.

Backus, J. & Hundley, T. C. (1966). Wall vibrations in flue organ pipes and their effect on tone. *The Journal of the Acoustical Society of America*, *39*(5A), 936–945.

Bamberger, A. (2004). Operation of flutes at low pitch investigated with PIV. *Proc. Int. Symp. on Musical Acoustics*, 243–245.

Boersma, P. & Kovacic, G. (2006). Spectral characteristics of three styles of Croatian folk singing. *The Journal of the Acoustical Society of America*, *119*(3), 1805–1816.

Breathnach, B. (1971). *Folk Music and Dances of Ireland*. London: Ossian.

Coltman, J. W. (1971). Effect of material on flute tone quality. *The Journal of the Acoustical Society of America*, *49*, 520.

Duggan, B. (2009). *Machine Annotation of Traditional Irish Dance Music*. PhD thesis, Dublin Institute of Technology, School of Computing, Dublin.

Hamilton, S. C. (1990). *The Irish Flute Player's Handbook*. Breac Publications, Eire.

Jančovič, P. & Köküer, M. (2007). Estimation of voicing-character of speech spectra based on spectral shape. *IEEE Signal Processing Letters*, *14*(1), 66–69.

Jančovič, P. & Köküer, M. (2009). Incorporating the voicing information into HMM-based automatic speech recognition in noisy environments. *Speech Communication*, *51*(14), 438–451.

Jančovič, P. & Köküer, M. (2011a). Automatic detection and recognition of tonal bird sounds in noisy environments. *EURASIP Journal on Advances in Signal Processing*, 1–10.

Jančovič, P. & Köküer, M. (2011b). Detection of sinusoidal signals in noise by probabilistic modelling of the spectral magnitude shape and phase continuity. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Prague, Czech Republic*, 517–520.

Keeler, J. (1972). The attack transients of some organ pipes. *IEEE Trans. on Audio and Electroacoustics*, *20*(5), 378–391.

Larsen, G. (2003). *The Essential Guide to Irish Flute and Tin Whistle*. Pacific, Missouri, USA: Mel Bay Publications.

Leino, T. (2009). Long-term average spectrum in screening of voice quality in speech: Untrained male university students. *Journal of Voice*, *23*(6), 671 – 676.

O'Neill, F. (1998). *O'Neill's music of Ireland*. Pacific, Missouri: Mel Bay Publications.

Robjohns, H. (2006). DPA 4090 & 4091 microphone reviews. "http://www.soundonsound.com/sos/aug06/articles/dpa.htm". [Online; accessed 02-March-2013].

Sergeant, D. C. & Welch, G. F. (2009). Gender differences in long-term average spectra of children's singing voices. *Journal of Voice*, *23*(3), 319 – 336.

White, P. (2001). Long-term average spectrum (LTAS) analysis of sex- and gender-related differences in children's voices. *Logoped Phoniatr Vocol.*, *26*(3), 97–101.

Widholm, G., Linortner, R., Kausel, W., & Bertsch, M. (2001). Silver, gold, platinum - and the sound of the flute. *Proc. of the Int. Symposium on Musical Acoustics, Perugia*, *1*, 277–280.

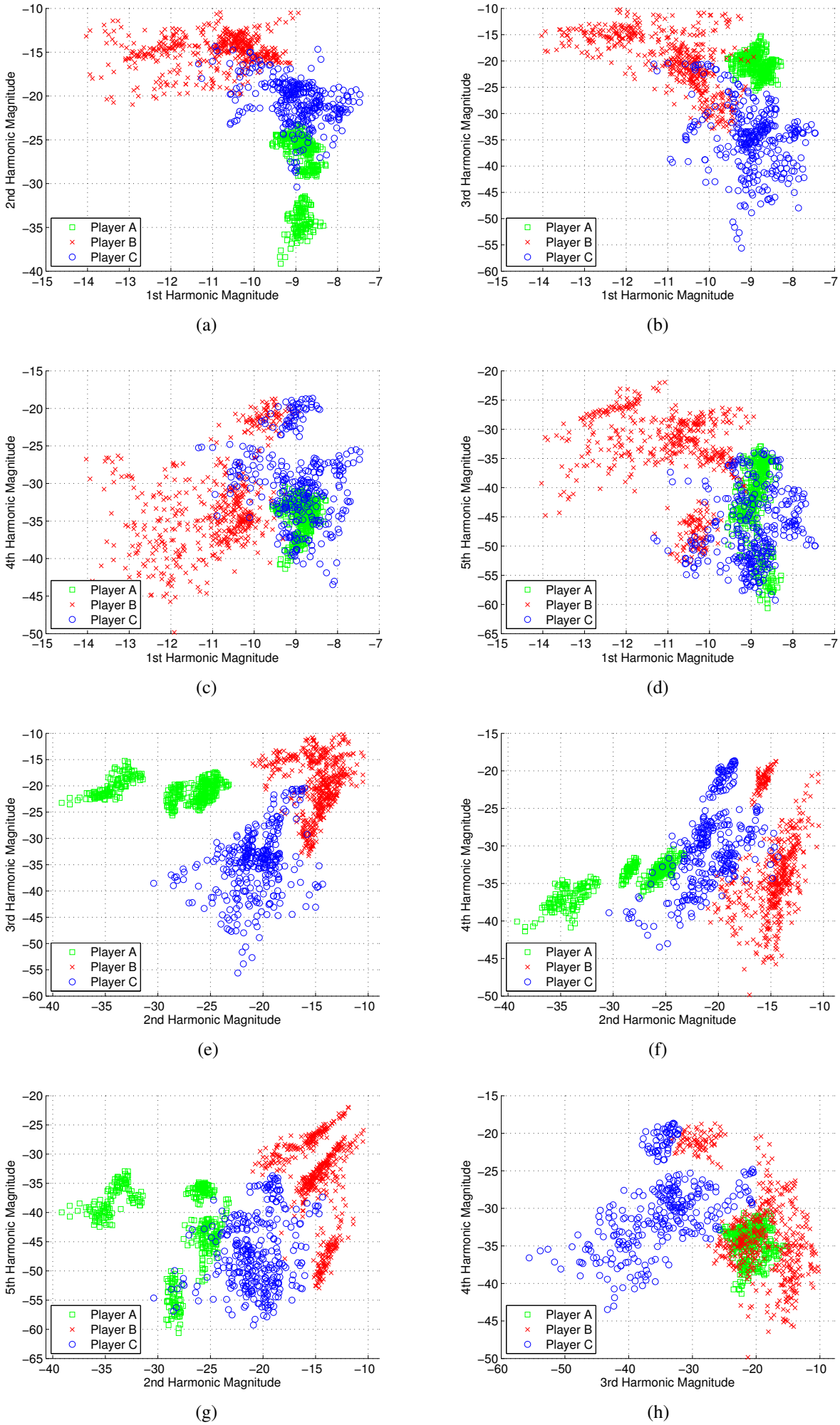
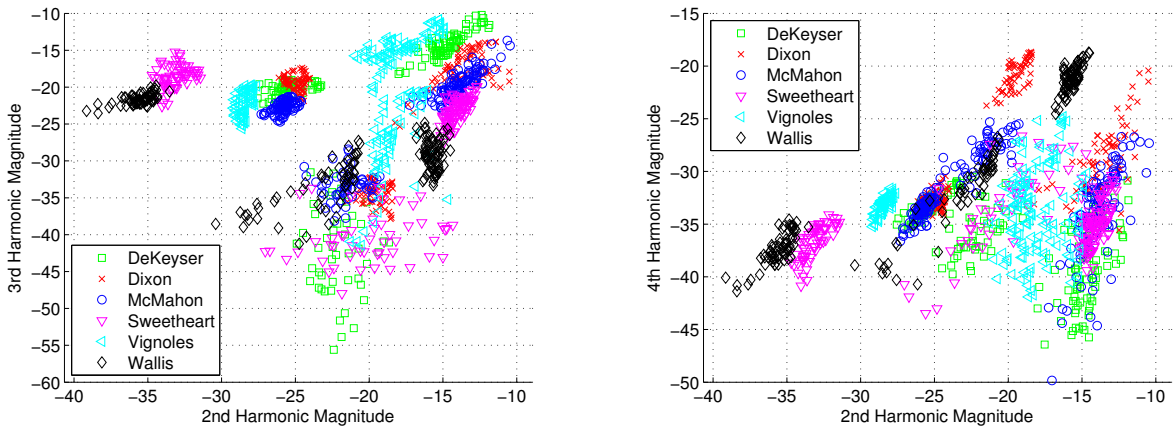


Figure 5: Scatter plots of the short-term magnitude values (in dB) at the harmonic peaks, indicated for individual players.



(a) (b)

Figure 6: Scatter plots of the short-term magnitude values (in dB) at the harmonic peaks, indicated for individual flutes.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

IDIOM-INDEPENDENT HARMONIC PATTERN RECOGNITION BASED ON A NOVEL CHORD TRANSITION REPRESENTATION

Emilios Cambouropoulos, Andreas Katsiavalos, Costas Tsougras

School of Music Studies, Aristotle University of Thessaloniki, Greece
emilios@mus.auth.gr, akatsiav@mus.auth.gr, tsougras@mus.auth.gr

ABSTRACT

In this paper, a novel chord transition representation (Cambouropoulos 2012) is explored in a harmonic recognition task. This representation allows the encoding of chord transitions at a level higher that individual notes that is transposition-invariant and idiom-independent (analogous to pitch intervals that represent transitions between notes). A harmonic transition between two chords is represented by a *Directed Interval Class (DIC) vector*. The proposed 12-dimensional vector encodes the number of occurrence of all directional interval classes (from 0 to 6 including +/- for direction) between all the pairs of notes of two successive chords. Apart from octave equivalence and interval inversion equivalence, this representation preserves directionality of intervals (up or down).

A small database is constructed comprising of chord sequences derived from diverse music idioms/styles (tonal music, different traditional harmonic idioms, 20th century non-tonal harmonic idioms). The proposed DIC representation is evaluated on a harmonic recognition task, i.e. we examine the accuracy of recognition of harmonic queries in this database. The results of the algorithm are judged by human music analysis experts.

It is suggested that the proposed idiom-independent chord transition representation is adequate for representing harmonic relations in music from diverse musical idioms (in equal temperament) and, therefore, may provide a most appropriate framework for harmonic processing in the domain of computational ethnomusicology.

1. INTRODUCTION

In recent years an increasing number of studies propose computational models that attempt to determine an appropriate representation and similarity measure for the harmonic comparison of two excerpts/pieces of music, primarily for music information retrieval tasks (Allali 2007, 2010; de Haas, 2008, 2011; Hanna et al., 2009; Paiement 2005; Pickens et al. 2002). Such models assume a certain representation of chords and, then, define a similarity metric to measure the distance between chord sequences. Chords usually are represented either as collections of pitch-related values (e.g., note names, MIDI pitch numbers, pitch class sets, chroma vectors, etc.), or as chord root transitions within a given tonality following sophisticated harmonic analysis (e.g. roman numeral analysis, guitar chords, etc.).

In the case of an absolute pitch representation (such as chroma vectors) transpositions are not accounted for (e.g., twelve transpositions of a given query are necessary to find all possible occurrences of the query in a dataset). On the other hand, if harmonic analytic models are used to derive a harmonic description of pieces (e.g., chords as degrees within keys or tonal functions), more sophisticated processing is possible; in this case, however, models rely on complicated harmonic analytic systems, and, additionally, are limited to the tonal idiom.

All the above models rely on some representation of individual chords. There are very few attempts, however, to represent chord *transitions*. For instance, de Haas et al. (2008, 2011) represent chord transitions as chord distance values adapting a distance metric from Lerdahl's *Tonal Pitch Space* (2001); however, a chord transition being represented by a single integer value seems to be an excessive abstraction that potentially misses out important information.

This paper explores a richer chord transition representation that can be extracted directly from the chord surface and that is idiom-independent (Cambouropoulos 2012). In the newly proposed chord transition representation, a harmonic transition between two chords can be represented as a *Directional Interval Class (DIC)* vector. The proposed 12-dimensional vector encodes the number of occurrence of all directional interval classes (from 0 to 6 including +/- sign for direction) between all the pairs of notes of two successive chords. As melodic intervals represent a melodic sequence in an idiom-independent manner, so does the DIC vector represent chord transitions in an idiom-independent manner. This means that, given a dataset comprising music pieces represented as sequences of chords, a given chord sequence query can be searched for directly in the chord progressions without any need for harmonic analysis.

In this study, a small database is constructed comprising of chord sequences (i.e. the main chord notes that form the underlying harmonic progression of any piece without root/key knowledge) derived from diverse music idioms/styles. More specifically, we include standard chord progressions from Bach chorales and along with harmonic progressions from modal Greek rebetiko songs, polyphonic songs from Epirus, Beatle songs and non-tonal pieces by B. Bartók, O. Messiaen, C. Debussy, E. Satie.

The proposed DIC representation is evaluated on a harmonic recognition task, i.e. we examine the accuracy of recognition of harmonic queries in the above database.

Both the query sequence and the chord progressions in the dataset are converted to DIC vectors and exact matching for recognition is employed (approximate matching is also considered). The results of the algorithm are judged by human music analysis experts.

In the first section below the new Directed Interval Class (DIC) representation is introduced and some of its potentially useful properties are highlighted. Then, the DIC representation will be used as the basis for a preliminary test on a harmonic recognition task for different musical idioms. Finally, a brief discussion will summarise the importance of the proposed representation along with problems and shortcomings, and will suggest interesting new avenues for further exploration.

2. THE DIRECTED INTERVAL CLASS (DIC) CHORD TRANSITION REPRESENTATION

A novel chord transition representation is proposed. A harmonic transition between two chords can be represented as a Directional Interval Class (DIC) vector. The proposed 12-dimensional vector encodes the number of occurrence of all directional interval classes (from 0 to 6 including +/- sign for direction) between all the pairs of notes of two successive chords. That is, from each note of the first chord all intervals to all the notes of the second chord are calculated. Direction of intervals is preserved (+,-), except for the unison (0) and the tritone (6) that are undirected. Interval size takes values from 0-6 (interval class). If an interval X is greater than 6, then its complement 12-X in the opposite direction is retained (e.g. ascending minor seventh '+10' is replaced by its equivalent complement descending major second '-2').

The 12-dimensional DIC vector features the following directed interval classes in its twelve positions: 0 (unison), +1, -1, +2, -2, +3, -3, +4, -4, +5, -5, 6 (tritone). For instance, the transition vector for the progression I→V is given by the DIC vector: Q = <1,0,1,1,1,1,0,1,0,0,3,0> (which means: 1 unison, 0 ascending minor seconds, 1 descending minor second, 1 ascending major second, etc.) – see Figure 1, and further examples in Figure 2.

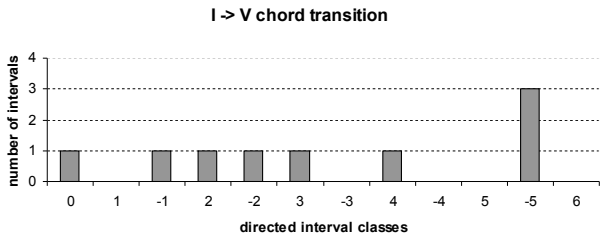


Figure 1 The DIC vector: <1,0,1,1,1,1,0,1,0,0,3,0> for the chord transition I→V depicted as a bar graph.

The DIC vector is unique for many tonal chord transitions. However, there are a number of cases where different tonal transitions have the same vector. For instance, the transitions I→V and IV→I share the same DIC vector as their directed interval content is the same; it should be noted, that, heard in isolation (without a tonal centre reference), a human listener cannot tell the difference between these two transitions.

The proposed DIC representation preserves directionality of intervals (up or down), and, therefore, it incorporates properties of voice leading. For instance, the DIC vector naturally accommodates chord transition asymmetry. If the two chords in a chord transition are reversed, the absolute values of intervals are retained; however, the directions of intervals are reversed. This way, the vectors, for instance, for the I→V transition and the V→I transition, are different (compare, DIC vectors of Figure 1 and Figure 2a-top).

It was initially hypothesised that the DIC vector (Cambouropoulos, 2012) uniquely determines the two chords that comprise the transition (except for cases when one of the two chords is symmetric, such as augmented chord, or diminished seventh chord). This is actually not true. Any specific chord transition has the same DIC vector with its retrograde inversion. This is an inherent problem in the DIC representation that introduces certain limitations in specific contexts. See next section for further discussion.

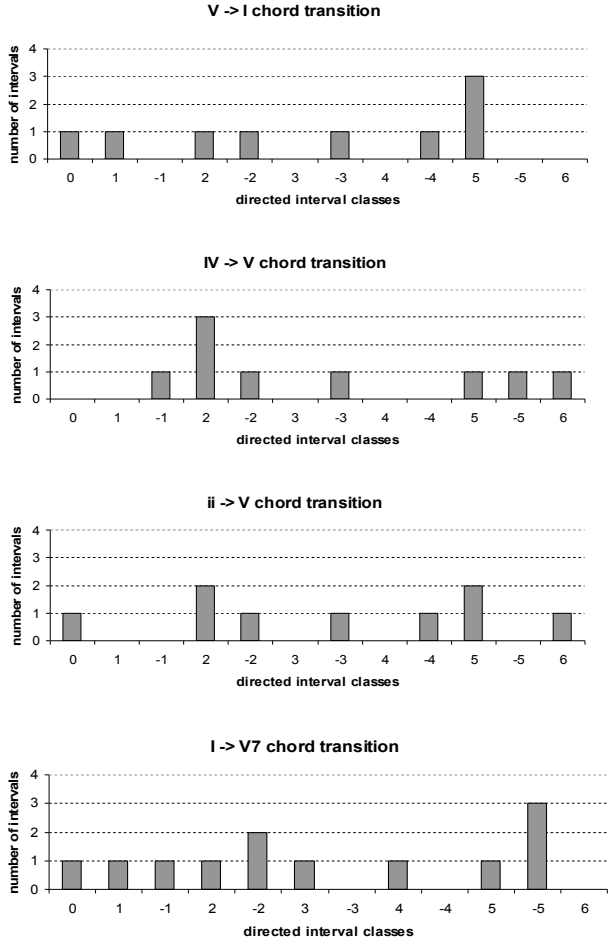


Figure 2 DIC vectors for four standard tonal chord transitions: V→I, IV→V, ii→V, I→V7.

3. HARMONIC RECOGNITION

In this study the proposed DIC representation is tested on a harmonic recognition task, i.e. we examine the accuracy of recognition of harmonic queries in a small diverse-style harmonic database. The results of the algorithm are judged qualitatively by human music analysis experts. The music database, the DIC-vector-based harmonic recognition prototype and results will be presented in the sections below.

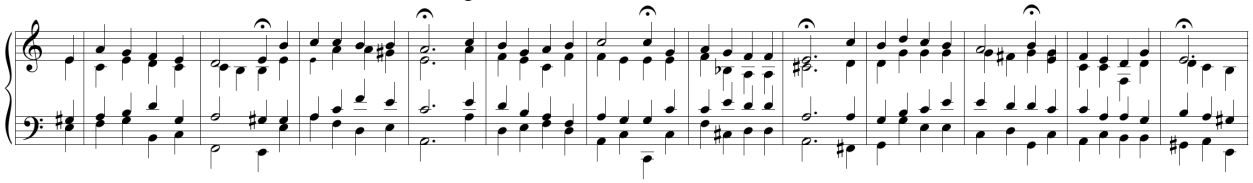
3.1 Test Dataset

A small database is constructed comprising of chord sequences (i.e. the main chord notes that form the underlying harmonic progression of any piece without root/key information) derived from diverse music idioms/styles. More specifically, this purpose-made collection comprises of 31 chord reductions of music pieces reaching an overall number of 957 chords (4 Greek modal rebetiko songs, 3 polyphonic songs from Epirus, non-tonal pieces by B. Bartók, O. Messiaen, C. Debussy, E. Satie, 11 Beatle songs, and 5 chorales by J.S.Bach). Below is a brief description of the main harmonic features of each style. Examples of chordal reductions are illustrated in Figure 3. The dataset comprises of the following chord sequences:

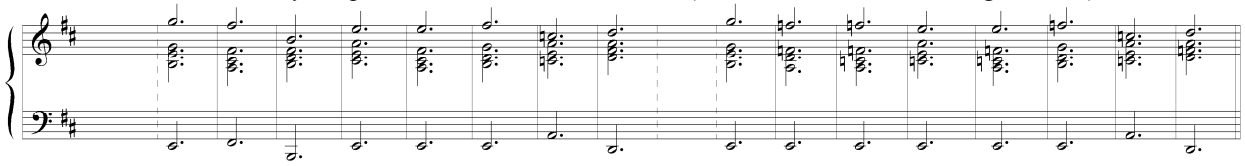
- *Five chorales* by J. S. Bach, after their rhythmical reduction to a quarter-note harmonic rhythm and the removal of non-harmonic tones. The following chorales were used (the numbers correspond to the Breitkopf Edition): 20, 80, 101, 138, 345. Two different harmonizations (80 & 345) of "O Haupt..." were used, one in D major and the other in A minor. This material typically represents tonal harmonic progressions of the Baroque period.
- Erik Satie's *Gymnopedie no 1*, reduced to chordal progressions (one chord per bar). This material represents diatonic modal harmony of the early 20th century and the modal interchange/modulation procedure, where different diatonic modes are used, either with the same or with different pitch centers. One characteristic difference of this idiom from the tonal harmonic idiom is that the chord progressions are more unrestricted and do not adhere to standard functional relations and progressional formulae (for further discussion and analyses see Tsougras 2003).
- Excerpts of music by Claude Debussy, reduced to chord progressions: selected chordal sequences (b. 1-5, b. 13-16, b. 17-21, b. 42-43) from *Nuages* (1st movement of *Nocturnes* for orchestra) and the opening bars (b. 1-14) from *Claire de Lune* (nr. 3 from *Suite Bergamasque* for solo piano). This material represents the fluent "impressionistic" harmony of the early 20th century, with its relatively free dissonances and frequent planning procedure (parallel harmony, either diatonic or real/chromatic).
- Olivier Messiaen's chordal sequence from the piano part of *Liturgie du Cristal* (1st movement of the *Quartet for the End of Time)*. This 17-chord sequence corresponds to the 17-part isorhythmic pattern of the piece (*talea*), and is part of the full 29-chord pattern that is in constant repetition (*color*). This material represents Messiaen's idiomatic modal harmony, based on his system of seven symmetrical chromatic modes of limited transposition.
- Béla Bartók's *Romanian Folk Dances* for piano nr. 4 *(Mountain Horn Song)* and nr. 6 *(Little One*, b. 1-16). These pieces are essentially harmonizations of original Romanian folk tunes from Transylvania, recorded and transcribed by Bartók himself. The original tunes are modal, either diatonic of chromatic, but the harmonizations create more complex sonorities and involve mixing of modes and symmetrical pitch structures. This material represents the initial stage of Bartók's "polymodal chromaticism" (a term used by the composer to describe modal mixing). For further discussion and analyses see Tsougras 2009.
- Three *Polyphonic Songs from Epirus* (Epirus is a region of northwestern Greece), reduced to vertical sonorities, and without the idiomatic glissandi or other embellishment types featured in the idiom. The reductions were produced from transcriptions made by K. Lolis (2006). This very old 2-voice to 5-voice polyphonic singing tradition is based on the anhemitonic pentatonic pitch collection - described as pc set (0,2,4,7,9) - that functions as source for both the melodic and harmonic content of the music. The songs chosen for the analysis are: *Την άμμο-άμμο πήγαινα* (4-voice, scale: **G**-Bb-C-D-F), *Έπιασα μια πέρδικα* (3-voice, scale: D-F-**G**-Bb-C), *Πέρασα 'πο 'να γιοφύρι* (4-voice, scale: G#-B-**C#**-E-F#).
- Four Greek *rebetiko songs*, reduced only to chord progressions. Each song's melody is based on a particular mode, called "dromos" (Greek "laikoi dromoi" have affinities with Turkish makams and frequently bear similar names, but they are quite different in their intervallic structure, since they are adapted to the equally-tempered scale and do not incorporate microtones). The songs chosen for testing were: *Με παράσυρε το ρέμα* by Vassilis Tsitsanis (D Ussak), *Απόψε φίλα με* by Manolis Chiotis (D hicaz), *Καϊξής* by Apostolos Chatzichristos (D natural minor) and *Πασατέμπος* by Manolis Chiotis (D hicaz/hicazkar). This material represents some cases of the idiomatic modal harmony of Greek "rebetiko". This modal harmony roughly results from the formation of tertian chords above structural pitches of the current mode and the use of certain cadence formulae for each mode.
- Eleven Beatles songs, reduced to chord sequences (expressed in conventional chord symbols, e.g. Cm, C^{M7}, etc). The harmony of these popular songs incorporates a number of diverse influences (which range from blues and rock n' roll to pop ballad and classical songs), together with Lennon & McCartney's original progressions and harmonic style. The chosen material, which attempts to represent most of these influences and originalities, comprises the following songs: *All my loving, From me to you, She loves you, A hard day's night, Help, Michelle, Misery, Because, Yesterday, Penny Lane, Strawberry fields forever*.

The musical style that is more likely to have unique harmonic progressions is Messiaen's: the highly chromatic 5-note to 7-note sonorities (see Figure 3) constitute a highly individual harmonic idiom, and no common elements can be found among the other styles in the present data. Another quite unique style is the Polyphonic Epirus singing, with its purely pentatonic harmonic content and unresolved dissonances. The other styles are mainly based on tertian harmony, with Debussy's and Bartók's styles containing extensions or deviations. Only Bach's style is explicitly tonal, i.e. it is based on standard harmonic functions (mainly of the type S-D-T, especially on cadences), while the other tertian harmonic styles are idiomatically modal.

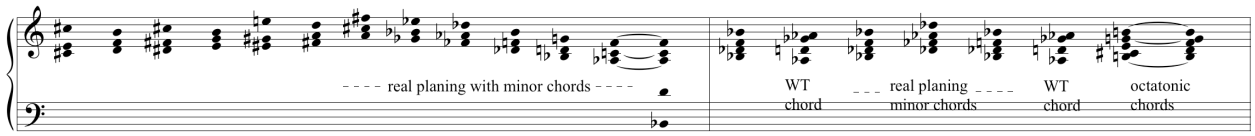
Reduction of Bach's chorale nr. 345 *O Haupt...*



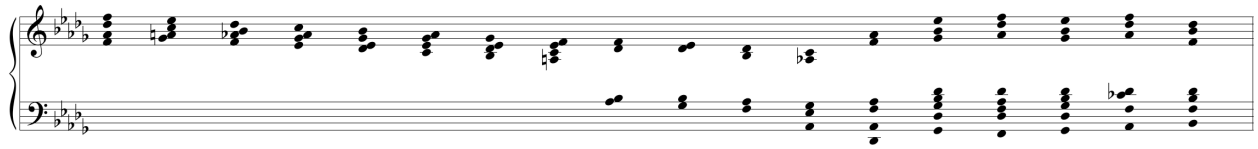
Reduction of Erik Satie's *Gymnopedie no 1*, b. 32-39 and 71-78 (for further details see Tsougras 2003)



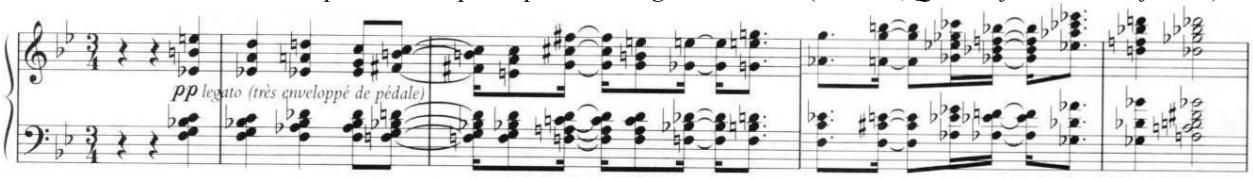
Claude Debussy, *Nuages* chord progressions (b. 13-16 and b. 42-43):



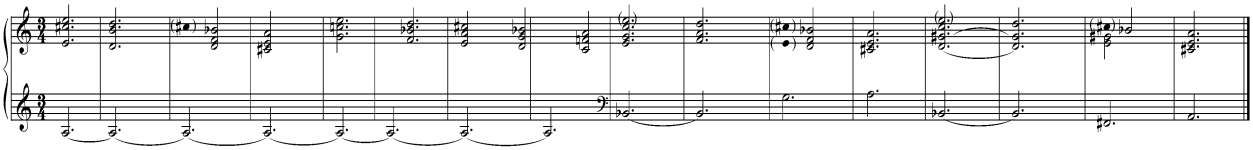
Reduction of Claude Debussy's *Clair de Lune:* (b. 1-14):



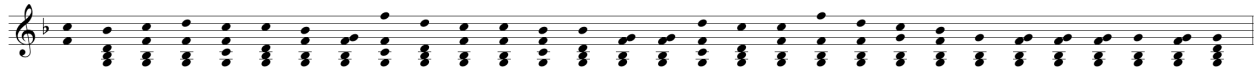
Olivier Messiaen's chordal sequence from piano part of *Liturgie du Cristal* (1st mov, *Quartet for the End of Time*)



Béla Bartók's *Romanian Folk Dance nr. 4* reduction (b. 3-16) (for further details see Tsougras 2009)



Reduction of Polyphonic Song from Epirus *Tin ammo-ammo pigena*



Chord sequence of rebetiko song *Pasatempos* by Manolis Chiotis (D hicaz/hicaskar):

Cm-Cm^{6}-Gm-D-Gm-Gdim-D-Cm-D-A-D-Gm-Cm-D^{7}-Gm-Cm-Eb-D

Chord sequence of song *A hard day's night* by the Beatles:

C-F-C-Bb-C-F-C-Bb-C-F-G7-C-F7-C-Em-Am-Em-C-Am-F7-G7-C-F-C-F7-C-F-C-F7-Bb-C

Figure 3 Examples of chord sequences from diverse harmonic idoms contained in the test dataset.

3.2 Harmonic Recognition Model

In order to test the potential of the DIC representation a simple computer application (in Java) was devised. The user inputs a harmonic query (sequence of chords) and the system finds exact matches in a given database of chord reductions. Both the query sequence and the chord progressions in the dataset are converted to DIC vectors and, then, exact matching for recognition is employed. Approximate matching is currently available by use of wild cards (i.e. DIC vector entries can be replaced by wild cards); δ & γ approximate matching is also possible but not explored in current study. The user interface is depicted in Figure 4; the exact positions of each match are listed in a side window (not depicted in Figure 4).

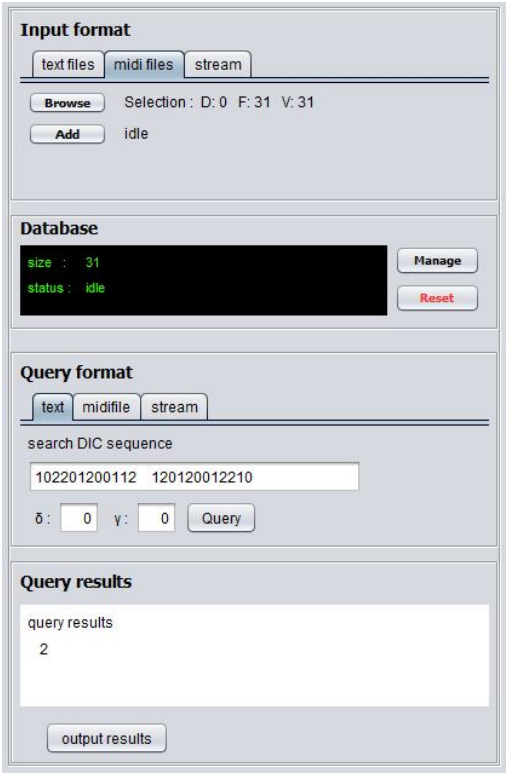


Figure 4 User interface of DIC-vector-based harmonic matching prototype

3.3 Results & Discussion

The harmonic recognition model can reveal individual harmonic elements that characterize the styles in our purpose-made database as well as common elements among them. We tested a large number of queries on the given dataset.

For this small dataset, relatively longer sequences comprising of four or more chords were uniquely identified in the correct position of the music piece from which they originated. For instance, we examined exhaustively Bach chorale 20 in terms of the longest repeating subsequence; the longest sequence found in at least one other piece was a 4-chord sequence (first four chords identified in position 26 of Strawberry Fields). Of course if the dataset is significantly extended we expect to find more occurrences of relatively longer harmonic sequences.

Below is a selection of specific harmonic progressions that were investigated and our comments on the results obtained:

- chorale style tonal cadence in major tonality: $ii^6{}_5$-V-I: found 5 times, only in Bach chorales (this is a characteristic cadence type of the chorale idiom)
- chorale style tonal cadence in minor tonality: $ii_o{}^6{}_5$-V-i: found 3 times, only in Bach chorales (this is a characteristic cadence type of the chorale idiom)
- major triad progressing to minor triad one perfect 5th down: found 21 times in Bach chorales, one Beatles song (Michelle) and three Rembetika songs (Απόψε φίλα με, Πασατέμπος, Με παρέσυρε το ρέμα)
- major triad progressing to major triad one perfect 5th down: found 100 times in Bach chorales, certain Beatles songs, Satie and certain Rebetika songs. This pattern is very common in styles based on the circle of fifths harmonic progression, namely tonal music and diatonic modal music. It is not encountered in chromatic modal styles, or in other idiomatic styles (Debussy, Bartok, Messiaen, Epirus songs). See further extended comments below.
- major triad progressing to major triad one major second higher: found 23 times in Beatle songs, rebetiko songs, chorales, Satie (and once in Debussy).
- major triad seventh progressing to major triad a perfect fifth lower (perfect cadence): found 45 times only in Bach and Beatle songs (and once in Debussy). Not encountered in the other non-tonal idioms.
- major triad with major 7th progressing to major triad with major 7th one perfect 5th down: found 8 times, only in Satie's *Gymnopedies*
- major triad progressing to minor triad one major 2nd down: found 1 time, in one Rebetiko song (Πασατέμπος). This progression is expected to be found more often if rebetiko dataset is enlarged especially for hicaz mode.
- minor triad progressing to major triad one major 2nd up: found 4 times, in one Beatles song (Michelle) and two Rebetika songs both in hicaz mode (Απόψε φίλα με, Πασατέμπος).
- minor triad progressing to minor triad one minor 3rd down: found 4 times, only in Debussy's Nuages (as part of real planing with parallel minor chords). Most chord sequences tested on Nuages were unique to this idiomatic harmonic language.
- the third chord transition Debussy's Clair de Lune is identified three times in this piece (positions 3, 11, 16 in Figure 3). Many transitions are unique in this piece.
- The vast majority of chord transitions in the polyphonic songs from Epirus are identified only in this music style. For instance, the second chord transition in the reduction of Polyphonic Song from Epirus "Tin ammo-ammo pigena" (G-Bb-D → G-Bb-F-C) is identified twice in this song (position 2 and 9)

- Most chord transitions in the pieces by Satie, Debussy, Bartók and Messiaen are uniquely identified in the respective pieces and are characteristic of the specific harmonic styles (always in the context of this small database).

Overall, the harmonic recognition model behaves as expected, and has sufficient distinctive power to discern the harmonic individualities of these different harmonic languages. Almost all of the harmonic queries were correctly detected (see one problem below) and all queries of at least three-chords length were identified without mistakes.

It is worth mentioning that the model is capable of finding repeating harmonic patterns even though pieces are in different keys (transposition invariance). Additionally, it is reminded that the system has no knowledge of the different kinds of harmonic systems (tonal, modal, chromatic, atonal, etc.), and is therefore interesting that it detects correctly any kind of harmonic query in diverse harmonic idioms.

A very interesting observation is the following: *the harmonic recognition model is equally successful/accurate when only the first five vector components are used.* We tested all the above queries using only the first five vector entries [0, 1, -1, 2, -2] (out of the 12) which correspond to *Unison* (i.e. common pitches between two chords) and *Steps* (i.e. ascending and descending semitones and tones); the resulting matches were the same in every case.

These small intervals may be thought as being mostly related to voice-leading as it is standard practice to try to connect chords avoiding larger intervals (using common notes and small step movements). The reduction of the DIC vector to a 5-component subvector, increases cognitive plausibility of the proposed representation. Although no cognitive claims are made in this paper, we just mention that representing the transition between two chords as the small intervals that link adjacent pitches (being potentially part of individual harmonic voices) affords this representation potential cognitive validity. This has to be explored further in future research. In any case, this reduced vector results in better computational efficiency.

As we were testing the DIC vector representation on the specific dataset, an important problem arised. There were certain special cases in which different chord successions were matched to the same vector. The most unsettling such occasion was the finding of 100 instances of *major triad* progressing to *major triad* one perfect 5th down. A number of the matched instances were *minor triad* progressing to *minor triad* one perfect 5th down. How was this possible?

After further investigation we realised that a chord sequence shares the exactly same DIC vector with its retrograde inversion! For the above instance, the retrograde inversion of a *major triad* progressing to *major triad* by an interval X is a *minor triad* progressing to *minor triad* by the same interval X. This is an inherent property of the DIC vector which reduces its descriptive power, and may have serious ramifications for certain tasks.

In the case of harmonic matching, this is not a serious problem if the sequences sought for are longer that 3 chords. The reason is that the additional context of neighboring chords disambiguates the overall sequence. For instance, in the above example (finding 100 instances), if our sequence of major chords is preceded by a major chord one tone lower than the first chord, then we find only 12 instances of the sequence (most likely constrained to IV-V-I). Longer sequences will be even more unambiguous. In another example (fifth bullet above) the transition found in 23 instances will correspond most likely to a IV-V chord progression (even though it may correspond to a ii-iii transition between minor chords). If the first chord is preceded by a diminished chord a semitone lower, then the whole sequence of three chords is found only once in Bach chorale 20 (the sequence is vii_o-I-V/V). The specific context restrains the search drastically.

There exist ways to address this problem by refining the DIC representation (introducing new concepts) but the DIC representation would lose its simplicity. We conjecture that such refinement might be necessary especially if the DIC representation is to be extended for use in the audio domain. Such options are open for further investigation.

Finally, an important issue not explored sufficiently in this study is chord progression similarity, i.e. how similar two chord sequences are. As it stands a V-I transition and a V7-I transition are different (not matched) because their DIC vectors are not identical. Or a V-I transition is not matched to another V-I transition if, a note is missing such as the fifth of the first or second chords. Such relations can be captured if certain tolerances are allowed. For instance, if all entries of one vector are smaller than the corresponding entries of the other vector, and the sum of the differences is three or less, then sequences such as V-I and V7-I would be matched. The similarity relations between vectors is an open issue for further research.

In the current implementation wild cards can be inserted allowing certain tolerance in the matching. Disabling entries 6-12 in the vector, i.e. using only the first five components, did not seem to make a difference as mentioned earlier in this section. Trying out a more radical example, we queried the system with the vector 0**********2; this means we are looking for a chord succession that contains no common notes and there exist two distinct tritone relations between different notes of the chords. The system returned 7 instances: 4 in the rebetiko songs and one in *Michelle* by the Beatles that correspond to the transition of major chord to minor chord a tone lower (or the reverse), 1 in Bartok's *Romanian Folk Dance nr. 4* that corresponds to the transition of minor chord 7th to major chord a tone higher, and 1 in Debussy's first measures of *Nuage* that corresponds to the transition of a perfect fourth harmonic interval to a perfect fifth a semitone lower. Such experiments allow the investigation of certain similarity relations. δ & γ approximate matching is also possible but not explored in current study. More extended studies are necessary to determine 'meaningful' similarities between different transitions.

4. CONCLUSIONS

In this paper the Directed Interval Class representation for chord transitions has been used in a harmonic recognition task. A small database was constructed comprising of chord sequences derived from diverse music idioms/styles (tonal music, different traditional harmonic idioms, 20th century non-tonal harmonic idioms). A harmonic recognition application based on the DIC representation was developed and tested on the above dataset. As expected harmonic recognition accuracy was high especially for relatively longer chord sequence queries (longer than three chords). One of the most useful properties of the DIC representation is that it is transposition-invariant (independent of key). Some inherent limitations of the DIC vector were also presented and potential future improvements suggested.

It is suggested that the proposed idiom-independent chord transition representation is adequate for representing harmonic relations in music from diverse musical idioms (i.e., it is not confined to tonal music) and, therefore, may provide a most appropriate framework for harmonic processing in the domain of computational ethnomusicology.

5. REFERENCES

Allali, J., Ferraro, P., Hanna, P., & Iliopoulos, C. (2007). Local transpositions in alignment of polyphonic musical sequences. *Lecture Notes in Computer Science,* Volume 4726, pp. 26–38, Springer.

Allali, J., Ferraro P., Hanna P., & Robine M. (2010). Polyphonic Alignment Algorithms for Symbolic Music Retrieval. *Lecture Notes in Computer Science*, 2010, Volume 5954, pp. 466-482, Springer.

Cambouropoulos, E. (2012) A Directional Interval Class Representation of Chord Transitions. In *Proceedings of the Joint Conference ICMPC-ESCOM 2012,* Thessaloniki, Greece.

de Haas, W. B., Wiering, F. and Veltkamp, R. C. (2011). A Geometrical Distance Measure for Determining the Similarity of Musical Harmony. Technical Report UU-CS-2011-015, Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands.

de Haas, W. B., Veltkamp, R. C., and Wiering, F. (2008). Tonal pitch step distance: A similarity measure for chord progressions. In *Proceedings of the 9th International Conference on Music Information Retrieval* (ISMIR), pp. 51-56.

Hanna, P., Robine, M., and Rocher, T. (2009). An alignment based system for chord sequence retrieval. In Proceedings of the 2009 Joint International Conference on Digital Libraries, pages 101-104. ACM, New York.

Kuusi, T. (2005) Chord span and other chordal characteristics affecting connections between perceived closeness and set-class similarity. *Journal of New Music Research,* 34(3), 257-271.

Lerdahl, F. (2001). *Tonal Pitch Space*. Oxford University Press, Oxford.

Lolis, K. (2006) *Polyphonic Song from Epirus (Ηπειρώτικο Πολυφωνικό Τραγούδι)*. Ioannina. (In Greek)

Paiement, J.-F., Eck, D., and Bengio, S. (2005). A probabilistic model for chord progressions. In *Proceedings of the 6th International Conference on Music Information Retrieval* (ISMIR), pp. 312-319, London.

Pickens, J. and Crawford, T. (2002). Harmonic models for polyphonic music retrieval. In *Proceedings of the 11th International Conference on Information and Knowledge Management*, pp. 430-437. ACM, New York.

Tsougras, C. (2003). "Modal Pitch Space - A Theoretical and Analytical Study". Musicae Scientiae 7/1, p. 57-86.

Tsougras, C. (2009). Analysis of Early 20th-century Chromatic Modal Music with the use of the Generative Theory of Tonal Music - Pitch Space and Prolongational issues in selected Modal Idioms. Proceedings of the *7th Triennial Conference of the European Society for the Cognitive Sciences of Music*, (12-16 August 2009, Jyväskylä, Finland), J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, P. Eerola (eds), p. 531-539.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

VARIABILITY AS A KEY CONCEPT: WHEN *DIFFERENT* IS *THE SAME* (AND VICE VERSA)

Stéphanie Weisser
Musical Instruments Museum
Montagne de la Cour 2
B-1000 Brussels (Belgium)
s.weisser@mim.be
stephanieweisser@gmail.com

Didier Demolin
Gipsa-Lab
Université Stendhal
1180, avenue Centrale BP25
38031 Grenoble Cedex 9(France)
didier.demolin@gipsa-lab.grenoble-inp.fr

ABSTRACT

Among the numerous traits characterizing non-Western musical performances, the variability of scales has intrigued researchers. Especially in a context of development of computer-assisted tools for the study of scales, it is important to take this variability into consideration, as it is significant regarding the way pitches are organized and conceptualized within a musical system. In the secular repertoire of the Amhara of Ethiopia, the intervals sizes present certain variability. This paper will investigate the variability of the pentatonic anhemitonic scale *tezeta*, by analyzing the intervals constituting this scale in different contexts: 1. in performances of the song *Tezeta* by 5 different musicians; 2. in performances of exercises preparing the musician to perform the song; 3. in inversions of other scales (*gelbatch* technique) used when a musician does not have the opportunity to switch to the 'normal' *tezeta* scale. These three different contexts of performance allow the understanding of the variability and the rules governing the different ways *tezeta* is performed.

1. INTRODUCTION

One of the major challenges an ethnomusicologist has to face in his/her research is to deal with variability. The scientific investigation of musical performances needs, as noted by several scholars already, to take into consideration the important variability in realization (compared with the Western relatively "fixed", standardized way of building musical instruments, organizing the music and selecting the pitches). Arom (1985, vol. I: 210-254), for example, developed the concept of *pertinence* in order to properly account for the variability of the pitches he encountered. In doing so, he hypothesized that the important relationship between variable musical features is not a relationship of *same*, but a relationship of being *culturally equivalent*.

Such a concept is a key to understand many musical performances, as well as an important factor to take into consideration for automated extraction of information from non-Western musical streams. In this paper, the case-study under consideration focuses on a 'classical' research theme in ethnomusicology: the musical scale.

2. SCALES: MORE THAN A SET OF INTERVALS

Musical scales are "the basis for all melodic construction and (...) all melodies are concrete manifestations of a musical scale."(Arom, Fernando & Marandola 2007, 107). We consider here, with Simha Arom (cited in Marandola 1999: 110), that a scale is "an ensemble comprising a finite number of discrete units extracted from the sonourous continuum – the degrees [...]". Fernando (2007, 946) even precises "opposing units" and differentiates, as several other ethnomusicologists, the scale and the mode : the scale being a finite group of *all* degrees of a musical culture, whereas the mode is a selection of the specific degrees needed for a specific performance, including a hierarchization between the selected degrees.[1] Chailley (1985, cited in Fernando 2007) introduced a conceptually useful intermediary level between the scale and the mode: the system, defined as the "selected sounds", without consideration of hierarchical ordering or polarity.

Fernando (2007) considers that three different "types" of pitch organization exist: 1. a fixed set of intervals with a reference pitch (such as the Western tempered system); 2. a system of untempered archetypal sets of intervals, sometimes performed with variations and often based on a conscious theorization (such as the Indian and Arab systems) ; 3. a complex, dynamic model in perpetual (re)construction and actualization, within the limits established by the collective norm. It can therefore be stated that an intervallic measurement will take on a different significance according to the type of pitch organization used in the repertoire/musical culture in question.

Especially in the absence of a fixed reference pitch, the intervallic distance between the degrees is of essential importance. Since the end of the 19th century, most measurements of the intervallic distance between the realizations of two different degrees in a musical performance has been conducted using A. J. Ellis' subdivision of the octave into 1200 cents. However, in several cases, ethnomusicologists faced and continue to face various situations regarding the realizations of

[1] According to this differentiation, the Western tempered *scale* comprises 12 semitones, whereas the major *mode* comprises only 7 degrees, with a "most important" note, the tonic. In this paper, in order to avoid confusion however, the *scale* will designate an "intervallic pool" – the most general level possible.

intervallic distances between degrees: in several contexts, important variations can be observed such as in Ouldémé flute music of North Cameroon, for example, the size of a culturally equivalent interval can vary from 120 to 320 cents (Arom, Fernando & Marandola 2007: 115).

In non-western musical contexts, it is also important to consider a specific performance as one realization of a theoretical model, which could possibly never actually materialize in performances [Fernando 2007: 974]. Ethnomusicologists are therefore interested in unveiling this model and the rules that govern the musical realization of the model. To that end, a corpus of several performances of a same piece played by different musicians (or by the same musician at different occasions) is the basis of the analysis.

3. THE ETHIOPIAN SECULAR MUSIC (*ZEFEN*)

3.1 Investigating Amhara secular scales: a political history

In Ethiopian Amhara music, scales used in secular music (*zefen*) remain a controversial issue. As opposed to the music related to Christianity (*zema*), the *zefen* has a ambiguous status in Amhara society. The professional secular musicians, the *azmari*, have long been considered suspiciously by the rest of the Amhara society and have remained secluded (Bolay 2004, Kebede 1975).

Figure 1. *Azmari* playing in a *tedj bet* [traditional beer bar] in Lalibela. Photo: S. Weisser, 2004.

According to recent research, this status has started to slowly change in the second half of the 20th century. The founding of national insitutions such as the *Hager Fikr* Theater and the Orchestra Ethiopia gave visibility to these musicians and to the music they performed. The *Derg* Regime (1974-1991) tried to remodel the traditional organization of secular music performances and placed, for political reasons, secular musicians in the spotlight – although under strict watch (Betreyohannes, personal communication, 2012).

The first serious attempt to describe the pitch organization of their music dates also from half a century ago. Michael Powne published in 1968 the first detailed study on Ethiopian music, in which was included a "description" of the musical scales used by the *azmari*. However, as he specifies himself, most of his work was conducted with the National Folklore Orchestra of the Haile Selassie Theater. Information collected in this context might differ from the traditional practices and reflect a distinctive musical system: "the Haile Selassie Theater group was recruited in 1946 by the Municipality of Addis Ababa with the main objective to play Ethiopian songs by soloists accompanied by a modern orchestra (…). Thus began the process of giving a modern setting to the folk traditions of the country (…) (official website of the institution, http://www.mysc.gov.et/National%20Theater.html).

Powne and most of the researchers after him used therefore Western staff system (with tempered intervals) in order to characterize Amhara secular scales, even though several mentions were made regarding the inadequacy of such system (namely by Powne himself). However, they are still used (see namely Abate 2009). Such transcriptions have an over-simplifying and over-standardizing effect on the way we understand Amhara secular scales. There is no doubt that *azmari* can play traditional songs with tempered intervals, especially when they perform with Western instruments (such as keyboard, electric guitars, etc.) but the analysis of the repertoires played with traditional instruments only and recorded respectively in 1939, 1976 and the 2005 (see Weisser 2012, Weisser and Falceto 2012) shows important variability in performance and clear deviation from the Western intervals.

Literature focusing on Amhara music – scientific as well as those targeting the "general audience", such as CD liner notes or concert presentations – usually holds that Amhara secular scales [*keniet*] are *Tezeta*, *Bati*¸ *Anchihoye* and *Ambassal* (e.g. Powne 1968, Kebede 1971, Kimberlin 1976). According to Kebede (1977, 385), "[t]he term *Kignit* (…) is commonly used by *azmariwoch* (plural of *azmari*) in reference to the tuning of secular musical instruments. *Kignitoch* (*Kignit*, sing.) consist of relative pitches that adhere to the pentatonic structure of melodic patterns; they also include both equidistant and non-equidistant forms (…). In addition, the term *Kignit* has a clear reference to song (poetry and melody) and to a system of tuning instruments to relative pitches found

in the melodic patterns of the major song genres. Thirdly, the same term also applies to tuning instruments through a process of improvisation and variation".

Even though Teffera (1999, 19) noted that not all *azmari* use nor know this term in several regions, it will be used here to designate the intervallic organization of *azmari* performances. It is plausible that the set of four *kenietoch* that is now commonly adopted as a theoretical framework were formalized in the second half of the 20th century in a context of rationalization and modernization, even though other intervallic settings were and are still used. So the attempt to 'westernize' the intervals seems to have come along with an attempt to simplify and reduce the intervallic settings used by the musicians. Fortunately, Amhara musicians are proficient enough and continue until today to perform the rich, complex and interesting tradition they inherited, while constantly expanding their musical palette.

3.2 Studying the *keniet*

In order to study the *keniet* used by the musicians, five professional Azmari musicians were recorded in Addis Ababa in 2004. All four were asked to play the same set of ten songs, identified by their 'names' (usually the refrain's first word). This set comprised the four 'canonic' songs, which supposedly provided the name for the four *keniet*. In order to differentiate the 'scale' from the song, the term *zefen* [song] or *keniet* ['scale'] will be systematically added to the name of these four songs.

The five recorded *azmari* play the single-fiddle *masinqo* and four of them are also singers. Sound extracts were taken from parties of the recording with *masinqo* alone, in order to collect all of the instrumental degrees used in a performance. However, such collection was, for technical reasons, only performed for the main tessitura: *azmari* players are able to produce pitches one octave higher with a specific bow technique and pressure, but these sounds are usually very noisy (generating troubles in pitch calculation). However, this technique shows that the pitch setting in secular Amhara music can be considered as octave-equivalent.

In order to investigate the variability in more detail and determine whether it is due to differences between players or if it plays a significant part in the performance of *zefen*, it was chosen to focus on the intervals constituting the *tezeta zefen* only. Interval measurements were made in the beginning and the end of the performed song for each musician, constituting a set (called corpus 1) of 50 intervals.

Figure 2. Traditional chordophones played by *azmari*: the one-stringed fiddle *masinqo* (left) and the six- or five-stringed lyre *krar*. Photos: S. Weisser, 2006.

In order to complete this investigation and determine whether this variability is due to differences between players, a second corpus was constituted, comprising performances of *tezeta keniet* performed several times by the same musician. As one of the *azmari* was recorded in several occasions (in the context of instrumental classes), the information and sounds collected from these recordings were included in the corpus. The sounds collected from these sessions comprise exercises called 'positions': they are meant to familiarize the fingers with the positions required for the correct performance of a song. In this study, nine sets of *tezeta* intervals performed by the main informant at different occasions were included in the corpus. This second corpus comprises 80 intervals (corpus 2).

Figure 3. Andalkachew Yihune, *azmari* and main informant, during a *masinqo* lesson. Photo: S. Weisser, 2005.

Finally, another set of intervals was studied in order to complete the study: the ones provided by a specific playing technique called *gelbatch*: with this technique, it is

possible to play a song usually played in one *keniet* by using the intervals of another, only with a reorganized order (inversion of the invervals). This technique is used by musicians in performance when they do not have the time to change from one *keniet* to another (for example, when songs are following each other without breaks). A performance of *gelbatch* was made by the main informant using the 6-strings lyre *krar*, an instrument also used – although more rarely these days – by *azmari*. *Tezeta* performed based on the three other *keniet*'s intervallic settings (*anchihoye*, *ambassal* and *bati*) constitute corpus 3 (15 intervals). Kimberlin (1976, 70-2) details this technique for the *masinqo*, explaining that the term is used when the starting pitch is shifted from the open pitch to another degree[1].

Pitch measurements were made with three pieces of software: *Audiosculpt* (Ircam), *Praat* and the *mirpitch* function of the MIRToolbox. The intervals were calculated in cents. Interval 5 was calculated by considering as second pitch the higher octave of the open string, in order to keep this interval's size within the same order of magnitude as the others to ease comparisons.

3.3 Results

Analysis of corpus 1 (*Tezeta Zefen* performed by 5 different musicians) shows that variability of intervals is rather limited (average RSD reaching 8%, see Table 1). Comparison with other songs (Weisser 2012) has shown that such value is among the lowest ones. Intervals 1, 3 and 5 are most "stable" '(RSD 5-6%); analysis of the range shows that intervals 3 and 5 and are significantly bigger than the other intervals.

	Interval 1	Interval 2	Interval 3	Interval 4	Interval 5
Musician 1 (beginning)	220	192	316	188	283
Musician 1 (middle)	222	184	301	164	329
Musician 1 (end)	233	179	311	179	297
Muscian 2 (beginning)	196	147	348	201	308
Musician 2 (end)	196	199	316	155	335
Musician 3 (beginning)	213	211	290	219	267
Musician 3 (end)	214	204	289	198	295
Musician 4 (beginning)	207	183	320	178	311
Musician 4 (end)	215	194	291	182	318
Musician 5 (beginning)	220	175	334	157	313
Musician 5 (end)	215	194	291	182	318
Mean	214	187	310	182	307
RSD	5%	9%	6%	11%	6%
Range	37	65	58	64	68
Max	233	211	348	219	335
Min	196	147	289	155	267

Table 1. Results of the analysis of corpus 1 (Intervallic setting in the performances of *Tezeta Zefen* by 5 musicians). Intervals are expressed in cents.

Analysis of intervals of corpus 2 (*masinqo*'s 'Positions', see Table 2)) show that even though the musician, instrument and conditions of recordings were rather similar, the variability is twice as significant as in corpus 1: averaged RSD reaches 14%. Intervals 2 and 4 vary the most, whereas 5 seem to be the most stable. Measurements of intervals 1, 2 and 4 share a rather large common "area": intervals from c. 169 to c. 246 cents. Intervals 3 and 5 also share an area from c. 269 to c. 318 cents.

	Interval 1	Interval 2	Interval 3	Interval 4	Interval 5
Class 1	227	161	328	195	290
Class 2	192	197	316	227	269
Class 3	192	197	316	227	269
Class 4	181	201	343	178	297
Class 5	187	225	303	202	282
Class 6	189	330	244	160	277
Class 8	189	330	244	160	277
Class 9	169	152	361	199	318
Class 9-2	212	223	312	171	282
Class 11	246	158	341	160	296
Class 11-2	201	195	328	189	287
Class 12	219	166	340	173	301
Mean	200	211	315	187	287

[1] Kimberlin however states that this shift of referent pitch from open string to the 5th degree only applies between pairs of scales: *tezeta-bati* on one hand and *anchihoye-ambassal* on the other – contrarily to the *krar*.

RSD	11%	29%	12%	13%	5%
Range	77	178	118	67	49
Max	246	330	361	227	318
Min	169	152	244	160	269

Table 2. Results of the analysis of corpus 2 ('Positions'). Intervals are expressed in cents.

Analysis of corpus 3 (Table 3) shows that the range reaches maximum values for intervals 3 and 1 (over a tempered semitone), whereas interval 5 present the smallest dispersion (range of 49 cents), followed by interval 4 (77 cents) and 2 (85 cents).

	Interval 1	Interval 2	Interval 3	Interval 4	Interval 5
Tezeta on *Ambassal 'Major'*	154	208	260	191	388
Tezeta on *Anchihoye 'Major'*	130	190	398	143	339
Tezeta on *Bati 'Major'*	239	123	357	114	367
Range	109	85	138	77	49
Max	239	208	398	191	388
Min	130	123	260	114	339

Table 3. Results of the analysis of corpus 3 (*Tezeta* obtained with the'*Gelbatch*' [inversion] technique). Intervals are expressed in cents. Because of the small number of samples, mean and RSD are not calculated.

4. DISCUSSION

These results seem at first confusing and even contradictory: why does the intervallic setting of 'positions', an exercise specifically meant to improve the precision of the performance of the intervallic setting, show more variability than the 'real' performances of songs? And how come the intervallic settings found in *gelbatch* show such important range (up to half the size of the averaged interval), without challenging the 'identity' of the scale?

In order to understand these results, it is important to investigate the nature of the information extracted from the intervallic settings measured in these cases: the results obtained from the analysis of *gelbatch* can be considered as the *limits* of the system. Indeed, *tezeta* produced with *gelbatch* technique is *acceptable*, whereas *tezeta* produced in 'normal' configuration is *preferable*. In a similar way, the intervallic settings from 'Positions' is rather abstract (even though it was usually played in preparation of the performance of the song), whereas those obtained from performances of the song (*Tezeta Zefen*) correspond to a specific realization of the *keniet*. All these results are therefore related different musical objects, corresponding to diverse levels of conceptualization.

A look at these results helps us understand more about the conceptualization of the system: analysis of the "limits" of the system clearly shows that intervals 1, 2 and 4 are smaller than c. 240 cents, whereas intervals are clearly bigger (between c. 260 cents and c. 390). This differentiation between intervals 1, 2 and 4 on one hand and 3 and 5 on the other can also be observed in the two other corpuses. The analysis of the 'Positions' corpus shows that the "small" intervals have usually a minimal size of c. 150 cents. It seems therefore that intervals smaller than 150 cents are admitted (since they appear in in the *gelbatch* corpus), although not preferred, since they are not performed in the 'Positions' and songs' corpuses. Finally, the analysis of the results of the songs' corpus shows that the combination of the degrees into a melody and the addition of the singing voice exert a standardizing effect on the intervallic setting. Variability diminishes (without disappearing), especially for interval 2.

It is now possible to consider that the relatively larger size of intervals 3 and 5 seem to be the defining criterion of the *tezeta* setting. On the contrary, intervals 1, 2 and 4 are usually smaller. It is interesting to note, however, that a bigger interval 2 (in the 'Positions' corpus, it can reach 330 cents) does not alter the '*tezeta*-nature' of the intervallic setting. However, such significant values are not reached by interval 2 in the 2005 'real' performances. It however explains why some intervallic settings of songs performed with a wider second interval (such as *Ye Enfras Abeba*, see Weisser 2012) is considered to be a song in *tezeta keniet*. It also explains why the songs performed with another chordophone, the *bagana*, are considered to be in *tezeta keniet* with either a wider second interval (most of the time) or a smaller second interval (less frequent, see Weisser 2005).

Comparison of analysis of the 2005 performances with another set of measurements (ten musicians performing the *tezeta zefen* recorded and analyzed by Kimberlin in 1976) shows that these 'rules' emerging from this analysis do also apply. It also shows that the variability of the intervals in Kimberlin 1976's corpus is more important than in the corpus 1. Two hypotheses can be formulated: 1. The variability is more important in 1976 performances because the number of musicians considered is more important than in corpus 1 (10 versus 5). This would imply

that variability would be directly in proportion with the number of musicians/performances: the more performances, the more variability. 2. The second hypothesis is related to the time lapse separating the two studies (about 30 years): the modification of variability would be a symptom of standardization of the intervallic settings, perhaps due to the generalization of western-tempered instruments (including keyboards and software-generated sounds).

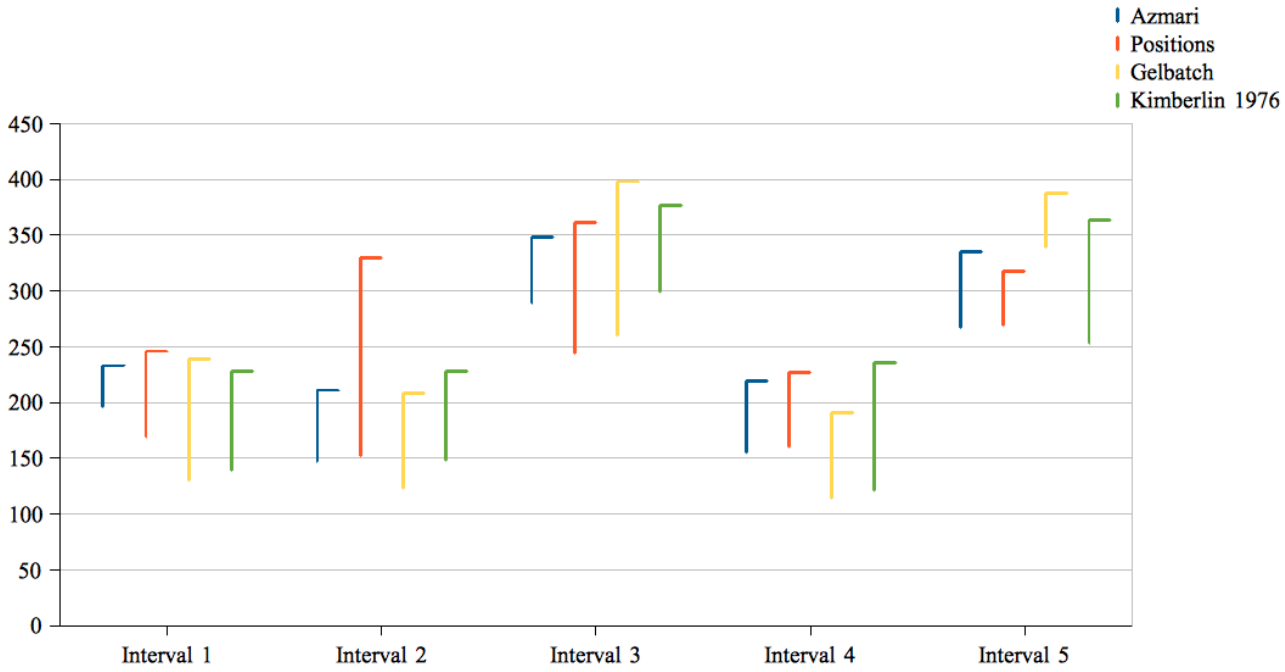


Figure 4. Maximal and minimal values (in cents) for intervals in corpus 1 (*Tezeta Zefen*), 2 (*'Positions'*), 3 (*Gelbatch*) and Kimberlin's.

An argument stands for the first hypothesis: the general degree of variability (averaged RSD for the intervals) is globally identical (about 14%) between corpus 2 (12 recordings) and Kimberlin's corpus. However, such a vast question requires however a much larger corpus of study to be solved with a reasonable degree of certainty.

5. CONCLUSION

Such an analysis shows that computer-assisted tools for analysis of scales requires several technical possibilities: batch processing for corpus of recordings, as well as tools for the investigation of variability (statistical, user-criteria based) are needed. Indeed, even if tools do perform accurate calculation of intervallic distances of performed degrees, the interpretation of the results, the formulation of hypotheses and sometimes experimentation are still an inavoidable step, and – so far – not yet automatizable, in the process of understanding the rules underlying the pitch organization of musical performances in non-Western contexts.

6. REFERENCES

Abate, E. (2009). Ethiopian Kiñit (scales). Analysis of the formation and structure of the Ethiopian scale system. In *Proceedings of the 16th International Conference of Ethiopian Studies* (vol. 4, pp. 1213-1224).

Arom, S. 1985. Polyphonies et polyrythmies instrumentales d'Afrique Centrale, Paris: Selaf, 1985, 2 vol.

Arom, S., Fernando, N. & Marandola, M. (2007). An Innovative Method for the Study of African Musical Scales: Cognitive and Technical Aspects. In *Proceedings of the Sound and Music Computing Conference (SMC'07)* (pp. 107-116).

Bolay, A. (2004). Les poètes-musiciens éthiopiens (azmari) et leurs constructions identitaires. Des marginaux qui aspirent à la normalité. *Cahiers d'études africaines, 176/4*, 815-839.

Fernando, F. (2007). Echelles et modes : vers une typologie des systèmes scalaires, In Jean-Jacques Nattiez (ed.), *Musiques,*

une encyclopédie pour le XXIème siècle, vol. 5. L'unité de la musique, Paris: Actes Sud, 945-979

Kebede, (1971). The Music of Ethiopia. Its Development and Cultural Setting. PhD Diss., Wesleyan University, Ann Arbor: Xerox University Microfilms.

Kebede, A. (1975). The “Azmari", Poet-Musician of Ethiopia. *The Musical Quarterly, 61/1*, 47-57.

Kebede, A. (1977). The Bowl-Lyre of Northeast Africa. Krar: The Devil's Instrument. *Ethnomusicology, 21*/3: 379-395.

Kimberlin, C. T. 1976. Masinqo and the nature of Qeñet. PhD diss., University of California, Los Angeles, Ann Arbor: Xerox University Microfilms.

Marandola, F. (1999). L'apport des nouvelles technologies à l'étude des échelles musicales d'Afrique centrale. *Journal des africanistes*, *69/2*, 109-120.

Powne, M. (1968). Ethiopian Music. An Introduction. London-New York-Toronto: Oxford University Press.

Teferra, Timkehet. 1999. Wedding Music of the Amhara in the Central Highland of Ethiopa. PhD Dissertation, Humboldt University of Berlin (Philosophical Faculty III, Studies of Culture and Art).

Weisser, S. (2005). Etude ethnomusicologique du bagana, lyre d’Ethiopie. PhD Dissertation, Université Libre de Bruxelles.

Weisser, S. (2012). Opening Pandora’s box: Investigating *Azmari*’s scales. In *Proceedings of the First International Conference on Azmari. From Ambivalence to Acceptance* (in preparation).

Weisser, S. & Falceto, F. (2012). Investigating qeñet in Amhara secular music performances. History and scales: an interdisciplinary study. In *Proceedings of 18th International Conference of Ethiopian Studies* (ICES18) (in preparation).

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

TRADITIONAL ASYMMETRIC RHYTHMS: A REFINED MODEL OF METER INDUCTION BASED ON ASYMMETRIC METER TEMPLATES

Thanos Fouloulis
Dept. of Music Studies,
Aristotle Univ. of Thessaloniki
thanos.fouloulis@gmail.com

Aggelos Pikrakis
Dept. of Computer Science,
University of Piraeus
pikrakis@unipi.gr

Emilios Cambouropoulos
Dept. of Music Studies,
Aristotle Univ of Thessaloniki
emilios@mus.auth.gr

ABSTRACT

The aim of this study is to examine the performance of an existing meter and tempo induction model (Pikrakis et al, 2004) on music that features asymmetric rhythms (e.g. 5/8, 7/8) and to propose potential improvement by incorporating knowledge about asymmetric rhythm patterns. We assume knowledge of asymmetric rhythms in the form of metric templates (consisting of both isochronous and asymmetric pulse levels). Such knowledge seems to aid the induction process for symmetric/asymmetric rhythms and thus improve the performance of the aforementioned model.

1. INTRODUCTION

In recent years a number of meter induction and beat tracking models have been implemented that attempt to identify perceptually pertinent isochronous beats in musical data. Such models assume an isochronous tactus within a certain tempo range (usually centered around the spontaneous tempo). The performance of such systems is usually measured against musical datasets drawn from Western music (e.g. classical, rock, pop, jazz) that features almost exclusively symmetric rhythmic structures (e.g. 3/4, 4/4, 6/8) (Mckinney et al. 2007; Dixon 2007; Davies et al. 2009). The tactus of asymmetric/complex musical rhythms, however, is non-isochronous; for instance, a 7/8 song is often counted/taped/danced at a level 3+2+2 (not at a lower or higher level). Such models fail to identify asymmetric beat levels (Fouloulis et al, 2012).

Musical time is commonly organized around a (hierarchic) metrical structure of which the most prominent level is the beat level (tactus) (Lerdahl and Jackendoff, 1983). Such a metric structure facilitates the measurement of time and the categorical perception of musical temporal units (durations, IOIs). In western music, an isochronous beat level is almost always assumed; any divergences from isochronous beat are treated as 'special cases' or even 'anomalies'.

A central assumption of this paper is that the beat level (tactus) of metrical structure need not be isochronous. It is asserted that metrical structure is learned implicitly (through exposure in a specific idiom), that it may be asymmetric and that the tactus level itself may consists of non-isochronous units. It is maintained that an acculturated listener may use spontaneously an asymmetric tactus to measure time, as this is the most plausible and parsimonious way to explain and organize rhythmic stimuli within specific musical idioms.

Rhythm and pitch share common cognitive underlying mechanisms (Parncutt, 1994; Krumhansl, 2000). Asymmetric structures are common in the pitch domain. Major and minor scales, for instance, are asymmetric. Listeners learn pitch scales through exposure to a specific musical idiom, and then automatically organize pitch and tonal relations around the implied asymmetric scales. Asymmetric scales are actually better (cognitively) than symmetric scales (e.g. 12-tone chromatic scale or whole-tone scale) as they facilitate perceptual navigation in pitch/tonal spaces. It is, herein, assumed that asymmetric beat structures may arise in a similar fashion to asymmetric pitch scales, and may organize certain rhythmic structures in an accurate and more parsimonious manner.

In more formal terms, the kinds of asymmetric beat structures mentioned in this study may be described as series of repeating asymmetric patterns consisting of long (three's) and short (two's) units. Such asymmetric patterns are 'sandwiched' in between a lower isochronous sub-beat level (commonly at the 1/8 duration) and a higher isochronous metric level (e.g. 5=3+2 or 7=3+2+2) (Fouloulis et al, 2012). Such hierarchic metric structures are considered in this paper as a whole rather than a number of independent isochronous and asymmetric pulse levels.

2. METER AND TEMPO INDUCTION MODEL

2.1 Original model architecture

In this study we examine a potential improvement in the performance of an existing model (Pikrakis et al, 2004) that focuses on meter and tempo extraction on polyphonic audio recordings. The existing version processes audio in non overlapping long-term windows while using an inner moving short-term window to generate sequences of feature vectors considering energy and mel frequency cepstral coefficients (MFCCs) (Figure 1). For every long-term window a Self Similarity Matrix (SSM) is formulated based on the assumption that its diagonals can reveal periodicities corresponding to music meter and beat. Calculating the mean value of each diagonal and plotting it against the diagonal index each audio segment reveals a "rhythmic signature" that can be further analyzed in order to infer the actual beat and meter. Two different ranges of SSM diagonal indices in this "rhythmic signature" are considered suggesting that beat and meter candidates are lying within respectively.

The original model relies on two criteria to associate certain periodicities to music meter and tempo. In the first criterion beat candidates are selected as the two neighbouring local minima that possess larger values. Meter candidates are validated in relation to beat candidates according to the accepted set of music meters under investigation. Calculating the sum of corresponding mean values for every pair, the music meter of a segment can be determined as the one that exhibits the lowest value. The second criterion differentiates in that it takes into account the slope (sharpness) of the valleys of each pair and not just their absolute values.

The meter of the whole audio is selected taking into account its frequency of occurrence through histograms that are formed using the calculated meter values per segment. Tempo estimation process, based on previous results about beat lag, is jointly extracted per long-term segment or as average for the whole audio.

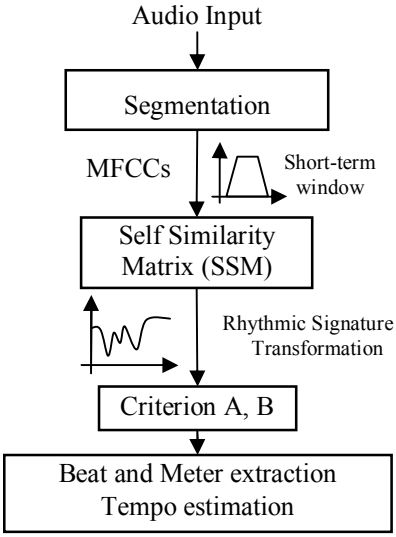


Figure 1. Overview of the architecture of the original meter and tempo induction model.

2.2 Refined model architecture

The main motivation behind the refined model relies on the assumption that meter induction can be assisted by querying an audio recording against known metric templates. Knowledge about metric structure is incorporated into the model by including a set of both symmetric and asymmetric templates in a form of a template library (Figure 2). During the induction process and for a given tempo hypothesis each "rhythmic signature" of an audio recording can be evaluated in turn with the contents of the template library so that we can conclude to the most prominent one.

2.3 Template Generation

Templates were generated for the following time signatures 2/4, 3/4, 4/4, 5/8 (3+2), 6/8, 7/8 (3+2+2), 8/8 (3+3+2) and 9/8(3+2+2+2) using MIDI and audio drum sounds for a reference tempo of 260bpm (1/8). The corresponding audio files were then transformed into their respective template "rhythmic signatures" by using the same procedure as before. In Figure 3 "rhythmic signatures" of 7/8 and 5/8 templates on a tempo of 260bpm (1/8) are presented. The lowest local minima (valleys) on these templates match strong periodicities and can be considered as meter candidates. The distance in the x-axis D_t between two successive meter candidates is generally altered accordingly to tempo changes and is utilized during the induction process in order to scale the template according to the calculated tempo.

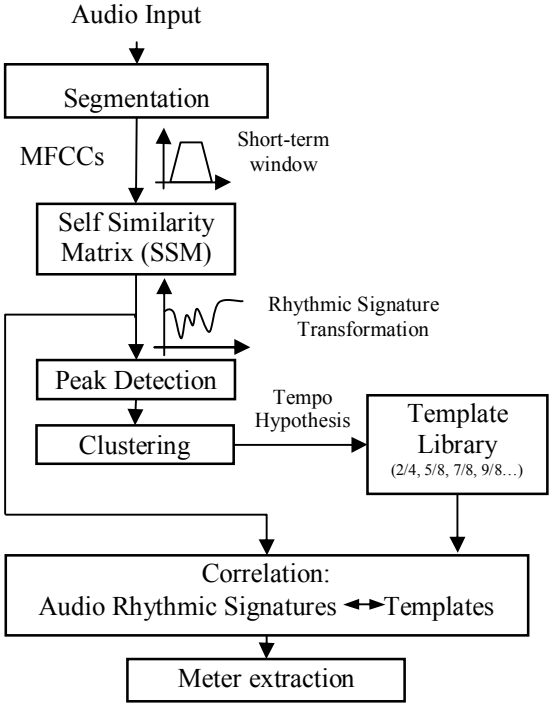


Figure 2. Overview of the architecture of the refined meter and tempo induction model.

3. IMPLEMENTATION DETAILS

The refined system keeps the initial audio processing steps of the original model but we refer to them anyway for the sake of comprehension. In the first step audio recordings are processed on a segment by segment basis in non overlapping long-term windows of 10s. Sequences of feature vectors are extracted using a "chroma based" variation of standard MFCCs, which yields significantly better results by emphasizing beat and meter candidates. This approach instead of assuming equally spaced critical band filters in the mel scale makes use of a critical band filter bank consisting of overlapping triangular filters, that is aligned with the chromatic scale of semitones (starting from 110 Hz and reaching up to approximately 5KHz) (Pikrakis et al, 2004).

Feature vectors in each long-term window are extracted by means of a short-term processing technique. The values for the length, w_s and hop size h_s of the short-term window were chosen as 100ms and 10ms respectively. Then, the sequences of feature vectors are utilized to form self-similarity matrices (SSM), using the Euclidean

function as a distance metric, in order to reveal the dominant periodicities inside each segment. This can be achieved by computing the mean value B_k for each diagonal k and plot the value against the diagonal index. Local minima in this curve correspond to strong periodicities that are prominent in the specific time frame. We can consider the function B(k) as the "rhythmic signature" of the long-term segment from which it is extracted.

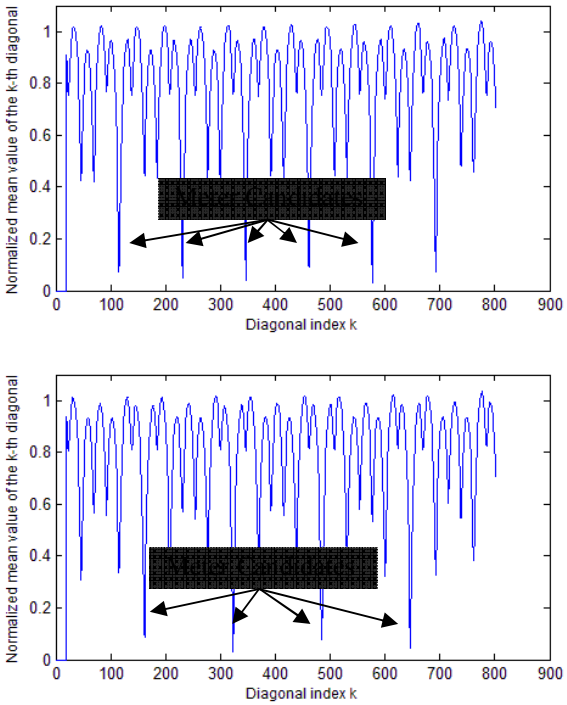


Figure 3. "Rhythmic signatures" for a 5/8 (top) and a 7/8 (bottom) template.

3.1 Peak detection - smoothing

Each "rhythmic signature" is then processed using a peak detection algorithm to extract the diagonal indices k that correspond to the most salient local minima (valleys). The peak detection algorithm uses the first derivative of the signal and relies on the fact that the first derivative of a peak has a downward-going zero-crossing at the peak maximum. To avoid picking false zero-crossing due to the noise we use a technique that initially smooths the first derivative of the signal using a rectangular window, and then it takes only those zero crossings whose slope exceeds a certain pre-determined minimum (slope threshold). The smoothing algorithm simply replaces each point in the signal with the average of m adjacent points defined by smooth width. For example, for a 3-point smooth (m = 3) (O'Haver, 2013):

$$S_j = \frac{Y_{j-1} + Y_j + Y_{j+1}}{3} \quad (1)$$

where S_j the j-th point in the smoothed signal, Y_j the j-th point in the original signal, and n is the total number of points in the signal.

3.2 Clustering valleys

In order to account for light tempo changes and also slight deviations from strict metronomical performance we cluster the detected valleys by using the notion of valley bins. Each bin is defined by a diagonal index mean value m_b and a tolerance window e. Each time a new valley is assigned to a relative bin the bin mean value m_b is updated. The time equivalent T_k for a local minimum k is $T_k = k * \text{step}$ (Pikrakis et al, 2004) where step is the short-term step of the moving window (10 ms for our study). In this work the width of the tolerance window was defined to be 8*step =80ms.

Valleys are weighted by taking into account their frequency of occurrence in the sequence of "rhythmic signatures", their slope and their amplitude. This relies on the assumption that meter periodicities are prominent in the majority of the "rhythmic signatures" and exhibit steeper slopes and narrower valleys. Therefore, bins which are more populated and contain sharper valleys are discriminated. The next step is to pick the two most important valleys bins that have successive mean values m_b. If the previous assumption is right and those two successive valley bins correspond to meter candidates then the distance D_s in x-axis between them can be compared to the corresponding distance D_t of each template.

This comparison determines the stretching/expanding factor f_t for each template that is needed to compensate for the tempo difference between the tempo of the real audio file and the reference tempo (260bpm - 1/8) that was specified during template generation. The product of this step is to conclude in tempo hypothesis using factor f_t and then perform a "time scaling" for each template of the template bank.

3.3 Meter extraction

The final step of the algorithm performs a correlation analysis between each "rhythmic signature" of the audio and every time-scaled template. In particular, each template is slided through every rhythmic signature and a correlation coefficient is calculated. Finally, the template for which the correlation coefficient has a maximum value is considered as the winner. The results for a 7/8 and a 9/8 song are presented in figure 4.

4. RESULTS AND DISCUSSION

In a previous study (Fouloulis et el. 2012) we tested the original version of the model (Pikrakis et al. 2004) against a set of Greek traditional songs that featured mostly asymmetric rhythms with time signatures of 2/4, 3/4, 5/8, 6/8, 7/8, 8/8, 9/8, 10/8 and 11/8. The majority of the songs were derived from educational material and most of them start with an introductory rhythmic pattern in order to indicate the correct way of tapping/counting.

In this study we used a similar set of 30 Greek traditional songs and examined the model's performance after incorporating templates with time signatures of 2/4, 3/4, 4/4, 5/8, 6/8, 7/8, 8/8, and 9/8 (Table 1). The preliminary results are encouraging, indicating that this architecture may prove to be quite effective and may assist the induction process. In general the model seems to retain its sig-

nificant behavior in processing non-symmetric meters but some more tweaking is needed in order to further improve performance.

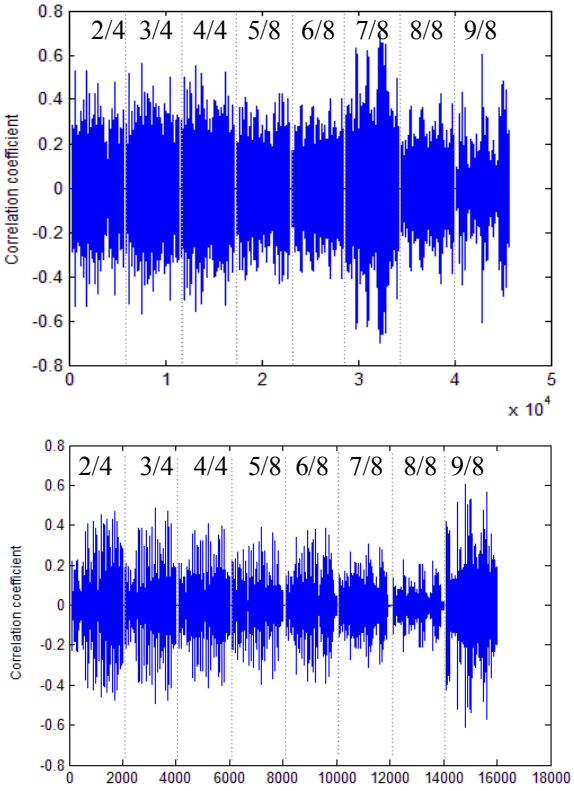


Figure 4. Correlation analysis for a 7/8 (top) and a 9/8 song (bottom).

In cases (tracks 19, 20, 21, 26, 28 and 29) when the original model tended to designate as more dominant periodicities the ones that referred to a span of two measures the refined model's output corresponds to the correct tempo and meter.

For tracks no. 4-6, the refined model assumes a 6/8 meter instead of 3/4. This seems to be supported but the nature of the performance (Figure 5). For track 7 and 8, it indicates an asymmetric 8/8 while the notation of the song indicates an isochronous pulse; again this is due to the performance elements that introduce asymmetric features.

It is worth pointing out that the instances in which the algorithm falls into a wrong estimation are songs with too fast tempi (songs 10, 16, 18, 22 and 23). In all these cases the actual meter value resides in the correlation plot but with a lower peak.

5. FURTHER RESEARCH

The architecture presented above still has many open issues that need to be explored. First of all it is necessary to evaluate its performance using a larger data set. Secondly, the results could probably be improved if further musicological/cogntive knowledge is incorporated. For example, constraints about tempo hypotheses that exceed some limits (e.g. too slow or too fast rates) could be integrated. Additionally, a wider range of more refined templates can be generated (by assigning a variety of sounds to the various metric midi templates), allowing a more effective discrimination between different metric structures for a given tempo.

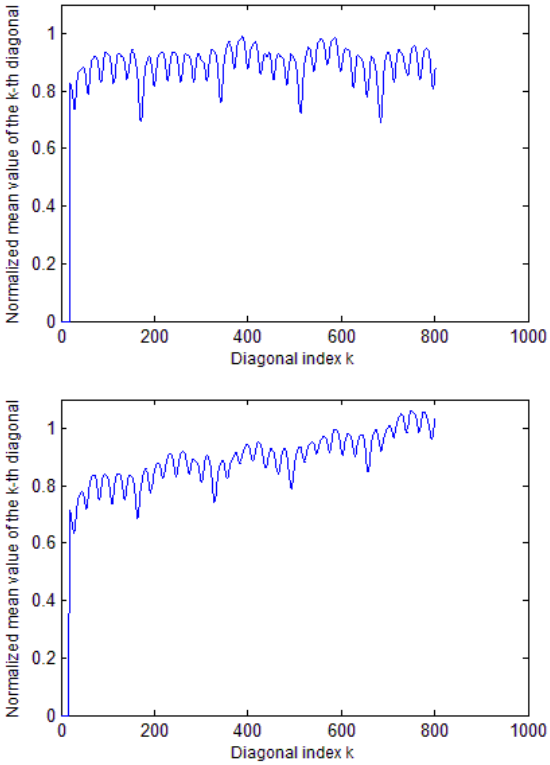


Figure 5. "Rhythmic signatures" from one segment of song no. 14 (top) and song no. 6 (bottom). Patterns seem to support the fact that even if song no. 6 is notated as 3/4 it can be considered as 6/8 due to performers' musical idiom.

6. CONCLUSION

In this study we investigate a potential improvement in the performance of an existing model (Pikrakis et al, 2004) in inducting meter from audio recordings of folk music by embedding knowledge about asymmetric/complex rhythmic structures. Templates of common asymmetric rhythm patterns were generated and then imported into the system. The preliminary results in this ongoing research are very encouraging, indicating that this architecture may prove to be quite effective and can assist the induction process.

7. REFERENCES

Cambouropoulos, E. (1997) Musical Rhythm: Inferring Accentuation and Metrical Structure from Grouping Structure. In *Music, Gestalt and Computing - Studies in Systematic and Cognitive Musicology*, M. Leman (ed.), pp. 277-293. Springer, Berlin.

Davies, M. E. P., Degara N. & Plumbley M. D. (2009). Evaluation Methods for Musical Audio Beat Tracking Algorithms. *Technical Report C4DM-TR-09-06*, Queen Mary University of London, Centre for Digital Music.

Dixon S. (2007). Evaluation of the audio beat tracking system BeatRoot. *Journal of New* Music *Research*, 36(1):39-50.

Fouloulis, A., Pikrakis, A., & Cambouropoulos, E (2012). Asymmetric beat/tactus: Investigating the performance of beat-tracking systems on traditional asymmetric rhythms. In *Proceedings of the Joint Conference ICMPC-ESCOM 2012*, Thessaloniki, Greece. 23-28 July 2012.

Krumhansl, C.L.(2000) Rhythm and pitch in music cognition. *Psychological Bulletin* 126, pp. 159–179 .

Lerdahl, F. and Jackendoff, R. (1983) *A generative Theory of Tonal Music,* The MIT Press, Cambridge (Ma).

Mckinney, M. F., Moelants, D., Davies, M. E. P., and Klapuri, A. (2007). Evaluation of audio beat tracking and music tempo extraction algorithms. *Journal of New Music Research*, 36(1):1–16.

O'Haver, T C., (2013). A Pragmatic Introduction to Signal Processing. Retrieved from: http://terpconnect.umd.edu/~toh/spectrum/IntroToSignalProcessing.pdf

Parncutt, R. (1994) Template-Matching Models of Musical Pitch and Rhythm Perception. *Journal of New Music Research,* 23:145-167.

Pikrakis, A., Antonopoulos, I., & Theodoridis, S. (2004). Music meter and tempo tracking from raw polyphonic audio. *Proceedings of ISMIR 2004*, Spain

Table 1. Meter and tempo induction results of the original and refined model.

					Refined Model with embedded templates		**Original model**	
	Song's Name	**Time Signature**	**Tempo**	**Meter**	**Calc. Tempo**	**Calc. Meter**	**Calc. Tempo**	**Calc. Meter**
1	Sousta Rodou	2/4	144 (1/4)		**141 (1/8)**	**2/4**	285	4:1
2	Mpalos	2/4	82 (1/4)		**84 (1/8)**	**2/4**	171	4:1
3	Ehe geia panagia (Hasapiko)	2/4	130 (1/4)		**258 (1/8)**	**4/8**	260	4:1
4	Tsamikos	3/4	98 (1/4)		**196 (1/8)**	**6/8**	98	3:1
5	Apopse mavromata mou	3/4	104(1/4)		**209 (1/8)**	**6/8**	206	6:1
6	Valtetsi	3/4	108(1/4)		**218 (1/8)**	**6/8**	214	6:1
7	Armenaki	4/4	180(1/4)		**357 (1/8)**	**8/8**	181	4:1
8	Louloudi ti marathikes	4/4	127(1/4)		**256 (1/8)**	**8/8**	260	8:1
9	Zagorisios -Kapesovo	5/8	94 (1/8)	**2-3**	**96 (1/8)**	**5/8**	97	2:1 or 5:1
10	Mpaintouska Thrakis	5/8	420 (1/8)	**2-3**	**250 (1/8)**	**3/8**	83	4:1
11	Dio palikaria apo to Aivali	5/8	239(1/8)	**3-2**	**241 (1/8)**	**5/8**	-	-
12	Esvise to keri kira Maria	5/8	249(1/8)	**3-2**	**243 (1/8)**	**5/8**	-	-
13	I Kiriaki	5/8	300(1/8)	**3-2**	**293 (1/8)**	**5/8**	-	-
14	Itia	6/8	201(1/8)		**202 (1/8)**	**6/8**	208	6:1
15	Enas aitos kathotane	6/8	209(1/8)		**209 (1/8)**	**6/8**	206	6:1
16	Perasa ap´tin porta sou	7/8	264(1/8)	**3-2-2**	**74 (1/8)**	**2/8**	130	7:1
17	Tik Tromakton Pontos	7/8	488 (1/8)	**2-2-3**	**499 (1/8)**	**7/8**	73	2:1
18	Mantilatos Thrakis	7/8	483 (1/8)	**2-2-3**	**199 (1/8)**	**3/8**	69	2:1 or 3:1
19	Mantili Kalamatiano	7/8	273 (1/8)	**3-2-2**	**273 (1/8)**	**7/8**	132	7:1
20	Milo mou kokkino	7/8	268 (1/8)	**3-2-2**	**265 (1/8)**	**7/8**	133	7:1
21	Na diokso ta synefa	7/8	266 (1/8)	**3-2-2**	**266 (1/8)**	**7/8**	130	7:1
22	Oles oi melahroines	8/8	381 (1/8)	**3-3-2**	**83 (1/8)**	**2/8**	193	4:1
23	Dyo mavra matia agapo	8/8	396(1/8)	**3-3-2**	**97 (1/8)**	**2/8**	200	4:1
24	Marmaromenios vasilias	8/8	198(1/8)	**3-3-2**	**195 (1/8)**	**8/8**	-	-
25	Feto to kalokairaki	9/8	136(1/8)	**2-2-2-3**	**139 (1/8)**	**9/8**	139	9:1
26	Karsilamas	9/8	256 (1/8)	**2-2-2-3**	**255 (1/8)**	**9/8**	130	9:1
27	Amptaliko neo	9/8	104 (1/8)	**3-2-2-2**	**246 (1/8)**	**9/8**	61	9:1
28	Tsiourapia Makedonias	9/8	276 (1/8)	**2-2-2-3**	**296 (1/8)**	**9/8**	109	9:1
29	karsilamas - Ti ithela	9/8	288 (1/8)	**2-2-2-3**	**290 (1/8)**	**9/8**	146	9:1
30	Ela apopse stou Thoma	9/8	185 (1/8)	**2-2-2-3**	**184 (1/8)**	**9/8**	96	9:1

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

INVESTIGATING NON-WESTERN MUSICAL TIMBRE: A NEED FOR JOINT APPROACHES

Stéphanie Weisser
Musical Instruments Museum
Montagne de la Cour 2
B-1000 Brussels (Belgium)
s.weisser@mim.be
stephanieweisser@gmail.com

Olivier Lartillot
Finnish Centre of Excellence in
Interdisciplinary Music Research
University of Jyväskylä
olartillot@gmail.com

ABSTRACT

This paper investigates the specific buzzing quality of the sounds of the Hindustani plucked lute *sitar*. The *sitar*'s wide curved bridge *jawari* is responsible for the production of these sounds and the precise adjustment of the curve is very important for the musicians. Two settings are clearly differentiated by the players: *khula* [open] and *bhand* [closed]. However, the description of the *jawari* effect is quite complex, namely because the sympathetic strings *taraf* are also equipped with a curved bridge (and therefore contribute to the global sound quality) and because of the open rhythmic strings *cikari*, tuned to specific pitches and acting as complementary sympathetic strings when particular notes are played by the main strings. This paper confronts the analysis of the musicians' discourse regarding the *jawari* with the results of new timbre descriptors developed for this specific case-study, showing that an interdisciplinary approach to that research question is fruitful.

1. INTRODUCTION

Varied types of musical timbres are used in musical expression. If Western art music of the 18th and 19th century privileged "clear" sounds, several musical cultures (including the European Renaissance art music) use buzzing sounds. However, investigation of these sounds is quite scarce in scientific literature: musicologists do not have the conceptual and practical tools to tackle this issue, and information scientists seem to be unaware or uninterested by them. Nevertheless, joint investigations are fruitful. In this paper, an ongoing collaboration conducted by an ethnomusicologist and computer scientist for music analysis will be presented and discussed, focusing on the Hindustani plucked lute *sitar*.

2. THE HINDUSTANI SITAR

The Hindustani *sitar* is a plucked fretted lute (Figure 1). It can comprise either 7 main strings and 3 rhythmic drone strings (*kharaj pancham sitar*) or 3 main strings and 4 drone rhythmic strings (*gandhar pancham sitar*). Sympathetic strings are 11 to 13. All the strings (main and sympathetic) are metallic: the first and fifth are always of tempered steel, the second of copper or phosphor bronze, and the others of brass or steel. The *taraf* are made of thin steel. The player plucks the strings with a metallic plectrum called *mizrab*, inserted on his/her right index.

Figure 1. Supratik Sengupta in concert in the Musical Instruments Museum (Brussels, 7 March 2012. Photo: S. Weisser).

Among the Hindustani classical chordophones, the *sitar* can be characterized by the presence of two specific devices: the curved wide bridge *jawari* and the sympathetic strings *taraf* (Weisser & Demoucron 2012). These devices play a considerable part in the sound production, providing an 'echo-like'effect (for the sympathetic strings) and a very noticeable sound quality (for the bridge). These effects are even combined, as the sympathetic strings are also equipped with their own *jawari* (Figure 2). This combination of devices makes the instrument rather unique among the classical chordophones: other instruments (such as the *sarangi*, the *rudra vina* and the *sarasvati vina* are equipped with either *taraf jawari* or main string *jawari*, but not both).

Figure 2. The two open *jawari* of the *sitar* pictured in figure 1. The difference between open and closed *jawari* is not visible to the naked eye (Kolkata, 16 April 2011. Photo: S. Weisser).

The two *jawari* of the *sitar* can be set in different configurations: *khula* ("open"), *bhand* ("closed") or more rarely *gol* ("round [intermediary]"). The choice of either one of these configurations is made according to the musician's *gharana* (aesthetic school of playing) and/or his or her personal taste.

3. "DOING" THE JAWARI: BRINGING THE LIFE

The desired configuration is attained by a series of complex operations (Marcotty 1974) consisting in skillfully filing with sandpaper specific areas of the upper part of the wide bridge (Figure 3), usually made of ivory, bone, ebony or nowadays synthetic material.

Figure 3. *Sitar* Maker Barun Roy shows the location of the area to file in order to close the main strings' *jawari* on an unfinished *sitar* (Kolkata, 11 April 2011. Photo: S. Weisser).

The *jawari* maintenance is usually performed every few months: it depends on how much the instrument is played. Indeed the pressure of the strings on the bridge impacts the shape of the upper surface of the bridge, responsible for the production of the sound. Instrument specificities make that pressure important, as the strings are placed high above the handle and the fingers (index and major) press down the strings in the desired fret. The extensive use the *meend* technique, consisting in pulling the string to modifiy the pitch and produce a gliding sound, contributes to this pressure.

Musicians consider this *jawari* adjustment as very important: even if not all of them perform it themselves (it is often be done by a maker), it is, in the latter case, conducted in close collaboration with the instrument's owner.

Several studies (or a brief review, see Vyasarayani, Birkett and McPhee 2009, 3673) have analyzed and modeled the way it impacts the vibratory behavior of the string. Auditory roughness has been shown to accurately account for the buzzing quality of the sound of a similar device placed on the Indian drone lute *tampura* (Vassilakis 2005).

However, the situation appears to be slightly different in *sitar*: the *tampura*'s *jawari* is usually set in an extreme *khula* position, providing very buzzing sounds. First, the *sitar*'s *khula jawari* is never that open. Second, the *taraf*, being also equipped with a *jawari* bridge, contribute also to the buzzing quality of the global sound. It is important to note, moreover, that the *taraf*'s sounds appear after the attack of the playing strings, impacting the decay part of the sound. Third, contrarily to the *tampura*, the *sitar* is not played with open strings only (although it is the case for some of the *sitar*'s strings). Finally, the *sitar* is played with nuances of intensity and duration (from very long and soft sounds in the introductory unmetered part *alap* to very fast and loud sounds of the climatic ending part *jhalla*), inducing a variety of timbres. All these specificities make the fine-tuning of the *sitar*'s *jawari* even more complex to describe and to quantify.

4. MATERIAL AND METHODS

This study is based on information and sounds gathered during a fieldwork which took place in Kolkata (India) in 2011. Interviews and recordings were conducted in the ITC *Sangeet Music Academy*, with the active collaboration of the professors and the students. ITC-SRA is an institution favoring the traditional way of teaching (*guru-sushiya parampara*, the master-disciple pedagogical relationship) and gathers numerous advanced students as well as renowned professors.

In order to analyze the *khula* and *bhand* settings, different types of data have been confronted. Sounds were collected during an experiment consisting in recording a musician playing isolated sounds with different nuances of intensity and plucking strength ('soft', 'normal', 'hard') with a *khula jawari sitar* and with a *bhand jawari sitar*. These instruments were tuned to the same *rag* and played by the same musician.

Data extracted from semi-structured interviews of musicians have been classified into categories according to content. The information collected in Kolkata was completed by other interviews conducted in Belgium and in the Netherlands with professional *sitar* players met during their European tour or living in these countries. In total, statements have been collected from 13 musicians (masters, advanced students and professional): 9 *sitariya*

[*sitar* players], 3 *sitar* makers, 1 vocalist and 1 *sarodiya* [*sarod* player]. The last two informants were added because of their expertise: the vocalist adjusted the *jawari* for 4 *tampura* in the context of an experiment. The *sarodiya* [*sarod* player], the famous Pandit Buddhadev Das Gupta, also teaches *sitariya* in the Academy and has an exhaustive knowledge regarding Hindustani classical music.

5. MUSICIANS' WORDS

The data extracted from the musicians' interviews show that identification of *khula jawari* is rather consensual: most musicians refer to Ravi Shankar's or the *tampura*'s sound as the archetypes of *khula jawari*. However, the identification of *bhand* and *ghol jawari* is ambiguous: according musicians, Vilayat Khan and Nikhil Bannerjee's adjustment are either considered as *bhand* or *ghol*. It can also be noted that the latter term is more rarely found in musicians' discourses. As noted by one of the informant (Pandit Partho Chatterjee), the terms are often misused and with an excess of generalization.

Several musicians insist on the concept of balance when it comes to *jawari*: balance is looked for at different levels. The instrument must generates sounds that are not identical but 'balanced' whether being produced with open strings or not (Barun Roy, maker), and whether playing high-pitch or low-pitch notes. Strict identity is not looked for, but excessive difference is not accepted (Pandit Ashok Pathak).

A second type of balance is related to the contribution of the *taraf*: *taraf* should not sound too loud and buzzing and 'come in the way' of the main strings' sounds (Barun Roy, maker). As the *taraf* are not controlled directly by the musicians, the fine-tuning of their *jawari* is the only way to adjust them.

A third type of balance must be achieved between roundness and clarity of the main strings'sounds: even if the *jawari* 'curves' the sound, an excess of *jawari* "jumbles up" the notes, generating confusion in the melodic line. On the other way, not enough *jawari* in sound makes the latter loose its "*sitar* quality": "it becomes a guitar" (Pandit Buddhadev Das Gupta).

All these types of balance directly impact the *jawari*'s fine-adjustment. Moreover, the maker Barun Roy added the importance of a balance between the *tumba* [gourd], *tabli* [soundtable] and *jawari*. According to him, what he calls 'hard sitar', characterized by thicker *tabli* and a generally bigger size cannot be equipped with a *khula jawari*: the resulting sound would be too nasal. On the contrary, 'soft *sitars*' (smaller, lighter and thinner) must have a slightly open *jawari* to provide the necessary loudness.

Even if musicians' discourses comprise contradictory information (the most divergent opinion regards the effect of the openness of the *jawari* on the duration of the resulting sound), several general statements can be found regarding the *jawari*. The point of *jawari*, whatever its fine-tuning, is to bring brightness, loudness and duration to the sound. *Jawari* also improves the pitch perception and therefore helps to play in tune. *Khula jawari* sounds are described as "lively, treble, sharper, warmer", whereas *bhand* is usually qualified as "rounded" and "sounding from the inside". Musicians concur also in the following opinions: *khula jawari* sounds are easier to produce than *bhand jawari* sounds and the *jawari* effect is an integral part of the sound: as put by Pandit Ashok Pathak, "without the *jawari*, the sound is nothing".

The divergences in musicians' evaluation and discourse regarding the *jawari* adjustment can be explained when considered as a question of degree: musicians do not always differentiate between open/closed *jawari* and 'over-open/closed': for example, an informant (Pandit Sanjoy Bhandhopadhyay) states that an open *jawari* increases the duration. However, another informant (maker Barun Roy) details: "a *jawari* too open provides a sound that is too short". It seems therefore that a *jawari* is adjusted within specific limits and that balance is, again, central.

6. NEW DESCRIPTORS

Based on the set of sounds recorded in Kolkata (2011), new descriptors have been conceived in order to better describe the timbre characteristics of *sitars*, and especially to distinguish between open and closed *jawari*. As only little specific information was contained in the musicians' discourse, criteria defining the sound quality had to be found in other sources. Even though most research focus on the modeling of the vibrating behavior of the string, some have noted specific sonorous characteristics:

1. Descending formants (not always heard on *sitar*, more frequently in *tampura* (Siddiq 2012; Bertrand 1992, Cuesta and Valette 1993)

2. Complex envelopes of the partials (Siddiq 2012)

3. Numerous beats are present on numerous partials (Bertrand 1991, Schmitt 2000).

Eventually, the general increase of number of intense overtones in the sounds has been observed by C.V. Raman as early as 1922.

Roughness analysis

Vassilakis (2005) has shown that the presence of numerous intense overtones can be linked with roughness. However, calculation of roughness is usually performed either as average or by frame. Since notes played on *sitar* present a complex temporal evolution, instead of computing a single average roughness for each note, we propose to describe temporal aspects of the roughness curve.

Plomp and Levelt (1965) have proposed an estimation of the sensory dissonance, or roughness, related to the beating phenomenon whenever pair of sinusoids are closed in

frequency. An estimation of the total roughness consists in estimating the dissonance provoked by each pair of sinusoids in each frame of sound. In other words, peaks are extracted in each frame of the spectrogram, and dissonance is estimated for each possible pair of peaks, and the average of all the dissonance between all possible pairs of peaks is finally computed (Sethares, 1998).

Figure 4a shows the roughness curve obtained using Sethares' approach implemented in the *mirroughness* operator included in *MIRtoolbox* (Lartillot and Toiviainen, 2007). We propose a new improvement of this roughness model. In Sethares' approach, for each pair of peaks detected in the spectrogram, Plomp-Levelt's predicted roughness value (which is estimated simply based on the relative frequencies of the two peaks) is multiplied by the amplitude of *both* peaks. We suggest an alternative where we multiply instead by the *minimum* of those two peaks. Why? Because the beating phenomenon between these two frequencies would not be influenced by the extra amount of energy on the peak of stronger intensity.

Figure 4b shows the roughness curve obtained using the proposed variant, which is available as a new '*Min'* option of *mirroughness* in the new version 1.5 of *MIRtoolbox*. This new roughness curve emphasizes better the part of roughness that appears after the initial attack of the note, and that seems closely related to the timbral particularities of sitars.

Figure 5 shows roughness curves for a same pitch (Pa played on high Ma string), with different dynamics and comparing closed and open jawari. We can observe how closed jawari leads to a particularly damped roughness evolution, while with open jawari the roughness is perceived for a longer time.

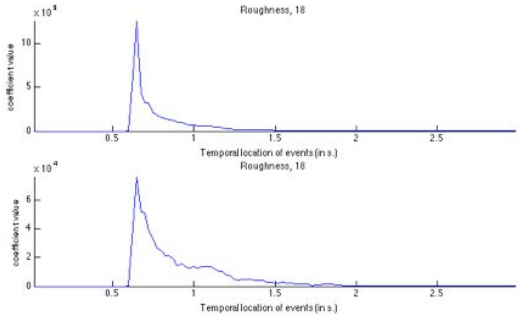


Figure 4. Roughness curve computed using the traditional Sethares method (top), and our proposed modified version (bottom).

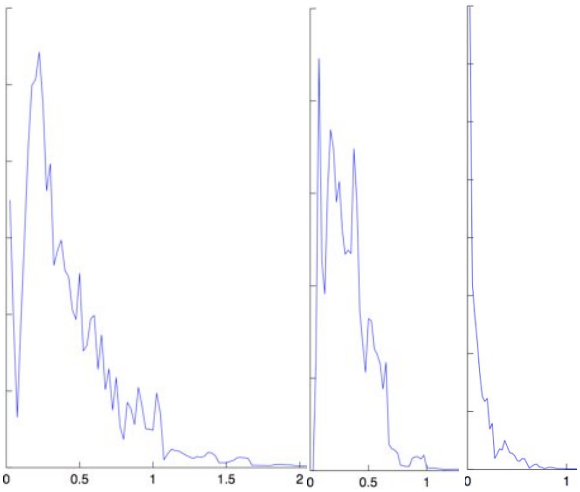


Figure 5. Temporal evolution of roughness while playing a Pa on high Ma string, played with open jawari, with soft (left) and medium (middle) strength, and played again with medium strength but with closed jawari (right).

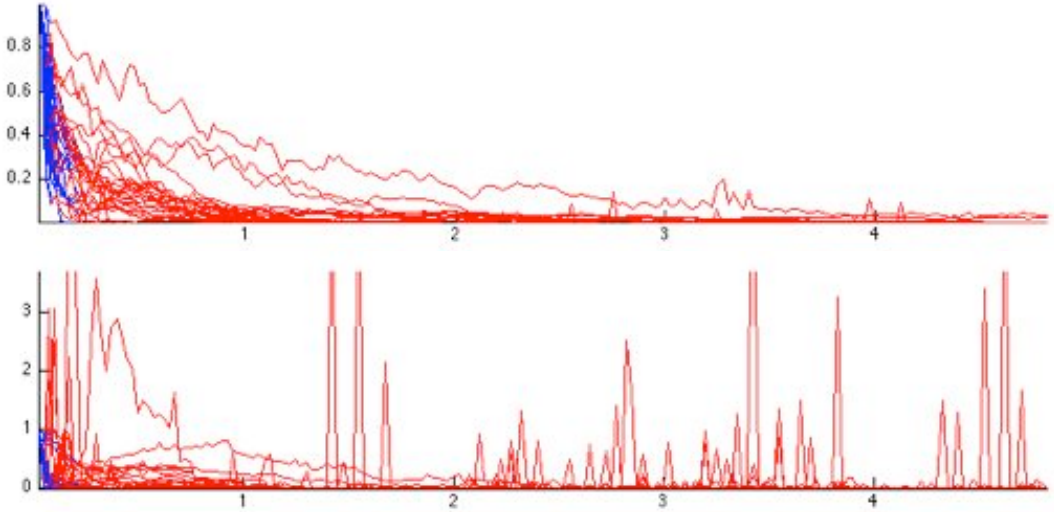


Figure 6. Superposed roughness curves of all the note recordings related to the closed jawari (top) and to the open jawari (bottom). All sounds start at time t = 0 s (on the X-axis) and with an initial energy amplitude normalized to 1 (on the Y-axis). The initial decrease of energy is represented in blue, and the curve turns red – and remains red until the end of the curve – when a new increase of roughness amplitude is detected.

In Figure 6, we have superposed all the roughness curves related to all the notes played[1] on open jawari on one hand, and to closed jawari on the other hand. We can notice that for open jawari, the roughness curve sometimes features an important increase after the note has been struck, with a delay ranging from a few milliseconds to half a second. At the apex of such reemergence of roughness, the actual roughness magnitude can sometimes be a few times higher than the maximal roughness at the attack of the note. For closed jawari, on the contrary, the roughness shows basically a decreasing trend, with some reemergence sometimes, but whose maximal roughness is much lower than the miaximal roughness magnitude during the attack of the note.

We introduce one way of describing the roughness curve called *cumulated roughness*, defined as the integral over time of roughness, indicates how much roughness sustains over time. The table below gives the basic statistics of cumulated roughness for both the set of notes recorded with a closed jawari, and with an open jawari.

Jawari	Min	Max	Mean	Median
Open	1.5	50.8	12.3	7.6
Closed	1.8	48.1	10.2	8.3

By just looking at the overall summation of roughness over time, open and closed jawari does not look very different.

In addition, through a basic analysis of the roughness curve, we detect any positive deviation from a predicted standard attenuation of roughness, indicating resurgence due to the non-linear phenomena related to *jawari* and *taraf*. This leads to an estimation of the increase of cumulated roughness due to that resurgence, compared to the standard attenuation, called *resurgent cumulated roughness*. The table below gives the basic statistics of resurgent cumulated roughness for both the set of notes recorded with a closed jawari, and with an open jawari.

Jawari	Min	Max	Mean	Median
Open	.04	45.2	10.0	6.0
Closed	.41	45.0	7.6	4.8

We can see now more clearly that a larger part of the roughness is expressed during the roughness reemergence phase for open jawari than for closed jawari.

Formant analysis

Other common timbral descriptions, such as brightness, spectral centroid or roll-off, give single descriptions of the whole spectral distribution. We propose instead to emphasize the decomposition of the spectral distribution into "formants": the nasal characteristic of *sitar* sound can be related to the energy on the formants higher than the first formant related to f0 and f1.

To detect formants, we use a standard method in voice analysis based on linear predictive coding, that we adapt a little in order to generalize it to a use outside the voice context. In particular, we avoid performing the pre-emphasis filtering as it is motivated by considerations related to voice and telecommunication. For the linear prediction, instead of fixing the model order to 8, which is tuned to the search for the three vocal formants, we tried we higher model ordered and found 12 as a good compromise for a rich analysis without too many irrelevant formants. Suitable formants are selected based on their frequency locations and their bandwidth. Common methods in voice processing choose 90 Hz for lower frequency threshold and 400 Hz for maximal bandwidth. In our experiments, trying larger bandwidths such as 500 Hz enables to extract a larger number of interesting formants.

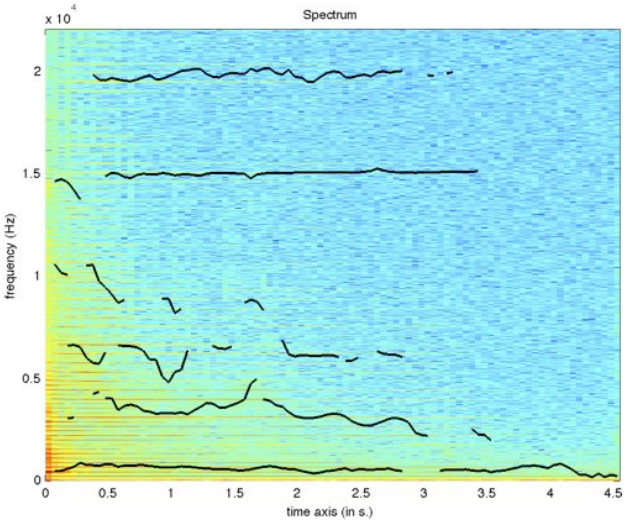


Figure 7. Formants detected on an analysis of a Pa played on a Ma string with an open jawari.

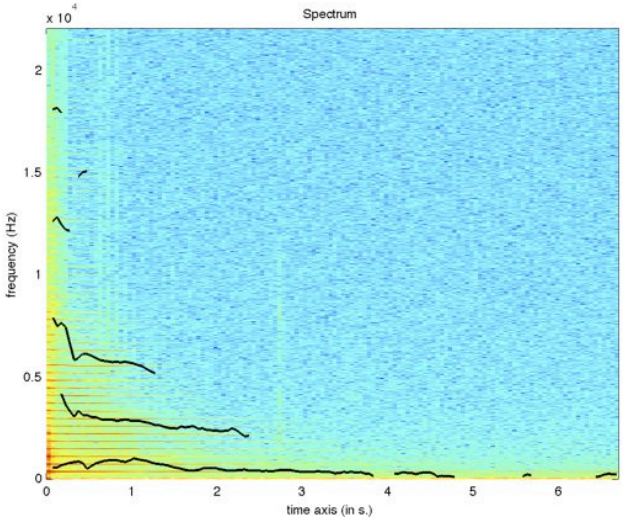


Figure 8. Formants detected on an analysis of a Ga played with moderate strength and with closed jawari.

[1] This experiments was done with a sound database somewhat different from the one used in the other analysis carried out in this paper. That set of recordings comprises sounds played on a 'soft sitar' with khula jawari played by an advanced student of the Academy.

Figures 7 and 8 show examples of interesting formants founds in the recordings. We can notice the typical descending formants as well as more complex fluctuations.

Detecting resurgence of energy on particular frequency regions

Energy resurgence can also be detected directly from the spectrogram, through a detection of local minima along the temporal evolution of each partial, leading to the segregation of resurgent parts in the spectrogram.

More precisely, the spectrogram is expressed in decibel, and each track in the spectrogram corresponding to a specific frequency is analyzed separately. Each track starts at the exact moment of the attack time where the energy is maximal. In order to focus on resurgence regions of sufficiently long time span (such as at least a few hundreds of milliseconds), and to filter out more spurious increase of roughness, each track is first filtered using an Infinite Impulse Response (IIR) low-pass filter. The filter is initialized so that it immediately tracks the decreasing envelope of the release phase appropriately from the start.

From the filtered track, we simply detect whenever the energy starts increasing, initiating a new resurgence phase. The resurgence phase terminates when the energy decreases back to the value at the beginning of the phase. To each resurgence phase is associated a strength, equal to the maximum different of magnitude within the phase. Resurgence phases that are too short (less than 10 ms) or with insufficient increase of magnitude (less than 10 dB in total) are discarded.

This analysis is performed in each frequency of the spectrogram separately, but a series of resurgence phases on similar temporal span can sometimes be detected for successive frequencies of the spectrogram, leading to convex frequency-time regions of resurgence of energy. Figures 9 and 10 show examples of resurgence phase detected by the algorithm.

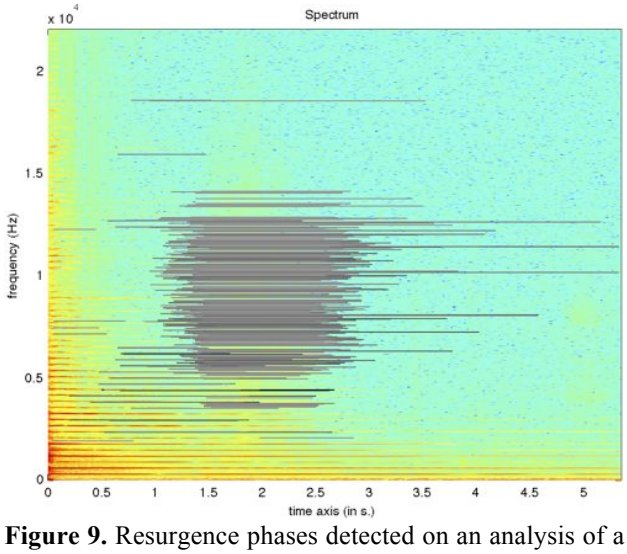


Figure 9. Resurgence phases detected on an analysis of a Sa played with moderate strength and with closed jawari.

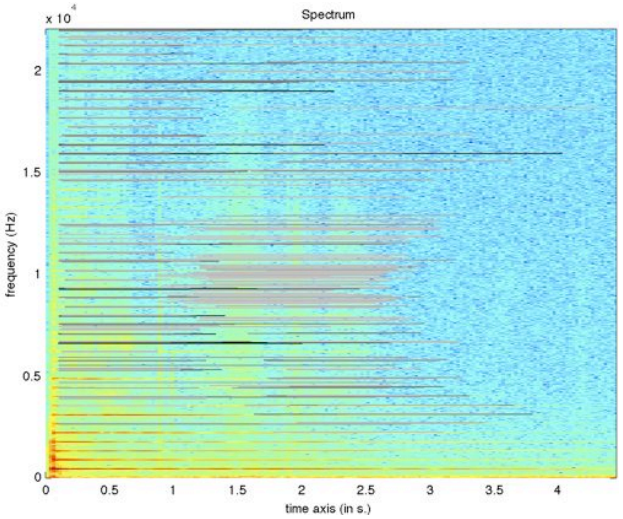


Figure 10. Resurgence phases detected on an analysis of a Pa on high Ma string played with moderate strength and with open jawari.

7. CONCLUSION

The timbre descriptors defined above show important variance within each type of adjustment of the *sitar* (with *open* or *closed jawari*): in particular, for specific pitch heights, particular resonances with sympathetic and extra strings can be observed. For this reason, instead of comparing global averages of the descriptors, the analysis of individual notes, systematically played with different nuances, are represented in multi-dimensional parametric spaces (corresponding to the different descriptors), so that the influence of these factors in the discrimination between open and closed *jawari* can be observed.

8. REFERENCES

Bertrand, D. (1992). *Les chevalets "plats" de la lutherie de l'Inde*, Paris: Editions de la Miason des sciences de l'homme, coll "Recherche, musique et danse".

Cuesta, C. & Valette, C. (1993). *Mécanique de la corde vibrante*, Hermes Science Publications.

Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction from Audio. In *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*.

Marcotty, T. (1974). Maintenance of the instrument. Djovari: giving life to the sitar. In Junius, M. M., *The sitar. The Instrument and Its Technique*, Wilhelmshaven Heinrichshofen's Verlag, 84-97.

Plomp & Levelt. (1965). "Tonal Consonance and Critical Bandwidth." *Journal of the Acoustical Society of America*.

Raman, C.V. (1922). On some indian stringed instruments. In *Proc. Indian Assoc. Adv. Sci.* 7 (pp. 29–33).

Schmitt, T. (2000). *Caractérisation des propriétés acoustiques des instruments à chevalet plat : cas de la* rudra vina, Rapport de stage [Unpublished Internship Report], Lyon: Ecole Centrale.

Sethares, W. A. (1998). *Tuning, Timbre, Spectrum, Scale*, Springer-Verlag.

Siddiq, S. (2012). A Physical Model of the Nonlinear Sitar String. *Archives of Acoustics, 37(1)*, 73–79.

Vyasarayani, C. P., Birkett, Stephen & McPhee John. (2009). Modeling the dynamics of a vibrating string with a finite distributed unilateral constraint: Application to the sitar. *Journal of the Acoustical Society of America, 125*(6), 3673-82.

Vassilakis, P. N. (2005). Auditory roughness as a means of musical expression. *Selected Reports in Ethnomusicology, 12*, 119-144

Weisser, S. and Demoucron, M. (2012). Shaping the resonance. Sympathetic strings in Hindustani classical instruments. In *Proceedings of the 163d Conference of the Acoustical Society of America (Acoustics 2012 Hong-Kong)*.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

MELODIC CONTOUR REPRESENTATIONS IN THE ANALYSIS OF CHILDREN'S SONGS

Christina Anagnostopoulou
Department of Music Studies,
University of Athens
chrisa@music.uoa.gr

Mathieu Giraud
CNRS, LIFL,
University of Lille
mathieu@algomus.fr

Nick Poulakis
Department of Music Studies,
University of Athens
npoulaki@music.uoa.gr

ABSTRACT

Children's songs is a particular musical genre related to folk music, with its own musical characteristics. This paper sets out to explore melodic contour in children's songs from seven different countries/nations across Europe. We look for distinctive contour patterns which differentiate the songs of each country. For pattern representation we use different viewpoints related to melodic contour, two of which also relying on beat information. Preliminary results are presented, and some initial observations regarding the patterns found, the representations used, and the genre as a whole, are discussed.

1. INTRODUCTION

Music plays a fundamental part in children's everyday lives: Children socialise with each other, reveal their emotions and entertain themselves through music. Their musical interactions can be surprisingly rich, varied and musically interesting. During the second half of the 20th century, the increased interest on children's study as individuals brought attention to their musicality from a cultural point of view, namely their expressive musical creativity as well as their responses and reflections to music (Small 1977, 1998; Blacking 1973). In this framework, it is a well-established fact that the notion of "song" plays a fundamental role in both children's educational and performance practices (Opie & Opie 1985). These simple in form and content songs usually share common musical (melodic and rhythmic) cross-cultural characteristics based on primal music fundamentals like the child's voice and gestures as well as the motions during the games played (Romet 1980). Research on children's musical songs however, apart from a few examples (such as Campbell 1991, 2010), has so far largely focused on social and educational perspectives, often ignoring the analysis of the music itself.

Melodic contour is a particularly significant musical feature relating to melodic shape, which has been studied from several perspectives: Music analysis and composition (e.g. Adams, 1976), semiotics (e.g. Seeger, 1960), music cognition (e.g. Lindsay, 1996), mathematical music theory (e.g. Buteau and Mazzola, 2008), and various computational approaches which take contour as a feature of computational music analysis (e.g. Kranenburg et al., 2011; Conklin 2010). In terms of representation, contour can be studied either as a continuous function (e.g. Muellensiefen and Wiggins, 2011), or described symbolically with discrete events at various levels of abstraction. One of the most elegant representations and implementations of melodic contour has been proposed by David Huron (1996), where 9 melodic shape descriptions have been applied to the analysis of melodic phrases of Western folk songs. Huron computes the shape type by comparing the pitch of the first note of the phrase, the average of all in-between pitches, and the pitch of the last note of the phrase.

Symbolic approaches to contour, however, like Huron's, do not usually take into account information related to rhythm and meter, which in some cases may be important for the characterisation of basic melodic shapes in the further analysis of a musical piece. In this study we propose a multi-level representation for melodic contour which takes into account beat information in order to describe melodic phrases. We then set out to explore patterns of melodic contour in children's songs. We believe that the choice of melodic contour might be a particularly appropriate level of representational abstraction for the analysis of children's songs in order to see the general melodic shapes that are predominant and characteristic. We look for distinctive patterns (Conklin, 2010) which characterise the songs of each country, as opposed to the songs of other countries.

The rest of the paper proceeds as follows: The next section describes the musical corpus of children's songs. Section 3 describes the methodology employed, including the definition of the contour representations chosen and the discovery of distinctive patterns. Section 4 presents some preliminary results found, and the paper ends with a discussion on the results, as well as pointers to future research.

2. THE CORPUS

A total of 110 traditional children's songs was collected, from seven different countries/nations across Europe: Catalunya (15 songs), England (15 songs), France (15 songs), Greece (20 songs), Spain (15 songs), Sweden (10 songs), and Turkey (20 songs). We selected those songs that seemed the oldest and more traditional for each country, based on information given to us by native speakers, and in this study we have not distinguished between songs made for children and songs made by children.

Each song was encoded into MIDI, and segmented into phrases based on the songs' lyrics, with a segmentation point at the end of each lyrics' phrase, giv-

ing a total of 505 segments. When more than one alternatives were possible, the segmentation giving the smallest possible units was chosen.

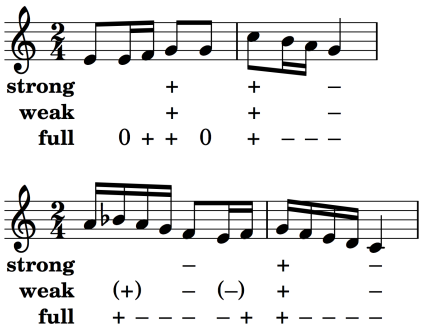


Figure 1. The three contour patterns on the first two segments of the Greek song "Πάνω στην κούνια".

3. METHODS

3.1 The contour representations

Each phrase in the corpus is represented using three pattern representations related to melodic contour. Two of them take into account the beat information (see also Figure 1) :

- *a full contour pattern,* which includes the melodic direction between each consecutive pitch of the phrase and its previous one (starting with the second note).
- *a strong contour pattern,* relying on pitches on the beats, including upbeats that might exist in the score, which are considered to be important for the overall shape of a phrase. To compute the strong contour pattern, we simply discard all notes not occurring on a strong beat (except upbeats), and compute the full contour on this reduced score.
- *a weak contour pattern*, that includes the strong contour, and, enclosed in parentheses, all contour changes of value on notes between the beats that were not present in the strong shape.

Note that the weak contour pattern can be different than the full contour, as the contour change between the beats does not influence the strong contour (for some examples see Figure 1).

The reason for trying out different contour representations is that we believe that melodic arch shapes, such as Huron's (1996), are better based on a reduced score, that is on the main notes of a melody, which in most cases in this corpus are found on the beat. At the same time, the information on surface melodic patterns is just as important, especially when looking at the similarity between songs within a musical style or across styles, so we included a full contour representation in our experiments.

The three types of contour patterns (full, strong, weak) were computed on all 505 segments. For each contour type, when duplicate patterns found in the same song, only one occurrence was kept. This computation gave 92 different strong contours, 211 different weak contours, and 279 different full contours. As a reference, we also classified all the segments with Huron's original nine contour classes, comparing the first pitch, the last pitch, and the average of all other pitches.

3.2 Discovery of distinctive contour patterns

In our attempt to discover distinctive patterns in the corpus, we followed the approach described in (Conklin, 2010). A distinctive pattern is one that is overrepresented in a corpus compared to an anti-corpus. This can be asserted by computing the likelihood ratio between the observed probabilities in a corpus and an anti-corpus.

For each pattern and each one of the 7 countries, we thus computed the probability of appearance in the corpus (songs from that country) and the one in the anti-corpus (songs from other countries). We kept patterns with at least ten occurrences in all the corpus and that were at least 2-fold overrepresented in a country compared to other countries.

4. RESULTS

Table 1 lists all distinctive patterns found in the way described above. This resulted to four patterns for strong contour and five patterns for weak contour.

Pattern	Contour	Country
[–,–]	strong weak	Turkey Turkey
[+,+,–]	strong weak	Catalunya Spain
[–,–,+]	weak	Turkey
[+,–,+]	strong weak	Greece Greece, Sweden
[–,+,–]	strong weak	France France

Table 1. Distinctive contour patterns found in the corpus.

With the same criterion (2-fold overrepresented, at least 10 occurrences), no pattern with the full contour or with Huron's melodic shapes was found distinctive. Lowering the threshold to 5 occurrences, the pattern [–,–] was also reported as distinctive, as well as three Huron's melodic shapes – however, in all these cases, the low number of occurrences prevented us from further analysis.

We observe that we have three types of distinctive patterns found: With one direction, with two directions, and with three directions. With one direction we have the straight downward motion ([–,–]), with a particularly high statistical significance for Turkey. With two directions we have the [+,+,–] and [–,–,+], instances of a convex and of a concave shape. With three directions we have [+,–,+] and [–,+,–]. Figure 2 shows instances of the pattern [–,–], Figure 3 instances of the pattern [+,+,–] and Figure 4 instances of the pattern [+,–,+]. Since all phrases in children's songs are short, there were no longer patterns found with more than ten occurrences in the current corpus.

5. DISCUSSION AND FUTURE WORK

The pattern [–,–] (Figure 2) shows a very high statistical significance for Turkish songs in all types of contour representations. Looking at the instances found, we observe that often these patterns are located at the end of the songs, denoting a cadence, or at an even number of a phrase, again denoting a (smaller) cadence. There are

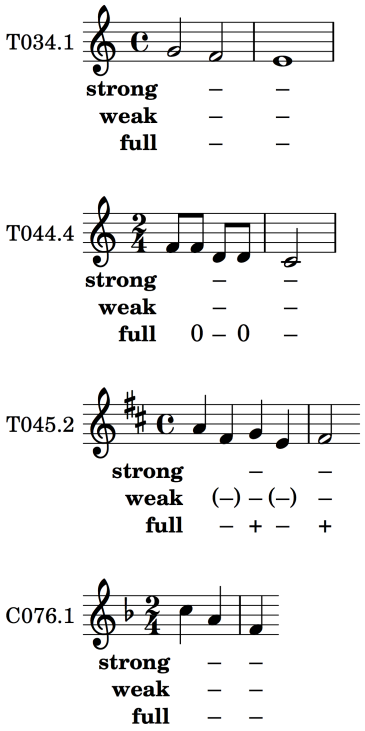


Figure 2. Strong contour [–,–] . This contour is five time over-represented in Turkish songs (9.1% of the Turkish segments) than in the rest of the corpus (1.8%).

however cases where this pattern is a first phrase, something we believe might be unique for Turkish songs. Further work could explore whether this overrepresentation could be explained by some linguistic or intonation characteristic of the Turkish language.

Interestingly there were no distinctive patterns found in English songs. All patterns found, apart from one, belong to countries of the Middle (France) or South Europe (Turkey, Spain, Greece, Catalunya), a fact that points to their special musical style. The strong pattern [+,+,-] appeared in both Catalan and Spanish songs, and shows one important similarity between the two.

It is noted that the differences between strong and weak contour patterns were not as big as expected: as expected, most patterns which appear in the strong contour representation also appear in the weak, but no weak contour with contour modifications between the beats was found as distinctive. This could be because the corpus is not large enough to gather songs on these contour modifications.

As expected, full contour patterns with a high significance were very few, in fact only one found by lowering the threshold. Our contour representations on the reduced score show thus firstly that it is possible to spot distinctive patterns in these representations, often shared across countries, and secondly that the instances of phrases found share some obvious melodic similarities which would have not been captured otherwise. We thus argue that the strong contour representation may be a good compromise between a very abstract representation (such as the Huron's melodic shapes) and a very detailed representation (such as the full contour or the continuous functions described by Muellensiefen and Wiggins (2011)).

The strong distinctive patterns found, apart from one, end up in a downward melodic motion. Even when one looks at the cases of the strong pattern [+,–,+], which is the exception, one can see that in the full version of its instances sometimes a downward motion can be detected. Also, these patterns [+,–,+] tend to have a counter-phrase following, which ends on a – . The trend to end on a – can be partly explained because all the songs are strongly tonal (whatever tonal means in each case), and a downward motion at the end of the phrase often reaches the tonal centre.

One question that might arise would be on the criteria used to create the reduced score upon which the strong contour patterns were calculated. Notes on the beat can sometimes be suspensions or other melodic embellishments that are not the main note of the melody in a Schenkerian sense. This can indeed be the case, but in the particular corpus analysed here this phenomenon is rare.

In general, it can be observed that the songs analysed are based on short phrases, simple melodic lines, symmetries, circularities and repetitions. This can be po-

tentially because musical learning, recalling, singing or playing needs to be enhanced. The ascending and descending directions of the melodies may have been shaped by various linguistic and kinetic factors, often simulating gestures, which relate to the context of musical performance of the songs. Further work is needed in order to compare children's songs as one corpus, with other folk song corpora.

Future work also includes the study of which contour shapes come at the various positions in the song (first, second phrase, etc). It also includes studying of sequences of shapes, to see what types of shapes follow each other. For example, by looking at phrases where the [+,–,+] pattern occurs, we have noticed that the next phrase ends with a downward motion. Inclusion of more features to describe the songs is also planned.

A systematic analysis of children's songs of various cultures can contribute towards an awareness of this music as a special genre with its own characteristics, and in viewing children as conscious musicians – especially if the approach takes children's music as its starting point.

6. REFERENCES

Adams, C.R. (1976). Melodic Contour Typology. *Ethnomusicology*, 20(2), 179-215.

Blacking, J. (1973). “How Musical is Man?” University of Washington Press.

Buteau, C. & Mazzola, G. (2008): Motivic Analysis Regarding Rudolph Réti: Formalization Within A Mathematical Model. *Journal of Mathematics and Music,* 2(3), 117-134.

Campbell, P. (2010). "Songs in Their Heads: Music and Its Meaning in Children's Lives". Oxford University Press.

Campbell, P. (1991). The Child-Song Genre: A Comparison of Songs by and for Children. *International Journal of Music Education,* 17, 14-23.

Conklin, D. (2010). Discovery of distinctive patterns in music. *Intelligent Data Analysis,* 14(5), 547-554.

Huron, D. (1996). The Melodic Arch in Western Folk Songs. *Computing in Musicology*, 10, 3-23.

Lindsay, A.T. (1996). Using contour as a mid-level representation of melody. Master's Thesis, MIT 1996.

Müllensiefen, D, and Wiggins, G. (2011). Polynomial functions as a representation of melodic phrase contour. *Systematic Musicology Empirical and Theoretical Studies*, 63-88, Peter Lang.

Opie, I. & Opie, P. (1985). The Singing Game. Oxford University Press.

Seeger, C. (1960). On the moods of a music logic. *Journal of the American Musicological Society*, 13, 224-261.

Romet, C. (1980). The Play Rhymes of Children: A Cross Cultural Source of Natural Learning Materials for Music Education. *Australian J. of Music Education,* 27, 27–31.

Small, C. (1977). Music, Society, Education. Calder.

Van Kranenburg, P., Biro, D.P., Ness, S., Tzanetakis, G. (2011). 'A computational investigation of melodic contour stability in Jewish Torah Trope Performance Traditions. *Proceedings of ISMIR 2011*.

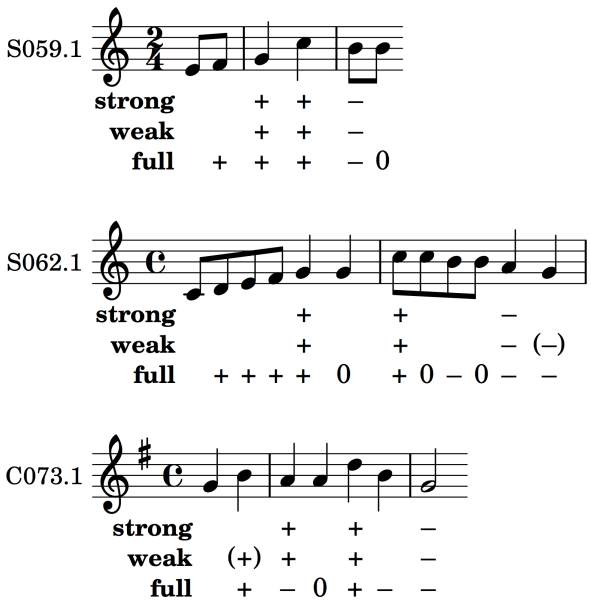


Figure 3. Strong contour [+,+,–], represented 2.4 times more in the Catalunya and Spain songs than in the rest of the corpus.

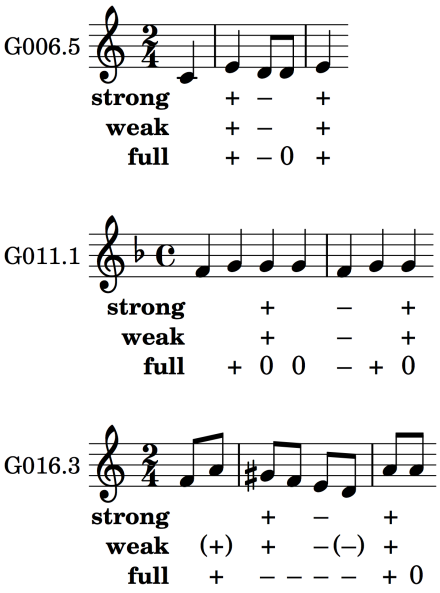


Figure 4. Strong contour [+,–,+], represented 4.8 times more in the Greek songs than in the rest of the corpus.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

AN ORIGINAL OPTICAL-BASED RETRIEVAL SYSTEM APPLIED TO AUTOMATIC MUSIC TRANSCRIPTION OF THE MAROVANY ZITHER

Dorian Cazau, Olivier Adam
Université UPMC Paris 6
Equipe Lutheries Acoustique Musicale (LAM)
cazau@lam.jussieu.fr

Marc Chemillier
Ecole des Hautes Etudes en Science Sociale
Centres danalyse et de mathmatique sociales
chemilli@ehess.fr

ABSTRACT

In this work, we introduced an original optical-based retrieval system dedicated to the music analysis of the *marovany* zither, a traditional instrument of Madagascar. From a humanistic perspective, our motivation for studying this particular instrument is its cultural importance due to its association with a possession ritual called *tromba*. The long-term goal of this work is to achieve a systematic classification of the *marovany* musical repertoire in this context of trance, and to classify the different recurrent musical patterns according to identifiable information. From an engineering perspective, we worked on the problem of competing signals in audio field recordings, e.g., from audience participation or percussion instruments. To overcome this problem, we recommended the use of a multichannel optical recording, putting forward technological qualities such as acquisition of independent signals corresponding to each string, high signal to noise ratio (high sensitivity to string displacement / low sensitivity to external sources), systematic inter-notes demarcation resulting from the finger-string contact. Optical signal characteristics greatly simplify the delicate task of automatic music transcription, especially when facing polyphonic music in noisy environment.

1. INTRODUCTION

The *marovany* is a tall zither in the form of a rectangular box built from recycled wood products. The metallic strings, measuring up to 1 m 20, and mostly coming from brake cables type motorcycle, are stretched on each side of the box. They are nailed at each end on an easel, made of wood or metal, and are raised by battens whose places along a string determines its pitch. Musically, each set of strings forms, like the famous tubular zither *valiha*, an alternating diatonic scale. The repertoire of the *marovany* consists of a succession of melodic phrases most often played in arpeggios. There is no vertical writing properly speaking in this music, excepting a few punctual chords. However, the high tempo at which notes are played, combined with the facts that strings are barely muted within a musical phrase and that they often resonate with each other, confers to this music a complexity of analyse comparable to that provided by polyphonic music. Wood type, sizing and number of strings of a zither are not fixed. One can indeed find zithers made of light or heavy weight wood, measuring from 1 to 2 meter in length, possessing from 8 to 12 strings on each side, with battens also ranging from 2 cm to 0.5 cm in height (it is known that bringing strings closer to the soundboard produce more powerful sounds, according to an effect called *mafo be*, "louder"). The zither used for our study, whose photo is given in the figure 1, possesses 13 strings on each side, with a pitch range covering more than two octaves.

In addition to its pure musical interests, the study of the *marovany* zither presents a major interest as its take part in a possession cult called *tromba*. This particular type of trance is "musically induced" in that the possessed person is stimulated by the music played. In the context of trance *tromba*, the mode of playing of the *marovany* mainly consists of a quick succession of melodic motifs, progressively transformed through multiple obsessing repetitions. This instrument is often accompanied with a rattle called *kantsa*, built from recycled cans filled with grains and nailed to a wooden handle, which constitutes with hand-clapping the rhythmic base of this music. This social context of trance associated with the *marovany* zither likely determines its musical repertoire. Indeed, music related to possession cults often carries identifiable information able to make some complex connections between music and symbolic extra-musical entities (Rouget, 1980). A large part of the *marovany* repertoire may therefore take the form of an association table between musical formulas and certain divinities (Chemillier, 2000). Another functional aspect of the *marovany* during *tromba* is that it greatly participates to the collective effervescence conductive to the trance itself, through specific uses and progressions of musical patterns. In order to better understand these two aspects of the link existing between *marovany* zither and trance *tromba*, an analyse of its repertoire must be performed, with a systematic inventory of musical patterns and air/divinities associations. Within a larger picture, studies dealing with neurophysiological mechanisms within trance events (Dianteill & Hell, 2008; Hell, 2008) could benefit from precise musical data provided by our system.

Malagasy zithers, and more particularly the *valiha* (made of bamboo), that is considered as the national instrument of Madagascar and from which the *marovany* is derived, have already been subjected by numerous ethnomusicological researchers (Razafindrakoto, 1999; Domenichini, 1984). However, as far as the authors know, there is currently no large-scale systematic analysis and classification of its repertoire, based on precise musical (rhythmic, modal, structural properties) and extra-musical (functional roles, symbolic content) criteria. To provide a deep insight into musical functionalities of the *marovany* in the context of trance *tromba*, investigations on the field must be undertaken, which systematically monitor and characterize mu-

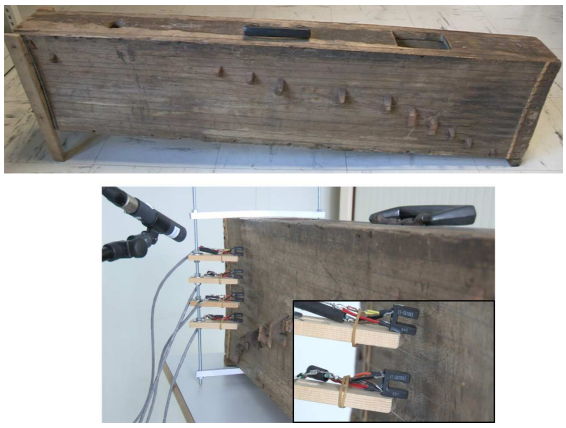


Figure 1: Photo of a zither *marovany* (on the top) and of the implemented optical-based system, with the details of sensors in the close-up (on the bottom)

sical patterns to statistically evaluate their occurrences over different trance sessions (e.g. for the recurrence or structural roles of certain musical motives), and to draw correlations between musical, behavioral and symbolic information over the time of a trance (e.g. for the way musical formulas are renewed and the impact it has on the possessed). A systematic study of such concordances should allow to establish the catalogue raisonné of the common repertoire of the *marovany* zither players in context of trance *tromba*. The automation of this process is made imperative as a manual transcription may be cumbersome, considering that trance sessions can last several hours and that there is no manuscript support for this music. Also, the complexity of this transcription (due to speed of playing, polyphonic characteristics, noisy environment) implies a great variability in hand-made results, making them prone to errors without possible estimation of their quality. Standard audio-visual devices recording each trance(e.g. see an excerpt of a trance video and other audio-visual material about the *marovany* on the web page of Chemillier (2012)), which allows the analysis of the behavioral indices mentioned above, do not provide an optimal support for music transcription of *marovany* music, as they exhibit competing signals within a noisy environment. To remedy this problem, this paper presents in section 2 an optical-based retrieval system dedicated to in situ recordings of musical airs of the *marovany* zither. This system was further integrated to an acquisition and processing chain aiming to perform music information automatic retrieval, presented in section 3.

2. CONCEPTION OF AN OPTICAL-BASED RETRIEVAL SYSTEM

Several constraints must have been considered in the choice of the recording system. Intrinsic constraints to the *marovany* mode of playing firstly, including its speed, the different modes of attack and string muting, the polyphonic sequences (due to intermittent chords and mutual resonances induced by the strings). In addition to that we have exterior constraints, such as external sound sources (mainly the rattle, hand-clapping of the audience, vocal interjections of the possessed), environmental (high humidity and heat) and technical (unreliable electrical sources) conditions. It is then preferable to avoid using too sensitive and preamplified systems (e.g. 48 V phantom powering), which could degrade very quickly. The accumulation of these constraints make the overall audio signal hard to acquire and process. The optical-based system of music acquisition described in the following has been conceived to optimize the task of automatic music transcription, attempting at best to comply with all these constraints.

Optical-based systems have already found various applications, such as metrological measures of string displacement (Seydoux, 2012; Chabassier, 2012) or a MIDIfication [1] of a piano through the Moog piano-bar technology (Mowat, 2005; Assayag & Bloch, 2008). Our system, illustrated in the figure 1, is closer from this second application, although distinguishing itself through the desire to integer as accurately as possible a great number of physical parameters characterizing the sound quality of the instrument. The selected optical sensor are slotted optical switches consisting of an infrared emitting diode and an NPN silicon phototransistor. It has a fork design, with the string placed between the two branches as illustrated in the close-up of the figure 1. On one side, the light-emitting diode (LED) emits a light beam whose diameter is 0.5 mm and wavelength 940 nm. On the other side, the phototransistor has a peak of sensitivity at 850 nm. When the string passes through the laser it modulates the output current of the sensor, accordingly to the surface of the laser shadowed by the string. In order to maximize the dynamic of the optical signals and obtain sharp transient attacks, the narrowest possible diameter for the laser is used. Such sensor then acts as a digital switch with a robust sensitivity to string displacements. An enhanced low current roll-off is used to improve contrast ratio and immunity to background irradiance. The power pack of the optical sensors needs a continue tension of 5 V, and is thermally isolated, which makes it well aligned with field conditions. Two compact portable digital recorder ZOOM R16, allowing the recording of 2 x 8 tracks simultaneously, are used for data acquisition of optical signals. Those are directly saved in the Cubase software. A synchronous audio signal of reference is also recorded with a microphone Neumann KM 184 mt. Sampling frequency for all recordings is 44.1 kHz, with 16 bits.

Each string is equipped with an optical sensor to get individual signals. Two constraints have been taken into account in the placement of these sensors, attached on a vertical bar exterior to the instrument (see figure 1). On one hand, the system must not be too cumbersome and disturbs the playability of the instrument. On the other hand, the measuring point of string displacement may create a bias in the amplitude measure. Indeed, as this displacement consists of a superposition of vibratory modes, de-

[1] Acronym meaning the in-situ conversion of an acoustic instrument into its homologous MIDI.

fined as a succession of nodes and anti-nodes, if a sensor is placed on a modal node the energy contribution of the corresponding mode is null. To answer these two constraints, the bar of sensors is positioned near the easel, in such a way that the playing zone is less disturbed and that the sensors are roughly placed on the ascending slope of the anti-node following directly the easel-related node common to all modes.

The figure 2 represents the spectrograms of the audio signal and four optical signals (after post-processing, see section 3) respective to four distinct strings, recorded on a traditional tune called *Sojerina*. As it can be seen, optical signals offer a high signal to noise ratio and sharp transients, both in the attack phases (end of plucking) and release (beginning of plucking with the contact finger-string). The independence of each string is well respected, each sensor detecting solely the vibration associated to its string. In addition to that, we have a good separability of successive notes of a same string, with inter-notes blanks resulting from the instants of finger-string contacts. Eventually, such a system of acquisition decomposes a multi-source audio signal into simple identifiable components, simplifying more particularly the complex analysis of a polyphonic sequence by processing individually several monophonic sequences. Another advantage of the optical technology is that it allows a straightforward conversion of string displacement to MIDI format files. Such signals also make easier their post-processing for analysis, and are well suited for applications to real-time.

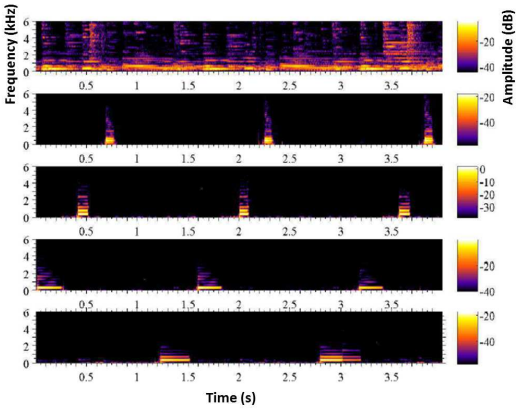


Figure 2: Spectrograms of the signal audio (on the top) and its decomposition in four optical signals (on the bottom) extracted from a *Midegana* air

We now present a short comparative study on the acoustic characteristics of audio and optical signals. After post-processing optical signals (see section 3), five acoustic descriptors have been computed on a set of thirty pairs of notes {audio;opical}, played separately and let in free oscillation until extinction. The used descriptors are defined as follows:

AT , the attack time (in s) is defined by the necessary time for the signal to reach 95 % of its maximal energy E_{max}

$$s(n = AT) = 0.95E_{max} \quad (1)$$

D , the physical duration of the signal (in s) will be defined as the time during which the signal energy remains between 5 % and 95 % of its maximal energy E_{max}

$$D = \{n/s(n) > 0.05E_{max} \ \& \ s(n) < 0.95E_{max}\} \quad (2)$$

E , the energetic level rms (in Pa) of a signal is defined by

$$E(k) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}|(x + kN)|^2} \quad (3)$$

computed for K successive frames of N samples ;

HD, the Harmonicity Detector (unitary value without dimension) is an indicator of harmonicity. The principle (Youngmoo & Whitman, 2002) is to automatically scan the spectral density of a signal with a comb filter whose fundamental frequency F_0 and varies within a given range of interest. When the valleys of this filter coincides with the peaks of an harmonic sequence for a particular F_0, their product will result in a very weak value which traduces the presence of an important harmonicity. Mathematically, we define it as

$$HD = min(\frac{E_{pond}}{E_{init}}) \quad (4)$$

with $E_{init} = \sum |Y(k)|^2$ and $E_{pond} = Filt(k, k_o)E_{init}$, where Filt is a comb filter defined as Filt = $2(1 - |cos(\frac{\pi F}{F_0})|)$.

The descriptors E and HD are evaluated relatively to arbitrary references respective to the types audio and optical. Absolute acoustic differences between these two types of signals are simply quantified through the operator $\Delta_D = |D_{Audio} - D_{Optical}|$, where D represents a given acoustic descriptor. The table 1 presents the results of this operator, showing that the distortional impact of optical-based acquisition mode (not really physically correlated to human auditory perception) is minor when considering temporal profiles. However, spectral content shows more significant differences. An observed tendency is that the harmonic structure of optical signals is stronger, which can be explained by the fact that a direct measure of string displacement privileges its fundamental frequency ans its harmonics in the observed vibratory behavior, minimizing the effect of coupling with the more complex modes of the soundboard. The difference on the amplitude may take important values when a string vibration excites strongly some of the soundboard modes, which allows a very efficient energy transfer from the string to the table.

Descriptors	At	AD	dB	HD
E(Δ_D)	0.0129	0.4366	0.23	0.26
$\sigma(\Delta_D)$	0.0028	0.11	0.15	0.09

Table 1: Average E and standard-deviation σ of the acoustic absolute differences between optical and audio signals for the descriptors At, AD, dB, HD

3. APPLICATION TO MUSIC INFORMATION RETRIEVAL IN THE CONTEXT OF TRANCE *TROMBA*

Information retrieval for music transcription can be classed into several levels: the low-level (pitch, note attack and duration), sufficient for the constitution of a partition, and the high-level (tonality, instrument recognition), which asks for more global and complex notions. Our chain of transcription consists of an acquisition data system (described above) and a processing part including analyse algorithms which will determine the durations, the pitches and the amplitudes of the played notes. These information will then be compiled into a MIDI file, which can be read and edited on any audio sequencer and score edition program.

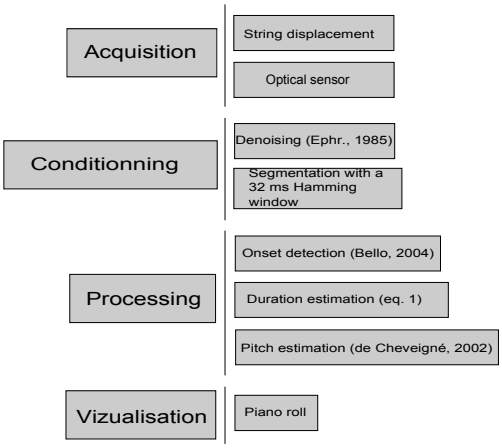


Figure 3: Block-diagram of the different functions constituting the detection and acoustic characterization algorithm

Once optical signals are properly acquired, their transcription does not pose any specific difficulties. Figure 3 represents a block-diagram of the different functions constituting the analyse chain, from the acquisition to the computational processings of the *marovany* note detection and acoustic characterization. The post-processing of optical signals is as follows. Because of memory concerns, sequences of 5 s are first imported in the software Matlab. An adaptive filtering (Ephraim & Malah, 1985) is then applied to optimize the signal to noise ratio, mainly deteriorated by parasite noise coming from electronics and mutual resonances of strings [2] This algorithm of denoising takes as inputs segments of noises, and allows their subtraction to the signal by minimizing a prediction error with a least-mean square optimization. A 0.049-s hamming window with a 0.005-s overlapping (that is 11.6 ms, providing a temporal resolution whose order of magnitude is similar to the time attack) scans the entire sequence. Each onset of notes is detected using a spectral difference which takes into account the phase increment, as introduced by Bello et al. (2004):

$$\hat{X}_{k,n} = |X_{k,n-1}|e^{j(2\phi_{k,n-1}-\phi_{k,n-2})} \quad (5)$$

with n the index of each window. As *marovany* sounds consist roughly of a superposition of short stationary sinusoids, the occurrence of an onset generates a peak in the prediction error defined by

$$r(n) = \sum_{n=1}^{N} |\hat{X}_{k,n} - X_{k,n}| \quad (6)$$

Windows for which this residual exceeds a fixed threshold are validated as onsets. From this detected onset, the descriptor E (eq. 3) is computed for the neighbouring windows to search the local maximum $E_{max}(i)$ associated with the note i, assuming this maximum is located near the onset, as expected for notes played by plucked string instruments. Then, E is computed on all the windows following the onset until the energetic value decreases below 5 % of $E_{max}(i)$, which may then be read as an adaptive note-specific energy threshold, or until another peak in the residual r is found. This estimation allows us to deduce the note duration (eq. 2), and its amplitude by averaging the energy over all windows within the note. We are not interested in the absolute amplitude of the notes, but only in their relative values within an air, in reference to a value determined by the MIDI gain. Once all notes are detected, the pitch is estimated within each window, using a robust algorithm derived from the autocorrelation method for pitch estimation (de Cheveigné & Kawahara, 2002). Figure 4 represents the evolution of a waveform signal processed through this algorithm.

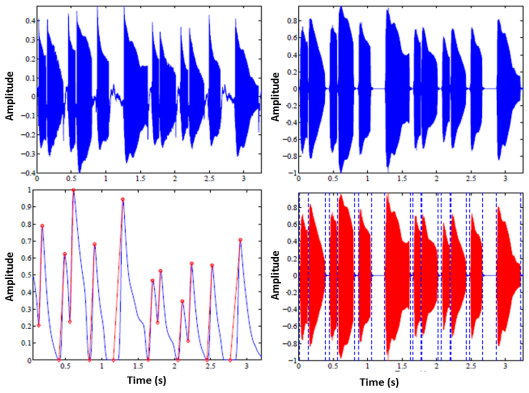


Figure 4: Evolution of a waveform signal processed through this algorithm. From left top to right bottom: original optical signal, denoised signal, residual r with location of onsets, and segmented signals.

[2] Although this acoustic phenomenon is considered as a disturbing noise in our situation of low-level transcription, it takes an important place in the definition of the instrument timbre.

This algorithm was evaluated on hand-labelled sequences taken from variants of *Midegana* and *Sojerina* airs (see below for details), containing information on the temporal location, duration, average amplitude and pitch. The tolerances for a correct estimation are fixed to 32 ms on the onset time and to 0.5 s for the duration. An application of the previous algorithm to the audio signal achieves performances between 50 and 60 % of correct note detection, whereas optical sequences provide satisfying results ($<$ 95 %).

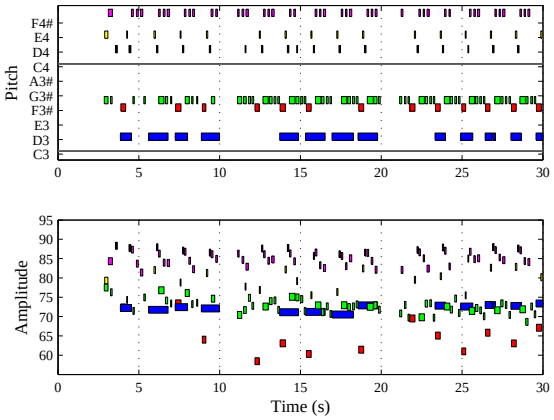


Figure 5: Piano roll of an automatic transcription of a *Midegana* air

We now propose two illustrations of this method through automatic transcriptions of two traditional Malagasy tunes, a *Midegana* and a *Sojerina* (audio files are available on the web page of Chemillier (Chemillier)). These tunes have been recorded in Madagascar and played by the musician Velonjoro. Technical problems on the spot rendered several sensors inoperative [3], and each variant of the airs has been partially reconstructed from different repetitions of a same sequence. In spite of these constraints of a manual intervention to resynchronize the tracks and a superposition of distinct loops, the signals sound satisfactory, with a good preservation of the rhythmic vitality of Velonjero's playing. Figures 5 and 6 show the piano rolls of these air variants. The piano roll is a means of representing graphically a MIDI file. On the vertical axis are the different notes, represented by rectangles either through their respective pitch (top graph) or their amplitude (bottom graph), and on the horizontal axis is the time. It is then easy to visualize the played notes over time.

The precision of the proposed automatic transcription system is well adapted to the speed of play of the musician and captures properly certain rhythmic structures characteristic of the *marovany* musical repertoire, as shortly explained now. From top to bottom, the left graph of figure 7 superposes as a function of time the automatic transcription performed on the *Sojerina* tune, with the rhythmic

[3] The main goal of the mission during which these recordings were done was to test a prototypical version of our acquisition system (Chemillier, 2012). The next one is planned for the summer 2013, and will benefit from a finalized and fully operational version of our transcription device.

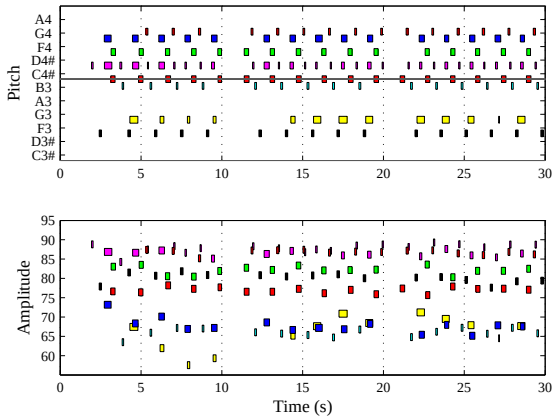


Figure 6: Piano roll of an automatic transcription of a *Sojerina* air

base given by hand clapping (in MIDI clap) and the rattle *kantsa* (in audio track). The right graph of figure 7 superposes against time their audio signals, from top to bottom: hand clapping, rattle, original record and MIDI transcription. Comparing audio tracks, we can see that the original audio and the MIDI file resulting from the automatic transcription are well synchronized, although it lacks a few notes in the MIDI file due to missing sensors as explained above. We find the contrametric character of this music (Chemillier et al., 2013), as we see that the rattle accent is slightly out of sync with the pulsation given by hand clapping, intervening more on the off-beats. More precisely, it falls on the second 8th note of a ternary subdivision of the pulsation (each pulsation is divided into three 8th notes). In the MIDI file, there are two obvious thirds: D-F# et C-E, and then a descending arpeggio G-R-B. As these two thirds and the G are most frequently placed on the rattle accent, and not on the hand clapping, the zither plays indeed out of the beat. This characteristic is confirmed through audio tracks where we can see both in the original and the transcribed MIDI file, that a stronger intensity is present in the two thirds, in coincidence with the rattle accents.

4. CONCLUSION AND PERSPECTIVES

An optical-based recording system applied to automatic music transcription of the *marovany* zither in context of trance *tromba* has been introduced in this work. The signals acquired from the optical-based retrieval system and post-processed present optimal characteristics for this task, with easily identifiable musical features extracted in a robust way. Among its conspicuous technological advantages, we can mention the high signal to noise ration, the multichannel output with independent signals corresponding to the played strings and the automatic demarcation between successive notes of a same string. Also, the low time-consuming computational method (as performing elementary operations directly on the acquisition buffers) is well suited for real-time analysis, allowing monitoring musical information simultaneously to events, and exploiting directly new findings on the field. In definitive, the mul-

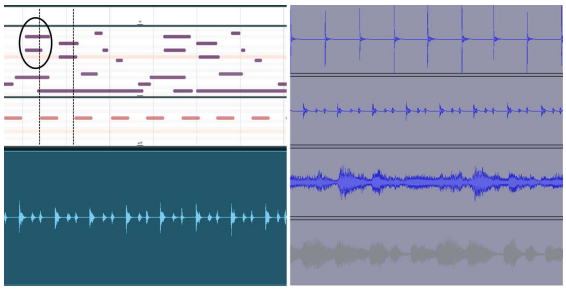


Figure 7: On the left, superposition against time of the MIDI transcribed file of a *Sojerina* air with the rhythmic base given by hand clacking (in MIDI clap) and the audio track of the rattle *kantsa*. The two vertical bars indicate two beats given by the MIDI clap, and the encircled part illustrates the contrametricity of this playing. On the right, superposition against time their audio signals, from top to bottom: hand clapping, rattle, original record and MIDI transcription.

tichannel output of such a device, insensitive to external sounds, offers an efficient alternative to the task of audio source separation, crucial to extract in-situ music information from the *marovany*. This system was conceived to meet the long-term demand of developing tools to perform a systematic classification of the repertoire of the *marovany* during trances, in a robust and automatic way.

Although the current interest in the *marovany* music deals with elementary musical information such as duration and pitch range of the notes, the need of reworking on audio data could be felt to integrate high-level information acoustic proprieties, including vibro-acoustic proprieties of the whole instrument, as the instrument timber, string shock modes, and the acoustic intensity. However, the optical system gives a complementary reliable support for following audio-based investigations, allowing a direct access to information simplifying complex problems linked to direct work on audio, such as the number of vibrating strings for studying the polyphony segments in a track.

Study of the *marovany* repertoire in trance *tromba*, founded on musical criteria, and complementing other behavioral indices observed with an audio-visual device and audio data, should bring original elements of investigation to the fascinating relationships between music and trance. Another application theme of such a system would be the Human-Machine musical interaction, through the OMAX improvisation IT environment (Nika & Chemillier, 2012) (developed from the OMax environment (Assayag et al., 2012) in collaboration with IRCAM). Future musical projects could involve malagasy musicians in this environment, using MIDI data from our retrieval system. Questions of a more aesthetic character (acceptability of musical formulas derived from a known repertoire, oral transmission of this skill, musical interest in the amplification, virtual reorchestration of a musical environment and real-time modifications of musical parameters) will be considered in future investigations following this direction.

5. ACKNOWLEDGEMENTS

This work was realized with the support of the French National Research Agency, in the framework of the project “IMPROTECH”, ANR-09-SSOC-068. This work largely comes from two Master projects realized in the LAM: *Transcription automatique de la cithare malgache* by Joachim Flocon-Cholet and *Etude de la cithare malgache en vue de la transcription automatique* by Mathilde Paul. Special thanks are given to Velonjoro who provided audio database, and Laurent Quartier for his high-quality technical support in the development of our transcription device.

6. REFERENCES

Assayag, G. & Bloch, G. (May 2008). Omax. the software improviser. In *Documentation version 2, IRCAM*.

Assayag, G., Bloch, G., Chemillier, M., Cont, A., & Dubnov, S. The omax project page, available at `omax.ircam.fr`.

Bello, J. P., Duxbury, C., Davies, M., & Sanders, M. B. (2004). On the use of phase and energy for musical onset detection in the complex domain. *IEEE Signal Processing Letters, 11*, 553–556.

Chabassier, J. (2012). *Modélisation et simulation numérique dun piano par modéles physiques.* PhD thesis, Ecole Polytechnique.

Chemillier, M. Web page on the marovany cithara : `http://ehess.modelisationsavoirs.fr/marovany/index.html`.

Chemillier, M. (1998-2000). Esthétique et rationalité dans les musiques de tradition orale. Technical report, Rapport de recherche au Ministére de la Culture.

Chemillier, M. (2012). Compte-rendu de mission sur les capteurs optiques - madagascar 29 juillet 13 août 2012, projet anr improtech. Technical report, Available on `http://ehess.modelisationsavoirs.fr/marovany/capteurs/`.

Chemillier, M., Pouchelon, J., André, J., & Nika, J. (2013). Le jazz, l’afrique et la contramétricité. To be published, Anthropologie et société

de Cheveigné, A. & Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. *Journal Acoustical Society of America, 111*, 1917–1930.

Dianteill, E. & Hell, B. (2008). Le possédé spectaculaire. In *Gradhiva, n8, p 4-5*.

Domenichini, M. (1984). *Stanley Dani*, chapter Valiha, (pp. 705–706, vol. 3). London:Macmillan.

Ephraim, Y. & Malah, D. (1985). Speech enhancement using a minmum-mean square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing, 33*, 443–445.

Hell, B. (2008). Négocier avec les esprits tromba mayotte. retour sur le théatre vécu de la possession. In *Gradhiva, n7, p. 6-23*.

Mowat, W. (2005). Bob moog piano bar : Midi output device for acoustic pianos. In *Sound On Sound, available at* `www.soundonsound.com/sos/mar05/articles/moogpianobar.html`.

Nika, J. & Chemillier, M. (2012). Improtek, integrating harmonic controls into improvisation in the filiation of omax. In *ICMC*.

Razafindrakoto, J. (1999). Le timbre dans le rpertoire de la valiha, cithare tubulaire de madagascar. *Cahiers dethnomusicologie*, *2*, 2–16.

Rouget, G. (1980). *La Musique et la transe. Esquisse d'une thorie gnrale des relations de la musique et de la possession*. Prface de Michel Leiris (Publi avec le concours du CNRS). Paris, Gallimard.

Seydoux, L. (2012). Mesure de dplacement de cordes avec des fourches optiques. Master's thesis, Mmoire de fin d'tudes, LAM (Laboratoire Acoustique Musicale).

Youngmoo, E. K. & Whitman, B. (2002). Singer identification in popular music recordings using voice coding features. In *ISMIR*.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

TRACES OF EQUIDISTANT SCALE IN LITHUANIAN TRADITIONAL SONGS

Rytis Ambrazevičius **Robertas Budrys**
Kaunas University of Technology, Kaunas, Lithuania
rytisam@delfi.lt budrys@super.lt

ABSTRACT

The study discusses features of musical scales in Lithuanian traditional vocal performances. Manifestation of the equidistant model of musical scale and its interplay with diatonic thinking are considered. Three samples representing three different ethnomusical dialects (54 songs, in total) are examined: the methods of diatonic contrast and clustering are applied in the processing of results of the acoustical pitch measurements. The evaluation of diatonic contrast shows that the song examples resembling the theoretical equidistant scale prevail over those resembling diatonics. The cluster analysis reveals certain groups of the scales characteristic of the examined dialects.

1. INTRODUCTION

1.1 Issue of the Ancient Greek modes

Question of equidistant scale seems to be closely connected to the phenomenon of so-called Ancient Greek (or Gregorian) modes, especially when modulations between the different modes are registered. The notion of Ancient Greek modes is quite frequent in the studies of Lithuanian folk music. It occupies a significant place in different classifications of modes. This idea was introduced to Lithuanian ethnomusicology (most clearly) in the first half of 20th century by Brazys (1920) and extended by Čiurlionytė (1938), Četkauskaitė (1998; the idea was supplemented with the concept of anchor tones as frameworks of tonal structures), and still is popular today. Earlier a great impact on setting this notion was made by Ukrainian researcher Sokalski (1888) which attempted to wedge all folk music into the firm frame of Greek tetrachords.

It might be suspected that the notion of Ancient Greek modes rests on the diatonic thinking of the researchers, who have no doubt that the original scales in folk music are diatonic. While certain examples of the original scales may really resemble the theoretical diatonic scales, the others can be essentially different. Then the collision of two emic scale systems – that of the cultural insider's (performer's) and outsider's (ethnomusicologist's) – results in the "aural ghosts", i.e., in the misperception of the peculiar original scales as Ancient Greek modes.

1.2 Issue of Equitonics

Recent acoustical measurements of musical scales in Lithuanian traditional singing have shown noticeable systematic deviations from 12TET. Ambrazevičius (2011) analyzed four sample repertoires representing quite a wide spectrum of Lithuanian vocal traditions. He concluded that, in general, none of the samples were consistent with 12TET (or other theoretical diatonic scales, e.g., Pythagorean or Just intonation) and such inconsistencies could not be explained by tolerable categorical zones of intonation, possible mistakes and imperfections, etc. However, these samples showed certain commonalities with the scales anchored on a framework of a fourth or fifth and filled in with "loosely-knit" (Grainger) intermediate tones (examples of such scales can be found in European folk music; see Grainger, 1908–1909; Sevåg, 1974; etc). Ambrazevičius suggested the model of roughly equidistant scale (equitonics) which, as purely theoretical frame, could explain the features or tendencies of the scales in traditional singing.

1.3 Aim of the Study

The aim of this paper is to identify some possible groups of different musical scales (featuring traces of the equidistant scale) and to verify the correspondence of those groups to the ethnomusical dialects (regions) or styles of Lithuania.

2. PROCEDURE

2.1 Samples

Three samples of songs recorded in the 1930s were selected. The samples represent different regions of Lithuania, respectively, Suvalkija (S), Žemaitija (Samogitia; Z), and Aukštaitija (A). The first two samples consist of monophonic songs (23 and 10 songs) while the third sample consists of polyphonic songs *Sutartinės* (21 song).

Sutartinės are Lithuanian polyphonic songs, typical to Aukštaitija region. Unfortunately, they vanished as an unbroken tradition in rural areas in the middle of the 20th century. Significant part of the *Sutartinės* can be considered as examples of *Schwebungsdiaphonie* ("beat diaphony") as majority of their intervals formed by the voices are seconds. Many of the *Sutartinės* are characteristic of quite pronounced modal structures centering on the nuclei and dissipating towards marginal pitches. The nuclei can be considered mostly as "double tonal centers" comprising two central pitches, with an interval of second in between, and belonging to two intertwining voices (parts).

In this paper, the upper pitch in the "tonal center" is denoted as the upper case letter I and the lower pitch is denoted as the lower case letter i. The scale degrees above I are denoted as II, III, IV, etc., and the degrees below i are denoted as ii, iii, iv, etc. In the case of monophonic songs, the conventional notation for scale degrees is used (i.e., 1st, 2nd, 3rd, etc.).

2.2 Acoustical Measurements

To evaluate the scales of songs, acoustical measurements of their recordings were made using PRAAT software. In the case of solo performances, pitches of all structurally important sounds of the songs were measured; grace notes were not considered because of the crude uncertainty of pitch. Perceived (integral) pitches of the tones were estimated from continuous tracks of "objective pitch" (log frequency) automatically transcribed by the software.

In the case of *Sutartinės*, spectra of the vocal dyads were considered. Two outstanding partials (belonging to different voices) in the spectrum of each dyad were selected. Most frequently they were the second harmonics. Then the fundamentals, corresponding pitches and intervals between the voices of the dyads were calculated.

The occurrences of scale degrees were averaged across every performance to obtain the averaged musical scale.

The recordings of solo performances were relatively long; most of the songs contained from 10 to 20 melostrophes. Only one (the first or the second) melostrophe of every song was chosen for the investigation. The recordings of the *Sutartinės* were not that long, thus almost all vocal dyads were considered. An exception was made for some dyads, e.g., because of their extremely poor recording quality which did not allow accurate measurement.

2.3 Diatonic Contrast

The index of diatonic contrast (DC) has been earlier introduced as the method of evaluation as to whether the scale is "more diatonic" or "more equitonic". More exactly, it defines "how much the scale is diatonic". The succeeding constituent intervals (i.e., the intervals between the adjacent scale degrees) are pooled into two groups of "narrow" (d_n) and "wide" (d_w) intervals. The following expression for DC is applied:

$$DC = \frac{\bar{d}_w}{\bar{d}_n} - 1 - \frac{2\sum_i \left| d_i - \bar{d}_{(n/w)} \right|}{N\bar{d}_n} \qquad (1);$$

where $N = N_n + N_w$ is the total number of intervals. The $\bar{d}_{(n/w)}$ means either $\bar{d}_n$ or $\bar{d}_w$, depending on the attribution of d_i.

The formula gives different values of DC depending on the certain grouping. The largest possible value is defined as the actual DC. The DC given by the formula above is normalized: if the value of the diatonic contrast equals 1, it means that the corresponding set consists of scale degrees separated by tempered whole tones and semitones. Zero for DC means ideal equitonics (equal intervals between the degrees; see Ambrazevičius, 2006).

The method of DC is intended to evaluate the overall asymmetry of an intervallic structure. It does not detect the individual differences between the scales, e.g., between the minor and major modes.

2.4 Cluster Analysis

Hierarchical cluster analysis can be applied for identification of groups of musical scales. Cluster analysis identifies "groups of individuals or objects that are similar to each other but different from individuals in other groups" (Norušis, 2011: 375). Agglomerative hierarchical clustering with squared Euclidian distance and averaged linkage is applied in our research.[1]

Intervallic structures of scales are considered as data for the evaluation of similarity. The intervallic structure can be represented in several ways. The basic representation is a set of relative pitches with regard to the first scale degree (which is thus normalized to 0; see Figure 1). This representation contains all intervallic information. Yet it has a certain disadvantage: similar scales represented in this way can be identified as different. For instance, if only pitch of the first degree is shifted, all other scale degrees become "sharp" or "flat". Other two representations are as follows: a set of intervals between adjacent scale degrees (Figure 2); a set of intervals of all possible pairs of scale degrees (Figure 3). Some tests on various scales showed that, in most cases, the third representation gives the results which approximate the subjective visual evaluations of similarity the best. Therefore cluster analysis was performed only with this kind of scale representation.

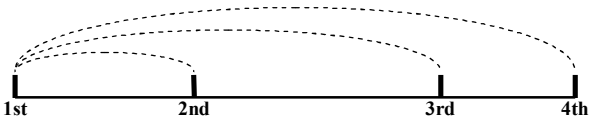


Figure 1. Scale representation, relative pitches with regard to the first scale degree.

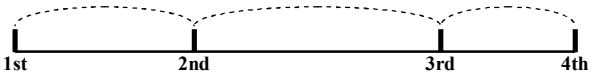


Figure 2. Scale representation, intervals between adjacent scale degrees.

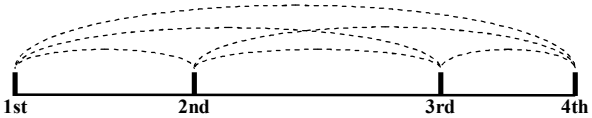


Figure 3. Scale representation, intervals of all possible pairs of scale degrees.

It should be also taken into account that scales with similar relative intervallic structures (and with only reasonable differences in absolute numbers) tend to be con-

[1] For details on hierarchical clustering, see, e.g., Tan, Steinbach, & Kumar, 2006: 515–526, Norušis, 2011: 377–388.

sidered as similar.[2] For instance, two versions of equidistant scales with different constituent intervals, nevertheless, are manifestations of the same principle of equitonics (DC for both scales equals 0). The same should be stated about two versions of diatonic scale (e.g., major): based on 12TET and slightly modified (evenly compressed or stretched). DC for both scales equals 1.

Therefore before the cluster analysis, the ranges of scales were normalized, i.e., the intervals between the lowest and the highest scale degrees were set to 1. This allowed the effect of "compressing" or "stretching" of otherwise similar scales from cluster analysis to be eliminated.

The disadvantage of cluster analysis is the requirement to classify objects with an equal amount of features. Musical scales vary in the number of scale degrees; therefore they have to be modified to meet the requirement. For instance, one should select only the scales comprising the same scale degrees; if some examples possess additional degrees above or below, the degrees should be excluded from the analysis. This means that the certain information is lost.[3]

Thus, for sample S, a set of scales truncated to 6 degrees (from 1st to 6th; 11 examples) was composed. All scales in sample Z had no less than 5 degrees (from 1st to 5th); therefore they all were included into cluster analysis (10 examples). For sample A, a set of scales truncated to 4 degrees (from ii to II; 14 examples) was composed. Also, a composite set from all samples containing 29 scales truncated to 5 degrees was prepared.

3. RESULTS

3.1 Diatonic Contrast

Evaluation of diatonic contrast showed that the scales of songs in all three samples are statistically more similar to the equidistant scale, with a special emphasis on Aukštaitija region (*Sutartinės* style; see Figure 4). The quartiles of DC (successively, Q1, Me, and Q3) are the following: (S) .32, .46, .56, (Z) .30, .45, .53, and (A) .16, .24, .34.

3.2 Cluster Analysis

Results of hierarchical cluster analysis are often presented in the form of a dendrogram (e.g., Tan, Steinbach, & Kumar, 2006: 515–516). There is no strict criterion for finding the best cluster solution and the optimal number of clusters depends on the goal of the investigation (e.g., Norušis, 2011: 377). In Figure 5, the 4 cluster-solution seems to be quite reasonable as the number of groups of scales is neither too large nor too small and the differences among clusters are sufficient. Figure 6 shows 6 degree-scales from sample S (represented as relative pitches in regard to the 1st scale degree) grouped according to the results of cluster analysis. Note that some groups consist of scales which seem to be quite different from each other. Actually some of them become similar if appropriate "stretching" is applied. Other relatively dissimilar scales fall into homogeneous groups due to the particular cluster solution (e.g., the scales of songs 4a and 12; compare Figures 5 and 6).

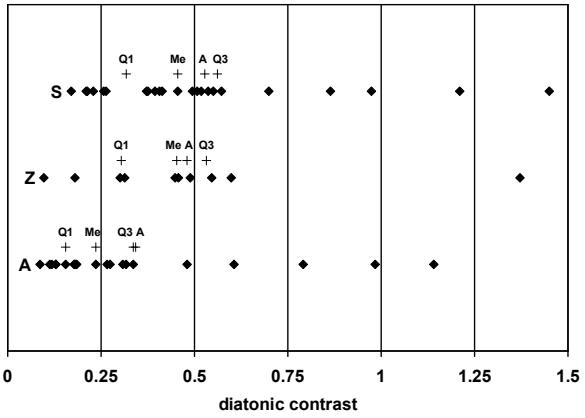


Figure 4. Diatonic contrast of scales in each sample (S, Z and A). A, Me, Q1, and Q3 denote, respectively, average, median, first and third quartiles of DC.

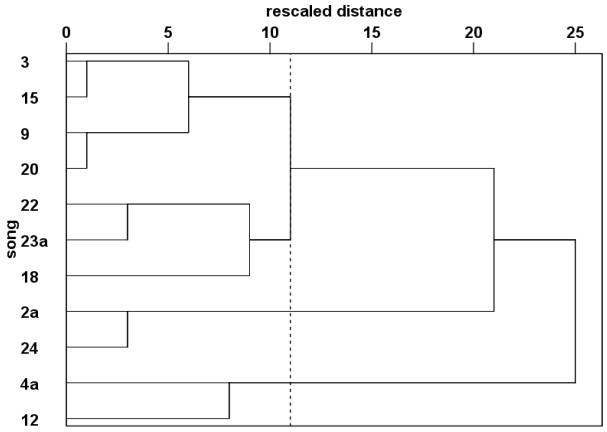


Figure 5. The dendrogram for 6 degree-scales from sample S. The Dotted line denotes a cutting point for 4 cluster-solution.

In sample S (Figure 6), the first and the largest group (4 examples) contains versions of roughly equidistant scales.[4] The second group (3 examples) includes the scales with intervals falling into two categories; approximately whole tones and semitones. These scales resemble the natural or harmonic major scales. If considering the excluded degrees, the scale of song 23a resembles the Mixolydian scale. The third group of scales (2 examples) is characteristic of the framework of a perfect fourth with equidistantly distributed intermediate degrees and additional fifth, and "lowered" sixth degrees (the mixture of equitonics and diatonics). The fourth group contains two "unclassifiable" scales with very unusual intervallic structures.

[2] Incidentally, even evolving scales, for instance, stretching out in the course of performance, are perceived by the performers as certain stable entities (Alexeyev, 1976: 49, etc.).

[3] This procedure of truncation is actually not as big a fault as it would seem at first glance. The marginal scale degrees are statistically quite rare (their occurrences are usually less frequent or even sporadic), thus their pitch averages are the least reliable.

[4] The grey points indicate the scale degrees excluded from the analysis; they are therefore not considered in this section.

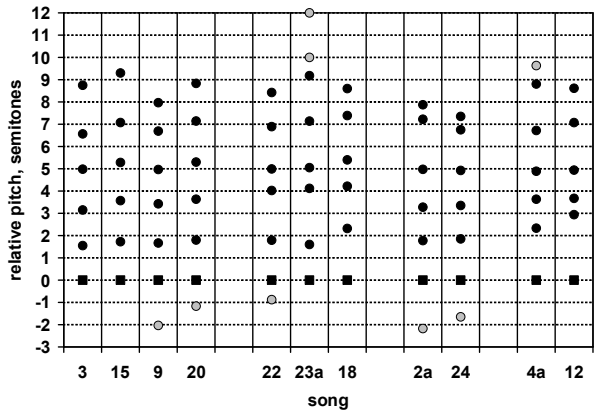


Figure 6. Groups of 6 degree-scales; sample S.

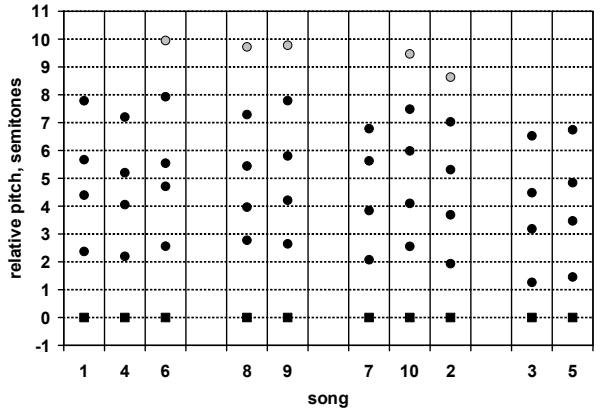


Figure 7. Groups of 5 degree-scales; sample Z.

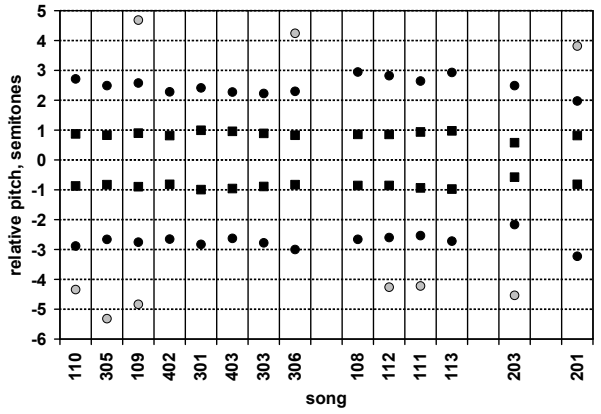


Figure 8. Groups of 4 degree-scales; sample A (*Sutartinės*). The scales are normalized so that the averages of the scale nuclei (degrees i and I) are set to zero.

Sample Z (Figure 7) is distinguished by scales with unusually wide intervals. Quite many of them are wider than 12TET whole tone; e.g., see the scales of songs 1, 4, 8, and 9. The scales in the first group (3 examples) have a narrow interval between the 3rd and 4th degrees and thus resemble stretched versions of major scale. It is somewhat difficult to describe and classify the other scales. At first glance, they seem quite different from each other inside the groups. Nevertheless, it can be concluded that the scales of songs 2 and 5 are two versions of rough equitonics, and the scale of song 7 resembles Lydian scale characterized by the narrow interval between the 4th and 5th degrees.

In the set of polyphonic songs *Sutartinės*, different scale degrees belong to only one of two voices and the aggregate scale is a result of two complementary subscales. Most of the *Sutartinės* scales are quite symmetric and equidistant (see Figure 8, the first and the second group of scales). The first group (8 examples) consists of scales with slightly narrower upper intervals (between the I and II degrees), while the second group (4 examples) has slightly narrower lower intervals (between the ii and i degrees). Single scales in the third and fourth groups have asymmetric diatonic structure, with quasi whole tones and semitones.

In general, the results show, that each sample splits into homogeneous groups and most of the groups have their unique intervallic structure.

The cluster analysis was also performed on the composite sample of the three regions to ascertain if there is any relation between the scale structure and particular region or style (monophonic vs. polyphonic). If the scales of particular region/style roughly fell into separate clusters, the relation would be confirmed. 4 cluster-solution was applied, as cluster 4 contains a single and very dissimilar scale (i.e. outlier) and other three clusters were expected to represent three different regions/styles. The results are shown in Table 1. At first glance, they do not provide solid evidence on the scales specific to particular regions/styles. However, certain shortcomings can be envisaged in such application of the method. For instance, the marginal scale degrees in the sample A used in the case of composite sample are mostly sporadic, thus the results are insufficiently reliable. In addition, the sample is too small to make statistically significant conclusions of the kind discussed. Thus probably, enlargement of the samples and revision of their combinations would help to reveal the factor of dialect (region/style) in the composite sample.

	Region		
Cluster	**A**	**S**	**Z**
1	4	8	3
2	1	3	5
3	1	1	2
4	0	1	0
Total	6	13	10

Table 1. A composite sample of 5 degree-scales formed of three region/style samples. Cross tabulation between the results of cluster analysis and the regions (cluster 4 contains single example-outlier).

4. DISCUSSION

Based on the results of evaluation of diatonic contrast, it can be stated that 65 percent of the analyzed song examples are more similar to the equidistant scale (DC < .5) whereas 35 percent resemble more diatonics than equitonics (DC > .5). The traces of the equidistant scale are not equally strong for the different dialects under investigation; the polyphonic *Sutartinės* from Aukštaitija show the strongest impact while the element of diatonics is relatively more pronounced in the monophonic songs from Suvalkija.

In general, the cluster analysis reveals the same regularities. Additionally, it can be stated that the examples from Žemaitija present, first, the most "chaotic" or diverse scales, and second, their intervals are significantly stretched compared to the conventional cases. On the contrary, the scales of many *Sutartinės* are even strikingly similar. Most probably, it results from the fact that the *Sutartinės* are polyphonic songs and precision of the intervals between simultaneous voices should be severely observed to reach the maximum "clash" (or roughness, in psychoacoustic terms). Perhaps, this attention to the relative precision of intonation explains the approximate equability of the interval sizes as well.

The different distinctness of the traces of equitonics correlates well with the historical position of the dialects: *Sutartinės* are considered as belonging to the most archaic musical relics and the songs in Suvalkija present relatively modern musical thinking.

Acknowledgments

We wish to thank the European Social Fund for their support of the study (Global Grant).

5. REFERENCES

Alexeyev, E. (1976). *Problemy formirovanija lada.* Moscow: Muzyka.

Ambrazevičius, R. (2006). Pseudo-Greek Modes in Traditional Music as Result of Misperception. In *ICMPC9. Proceedings of the 9th International Conference on Music Perception and Cognition. 6th Triennial Conference of the European Society for the Cognitive Sciences of Music. Alma Mater Studiorum University of Bologna, Italy, August 22–26* (pp. 1817–1822). Bologna: Bononia University Press.

Ambrazevičius, R. (2011). Concerning the Question of "Looseley-Knit" Roughly Equidistant Scale in Traditional Music. In A. Schneider and A. von Rushckowski, eds. *Systematic Musicology: Empirical and Theoretical Studies (Hamburger Jahrbuch für Musikwissenschaft, vol. 28)* (pp. 307–316). Frankfurt am Main: Peter Lang.

Brazys, T. (1920). *Apie tautines lietuvių dainų gaidas (melodijas)*. Kaunas: Švietimo ministerijos komisioneris „Švyturio" B-vė.

Četkauskaitė, G. (1998). *Lietuvių liaudies dainų melodijų tipologija.* Vilnius: Lietuvos rašytojų sąjungos leidykla.

Čiurlionytė, J. (1938). *Tautosakos darbai: Vol. 5. Lietuvių liaudies melodijos.* Kaunas: Lietuviu tautosakos archyvas.

Grainger, P. (1908–1909). Collecting with the Phonograph. *Journal of the Folk Song Society, 3*, 147-242.

Norušis, M. J. (2011). *IBM SPSS Statistics 19 Statistical Procedures Companion.* Upper Saddle River, NJ: Prentice Hall.

Sevåg, R. (1974). Neutral Tones and the Problem of Mode in Norwegian Folk Music. In G. Hilleström, ed. *Studia Instrumentorum Musicae Popularis III* (pp. 207–213). Stockholm: Musikhistoriska Museet.

Sokalski, P. (1888). *Russkaja narodnaja muzyka* <…>. Kharkov.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining.* Boston, MA: Addison-Wesley.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

WAVELET-FILTERING OF SYMBOLIC MUSIC REPRESENTATIONS FOR FOLK TUNE SEGMENTATION AND CLASSIFICATION

Gissel Velarde
Aalborg University
gv@create.aau.dk

Tillman Weyde
City University London
T.E.Weyde@city.ac.uk

David Meredith
Aalborg University
dave@create.aau.dk

ABSTRACT

The aim of this study is to evaluate a machine-learning method in which symbolic representations of folk songs are segmented and classified into tune families with Haar-wavelet filtering. The method is compared with previously proposed Gestalt-based method. Melodies are represented as discrete symbolic pitch-time signals. We apply the continuous wavelet transform (CWT) with the Haar wavelet at specific scales, obtaining filtered versions of melodies emphasizing their information at particular time-scales. We use the filtered signal for representation and segmentation, using the wavelet coefficients' local maxima to indicate local boundaries and classify segments by means of k-nearest neighbours based on standard vector-metrics (Euclidean, cityblock), and compare the results to a Gestalt-based segmentation method and metrics applied directly to the pitch signal. We found that the wavelet based segmentation and wavelet-filtering of the pitch signal lead to better classification accuracy in cross-validated evaluation when the time-scale and other parameters are optimized.

1. INTRODUCTION

One of the aims of folk song research is the study of melodic variations caused by the process of oral transmission between generations (van Kranenburg et al., 2009). Wiering et al. (2009) propose an interdisciplinary and ongoing process between human expertise, methods and models to understand melodic variation and its mechanisms. Classification models and methods dealing with such challenges define their representation and processing to be evaluated based on some ground truth. In this paper, we present our method based on wavelet-filtering and evaluate it on a collection of Dutch folk songs ("Onder de groene linde", Grijp, 2008), in which songs were classified into tune families according to expert similarity assessments, mainly based on rhythm, contour and motifs (Wiering et al., 2009; Volk & van Kranenburg, 2012).

The collection of folks songs that we study in this paper, is a monophonic collection of Dutch folk melodies encoded in MIDI files, so that we have pitches encoded as integer numbers, ranging from 0 to 127, and onsets and durations in quarter notes and subdivisions. In order to analyse these files via wavelets, we sample each melody as a one dimensional (1D) signal. Graphically, the melodic contour of 1D pitch signal can be drawn in a pitch over time plot, with the horizontal axis representing time in quarter notes, and the vertical axis representing pitch numbers. This contour representation of melodies has been linked to human melodic processing, using contour classes (Huron, 1996), interpolation lines (Steinbeck, 1982) and polynomial functions (Müllensiefen & Wiggins, 2011; Müllensiefen, Bonometti, Stewart & Wiggins, 2009). However, the contour representation does not give direct access to some aspects that are important for music similarity. Large-scale changes, like transposition of a melody lead to a completely different set of values although the melody is not substantially different. Similarly, small-scale changes like ornaments can lead to different pitch values even if the main essential shape of the melody is preserved.

Wavelet coefficients are obtained as the inner product of a 1D signal and a wavelet (i.e., a short signal with zero average and defined energy). The wavelet is shifted along the time axis and for each time position a coefficient is calculated. This is equivalent to a convolution with the wavelet flipped along the time axis, and thus to a finite impulse response filtering of the signal. The wavelet can be stretched on the time axis, leading to coefficients at different time-scales, corresponding to different filters. This process can also be understood as comparing the melodic shape with the wavelet shape, so that the coefficients represent similarity values at different time-positions and time-scales. The process of producing a full set of wavelet coefficients for a signal is known as the wavelet transform (WT), of which there are different variants. The transformed signal is represented as a set of coefficient signals at different scales. We use the Haar wavelet, which is a function of time t that takes values of 1 if $0 \leq t < 0.5$, or $0.5 \leq t < 1$, and 0 otherwise.

We use the information of the wavelet coefficients to define and compare melodic segments. Local maxima of the wavelet coefficients occur when the inner product of the melody and the wavelet is maximal in that position. In the case of the Haar wavelet this occurs when there is a locally maximal change of pitch - averaged over half the length of the wavelet - in the melody. Therefore, we use the local maxima of wavelet coefficients to indicate segmentation points. If the found segments correlate with human structural perception and music theory, we assume that they can be used to classify melodies containing similar segments. A melodic fragment and its transposed version will be represented by the same wavelet coefficients (except for very beginning of the melody).

Musical similarity in folk music is a hard problem to define (Wiering et al., 2009). We can understand it as a partial identity, where entities share some properties that can be measured (Cambouropoulos, 2009). With wavelet-filtering we apply a process that selectively focuses on a specific time-scale. It is a preprocessing step before determining segment similarities, which we calculate based on distance metrics. In the following section we will discuss some computational models and methods that have been used to model melodic similarity in symbolic music representation and have been applied to classify folk melodies.

2. RELATED WORK

2.1 Modelling melodic variations

Computational models applied to modelling melodic variations in symbolic music representations of folk songs include string matching methods and multidimensional feature vectors to represent global properties of melodies (Hillewaere, Manderick & Conklin, 2009; Hillewaere, Manderick & Conklin, 2012; van Kranenburg, 2010). In origin and genre classification, global representations perform only slightly worse than string-based methods (Hillewaere et al., 2009 and 2012). However, methods based on global representation depend heavily on the choice of features, which can lead to reduce generalizability.

Van Kranenburg, Volk & Wiering (2013) showed that sequence alignment algorithms using local features prove successful in classifying folk song melodies to tune families defined by experts. Sequence alignment algorithms are used to quantify similarity of sequences by computing the operations needed to transform one sequence into another, by means of substitutions, insertions and deletions (Manderick & Conklin, 2012; van Kranenburg, 2010). Although van Kranenburg's (2010) method was very successful when used to classify melodies from the Dutch folk-song corpus into tune families, its representation requires 14 attributes for each note in a melodic sequence (see van Kranenburg, 2010, pp. 94-95), apart from the standard information that is encoded in MIDI format (pitch number, onset and duration), meaning that this approach might not be applicable for classification using MIDI files only. In the following section we present our method, which can be applied to any data set encoded in MIDI format, or any other format containing pitch, onset and duration information for each note in a melody.

2.2 Gestalt-based segmentation

Segmentation is a core activity for musical processing and cognition (Lerdahl & Jackendoff, 1983). In order to study this mechanism, some authors adapt concepts of visual processing to study musical processing. Cambouropoulos (1997, 2001) presents a segmentation model based on Gestalt principles of similarity and proximity, known as the local boundary detection model (LBDM). The LBDM computes a profile of segmentation strength in the range [0, 1], based pitch intervals, inter-onset-intervals and rests. When the strength exceeds a threshold, a segmentation point is introduced. (Cambouropoulos, 2001). We use the LBDM here as a baseline for our model.

2.3 The use of wavelets in the symbolic domain

Wavelet analysis has been applied to diverse time series datasets. A time series is a set of observations recorded at a specified time (Brockwell & Davis, 2009). The use of wavelets for time series processing and analysis can be found in different areas, i.e. meteorological (Torrence & Compo, 1998), political (Aguiar-Conraria, Magalhaes, Soares, 2012), medical (Hsu, 2010), financial (Hsieh, Hsiao, & Yeh, 2011). Wavelets are also well known in audio music information retrieval (Andén & Mallat, 2011; Jeon & Ma, 2011; Smith & Honing, 2008; Tzanetakis, Essl, & Cook, 2001), but they have been scarcely applied on symbolic music representations. The only example of wavelets applied to symbolic music representation, apart from our previous study (Velarde & Weyde, 2012), is presented by Pinto (2009), demonstrating that it is possible to index melodic sequences with few wavelet coefficients, obtaining improved retrieval results compared to the direct use of melodic sequences. The method used by Pinto can be exploited for compression purposes, whereas our method is used for structural analysis and classification.

3. THE METHOD

We extend the method introduced in Velarde and Weyde (2012) by exploring segmentation based on the information of the wavelet coefficients' local maxima, and evaluate it on the classification of folk tunes into tune families. Our previous study (Velarde & Weyde, 2012) showed good results in a different classification task using the 15 Two-Part Inventions by J. S. Bach.

3.1 Representation

We represent melodies as normalized pitch signals or by the wavelet coefficients of the pitch signals. Discrete pitch signals $v[l]$ with length L are sampled from MIDI files at a rate r (given in number of samples per quarter note), so that we have a pitch value for every time point, expressed as $v[t]$. Rests are replaced by the following procedure: if a rest occurs at the beginning of a sequence, it is replaced by the first pitch number that appears in the sequence, otherwise it is replaced by the pitch number of the last note that precedes it.

Normalized pitch signal representation (vr). We normalize pitch signals segments, by subtracting the average pitch in order to make the representation transposition-invariant. The normalization is applied after the segmentation.

Wavelet representation (wr). We apply the continuous wavelet transform (CWT) (Mallat, 2009), expressed in a discretized version as the inner product of the pitch signal $v[l]$ and the Haar wavelet $\psi_{s,u}[l]$, at position u and scale s:

$$w_s[u] = \sum_{l=1}^{L} \psi_{s,u}[l] v[l] \quad (1)$$

To avoid edge effects due to finite-length sequences (Torrence & Compo, 1998), we pad on both ends with a mirror image of the pitch signal (Woody & Brown, 2007). Once the coefficients are obtained, the segment that corresponds to the padding is removed, so that the signal maintains its original length.

3.2 Segmentation

Wavelet segmentation (ws). Local maxima of the wavelet coefficients occur when the inner product of the melody and the wavelet is maximal. This occurs with the Haar wavelet, when there is a locally maximal change of pitch (averaged over half the length of the wavelet) in the melody. We use local maxima of wavelet coefficients to determine local boundaries.

3.3 Classification

The melodic segments are used as the data points for classification. A melody is represented as a set of segments, and we use the *k*-Nearest-Neighbour (kNN) method for classification (Mitchell, 1997). We use two different distance measures: cityblock distance and Euclidean distance. We define the maximal length n of all segments to be compared and pad shorter segments as necessary with zeros at the end.

4. EXPERIMENT

In our experiment we address the question of how filtering the representation of melodic segments affects the folk tune family classification. We assumed that if segments represent meaningful melodic structures, they can be used to identify tunes belonging to a tune family and that some time-scales of the melodic contour might be more discriminative than others.

We ran the experiment[1] using the collection "Onder de groene linde" (Grijp, 2008). This collection is a high quality data set of 360 monophonic songs classified into 26 families according to field-experts' similarity assessments in terms of melodic, rhythmic and motivic content (Volk & van Kranenburg, 2012). The MIDI files of this collection are sampled into pitch signals with a sampling rate of 8 samples per quarter note (qn). We apply the CWT with the Haar wavelet using a dyadic set of 8 scales. Melodies are represented as normalized pitch signals (vr) or as the resulting wavelet coefficients (wr). Signals are segmented by the wavelet coefficients' local maxima (ws), or by the local boundary detection model (LBDM; Cambouropoulos, 1997, 2001) using thresholds from 0.1 to 0.8 in steps of 0.1. We explored the parameter space with a grid search testing all combinations of representations and segmentations: wavelet representation (wr), normalized pitch signal representation (vr), wavelet segmentation (ws), LBDM (LBDM) segmentation and 1 to 5 nearest neighbours. Segments are used to build classifiers from training sets and that are tested on unseen folk melodies. We evaluate the classification accuracy with cityblock and Euclidean distances in leave-one-out cross validation.

5. RESULTS

The results of the experiment can be seen in Figures 1 to 4. Alternatively, Tables 1 and 2 shows the best and worst classification values over all parameters for each combination of representation-segmentation, for each value of *k* in the kNN method, and for Euclidean and cityblock distance metrics. The results show that wavelet filtering of the melodies can improve classification performance compared to using the pitch signal directly. Independently of the segmentation method, wavelet representation proves to be more discriminative than pitch signals. For this corpus and experimental setup, we have used single time-scales and evaluated this melodic discrimination performance. The classification performance varies, obtaining best results at small scales and poor results at large scales, with exception of the largest scale which recovers its performance to some extent.

In terms of segmentation, it is possible to observe that shorter segments produce better results when used with wavelet representation. This is contrary to the results of the LBDM applied to pitch signals, where shorter segments produce worse results than larger ones. We observe an improvement towards threshold 0.4 and a gradual improvement towards the threshold of 0.8, which corresponds to larger segments, meaning that using the complete melodic sequences or a combination of complete melodies and melodic segments, can lead to better classification results when using pitch signals.

[1] The algorithms are implemented in MATLAB (The Mathworks, Inc) using the Wavelet Toolbox and the MIDI Toolbox for the implementation of the LBDM (Eerola & Toiviainen, 2004), and we use an update of Christine Smit's read_midi function (http://www.ee.columbia.edu/~csmit/matlab_midi.html, accessed 4 October 2012).

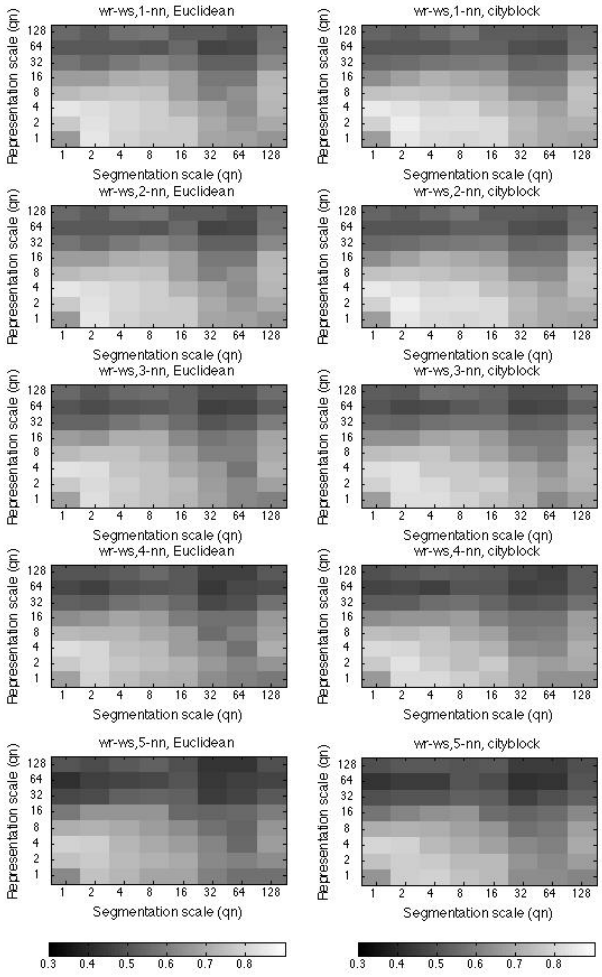


Figure 1. Accuracies for the combination of wavelet representation (wr) and wavelet segmentation (ws).

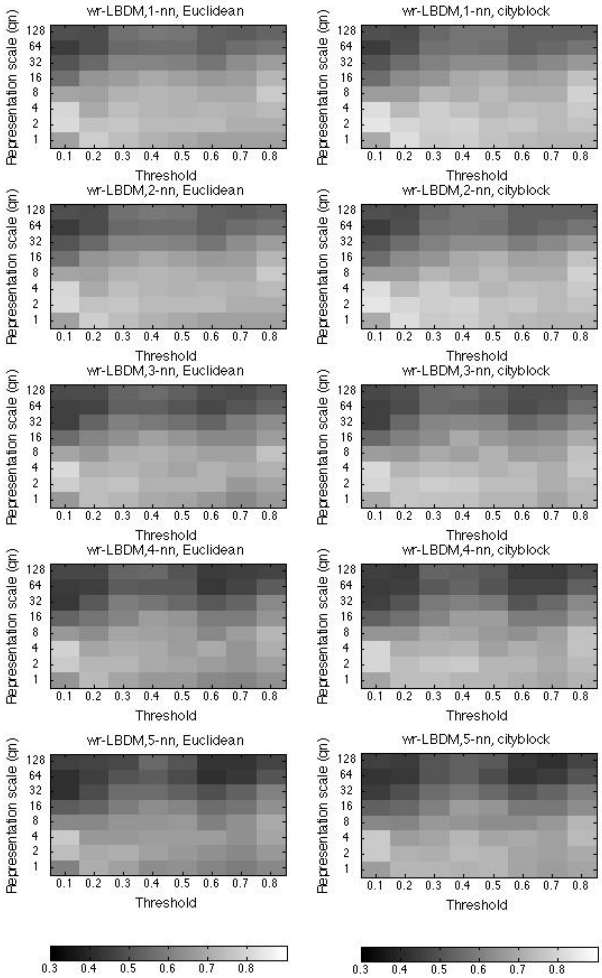


Figure 2. Accuracies for the combination of wavelet representation (wr) and local boundary detection model (LBDM).

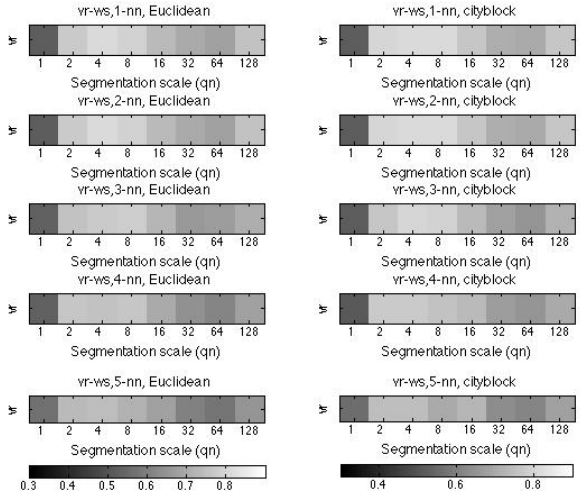


Figure 3. Accuracies for the combination of pitch signal representation (vr) and wavelet segmentation (ws).

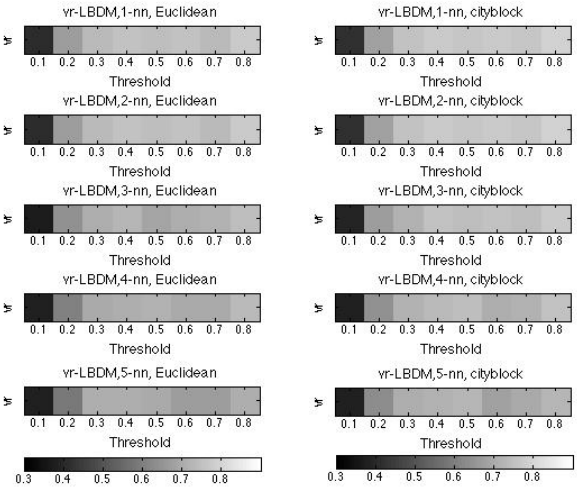


Figure 4. Accuracies for the combination of pitch signal representation (vr) and local boundary detection model (LBDM).

In general, similarity measured by cityblock distance proves more accurate than by Euclidean distance in pitch signals over time or wavelet representations, and the effect of using cityblock distance makes the difference between segmentation methods less important. The number of *k*-nearest neighbours shows that one or two neighbours produce the best results and when *k* increases further the accuracy decreases.

Euclidean distance						
represent.-segment.	Value	Nearest Neighbours				
		1	2	3	4	5
wr-ws	best	**0.8417**	**0.8417**	0.8306	0.8194	0.7917
	worst	0.4667	0.4667	0.4583	0.4333	0.4167
wr-LBDM	best	0.8111	0.8111	0.8083	0.7889	0.7694
	worst	0.4472	0.4472	0.4528	0.4333	0.4139
vr-ws	best	0.8083	0.8083	0.7806	0.7667	0.7444
	worst	0.5194	0.5194	0.5333	0.525	0.5639
vr-LBDM	best	0.7778	0.7778	0.7444	0.7333	0.7083
	worst	0.4111	0.4111	0.3722	0.3806	0.3806

Table 1. Classification accuracies best and worst values for each combinations using Euclidean distance.

Cityblock distance						
represent.-segment.	Value	Nearest Neighbours				
		1	2	3	4	5
wr-ws	best	**0.8556**	**0.8556**	0.8333	0.8306	0.7972
	worst	0.4833	0.4833	0.4639	0.45	0.4167
wr-LBDM	best	0.8417	0.8417	0.8083	0.8028	0.7778
	worst	0.4417	0.4417	0.4556	0.4417	0.4139
vr-ws	best	0.8139	0.8139	0.7972	0.7778	0.7472
	worst	0.5194	0.5194	0.5194	0.5139	0.5583
vr-LBDM	best	0.7889	0.7889	0.7778	0.75	0.725
	worst	0.4139	0.4139	0.3861	0.3778	0.3806

Table 2. Classification accuracies best and worst values for each combinations using cityblock distance.

6. DISCUSSION AND FUTURE DIRECTIONS

The best classification accuracies based on wavelet segmentation are only slightly better than the best accuracies obtained by the LBDM. The parameter exploration shows however, that wavelet segmentation performs better across different scales than the LBDM across different thresholds. Interestingly, these comparable methods meet the criteria of measuring local changes in melodic contour. While the LBDM measures the degree of change between successive values, the wavelet segmentation finds locally maximal falls of average pitch in melodies using different scales. The fact that small scales perform better than larger scales corroborates the findings of van Kranenburg et al. (2013) that local processing is most important in melodic similarity.

In terms of representation, wavelet-representation proves more discriminative than raw pitch signals. We assume that this is due to the transposition invariance of the wavelet representation and the emphasis on a specific time-scale.

Our best results are far less accurate than the results reported by van Kranenburg et al. (2013) using alignment methods on the same corpus. Our method uses only the information that is encoded in MIDI format (pitch number, onset and duration). It requires less encoded expert knowledge than the method used by van Kranenburg (2010), making it applicable to other corpuses of folk songs encoded in MIDI format or similar. In order to make a more reliable comparison, our method would need to include the expert based features used by van Kranenburg (2010). For instance, annotated phrase information seems to improve importantly the results obtained by sequence alignment algorithms. This information could be used to improve the scale selection. Also, our method uses only the information about contained segments, and not the order of the segments, leaving room for further work.

We used one default setup for the whole corpus, i.e. one best performing scale for all songs. In a future study, we are interested to address wavelet scale selection derived from individual songs' periodicities.

7. CONCLUSION

The main contribution of this research is the evaluation of wavelet-filtered signals for melodic segmentation and classification on a corpus of folk songs in MIDI format. Wavelet-filtering proves more discriminative than direct representation of pitch signals or pitch-time series. Segmentation by local maxima of wavelet coefficients performs slightly better than LBDM segmentation when processing at individual scales. Small scales perform better than large scales, indicating that local processing may be more relevant for melodic similarity in classification tasks.

The method presented here can be applied to other corpora and other symbolic formats that encode melodies. Possible ways to improve the classification performance of the method presented in this paper could be using alignment of wavelet representations of complete melodies, using selective combination of scales and exploring metrical information derived from songs' periodicities.

Acknowledgements

We thank Peter van Kranenburg (Meertens Institute, Amsterdam) for sharing the Dutch Tune Families data set. Gissel Velarde is supported by the Department of Architecture, Design and Media Technology at Aalborg University.

8. REFERENCES

Andén, J., & Mallat, S. (2011). Multiscale scattering for audio classification. In: *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), Utrecht, NL.: ISMIR*, pp. 657–662. Available online at http://ismir2011.ismir.net/papers/PS6-1.pdf

Aguiar-Conraria, L., Magalhães, P. C. and Soares, M. J. (2012), Cycles in politics: wavelet analysis of political time series. *American Journal of Political Science*, 56: 500–518. doi: 10.1111/j.1540-5907.2011.00566.x

Brockwell, P. & Davis, R. (2009). Time series: theory and methods. Springer series in statistics. Second edition.

Cambouropoulos, E. (1997). Musical rhythm: a formal model for determining local boundaries, accents and metre in a melodic surface. In: *M. Leman (Ed.), Music, Gestalt and Computing: Studies in Cognitive and Systematic Musicology*. Berlin: Springer, pp. 277-293.

Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its application in the study of expressive timing. In: *Proceedings of the International Computer Music Conference. San Francisco, CA: ICMA*, pp. 17-22.

Cambouropoulos, E. (2009). How similar is similar?. *Musicæ Scientiæ*. Discussion Forum 4B, pp. 7-24

Eerola, T. & Toiviainen, P. (2004). MIDI Toolbox: MATLAB Tools for Music Research. University of Jyväskylä. Available at http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/.

Grijp, L.P.. (2008). Introduction. In: L.P. Grijp & I. van Beersum (Eds.), Onder de groene linde. 163 verhalende liederen uit de mondelinge overlevering, opgenomen door Ate Doornbosch e.a./Under the green linden. 163 Dutch Ballads from the oral tradition recorded by Ate Doornbosch a.o. (Boek + 9 cd's + 1 dvd). Amsterdam/Hilversum : Meertens Instituut & Music and Words. pp. 18-27.

Hillewaere, R, Manderick, B., & Conklin, D. (2009). Global feature versus event models for folk song classification. In: *10th International Society for Music Information Retrieval Conference (ISMIR 2009), Kobe, Japan*. pp. 729-733. Available online at http://ismir2009.ismir.net/proceedings/OS9-1.pdf.

Hillewaere, R., Manderick, B. and Conklin, D. (2012) String methods for folk music classification. In: *13th International Society for Music Information Retrieval Conference (ISMIR 2012)*.

Huron, D. (1996) The melodic arch in western folksongs. *Computing in Musicology*, Vol. 10, pp. 3-23.

Hsieh, TJ. Hsiao, H.f., & Yeh, WC. (2011). Forecasting stock markets using wavelet transforms and recurrent neural networks: an integrated system based on artificial bee colony algorithm. *Applied soft computing*, Vol. 11 Issue 2. March 2011, pp. 2510–2525. Elsevier

Hsu, WY. (2010). EEG-based motor imagery classification using neuro-fuzzy prediction and wavelet fractal features. *J. Neurosci Methods*. 2010 Jun 15;189(2):295-302. doi: 10.1016/j.jneumeth.2010.03.030. Epub 2010 Apr 8.

Jeon, W., & Ma, C. (2011). Efficient search of music pitch contours using wavelet transforms and segmented dynamic time warping. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, pp. 2304–2307.

Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music.Cambridge, MA.: MIT Press.

Mallat, S. (2009). A wavelet tour of signal processing - The sparse way. Academic Press, Third Edition, 2009.

Mitchell, T. (1997). Machine Learning, (McGraw-Hill).

Müllensiefen, D., Bonometti, M., Stewart, L., and Wiggins, G. (2009). Testing Different Models of Melodic Contour. *7th Triennial Conference of the European Society of the Cognitive Sciences of Music (ESCOM 2009), University of Jyäskylä, Finalnd.*

Müllensiefen, D, and Wiggins, G. (2011). Polynomial functions as a representation of melodic phrase contour. In *A. Schneider & A. von Ruschkowski (Eds.), Systematic Musicology: Empirical and Theoretical Studies.* pp. 63-88. Frankfurt a.M.: Peter Lang.

Pinto, A. (2009). Indexing melodic sequences via wavelet transform. In: *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME'09*). pp. 882–885.

Smith, L.M., & Honing, H. (2008). Time-frequency representation of musical rhythm by continuous wavelets. *Journal of Mathematics and Music*, Vol. 2, No. 2, pp. 81–97.

Steinbeck, W. (1982). Struktur und Ähnlichkeit: Methoden automatisierter Melodieanalyse. Kassel: Bärenreiter.

Tzanetakis, G., Essl, G., & Cook, P. (2001). Audio analysis using the discrete wavelet transform. In: *Proc. WSES Int. Conf. Acoustics and Music: Theory and Applications (AMTA 2001), Skiathos, Greece*. Available online at http://webhome.cs.uvic.ca/~gtzan/work/pubs/amta01gtzan.pdf

van Kranenburg, P. (2010) A Computational Approach to Content-Based Retrieval of Folk Song Melodies. [S.l.] : [s.n.], 2010. Full text: http://depot.knaw.nl/8400

van Kranenburg, P., Garbers, J., Volk, A., Wiering, F. Grijp, L.P. and Veltkamp, R. (2009). Collaboration perspectives for folk Song research and music information retrieval: The indispensable role of computational musicology, *Journal of Interdisciplinary Music Studies*. (2009), doi: 10.4407/jims.2009.12.030

van Kranenburg, P., Volk, A., & Wiering, F. (2013): A Comparison between Global and Local Features for Computational Classification of Folk Song Melodies, *Journal of New Music Research*, 42:1, 1-18

Velarde, G., & Weyde, T. (2012). On symbolic music classification using wavelet transform. In: *International Conference on Applied and Theoretical Information Systems Research, Taipei, Taiwan*.

Volk, A. & van Kranenburg, P. (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, November 2012 vol. 16 no. 3, pp. 317-339.

Wiering, F., Veltkamp, R., Garbers, J., Volk, A. and van Kranenburg, P. (2009). Modelling Folksong Melodies. *Interdiciplinary Science Reviews*, Vol 34, No. 2-3, 154-171.

Woody, N. A. & Brown, S. D. (2007) Selecting wavelet transform scales for multivariate classification. *J. Chemometrics*, 21: 357–363. doi: 10.1002/cem.1060

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

A MORE INFORMATIVE SEGMENTATION MODEL, EMPIRICALLY COMPARED WITH STATE OF THE ART ON TRADITIONAL TURKISH MUSIC

Olivier Lartillot
Finnish Center of Excellence in Interdisciplinary Music Research
olartillot@gmail.com

Z. Funda Yazıcı
Istanbul Technical Univ., Dept. Music Theory and Musicology
fuyazici@gmail.com

Esra Mungan
Boğaziçi Univ., Istanbul Psychology Department
mungan@boun.edu.tr

ABSTRACT

We present a new model for segmenting melodies represented in symbolic domain based on local discontinuity. Based on a discussion of the limitations of two major models, Tenney and Polansky's model and LBDM, we propose to alleviate the general heuristics ruling boundary detection in order to allow a large set of relevant boundary candidates. We discuss also about the limitations of combining different musical dimensions (pitch, onset, rest) altogether. The new proposed model develops heuristics specific to each musical dimension, and can also predict the temporal location of the onset- and rest-based boundaries. The three segmentation models are compared to listeners' segmentation decisions collected through an empirical experiment. Our experimental data show a high degree of accordance in segmentation locations between musicians and non-musicians. We compared the responses of the participants with the predictions of our proposed model as well as with the LBDM and Tenney and Polanskys model. The results, in general, show that the proposed model offered the best congruence with listeners indications.

1. INTRODUCTION

Much research has been carried out on the computational modeling of music segmentation, aiming at predicting how the musical discourse can be decomposed into a succession of small parts. In this paper, we discuss two major models, Tenney and Polansky's model (Tenney & Polansky, 1980) and LBDM, (Cambouropoulos, 2006) showing some important limitations, and proposing ways to overcome them, leading to a complete and original segmentation model. The models implemented in this paper – the two previous models presented in section 2 as well as the new model introduced in the section 3 – have been integrated in *The MiningSuite* (Lartillot, 2011). The visualization of the results, as shown in Fig. 1, can also be obtained using this toolbox.

2. CURRENT SEGMENTATION MODELS AND THEIR LIMITATIONS

In this section, we introduce two famous segmentation models and demonstrate some of their main limitations.

2.1 Tenney and Polansky

2.1.1 Description

The model consists simply in segmenting at all local maxima in the series of successive intervals $I_1, I_2, \ldots, I_N$. This means that boundaries are assigned at each interval I_k that is bigger than *both* its immediately previous and next intervals, i.e. if

$$I_k > I_{k-1} \text{ and } I_k > I_{k+1} \quad (1)$$

The series of intervals can be defined in various ways:

- the series of pitch intervals $I_1^P, I_2^P, \ldots, I_N^P$
- the series of inter-onset intervals $I_1^{IO}, I_2^{IO} \ldots, I_N^{IO}$
- a series of intervals $I_1, I_2, \ldots, I_N$ where each interval is a weighted summation of the pitch interval and the inter-onset interval:

$$I_i = w^P.I_i^P + w^{IO}.I_i^{IO} \quad (2)$$

The segments defined by the boundaries are called *clangs*. In a successive step in the Tenney and Polansky hierarchically recursive model, each clang intervals are defined between clangs, so that the series of clangs form a new series of "intervals", that can also be segmented in the same way, leading to segments of second order (called *segments*), and so on.

2.1.2 Example

An example of analysis is given in figure 1. The clang boundaries given by Tenney and Polansky model are shown at the bottom of the piano-roll representation by the following conventions:

- Clang boundaries based on pitch intervals are represented with red circles.
- Clang boundaries based on inter-onset intervals are represented with blue stars.

The same example is shown in a score in figure 2. The segment boundaries given by Tenney and Polansky model are represented in the third line (TP) under each stave, using the following conventions:

- Segment boundaries based on pitch intervals are represented with red triangles.
- Segment boundaries based on inter-onset intervals are represented with blue triangles.

2.1.3 Limitations

The condition used for boundary detection, shown in formula 1, does not allow to detect boundary if the longer interval (which should have been considered as a boundary) is followed by another interval of same (or longer) duration.

Combining pitch and inter-onset intervals altogether, as formalized in formula 2 is an operation that might require musicological and cognitive validation.

The model does not indicate *where* exactly the boundary might be located. Surely enough, a boundary decision based on pitch interval cannot be made before hearing the second pitch defining that interval. But what about inter-onset interval? Can't we already detect that the duration of the current note perceived is sufficiently long to indicate a boundary, before hearing the actual start of the next note?

Besides, if the pitch and inter-onset interval information is mixed, as is done in formula 2, then the precise location of boundary becomes even more difficult to make.

Boundaries are expressed in a purely binary fashion: either there is or there is not a boundary at each given interval. But what about the relative strength of each boundary?

Finally the hierarchically recursive model, with the definition of intervals between clangs and the detection of local boundaries along that series of meta-intervals, is something that has not been seriously discussed and cognitively validated so far.

2.2 Local Boundary Detection Model

2.2.1 Description

The Local Boundary Detection Model assigns a score (a *strength*) to each successive interval, based on heuristics that are somewhat related to Tenney and Polansky's approach, replacing however the binary logical of Tenney and Polansky's model with a continuous measure.

Instead of simply comparing whether a given interval is longer than previous and next interval, the relative difference between successive intervals is computed, called *degree of change*:

$$DC_{k-1,k} = \frac{\text{abs}(I_k - I_{k-1})}{I_{k-1} + I_k} \tag{3}$$

$$DC_{k,k+1} = \frac{\text{abs}(I_{k+1} - I_k)}{I_k + I_{k+1}} \tag{4}$$

The strength assigned to each interval is equal to the sum of the degrees of changes with respect to the previous and next intervals, multiplied by the amplitude of the current interval:

$$S_k = (DC_{k-1,k} + DC_{k,k+1}) \times I_k \tag{5}$$

A strong interval would therefore correspond to an interval that is particularly large, and particularly larger than *both* its previous and next intervals.

The resulting strength curve shows clear peaks at the location of boundaries. Besides, the strength value at that peak gives an indication of the structural importance of the boundary.

The strength curve can be computed for pitch intervals, inter-onset intervals as well as rests. Rests are defined as the duration of the interval between the end of the previous note and the beginning of the next note. If there is no silence between the two notes, the rest value is set to 0.

Finally a combined strength curve can be computed through a weighted summation of the individual strength curves:

$$S_k = w^P.S_k^P + w^{IO}.S^{IO} + w^R.S^R \tag{6}$$

2.2.2 Example

An example of analysis is given in figure 1. The LBDM model is represented on the piano-roll representation by the following conventions:

- The boundaries based on pitch intervals are represented with red dashed vertical lines and are located at the location of the note ending the interval. In fact, since the LBDM does not give an explicit selection of boundaries but gives a score to each successive intervals, a boundaries is represented for all notes in the pianoroll (except the first one). The strength associated with each boundary is indicated by the darkness of the line. In this way, boundary of lowest strength, of no interest, are hardly visible in the figure.
- In the same way, the boundaries based on inter onset intervals are represented with blue dashed vertical lines and are also located at the location of the note ending the interval. The boundary strength is shown using the same convention as above.
- The boundaries based on rests are represented with green *horizontal* lines at the center of the piano roll, spanning the temporal extent of the given rest. The boundary strength is shown using the same convention as above.

A selection of boundaries output by the LBDM model on that same example is shown on the score in Figure 2. The segment boundaries are represented in the second line (LBDM) under each stave, using the following conventions:

- Boundaries based on pitch intervals are represented with red triangles.
- Boundaries based on inter-onset intervals are represented with blue triangles. Strong boundaries are shown in dark blue while weaker boundaries are shown in light blue.
- Boundaries based on rest are represented with green triangles. Strong boundaries are shown in dark green while weaker boundaries are shown in light green.

2.2.3 Limitations

Similar to the Tenney and Polansky's approach, boundaries are supposed to locate to place where the interval is bigger than the previous interval *and* the next interval as well. As

a consequence, boundary is not detected if the longer interval (which should have been considered as a boundary) is followed by another interval of same duration.

The LBDM outputs a strength curve indicating the boundary strength related to each successive interval, but does not explicitly give a decisive conclusion whether or not a given interval is a boundary or not. Decision could be based on a comparison of each strength to a given threshold, but this make the decision arbitrary and highly depend on the choice of that threshold.

In LBDM, as we also noted in Tenney and Polansky's approach, by combining altogether the different dimension (pitch, onset and rest), we cannot locate more precisely the actual location of the boundaries. The combination of strength related to distinct dimensions raises issues related to musicological and cognitive relevance. In particular, by combining those individual strength curve, we obtained a curve that can sometimes indicates local maxima that did not exist in the individual curves.

3. PROPOSED SEGMENTATION MODEL

We saw that both Tenney and Polansky's model and the LBDM presuppose that boundaries are due to intervals that are bigger than both their previous and next intervals. We showed the limitation of such heuristics. We propose to generalize the approach by alleviating the boundary condition: A condition for local boundary less constrained that the one defined in TP (formula 1) could detect whenever a new interval is simply longer than its previous one:

$$I_k > I_{k-1} \tag{7}$$

We saw also that the question of precise location of boundaries were not addressed in the previous models. In fact, the study of this question seems to be highly dependent on the considered musical dimension: the location of inter-onset boundaries are based on particular aspects that are highly different from those relating to pitch boundaries, and same for rest boundaries. This supports the idea that boundaries related to different musical dimensions have to be treated independently. It turns out the model we propose in this section has particular aspects related to each different musical dimension.

Finally, we would like to take the advantages of both previous approaches, while overcoming their respective drawbacks: we propose to exhaustively show all possible boundary points, but in the same time assign a score to each boundary.

3.1 Pitch interval model

3.1.1 The particular case of unison intervals

In previous segmentation models, the pitch representation consists in the series of intervals between successive notes. We propose instead to filter out unison intervals because of their particularity: Unisons form series of notes of same pitch that necessarily form a coherent segment, and any non-unison interval following such segment would necessarily imply a boundary. Such segmentation is not very informative and might obstruct more interesting structural information. For that reason, we consider series of successive notes of same pitch as one single meta-note for the pitch-interval analysis, so that pitch-interval between successive meta-note are taken into account instead of unisons.

3.1.2 Distance threshold

In the pitch domain, the previous heuristics defined in formula 7 would means a boundary is assigned when the current pitch interval is longer than the previous pitch interval:

$$I_k^P > I_{k-1}^P \tag{8}$$

This would however lead to a large set of boundaries, and intervals that are quite similar – in particular of just one semitone difference, such as between a minor and major second – do not seem to give interesting boundary candidates. We propose therefore to impose a minimal threshold in the increase of pitch interval that would define boundary to be selected. The condition is therefore rewritten as follows:

$$I_k^P - I_{k-1}^P \geq \delta \tag{9}$$

One typical value of δ that we think of interest – and that is used in the version of the model presented in this paper – is one whole tone.

3.1.3 Boundary strength and location

The strength of the pitch-based boundary is defined as the different of pitch interval amplitude:

$$S_k^P = I_k^P - I_{k-1}^P \tag{10}$$

The boundary can be simply located at the location T_k^{on} of the onset of the new pitch (i.e., the one at the end of the current interval under study), since the interval is recognized as soon as the pitch is perceived:

$$T_k^P = T_k^{on} \tag{11}$$

3.2 Onset expectation model

3.2.1 Simple model

A boundary is assigned whenever the current inter-onset interval is longer than the previous inter-onset interval:

$$I_k^{IO} > I_{k-1}^{IO} \tag{12}$$

In order to estimate the exact location of such boundary, we need to understand the underlying reason of the heuristic given in the previous formula. The previous inter-onset interval I_{k-1}^{IO} is the temporal interval between the onsets T_{k-2}^{on} and T_{k-1}^{on}. Why would in fact a smaller interval I_{k-1}^{IO} followed by a longer interval I_k^{IO} induce the perception of a boundary? We propose the idea that it might be related to the expectation of a regular succession of same duration, hence that the new interval I_k^{IO} would be equal to the previous interval I_{k-1}^{IO}. We predict therefore that a note would appear at time

$$T_k^{IO} = T_{k-1}^{on} + I_{k-1}^{IO} \tag{13}$$

If the new interval I_k^{IO} is longer than I_{k-1}^{IO}, it means that the actual onset T_k^{on} appears after the expected onset T_k^{IO}. For that reason, we propose to locate the boundary at that expected onset time.

The strength of the boundary is proportional to the duration of the previous interval I_{k-1}^{IO} and to the increase of duration between the previous and the new intervals:

$$S_k^{IO} = \log_2(I_{k-1}^{IO}) \times \log_2(I_k^{IO} - I_{k-1}^{IO}) \quad (14)$$

3.2.2 Multi-level model

All models considered so far only look at the relative difference between successive interval. What about longer scale structure?

If we suppose that the older intervals I_{k-2}^{IO}, I_{k-3}^{IO}, etc., are of same duration, then by definition there is no boundary between them, and the only boundary would appear at the new longer interval I_k^{IO}, which makes sense.

If on the contrary the older intervals were shorter, then they would be shorter than the currently previous I_{k-1}^{IO}, and so that there would have been already a boundary at location T_{k-1}^{IO} related to that increase of interval duration between I_{k-2}^{IO} and I_{k-1}^{IO}. [1]

If the older interval I_{k-2}^{IO} is instead longer than I_{k-1}^{IO}, two cases should be considered:

- If that older interval I_{k-2}^{IO} is also longer than the current interval I_k^{IO}, that simply means that the new boundary at T_k^{IO} closes a segment that was starting at onset T_{k-2}^{on}, i.e. the segment $[T_{k-2}^{on}, T_{k-1}^{on}]$ that has a granularity (the maximal distance between successive onsets within the segment) equal to I_{k-1}^{IO}.
- If that older interval I_{k-2}^{IO} is shorter than the current interval I_k^{IO}, this means that the new boundary given by the new interval I_k^{IO} closes not only that segment $[T_{k-2}^{on}, T_{k-1}^{on}]$ but also a larger segment $[(\ldots,)T_{k-3}^{on}, T_{k-2}^{on}, T_{k-1}^{on}]$ of larger granularity equal to I_{k-2}^{IO}.

This observation leads to an extension of the onset expectation model, where a given inter-onset interval I_k^{IO} can lead to several boundaries closing several segments that are imbricated one into another. In the above example, if the first boundary already defined by the simple model has location and strengths reexpressed as $T_{k,1}^{IO}$ and $S_{k,1}^{IO}$, then the new example just discussed lead to a new closing boundary of location:

$$T_{k,2}^{IO} = T_{k-1}^{on} + I_{k-2}^{IO} \quad (15)$$

and of strength:

$$S_{k,2}^{IO} = \log_2(I_{k-2}^{IO}) \times \log_2(I_k^{IO} - I_{k-2}^{IO}) \quad (16)$$

In other words, in the multi-level model, a given inter-onset interval can lead to a series of closing boundaries located at successive time $T_{k,i}^{IO}$ given by formula 15.

[1] We recall that this capability of applying successive "closing" boundary on successive intervals due to a progressive slowing down of duration is something that was not possible in the previous models based on the less constrained boundary condition given by formula 1, but that we made possible thanks to the new condition given by formula 7.

3.3 Rest model

In the rest domain, the general heuristics defined in formula 7 would means a boundary is assigned when the current rest is longer than the previous rest:

$$I_k^R > I_{k-1}^R \quad (17)$$

The strength of the rest-based boundary is defined as the different of rest amplitude:

$$S_k^R = I_k^R - I_{k-1}^R \quad (18)$$

The boundary can be located at the location where the rest reaches the duration of the previous rest:

$$T_k^R = T_{k-1}^{off} + I_{k-1}^R \quad (19)$$

3.4 Example

Figure 1 shows an example of analysis based on the proposed model. The model is represented on the piano-roll representation using the following conventions:

- The boundaries based on pitch intervals are represented with red diagonal lines that virtually cut the corresponding pitch interval. The strength associated with each boundary is indicated by the darkness of the line.
- The boundaries based on inter-onset intervals are represented with black vertical lines and are located at the temporal location predicted by the model. The boundary strength is shown using the same convention as above. The successive boundaries that are associated to a single inter-onset interval are linked together with a horizontal line at their top.
- The boundaries based on rests are represented with green *vertical* lines and are located at the temporal location predicted by the model. The boundary strength is shown using the same convention as above.

Figure 2 shows the same example of analysis represented on an actual score, on the first line ("New Model") below each stave, using the following conventions:

- Boundaries based on pitch intervals use red diagonal lines in the same way as in the piano roll representation explained above. Strong boundaries are shown with thick lines while weaker boundaries are shown with narrow lines.
- Boundaries based on inter-onset intervals are represented with blue triangles. Strong boundaries are shown in dark blue while weaker boundaries (of score between and) are shown in light blue.
- Boundaries based on rest are represented with green triangles. Strong boundaries are shown in dark green while weaker boundaries are shown in light green.

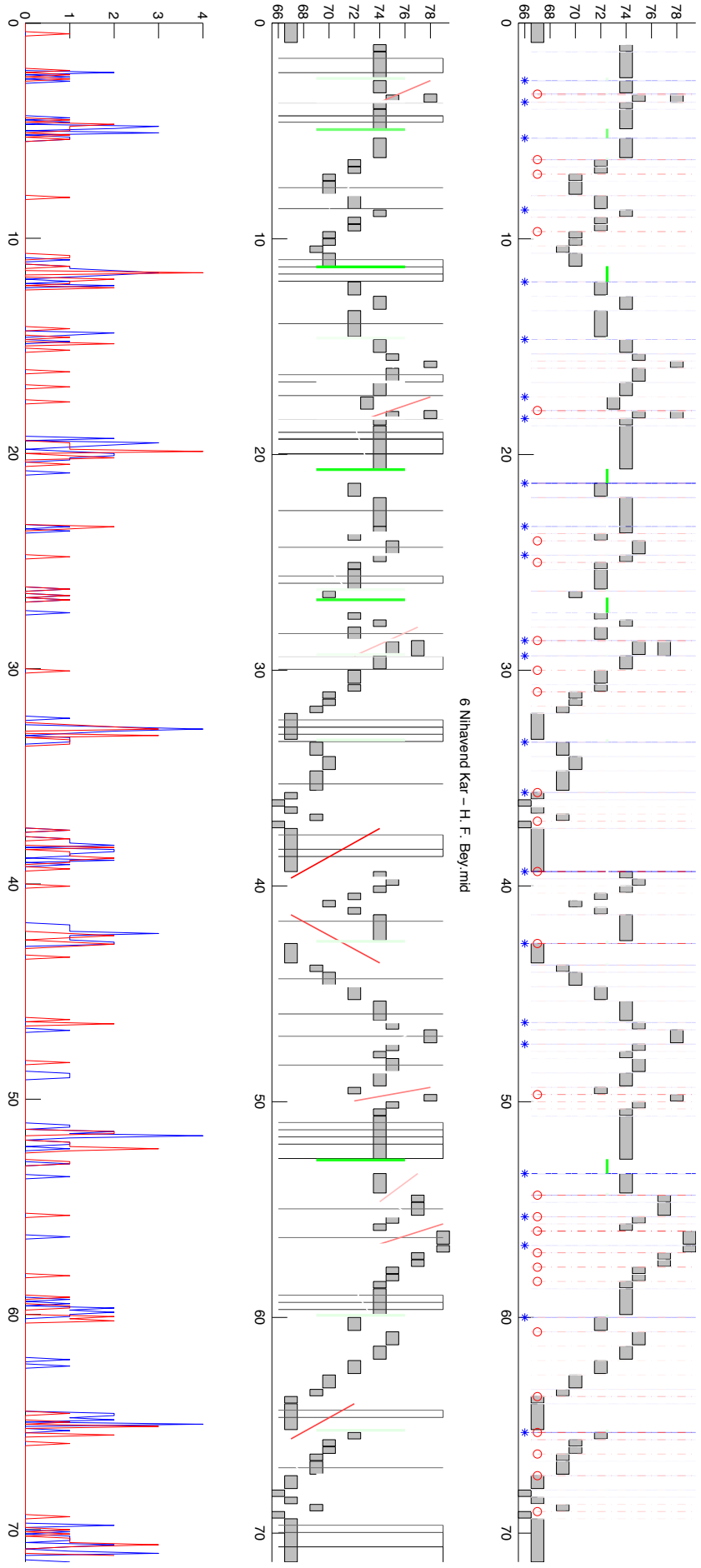


Figure 1: Analysis of Nihavend maqam in Kar form. Top: boundaries given by Tenney and Polansky's model and LBDM. Middle: boundaries given my our proposed model. Bottom: number of listeners' segmentation collected on successive 100-ms long period, with musicians in blue and non-musicians in red.

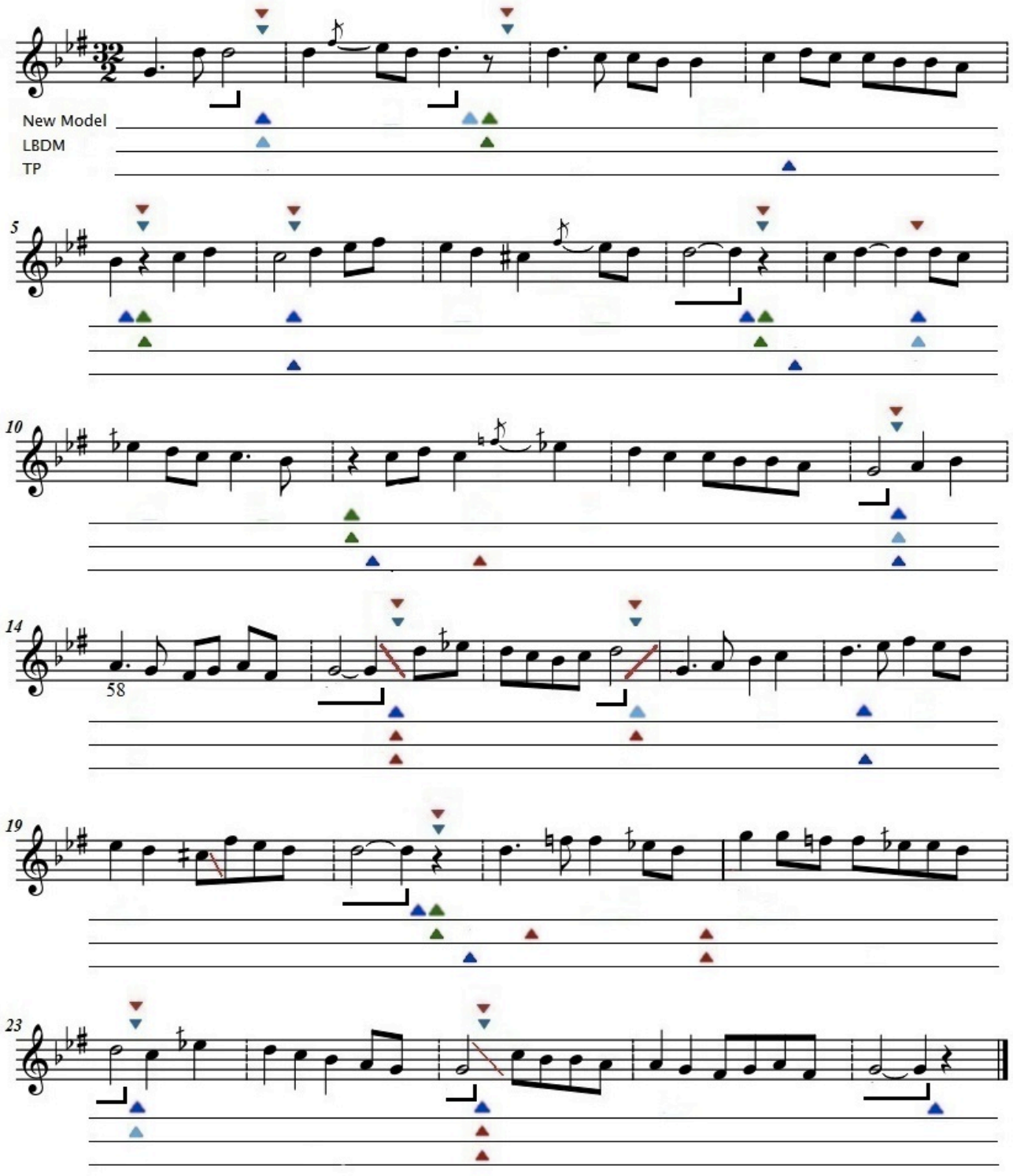


Figure 2: Analysis of same Nihavend maqam. Comparison of the boundaries given by the three computational models (Tenney and Polansky's model, LBDM and our new proposed model) with the listeners' segmentations.

4. EMPIRICAL COMPARISONS ON TRADITIONAL TURKISH MUSIC

The goal of the empirical part of this research was to see how makam-music trained (”musicians”) and untrained but culturally exposed (”non-musicians”) participants segment unfamiliar makam tunes of traditional Turkish art music. Furthermore, in the analyses, we looked at which of the three different computer models overlapped with our empirical data.

4.1 Music segmentation experiment

16 musicians and 14 non-musicians served as participants for this study. Musicians were undergraduate conservatory students with an average of 8 yrs of makam music conservatory training. non-musicians were university students with an average of 0.6 yrs of general (not makam) music training (ranging from 0 to 3 yrs). All non-musicians had to pass a melody discrimination test first to participate in the experiment. Except for one person, all other non-musicians reported to be listening to music that was not traditional Turkish music.

Ten musical excerpts were used, each of which had a duration ranging from 60 s to 75 s, with an average duration of 66 s. Excerpts were taken from the first measures of 10 different pieces that were written in five of the most common makams of Turkish traditional music (Hicaz, Nihavend, Saba, Ussak, Segah). The pieces were composed in various rhythmic patters of traditional Turkish music. All tunes were written via Mus2, a specific application for the notation of microtonal pieces which allows the user to play back the score with accurate intonation by using the sound samples of acoustic instruments in Turkish Music. Tunes were then recorded in the Qanun (a different type of zither) sound sample.

In traditional experimental set ups, participants only hear the melodies over headphones and then press a given key to mark a boundary. Typically, they are given two or three more trials to reattempt their segmentations. A major problem with this kind of a set up is that at each repetition participants simply do the task from anew, i.e., without benefitting from their earlier responses, except if they retain some kind of a memory for it. This is likely to incur a constant working memory load per trial, hence preventing any chances of improving their performance per repetition trial. Another potential handicap of the traditional segmentation set up is that one never knows which of the segmentation attempts is to represent the most accurate one. Participants best segmentations per melody could be their segmentation in the first trial, the second trial, or the third trial, or even worse, a mixed combination across trials. To deal with both the working memory load issue and the response accuracy issue described above, we decided to add a visual component to the task without, however, providing any visual information about the pitch and temporal/metric aspects of the tunes.

Participants listened to each tune four times, once for free listening and an additional three times to attempt their segmentations. For the segmentation trials, their task was to indicate all instances, at which they perceived a melodic boundary.

The experimental session consisted of 10 different makam tunes, each of which was repeated four times. Unlike the earlier phases, in which occasional interruptions were allowed, the experimental session was conducted in a strictly standardized fashion without any interruptions.

4.2 Analysis of the experiment data

Since Tune 9 and Tune 10 were accidentally skipped in five sessions with musicians and four sessions with non-musician, all following analyses were done on Tunes 1 to 8.

Except for one nonmusician, all other non-musicians mentioned that that their final (third) segmentations were their best segmentations. Musicians, too, predominantly reported their last segmentations to be their best ones.

Musicians and non-musicians segmentation locations in milliseconds for all ten tunes showed overall good convergences within, and more importantly, between-groups. There were fewer convergences for Tunes 9 and 10, most likely so because those two tunes were accidentally skipped for five musicians and four non-musicians, but maybe also because participants might have worn out towards the end of the study (though only one or two participant reported fatigue). A third possibility is that those two tunes were tunes that lacked salient boundaries.

Even a purely visual evaluation of musicians and non-musicians histograms per tune suggests a considerable overlap in the locations of the most frequently chosen segmentations. This was true for all remaining tunes as well.

A possible way of statistically testing the degree of overlap in segmentations between musicians and non-musicians is to calculate the unbiased variances (in seconds) of all the segmentation locations separately for each group and then calculate the unbiased variances for the combined segmentation locations of musicians and non-musicians (Table 1). If musicians and non-musicians have good convergences within themselves but not across each other, we shall expect the variances per group to be always smaller than the variance of both participant groups combined. If, on the other hand, musicians and non-musicians have strongly overlapping segmentations, we shall expect very similar variances across those three different data sets. Table 3 shows that the variances of the combined set are very similar to the separate variances of each group. Moreover, the variances of the combined data sets always fell in between the variances of the two separate sets. For some tunes, the smallest variances in segmentation locations were observed in musicians and for some tunes, in non-musicians.

4.3 Comparison between listeners reactions and computational predictions

We compared the responses of the participants with the predictions of our proposed model as well as with the LBDM and Tenney and Polanskys model. The aim of this comparison was to see how well these models predicted the perceptual boundaries. The results, being obtained very

Piece	Musicians	non-musicians	Combined
5	396.3 (141)	392.9 (128)	393.6 (270)
6	396.3 (141)	392.9 (128)	393.6 (270)
7	286.4 (125)	306.5 (108)	294.5 (234)
8	320.0 (122)	343.5 (118)	330.2 (241)

Table 1: Unbiased variances of segmentation locations in milliseconds (and degrees of freedom in parantheses) for musicians and non-musicians per tune. The last column shows the variance for both musicians and non musicians altogether.

recently, are given as a preliminary consideration due to lack of sufficient time. The comparisons for one specific piece (a Nihavend maqam in Kar form) are shown in both Figures 1 and 2. In Figure 2, the listeners' segmentation decisions are shown by triangles above the score, blue for musicians and red for non-musicians. Because the first segmentation level (in "clangs") in Tenney and Polansky's model gives too many segmentations (as shown in Figure 1), we only keep the second segmentation (in "segments"), as shown in the (TP) line in Figure 2.

The three models are tested individually for all monophonic makam tunes. The results, in general, show that the proposed model offered the best congruence with listeners indications. Tenney and Polanskys model gave the worst congruence with listeners' segmentation, except in tune (4) in which outperformed LBDM in predicting certain perceptual locations. The boundary locations suggested by LBDM with strength superior to 0.50 are coincided with some of the perceptual segmentations across five tunes (3, 4 and 6-8) whereas the perceptual locations indicated in the rest of the tunes can be related to LBDM segments of lower strengths on the whole.

Table 2 shows the total number of boundaries indicated by the participants and the number of boundaries predicted by the three computational models that overlaps with those of participants.

Piece	M	NonM	New	LBDM	TP
1	14	15	15	6	6
2	14	12	14	10	4
3	12	10	10	9	4
4	15	11	13	4	7
5	11	12	11	10	6
6	11	12	12	11	6
7	9	9	9	7	5
8	9	10	8	6	5

Table 2: Number of boundaries indicated by musicians (M) and non-musicians (NonM) compared to those predicted by our proposed model (New), LBDM and Tenney and Polansky's model (TP) that overlaps with the listeners' boundaries.

The initial analysis made on the perceptual groupings for all eight tunes shows that the inter-onset interval was the prominent dimension for all the participants in determining a potential boundary, which is highly correlated with the makam structure of the tunes. The inter-onset intervals perceived by the participants overall, with a few exceptional locations in certain tunes which were only perceived by musicians, correspond to the central or the dominant scale degrees of the related makam of each tune to a large extent, which supports the idea in our proposed model that the location of inter-onset boundaries are based on particular aspects that are highly different from those relating to pitch and rest boundaries. The central (G) and the domimant (D) degrees of Nihavend makam scale is displayed by the small horizantal lines belove the staves in Figure 2 as an illustration. Due to the real-time setting of the experiment, the boundaries of the participants have been relocated on the score in the post-experimental phase.

The future analysis of the segment locations of the participants may provide a deeper level of understanding in evaluating the performance of the three computational models in terms of their predictions. The previous studies showed that both the low-level perceptual processes mainly explained by Gestalt principles and the culture based knowledge that belongs to higher levels may play an important role in the perceptual grouping mechanisms of listeners, which would be depended on a complex interaction among themselves (Lartillot & Ayari, 2011). Although the aim of this research is not to investigate these interactions, the preliminary evaluations suggests a probable interaction.

5. ACKNOWLEDGMENTS

We would like to thank Mustafa (Ugur) Kaya, former graduate from the Bogazici University Psychology department and an current M.A. student in the Developmental Psychology track at Koc University, for being able to implement the experimental set up thanks to his superb expertise in computer programming and strong understanding of experimentation. Special thanks also goes to Taylan Cemgil, PhD, a colleague from Bogazici University, for helping out with some of the graphical representations in Matlab.

6. REFERENCES

Cambouropoulos, E. (2006). Musical parallelism and melodic segmentation: A computational approach. *Music Perception*, *23*(3), 249–269.

Lartillot, O. (2011). A comprehensive and modular framework for audio content extraction, aimed at research, pedagogy, and digital library management. In *130th Audio Engineering Society Convention*.

Lartillot, O. & Ayari, M. (2011). Cultural impact in listeners' structural understanding of a tunisian traditional modal improvisation, studied with the help of computational models tunisian traditional modal improvisation, studied with the help of computational models. *Journal of Interdisciplinary Music Studies*, *5*(1), 85–100.

Tenney, J. & Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, *24*, 205–241.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

COMPUTER-ASSISTED TRANSCRIPTION OF ETHNIC MUSIC

Joren Six, Olmo Cornelis

School of Arts
University College Ghent
Hoogpoort 64, 9000 Ghent - Belgium

joren@0110.be
olmo.cornelis@hogent.be

1. EXTENDED ABSTRACT

This research presents a system that eases the difficult, and time consuming task of transcribing ethnic music, especially if the pitch organization used in that music, is not well documented, or even unknown beforehand. The system analyses the music, suggests pitch organization automatically, and has features to assist transcription.

A system assisting pitch transcription should tackle the following challenges: it should be easy to repeat a small audio excerpt and to go from one loop to the next. Preferably, it should be possible to play loops slower than real-time, without affecting pitch, so that the transcriber can pick up on small details, and is able to follow quick passages. Another practical feature should be a way to visualize the main melody. The system should also provide a suggestion of the used pitch organization automatically. The transcriber also might want to check if the transcription is correct by performing the transcription. Therefore, an interface is wanted that allows musical performances in any tone scale.

Our system to transcribe pitch is based on Tarsos. Tarsos [1] is a modular software platform to extract and analyze pitch and scale organization in music. It is especially geared towards the analysis of non-Western music. Tarsos aims to be a user-friendly, interactive tool to explore tone scales and pitch organization in music of the world. An overview of Tarsos and its applications can be found in Six & Cornelis (2011); Six et al. (2013). Tarsos was mainly developed for analysis, but is now extended with features to assist transcription:

- A way to loop small audio excerpts, and to go from one loop to another easily has been built-in.
- A time stretching feature has been added, it allows to slow down audio playback without affecting pitch. The WSOLA (Verhelst & Roelands, 1993) time stretch algorithm has been implemented in TarsosDSP [2] . The feature allows transcribers to pick up on details in quick passages.
- The melograph shows the contour of the main melody. The contour depends on the pitch detection scheme chosen. Tarsos contains several pitch detection algorithms. In Figure1, the melograph can be found in the top pane.
- Pitch histograms and pitch class histograms are computed automatically. They suggest the pitch organization and can be used to extract pitch classes. In Figure 1, the pitch class histogram can be found at the bottom right.
- Tarsos contains a MIDI synthesizer that supports tuning dump messages, the synthesizer can be tuned to using any tone scale [3] . This means that a transcription can be played in the original tuning, e.g. to see if the transcription aligns well with the original material.

[1] Tarsos is available on http://tarsos.0110.be and is open source software. It runs on all major operating system with a recent Java Runtime.

[2] TarsosDSP is a Java DSP library which contains various practical audio processing algorithms, of which some are used within Tarsos. For easy re-use it is separated from the main Tarsos project and available on GitHub https://github.com/JorenSix/TarsosDSP

[3] mid (1996) defines how tuning can be done. In essence, it defines how to assign arbitrary pitch values, in cent, to 128 available keys.

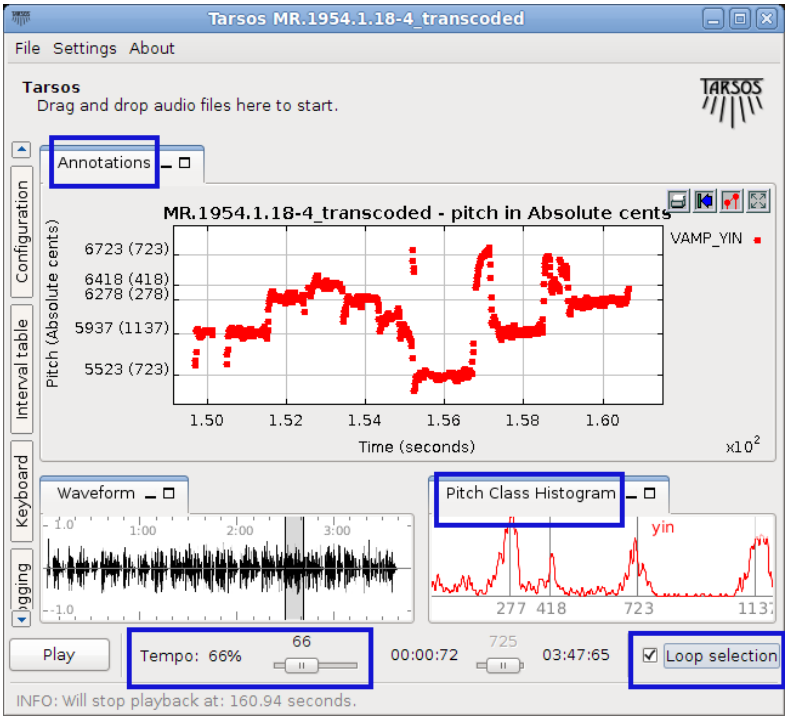


Figure 1: Transcription features in Tarsos, from upper-left to bottom right: The melograph with a pitch contour, the pitch class histogram of the current selection, the playback tempo, and the check box that determines if the current selection is looped.

1.1 Conclusion

Extensions to Tarsos have been presented, which allow Tarsos to assists in transcription of ethnic music, even when the pitch organization of the music is unknown beforehand. The system in its current form does not deal directly with timbral or rhytmical features, but is well suited for transcribing melodic material. It enables the transcriber to easily go from one, optionally time stretched, audio loop to the next. It has a visual representation of the main melody, and suggests pitch organization automatically. It also offers a way to play transcribed material on an automatically tuned MIDI synthesizer.

2. REFERENCES

(1996). *The Complete* MIDI *1.0 Detailed Specification*. MIDI Manufacturers Association.

Six, J. & Cornelis, O. (2011). Tarsos - a Platform to Explore Pitch Scales in Non-Western and Western Music. In *Proceedings of the 12th International Symposium on Music Information Retrieval (ISMIR 2011)*.

Six, J., Cornelis, O., & Leman, M. (2013). Tarsos, a Modular Platform for Precise Pitch Analysis of Western and Non-Western Music. *Journal of New Music Research (JNMR)*. Accepted for the upcomming special issue on computational ethnomusicology.

Verhelst, W. & Roelands, M. (1993). An Overlap-Add Technique Based on Waveform Similarity (WSOLA) for High Quality Time-Scale Modification of Speech. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 1993)*, (pp. 554–557).

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

AN CONTENT-BASED EMOTION CATEGORISATION ANALYSIS OF CHINESE CULTURAL REVOLUTION SONGS

Mi Tian, Dawn A A Black, György Fazekas, Mark Sandler
Centre for Digital Music, Queen Mary University of London
{firstname.lastname}@eecs.qmul.ac.uk

1. INTRODUCTION

Traditional Chinese music differs from Western music in many ways. Instruments, rhythms and melodic constructs all possess elements that are unique to Chinese music and its Asian ancestry. Differing music genres are clearly identifiable as well as their roots, regional and musicological origins, history and representation styles. In recent times, with the incorporation of many western music features, some music styles in China have made gradual evolvements while still maintaining most of their traditional elements and express very specific emotions within the social context. In this paper, the musicology of Cultural Revolution period songs [1] in Post-Liberation China is studied. The paper aims at casting a Chinese perspective and bias into the development of metrics and semantic tags for this certain music style and extracting and analysing the emotional features of it.

In recent decades after the foundation of the Peoples Republic of China, Chinese music (Shen, 2001) is developing in an ever-increasing pace. The music continues a rich traditional heritage in one aspect, while in another, it has been incorporating many new elements and emerging into a more contemporary and prosperous form. One of the most popular and representative genre in this era is the Cultural Revolution period songs, also referred to as Red Songs, which, based on a typical Chinese musical construct, references the marching rhythms and instrumentation of European music. In the particular background of political unrest, the spirit lifting rhythms of Cultural Revolution period songs impose strong and very specific emotional effects on contemporary audiences as well as later generations.

Due to the subjective nature of human perception, classification of the emotion in music is a challenging problem. Simply assigning an emotion class to a piece of rhythm could be problematic because people may hold different feelings for a song. Multiple research approaches for analysing and quantifying emotions related to music have been raised(Juslin et al., 2001; Kim et al., 2010). Emotions evoked in music is usually described either as a decomposition into a few basic emotions (happy, anger, etc.) or as a multidimensional space of valence, arousal, etc. Multiple categorial and dimensional emotion models have been raised in the MER studies. (Bischoff et al., 2009; Eerola et al., 2009; Han et al., 2009; Russell, 1980; Wang et al., 2011). Music Emotions can be influenced by multiple acoustic attributes such as timbre, harmony and tempo. Emotional representation of music is mainly derived from two channels: contextual text annotation (online documentation, social tags(Levy & Sandler, 2007; Laurier & Sordo, 2009) and lyrics) and content-based feature analysis. As suggested in , mean and standard deviation vale with a total feature-Some common utilised features for music mood recognition is given in Tale 1. This work utilises the audio-based feature analysis to conduct the emotion detection and human annotation to evaluate the results.

Krumhansl (2002)indicats that mode, intensity, timbre and rhythm are of great significance in arousing different music moods. The features extracted for this work are RMS, tempo, MFCCs, spectral centroid and beat histogram.

[1] http://academics.wellesley.edu/Polisci/wj/China/CRSongs/wagner-redguards_songs.html

2. PROPOSED METHOD

2.1 Feature Extraction

The Marsyas tool (Tzanetakis & Cook, 2002) was used in the process of extraction of features of MFCCs, spectral centroid and beat histogram. A dataset of 30-second music clips are gathered with manually annotation as the training and testing set. The attributes in the generated .arff file are later used in the Weka machine learning environment(Hall et al., 2009) for training the classifiers.

Type	Features
Dynamics	RMS energy, etc.
Timbre	MFCCs, attack slope, attack time, brightness, etc.
Harmonic	Harmonic deviation, key strength, pitch, etc.
Rhythm	Beat histograms, tempo, event density, etc.
Spectral	Chromagram, centroid and deviation, spread, skewness, etc.

Table 1: Common audio feature types in music mood analysis

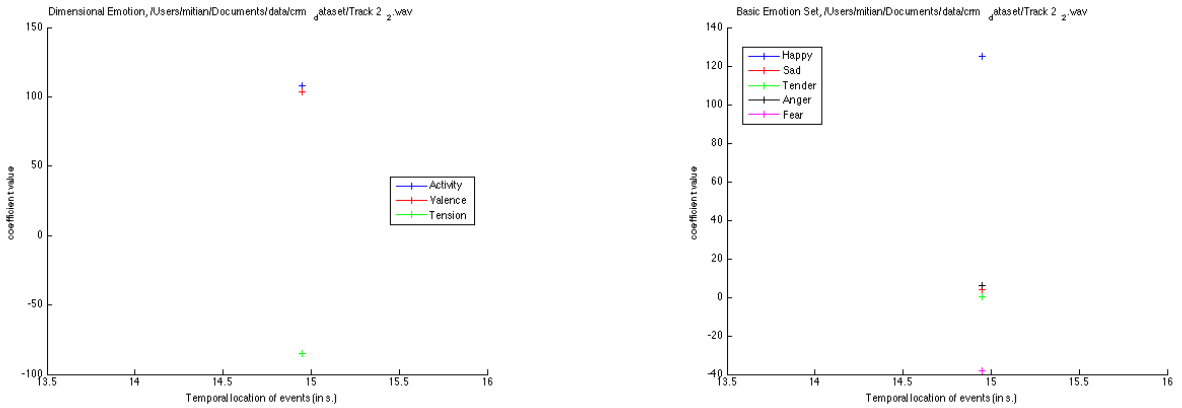


Figure 1: Results of the 'miremotion' function of an example music.

2.2 Listening Test

Listening test has been set in which participants are divided into two groups: those who have certain knowledge of the music background or understand the Chinese lyrics, and those who don't. listen to 30 clips of Chinese Cultural Revolution Songs that last 30 seconds. For each track, they are ask to rate each piece using the two dimensions: Valence (happy-sad continuum) and Arousal (excited-relaxed continuum). Each class or dimension is supposed to have values spanned in the interval [-5, 5].

3. EVALUATION

The results of this study is evaluated by human annotation. The list of music tags are cited from the MIREX Mood Tag Dataset[2]. Hu et al. (2009) detailed how the mood tag groups were described. A group of listeners with Chinese backgrounds participated in listening to those pieces of music and chose the from tags they think that can best describe the music itself. The most chosen five tags are: 'happy', 'zest', 'high spirits', 'desire', 'excitement', and generally reflects the spirit-lifting and arousing feature of this kind of music. We also used MIRtoolbox (Olivier Lartillot, 2007), a Matlab toolbox dedicated to the extraction of musical features from audio files, in which the emotion is decomposed into 5 basic emotion classes: happy, tender, anger and fear. The results generally indicates high value in the v-a space or 'happy' in the emotion class. One example is shown in Figure 1.

4. CONCLUSION AND FUTURE WORK

Due to time limit, the experiments proposed in the paper is still in progress. Future work includes further explore the applicability of the MIR feature extraction tools such as MIRtoolbox and MARSYAS and explore more suitable emotion description model of this kind of music discussed in the paper.

[2] MIREX 2012 Audio Tag Classification http://www.music-ir.org/mirex/wiki/2012:Audio_Tag_Classification#Mood_Tag_Dataset

5. REFERENCES

Bischoff, K., Firan, C. S., Paiu, R., Nejdl, W., Laurier, C., & Sordo, M. (2009). Music mood and theme classification-a hybrid approach. In *10th International Society for Music Information Retrieval Conference, number Ismir*, (pp. 657–662).

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In *Proceedings of 10th International Conference on Music Information Retrieval (ISMIR 2009)*, (pp. 621–626).

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*.

Han, B.-j., Ho, S., Dannenberg, R. B., & Hwang, E. (2009). Smers: Music emotion recognition using support vector regression.

Hu, X., Downie, J. S., & Ehmann, A. F. (2009). Lyric text mining in music mood classification. *American music*.

Juslin, P. N., Sloboda, J. A., et al. (2001). *Music and emotion*, volume 315. Oxford University Press New York.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., & Turnbull, D. (2010). Music emotion recognition: A state of the art review. In *Proc. ISMIR*.

Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, *11*(2).

Laurier, C. & Sordo, M. (2009). Music mood representations from social tags. *10th International Society for Music Information Retrieval Conference*.

Levy, M. & Sandler, M. (2007). A semantic space for music derived from social tags. *Proceedings of the International Conference on music Information Retrieval*.

Olivier Lartillot, P. T. (2007). A matlab toolbox for musical feature extraction from audio. In *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*.

Shen, S. (2001). *Chinese Music in the 20th Century (Chinese Music Monograph Series)*. Chinese Music Society of North America Press.

Tzanetakis, G. & Cook, P. (2002). Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*.

Wang, X., Chen, X., Yang, D., & Wu, Y. (2011). Music emotion classification of chinese songs based on lyrics using tf* idf and rhyme. In *12th International Society for Music Information Retrieval Conference, number Ismir*.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

INTRODUCING THE JAZZOMAT PROJECT AND THE MELOPY LIBRARY

Klaus Frieler, Martin Pfleiderer, Wolf-Georg Zaddach
Hochschule für Musik "Franz Liszt" Weimar

Jakob Abesser
Fraunhofer IDMT, Ilmenau

1. INTRODUCTION

In October 2012 a research project on the statistical analysis of jazz solos started operation at the HfM "Franz Liszt" in Weimar [1] . The aim of this contribution is to introduce this project, its goals and methods, and to place it in the context of jazz research, cognitive music psychology and computational (ethno-)musicology with special consideration of oral traditions in jazz and its implications for creativity research (cf. Pfleiderer & Frieler (2010)).

2. RESEARCH GOALS

The Jazzomat project consists of two main building blocks: The first is a considerably large database of (approx. 200) high-quality jazz solo transcriptions taken from a representative sample of jazz styles, the second are analytical tools, which will be used to investigate the creative design of jazz solos. The lack of a solid data base to confirm certain persistent but largely untested theories is one of the most important (but often neglected) problems in jazz research (cf. Pfleiderer (2004)). One well-known example for these theories, is the so-called "formulaic approach to improvisation". Thomas Owens (1974) demonstrated the existence of certain formulas in Charlie Parker's solos, but no thorough systematic examination for other improvisers—let alone across improvisers—has been carried out so far. Especially, the "orality" of formulas, i. e., the question if, how and to what extend they spread from one improviser to another (and perhaps nowadays by textbooks) is in the main focus of the project. This tasks amounts essentially to pattern mining in a large data set of solo transcriptions, based on the definition of cognitively adequate and instrument specific musical patterns (Conklin & Anagnostopoulou (2006); Lartillot & Toiviainen (2007); Meredith et al. (2002)). The transmission of patterns will be examined by tracing them across solos, which will provide insights in the practice of oral tradition in jazz.

3. DATABASE ASSEMBLY AND MUSIC REPRESENTATION

The database is being assembled using modern MIR tools (Songs2See, SmartScore) and notated transcriptions taken from the literature and other public sources. This material is meticulously cross-checked, rectified and enhanced by experienced jazz musicians and musicology students with the help of Sonic Visualiser. Beat tracks, chords and phrase structure are annotated manually. Since the focus lies on the syntactic and not on the expressive nor the semantic level, the improvisations are coded as metrical and harmonically annotated lists of pitches with precise onset and durations times but currently without loudness and timbre information, which, however, is planned to be included in the future.

4. THE MELOPY LIBRARY

The statistical approach to music cogition has gained some impetus during the last years (e. g. Huron (2006); Pearce et al. (2008); Müllensiefen et al. (2008)). The main idea of this approach is that human brains build up, extract and update probability distributions from incoming perceptual streams which leads to the generation of various expectations while shaping (re-)cognition. This is actually the base of the cultural background of a listener. Hence, while studying probability distributions of musical objects, conclusions on listeners' perceptions and cognitions can be drawn and adequate models can be built. Furthermore, probability distribution of musical elements give descriptions of music-cultural traditions and means to compare these. In our context, for example, the distribution of certain formulas and patterns might serve as a discriminating feature for individual and genre styles. Furthermore, creativity in general relies on the cultural-cognitive background and on preconfigured sets of cultural transmitted building blocks.

To this end, MELOPY, an open-source Python library, is currently under development (with a base system already in operation). The basic frame work is designed to be very general, which makes the library widely usable, e. g., in the field of computational music ethnology. The general philosophy of its music representation is not based on a musical score – such as, e. g., Music 21 (Cuthbert & Ariza (2010)), Humdrum (Huron (1995)) and many other systems–, but on physically measured sonic entities (tones), e. g., suited for ethnological field recordings and other non-scored music such as jazz, pop and rock. Once a comprehensive representation of musical entities is achieved, application of statistic methods is straight-forward, but the biggest challenge is the selection of suitable abstractions from the musical surface, a problem which will be addressed thoroughly in the project including auxillary lab experiment. To allow greatest flexibility, we will include a so-called "feature machine", which provides the user with

[1] "Melodisch-rhythmische Gestaltung von Jazzimprovisationen. Rechnerbasierte Musikanalyse einstimmiger Jazzsoli", DFG-PF 669/7-1

the possibility of defining features and arbitrary combinations of them in a highly modular fashion by connecting simple building blocks. This approach allows fruthermore-automated feature generation and selection for machine learning purposes, such as stilistic classifications. Furthermore, we plan to implement interoperability with existing computational system such as Music21 or Humdrum, enhancing the research options even further.

5. WEB-PLATFORM

Finally, we plan to set up a publicly available web platform prodiving access to our database which will not only contain our jazz data but preferably other melody corpora as well, e. g. the EsAC data (Schaffrath (1995)). The web site will allow users to download, listen, read, analyse, visualize and compare melodies in an easy-to-use way, and might serve as an educational tool as well [2] .

6. REFERENCES

Conklin, D. & Anagnostopoulou, C. (2006). Segmental pattern discovery in music. *Informs. Journal of Computing*, *18*, 285–293.

Cuthbert, M. S. & Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the 11th International Symposium on Music Information Retrieval*, (pp. 637–642).

Huron, D. (1995). *The Humdrum Toolkit. Reference Manual*. Menlo Park.

Huron, D. (2006). *Sweet Anticipation. Music and the Psychology of Expectation*. Cambridge: MIT Press.

Lartillot, O. & Toiviainen, P. (2007). Motivic matching strategies for automated pattern extraction. *Musicae Scientiae*, *11*(1 Suppl), 281–314.

Meredith, D., Lemström, K., & Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidmensional representations of polyphonic music. *Journal of New Music Research*, *31*(4), 321–345.

Müllensiefen, D., Wiggins, G., & Lewis, D. (2008). High-level feature descriptors and corpus-based musicology: Techniques for modelling music cognition. In Schneider, A. (Ed.), *Hamburger Jahrbuch für Musikwissenschaft: Vol. 25. Systematic and Comparative Musicology*, (pp. 133–156)., Frankfurt/M., Bern. P. Lang.

Owens, T. (1974). *Charlie Parker. Techniques of Improvisation*. PhD thesis, University of California, Los Angeles.

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2008). A comparison of statistical and rule-based models of melodic segmentation. In *Proceedings of the Ninth International Conference on Music Information Retrieval*, (pp. 89–94)., Philadelphia, USA. Drexel University.

Pfleiderer, M. (2004). Improvisieren – ästhetische Mythen und psychologische Einsichten. In Knauer, W. (Ed.), *Improvisieren. 8. Darmstädter Jazzforum 2003*, (pp. 81–99)., Hofheim. Wolke.

[2] A prototype of such a system is already implemented, and can be accessed under `http://jazzomat.hfm-weimar.de/meloworks`.

Pfleiderer, M. & Frieler, K. (2010). The jazzomat project. issues and methods for the automatic analysis of jazz improvisations. In R. Bader, C. Neuhaus, & U. Morgenstern (Eds.), *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology* (pp. 279–295). Frankfurt/M., Bern: Peter Lang.

Schaffrath, H. (1995). The essen folksong collection in kern format. In Huron, D. (Ed.), *(Computer Database)*, Menlo Park, CA.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

A PROBABILISTIC STUDY OF CULTURE-DEPENDENT NOTE ASSOCIATION PARADIGMS IN FOLK MUSIC

Zoltán Juhász

Research Centre for Natural Sciences of the HAS
P.O.B 49. Budapest H-1525, Hungary
juhasz@mfa.kfki.hu

1. INTRODUCTION

The regularities in a given musical piece cannot be interpreted without the knowledge of the whole cultural context in that the piece arose, and this requires the analysis of many pieces belonging to the given culture. Thus, many of the rules determining the musical expectations and predictions of listeners and musicians in a given culture can be identified by statistical analyses of the pieces themselves (Huron, 2006). It follows from the above argumentation that the cross-cultural study of certain musical characteristics - including the melody, the scales, pitch and rhythm distributions - can clarify some peculiarities of different cultures on the one hand, and highlight the universal features on the other hand (Stevens, 2004; Kolinski, 1961). In this work, we study some statistical characteristics of the pitch in different folk music cultures. According to this task, we represent the musical context of the pitch cognition by folksongs, and the musical cultures determining the context in the concrete melodies by folksong corpora.

2. METHODS

In order to reveal some special rules of melody evolution in different folk music cultures, we characterised the correlations of the notes by the conditional probabilities of their joint appearances in common melodies. Our main question is illustrated in Figure 1. Here, the reader can see at the first glance that the pitch distributions of the Chuvash and Bulgarian folksongs refer to a pentatonic and a diatonic Aeolian melody, respectively. The time rate values show that the Chuvash melody spends the most time at the octave (8), the fifth (5), the fourth (4) and the tonic (1), while the Bulgarian one at the supertonic (2), the tonic (1) and the subtonic (bVII). We ask if these common dominances of the octave, the fifth, the fourth and the tonic, as well as the supertonic, tonic and subtonic are regular events in Chuvash as well as Bulgarian folk music, or not? More generally: which mutual preferences of notes are characteristic in the melodies of different cultures? More exactly: we want to study the conditional probabilities indicating joint and dominant tone appearances within melodies. For instance, we calculate the probability of the event that a presence taking at least 5% of the total time of a melody of the fifth results in a similar presence of the third, fourth, sixth, etc. Thus, the meaning of the conditional probability $p_{i,j}$ can be expressed as follows: $p_{i,j}$ is the conditional probability of the event that – providing that the *i*th degree plays a dominant role in a melody of the culture - the *j*th one also plays a similar role. Accomplishing all these calculations, we get a quadratic, non symmetric matrix of size of the number of the degrees studied, containing the above mentioned conditional probability values. Although the temporal order of the degrees could also be characterised statistically, the definition of the conditional probabilities given in this work focuses only on the joint appearance of them, independent of their order.

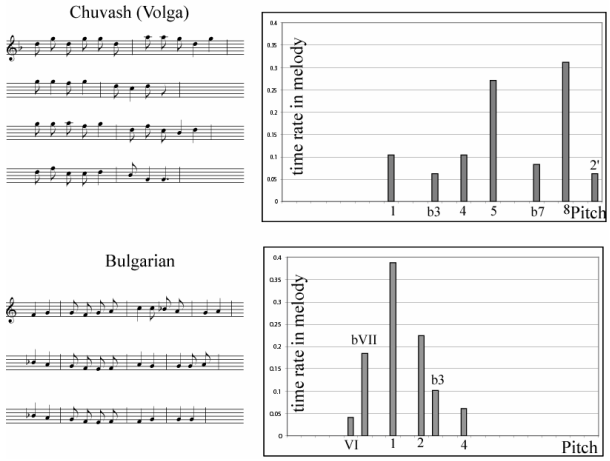


Figure 1. Pitch distributions of a Chuvash and a Bulgarian folksong.

In order to visualise the mutual "affinities" of the degrees, we elaborated a special version of the multidimensional scaling (MDS) algorithm (Borg, 2005). In MDS technique, the input data to be visualised are presented in a quadratic matrix containing some distance-like or similarity-like values between some objects. (For instance, the matrix can contain geographical distances between towns, or dissimilarity ratings of melodies, etc.) The aim of the algorithm is to represent the objects (towns or melodies) in a low dimensional space (often in a plane) with the requirement that the distances of the low dimensional points must optimally correspond to the input values.

3. RESULTS

Using the MDS technique, we investigated the degrees in the range of IV – 2' (being defined 1 as the tonic), thus we had D=22 notes, including all semitones in the given range. The "discrepancy" of the jth and ith degrees can be characterised by $q_{i,j} = 1 - p_{i,j}$, so the aim of our MDS algorithm was to arrange D=22 points in a low dimensional space with the requirement that their distances should optimally correspond to the above discrepancy values.

Two resulting arrangements are shown in Figure 2. In the Volga-Kama-Region (the database contains Mari, Chuvash, Tatar and Votiak folksongs, see Figure 2d), the central part of the graph is occupied by the tonic (1), fourth (4), fifth (5) and octave (8), being mutually very close to each other. It was discussed in a previous paper that there is a very close relation between the harmonic systems of these notes (1-4-5-8), therefore we call this group the "spectral basis" (Juhász, 2012). Supplementing these notes with the second and the sixth, being very close neighbours on the left side of the graph, we obtain a pentatonic scale which is called sol-pentatonic in Hungarian and Zhi in Chinese music theory (1-2-4-5-6-8). For instance, the G-A-c-d-e-g notes construct a sol-pentatonic (Zhi) scale (Yaxiong, 2008). VI and 2', also being very close to 2 and 6, are the extensions of this pentatonic octave. On the right side, the 1-4-5-8 "spectral basis" is completed by the missing notes of the "la-pentatonic" (Chinese "Yu") scale b7 and b3, while the close neighbour bVII is also an extension of this la-pentatonic octave (1-b3-4-5-b7-8). Since the folk music of the Volga-region is pentatonic, other notes become much less important, so their distances from the central part are much larger. The main lesson of the graph can be summarised in the statement that the tonal structure of this culture is based on the "spectral basis" (1-4-5-8), which is completed by further notes in order to obtain the sol-pentatonic Zhi as well as la-pentatonic Yu scales. Other pentatonic modes are also found in the Volga Region, but the two scales mentioned above fill the dominant role in this culture.

The Bulgarian graph shows a radically different picture in Figure 2a. Here, the notes are arranged mainly according to their degree, thus, the neighbourhood of the points corresponds to a second or semitone interval in the most cases. (As an illustration, the notes of the Aeolian scale are marked by bold letters.) The tonic-fourth-fifth-octave group also fits to this arrangement, thus, the points 1, 4, 5 and 8 are very far from each other. There is no reason to speak about a spectral basis or a dominant-subdominant system, because the figure clearly shows that the integration of the notes is based mainly on the pitch height in this culture. Anatolian and Hungarian systems in Figures 2b and 2c represent a continuous transition between the interval-based and the pentatonic note association paradigms.

For instance, comparing the Anatolian structure to the Bulgarian one, some connections and differences can be observed simultaneously (See Figures 2a and 2b). The notes of the spectral basis (1-4-5-8) are arranged along a more or less straight long line, and the sequence of the notes corresponds to their degree; these features remind us of the Bulgarian structure. However, an important difference is that the notes fitting to the major-like (Ionic) as well as minor-like (Aeolic) scales are separated on the left, as well as right sides of the spectral basis (2,3,6,7, as well as b2,b3,b6,b7).

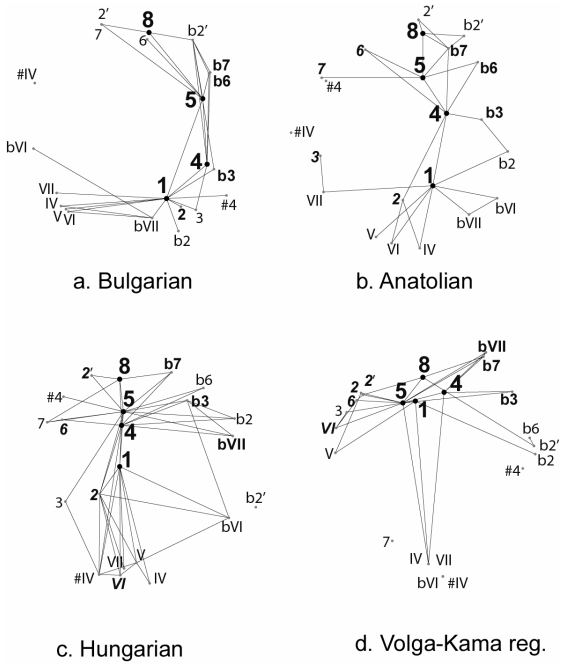


Figure 2. MDS plots of Bulgarian Anatolian, Hungarian and Mari – Chuvash-Tatar-Votiac (Volga-Kama) note affiliations.

Finally, we show two asymmetric note association paradigms preferring major-like as well as minor-like scales in Figure 3. Figure 3a shows that the asymmetry of the Dutch plot follows from the dominance of the major melodies in this database. For instance, the conditional probability of the presence of the minor third in a Dutch melody with the assumption that the same melody contains the fifth is 0.042, while the opposite case, i.e. the minor third "allures" the fifth, has the probability of 0.479. Due to this asymmetry, our MDS algorithm places the corresponding points rather far from each other, since one of the conditional probabilities is very small (the discrepancy is very large). The same values are 0.48 and 0.833 for the major third, so the point representing the major third gets much closer to the fifth. Similar numerical examples can illustrate that the asymmetry of the Romanian system mirrors the preference of minor-like scales in this culture.

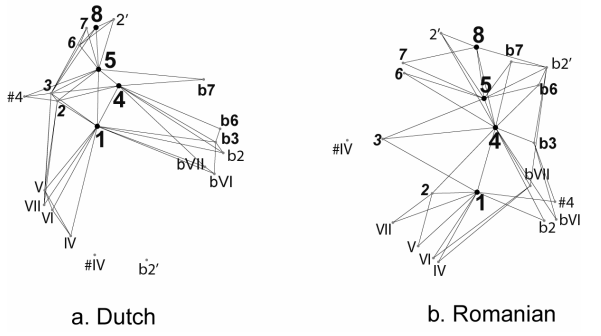


Figure 3. MDS plots of Dutch and Romanian note affiliations and the corresponding model.

4. SUMMARY

In the present work, our basic assumption was that the conditional probabilities characterising the common appearances of notes refer to a hidden, not formulated evaluation of the affinity between the different degrees in oral musical cultures. We described these systems of note affiliations by the conditional probability matrices, and visualised them using a multidimensional scaling algorithm. We found that the resulting graphs can indeed originate in culture-dependent paradigms being connected in some measure.

For instance, the Chuvash melody in Figure 1 suggests that the fifth and the octave are strongly related to the tonic and each other, while the Bulgarian song does not use them at all. In this latter case, the neighbouring notes seem to allure each other. In a previous work, we constructed a numerical model describing the culture-dependent relation of degrees as a weighted sum of two parameters depending on their interval distance as well as "spectral similarity" (calculated from the number of the common harmonic components of the notes) (Juhász, 2012). This model showed that the difference between the cultures represented here by the Bulgarian and Volga-Kama systems can be traced back to the dominance of the interval-based, as well as the spectral similarity-based note association paradigms.

5. REFERENCES

Borg, I, P. Groenen, P (2005). Modern Multidimensional Scaling: theory and applications (2nd ed.), Springer-Verlag New York, 2005.

Huron, D ((2006). Sweet Anticipation: Music and the Psychology of Expectation. Cambridge, MA: MIT Press, 2006.

Juhász, Z. (2012). A mathematical study of note association paradigms in different folk music cultures, *Journal of Mathematics and Music, Vol. 6, Issue 3, 2012*, pp 169-185

Stevens, C (2004). Cross-cultural studies of musical pitch and time, *Acoustical Science and Technology*, Vol. 25 (2004) , No. 6 pp.433-438

Kolinski, M (1961). 'Classification of tonal structures, illustrated by a comparative chart of American Indian, African Negro, Afro-American and English-American structures,' *Studies in Ethnomusicology*,S vol 1, 1961

Yaxiong, D (2008). Fundamentals of Chinese music theory and its cultural basis. Flaccus, Budapest, 2008,(in Hungarian).

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

THE CHURCHES' TUNING

Enric Guaus
Escola Superior de Música de Catalunya
Barcelona, Spain
enric.guaus@esmuc.cat

Jaume Ayats
Museu de la Música de Barcelona
Barcelona, Spain
jayatsa@bcn.cat

1. INTRODUCTION

In this study, we explore the relationship between the tuning of fundamental pitches obtained from different recordings of religious chants from oral tradition in the Pyrenees, and the acoustic properties of the space in which the recordings were made. These chants are never accompanied by other instruments that can help the tuning process. Our hypothesis is that the acoustics of the space is crucial for selecting the frequency of the base note of the chant. To explore the relationship between the tuning and the room, we analyze the frequencies involved in the recorded chants and compare the results with the room properties obtained by a set of acoustic measurements. Results suggest that the tuning frequencies are close to the measured room resonances.

2. MUSICAL ANALYSIS

The musical analysis focuses on Magnificat chants, assuming their similar structure will produce comparable results. Magnificat is the most appreciated chant in the mass by most of the singers (Ayats et al., 2011). Specifically, we focus on the pitch assigned to the first G note that, according to the structure of the Magnificat, fixes the tonality of the whole chant. Our goal is to determine whether Magnificat sung in specific spaces present similarities in their base pitches (i.e. G note). Pitch detection is performed using YIN (De Cheveigne & Kawahara, 2002) for Audacity. As most of the singers have no musical training, the detected base pitches are unstable and it is difficult to establish a unique frequency value. Because of that, in this work we use the mean of all the exact pithes for a given note.

Place	Year	Freq.(Hz)	Type
Enviny	2006	173	Magnificat
Enviny	2006	194	Magnificat
Enviny	2006	174	Magnificat
Llessui	2007	130	Magnificat
Particular house	1982	188	Magnificat
Enviny	2006	167	Magnificat
Gerri de la Sal	2007	145	Magnificat
València d'Àneu	2007	145	Magnificat

Table 1: Summary of Magnificat detected pitches recorded at different churches by different chanters.

3. ACOUSTIC MEASUREMENTS

We analyzed seven churches in the western catalan Pyrenees and four from the Hautes-Pyrénées in France. All of them are small churches with a small choir for the singers, usually made of wood, over the main entrance, in the opposite side of the altar. This is the typical internal distribution for early baroque decorated churches, even the origin of the church is older. Some of them are in good conditions while others are not. In general, churches which have been recently reconstructed do not preserve the original reverberation properties, changing wall surfaces, choir properties, etc. but they keep the overall structure, preserving the acoustic room mode distribution.

To establish the relationship between the singed chants and the acoustic properties of the churches, we focus on the measures of the Reverberation Time (RT60) and the Frequency Response function (FRF). We obtained a set of three measurements for each church, selected according to the communication paths during the ceremony: (a) priest to parishioners: both loudspeaker and microphone in the main space of the church, with the loudspeaker near the altar in a non symmetrical position with respect to the geometrical axes, and the microphone in the parishioner's area in a non symmetrical position with respect to the geometrical axes, (b) chanters to parishioners: the loudspeaker upstairs in the choir in a non symmetrical position with respect to the geometrical axes, and the microphone in the parishioner's area in a non symmetrical position with respect to the geometrical axes, and (c) chanters to chanters: both loudspeaker and microphone upstairs, in a non symmetrical position with respect to the geometrical axes.

4. DISCUSSION

RT60 values are in the expected range and shape, that is, reverberation times are moderate and decreasing as the frequency increases. A detailed analysis of the correlation between RT60 and base pitches doesn't show clear evidences, so, we conclude that RT60 does not influence the tuning process in churches (this conclusion is consistent with the fact that reconstructions have changed your original reverb parameters). On the other hand, we observe that there are some clusters of room resonances shared by different churches. Clusters emerged from at least 7 out of 11 churches show frequencies at $\bar{f}_1 = 31.5Hz, \sigma = 3.69$; $\bar{f}_2 = 136.3Hz, \sigma = 6.21$; and $\bar{f}_3 = 184.2Hz, \sigma = 7.40$. This means that they share some basic elements of con-

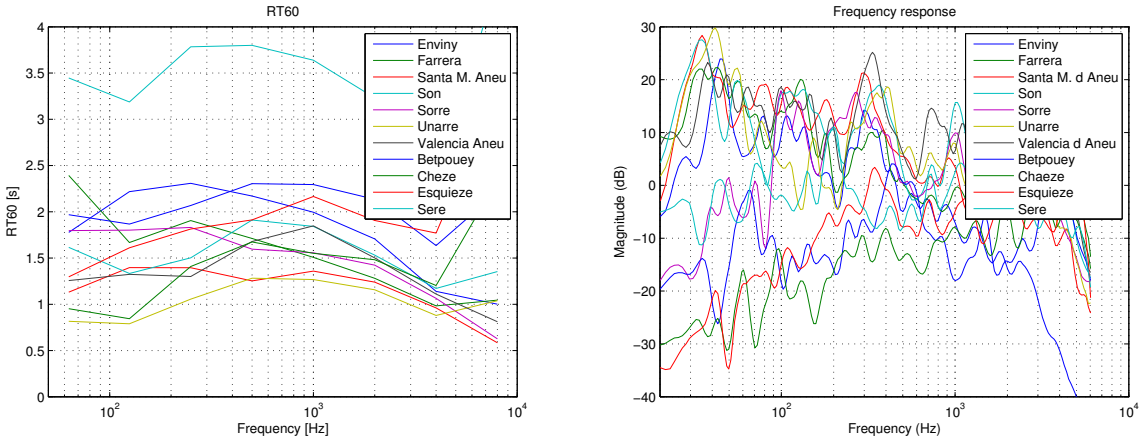


Figure 1: Summary of the (a) measured RT60 (in seconds) in octave bands, and (b) magnitude of the FRFs.

struction. According to the analyzed base pitches, we observe how the frequency resonance of the chants recorded at Llessui, Gerri de la Sal and València d'Àneu are close to the second cluster, and recordings made at Enviny are close to the third cluster. Far of being a mathematical demonstration, these results show there exists some evidence between these two factors.

5. CONCLUSIONS

In the previous section we presented the coincidence between the base pitch of the recordings and the center frequencies for the second and third clusters of room resonances. These frequency centers also coincide, in terms of musical notes, with the $D_2(f = 138.5Hz)$ and $G_2(f = 184.9Hz)$ with reference at $A_3(ref = 415Hz)$. This suggests that the two main tonalities performed in these chants are reinforced by the room. From that, we gess that, in fact, the churches at the Pyrenees are tuned near these frequencies. Moreover, the two mentioned musical notes also coincide with the "natural" modes of the iberian old bassoon instruments ("Per Natura" and "Per bemoll") which sometimes used to be included in the ceremony (Borràs, 2009). From the qualitative viewpoint, chanters from different geographic areas coincide explaining how "the church must thunder" as part of the musical performance. This implies that, above all, the chanter must feel comfortable with room resonances, and the resulting tuning slightly varies according to the specific space, as verified in this study. These results should influence some decisions taken in reconstruction processes of churches, preserving the acoustic properties as part of the heritage instead of using only visual criteria. Moreover, these results have also an impact in musicological studies providing new perspectives to some phenomena that can not be explained without the understanding of the environment (Ayats, 2011). Our future work is focused on the collection of old and new recordings to mathematically verify our hypothesis. Nevertheless, chanters with the required knowledge in the old traditions are difficult to find, and only few of them are in good enough physical conditions to sing for a recording session.

6. REFERENCES

Ayats, J., Costal, A., Gayete, I., & Rabaseda, J. (2011). Polyphonies, bodies and rhetoric of senses: latin chants in corsica and the pyrenees. *Transposition. Musique et sciences sociales*, *1*.

Borràs, J. (2009). *Baixó a la Península Ibèrica*. PhD thesis, Universitat Autònoma de Catalunya.

De Cheveigne, A. & Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, *111*(4), 1917–1930.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

DESCRIPTIVE RULE MINING OF BASQUE FOLK MUSIC

Kerstin Neubarth[1,2] **Colin G. Johnson**[2] **Darrell Conklin**[3,4]

[1]Canterbury Christ Church University, Canterbury, United Kingdom
[2]School of Computing, University of Kent, Canterbury, United Kingdom
[3]Department of Computer Science and Artificial Intelligence,
University of the Basque Country UPV/EHU, San Sebastián, Spain
[4]IKERBASQUE: Basque Foundation for Science, Bilbao, Spain

1. INTRODUCTION

As early as in the 1950s, Bronson (1959) proposed a computational approach to address typological or geographical questions about folk music, such as: "What are the characteristic differentiae of specific regions? Are there rhythmical preferences? Modal preferences? And to what degree of intensity?" One way to represent such regional or typological preferences computationally is using rules. Within predictive data mining of folk music, in particular classification, most studies (e.g. Bohak & Marolt, 2009; Hillewaere et al., 2009; Conklin, 2013) focus on global classification performance and do not report individual rules, which in isolation may not have high accuracy but nevertheless could provide answers to Bronson's questions. For classification of tune families, a recent study (van Kranenburg et al., 2013) indicates the discriminative power of features, but does not systematically specify feature values and the partitioning of the feature space. Applications of descriptive mining to folk music include subgroup discovery (Taminau et al., 2009) and distinctive pattern discovery (Conklin & Anagnostopoulou, 2011). These studies highlight musical characteristics – global features or interval patterns – that are over-represented in a geographical region or in a genre relative to their distribution in the total corpus. Metadata of a folk music collection has been mined to extract qualified associations between folk music genres and geographical regions (Neubarth et al., 2012).

Here the method of the previous study (Neubarth et al., 2012) is extended to mine for associations between musical content (global features computed from midi files) and folk music genres or regions. Association rule mining (e.g. Srikant & Agrawal, 1995) is adapted to identify different categories of rules, covering both over- and under-representation of content characteristics in genres or regions. The aim of our research is to extract rules which reflect musicological observations such as "Western melodies are largely in triple metre", "ballads rarely have unmeasured rhythms" or "the pentatonic system is not found in children's songs" (Sadie, 2001). While such statements are prominent in folk music surveys, they have not been captured by existing approaches to computational folk music analysis. The method proposed here explicitly distinguishes different relations between music content and genres or regions.

2. DATA

As a corpus we use the *Cancionero Vasco*, an iconic collection of Basque folk songs and dances. Its musicologically curated metadata includes information on the genre of a folk tune (e.g. wedding song or work song) and the geographical location where it was collected. The corpus thus offers an opportunity to analyse both song types or genres and regions, for the same corpus.

The digitised collection contains 1902 midi files. Of the 1902 tunes in the corpus, 1561 are annotated with a folk music genre and 1630 are annotated with a location. The metadata vocabulary consists of 50 genre labels and 2968 geographical labels. Both genres and geographical regions are hierarchically organised (Goienetxea et al., 2012).

Music content features were selected from an existing feature set (McKay & Fujinaga, 2006), such that the selected subset reflects content characteristics in folk music surveys (Sadie, 2001). These features were computed with jSymbolic (McKay & Fujinaga, 2006). Another three features were additionally implemented: pitch class entropy, interval perplexity and duration perplexity. Numeric features were discretised, using Weka (Hall et al., 2009), and transformed into string content items; discretisation bins are based on musicological statements (Sadie, 2001) and the distribution of feature values in the corpus.

3. METHOD

To analyse content–genre and content–region associations of different categories, we proceeded in four steps: identification of association categories; translation of the categories into association constraints; mining for association rules which meet the defined constraints; and post-processing of hierarchical rules.

Association categories were identified through a qualitative analysis of 25 reference articles on European folk music (Sadie, 2001). Statements linking music content and genres or regions were extracted and grouped according to similar meaning. This resulted in nine association categories. The categories were given an interpretation in terms of over- and under-representation. For example, a content feature can be over-represented in a genre or region with respect to its occurrence in other genres or regions (category *Primarily*), or a content feature can be under-

represented in a genre or region with respect to other content features in the same genre or region (category *Uncommon*). The category interpretations were translated into association constraints: specific combinations of rule templates (Klemettinen et al., 1994) and rule evaluation measures (Geng & Hamilton, 2006; Lenca et al., 2008).

During the mining, item sets are formed as pairs between a content item and a genre or region item. Pairs are evaluated against the constraints of each category. If a candidate pair meets the constraints, a rule is added to the results. For each rule a p-value is calculated according to Fisher's one-tailed exact test. The lower the p-value, the less expected is the number of encountered co-occurrences between a content item and genre or region, given their distributions in the entire corpus.

As the items are hierarchically organised, the discovered rules are partly redundant. In a post-processing step we thus prune or group more specific rules relative to their parent rules, depending on whether they confirm, specify or deviate from the parent rule (Liu et al., 2000).

The method outputs a structured list of qualified associations, which provide an overview of the corpus and highlight content–genre or content–region patterns for further musicological exploration.

4. RESULTS

Traditional association rule mining using the support/confidence framework (e.g. Srikant & Agrawal, 1995) usually yields a large set of rules in the following form:

> Araba $\rightarrow$ high pitch class entropy;
> $s = 26$; $c = 0.96$

In the corpus, 96% of the Araba tunes have high pitch class entropy (confidence c), and there are 26 such tunes (support s).

Association categories provide an additional qualification based on natural language and support different views, e.g. focusing on musical characteristics of a region or focusing on the regional distribution of content features, for example:

> Araba, high pitch class entropy: *Usually*
> (template $R \rightarrow C$; $c = 0.96$; $p = 0.00003$)
>
> Navarra, low pitch class entropy: *Primarily*
> (template $C \rightarrow R$; $c = 0.52$; $p = 0.006$)

The first rule describes a region and an aspect of its musical character (template $R \rightarrow C$): within the *Cancionero Vasco*, tunes from Araba usually have high pitch class entropy. The second rule highlights a content characteristic and its regional occurrence (template $C \rightarrow R$): more than half of the tunes with low pitch class entropy in the corpus (52%) is concentrated in one of the seven provinces, i.e. tunes with low pitch class entropy are primarily found among the songs collected in Navarra.

By exploiting the hierarchical organisation of the item vocabulary, the post-processing allows to distinguish general rules, contributing, specialised and deviating sub-rules, so that resulting rule sets are structured, for example:

> life-cycle songs, narrow intervals: *Present*
> (template $G - C$; $s = 251$; $p = 0.0998$)
>
> Contributing:
>
> > love songs, narrow intervals: *Present*
> > (template $G - C$; $s = 116$; $p = 0.8038$)
>
> Specialised:
>
> > lullabies, narrow intervals: *Usually*
> > (template $G \rightarrow C$; $c = 0.65$; $p = 0.00099$)
>
> Deviating:
>
> > wedding songs, narrow intervals: *Absent*
> > (template $G \rightarrow \neg C$; $c = 1.0$; $p = 0.0529$)

Average narrow intervals are present in life-cycle songs, and within life-cycle songs are found in love songs. While life-cycle songs *may* move in narrow intervals, lullabies *usually* move in narrow intervals: about two thirds of the lullabies in the corpus ($c = 0.65$) have average narrow intervals. In the *Cancionero Vasco* tunes with average narrow intervals are absent among wedding songs, which form a sub-genre of life-cycle songs: all the wedding songs in the corpus ($c = 1.0$) have average melodic intervals other than narrow ($G \rightarrow \neg C$).

5. CONCLUSIONS

Recent supervised approaches to computational folk music analysis have focused on comparing predictive methods (e.g. for region, genre or tune family classification, Conklin, 2013; Hillewaere et al., 2009; van Kranenburg et al., 2013). By contrast, we deliberately chose a descriptive approach, within a knowledge discovery paradigm, in order to extract understandable rules in response to musicological questions such as those formulated by Bronson (1959). Knowledge discovery advocates a high level of interaction between the computational design and appreciation of the application domain (e.g. Fayyad et al., 1996). In our study a qualitative analysis of folk music surveys informs the task specification, selection and discretisation of content features, definition of association categories and presentation of results.

This research extends earlier work on descriptive mining of folk music. In a previous study, categorised genre–region associations were extracted from the metadata of the *Cancionero Vasco* (Neubarth et al., 2012); here we demonstrate that the method can also identify musical characteristics of folk music genres and regions and thus link music content and metadata. The subgroup discovery study by Taminau et al. (2009) was restricted to one form of content–region rules, which corresponds to mining rules with one template ($C \rightarrow R$) and one evaluation measure (weighted relative accuracy). The association mining approach presented here provides additional flexibility to discover rules of different categories by combining several rule templates and evaluation measures. Finally, our analysis goes beyond previous research (Conklin & Anagnostopoulou, 2011; Neubarth et al., 2012) by taking into account the hierarchical structure of the metadata.

6. ACKNOWLEDGEMENTS

We thank Fundación Euskomedia and Fundación Eresbil, Spain, for providing the *Cancionero Vasco* for study.

7. REFERENCES

Bohak, C. & Marolt, M. (2009). Calculating similarity of folk song variants with melody-based features. In *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, (pp. 597–601), Kobe, Japan.

Bronson, B. H. (1959). Toward the comparative analysis of British-American folk tunes. *The Journal of American Folklore*, *72*(284), 165–191.

Conklin, D. (2013). Multiple viewpoint systems for music classification. *Journal of New Music Research*, *42*(1), 19–26.

Conklin, D. & Anagnostopoulou, C. (2011). Comparative pattern analysis of Cretan folk songs. *Journal of New Music Research*, *40*(2), 119–125.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge discovery and data mining: towards a unifying framework. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, (pp. 82–88), Portland, Oregon, USA.

Geng, L. & Hamilton, H. J. (2006). Interestingness measures for data mining: a survey. *ACM Computing Surveys*, *38*(3), 1–32.

Goienetxea, I., Arrieta, I. Bagüés, J., Cuesta, A., Leiñena, P., & Conklin, D. (2012). Ontologies for representation of folk song metadata. Technical Report EHU-KZAA-TR-2012-01, Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU. http://hdl.handle.net/10810/8053.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2009). The Weka data mining software: An update. *SIGKDD Explorations*, *11*(1), 10–18.

Hillewaere, R., Manderick, B., & Conklin, D. (2009). Global feature versus event models for folk song classification. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, (pp. 729–733), Kobe, Japan.

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., & Verkamo, A. I. (1994). Finding interesting rules from large sets of discovered association rules. In *Proceedings of the 3rd International Conference on Information and Knowledge Management*, (pp. 401–407), Gaithersburg, Maryland.

Lenca, P., Meyer, P., Vaillant, B., & Lallich, S. (2008). On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid. *European Journal for Operational Research*, *184*(2), 610–626.

Liu, B., Hu, M., & Hsu, W. (2000). Multi-level organization and summarization of the discovered rules. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2000)*, (pp. 208–217), Boston, MA.

McKay, C. & Fujinaga, I. (2006). jSymbolic: A feature extractor for MIDI files. In *Proceedings of the International Computer Music Conference*, (pp. 302–305), New Orleans, USA.

Neubarth, K., Goienetxea, I., Johnson, C. G., & Conklin, D. (2012). Association mining of folk music genres and toponyms. In *Proceedings of the 13th International Society of Music Information Retrieval Conference (ISMIR 2012)*, (pp. 7–12), Porto, Portugal.

Sadie, S. (Ed.). (2001). *New Grove Dictionary of Music and Musicians*. London: Macmillan.

Srikant, R. & Agrawal, R. (1995). Mining generalized association rules. In *Proceedings of the 21st VLDB Conference*, (pp. 407–419), Zurich, Switzerland.

Taminau, J., Hillewaere, R., Meganck, S., Conklin, D., Nowé, A., & Manderick, B. (2009). Descriptive subgroup mining of folk music. In *2nd International Workshop on Machine Learning and Music (MML 2009)*, Bled, Slovenia.

van Kranenburg, P., Volk, A., & Wiering, F. (2013). A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research*, *42*(1), 1–18.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

ON FINDING REPEATED STANZAS IN FOLK SONG RECORDINGS

Ciril Bohak, Matija Marolt
University of Ljubljana, Faculty of Computer and Information Science
{ciril.bohak, matija.marolt}@fri.uni-lj.si

ABSTRACT

In this paper we present our current work on an approach for finding repeated stanzas in folk song recordings. We improve our previous work, which relied on detection of vocal pauses to find repetitions, by relying less on prior knowledge of folk song collections. Instead, we calculate similarities between several chunks of the audio signal with the whole signal to obtain repetition patterns. We align the obtained similarity curves and calculate the average similarity curve that represents repetitions in the whole track. Distances between peaks in the obtained curve represent lengths of individual repetitions. Repetitions are aligned with the audio to yield the final segmentation.

1. RELATED WORK

Many approaches for segmenting and finding repeating parts in music recordings were developed in recent years. Most of them are using various audio features such as MFCCs or chroma vectors to calculate self-similarity matrices (Foote (1999)) and were developed for commercial music (Bartsch & Wakefield (2001); Goto (2006); Cooper & Foote (2002); Foote & Cooper (2003); Peeters (2002)). With increased interest in computational folk music analysis, several approaches for segmentation of these recordings were also introduced, based on algorithms such as DTW (Müller et al. (2009); Bohak & Marolt (2012)) and a special fitness measure (Müller et al. (2011)).

2. METHOD

Most of current segmentation methods are developed for commercial music and do not take into the account features of folk music such as inaccurate singing, presence of noise and tempo deviations. To find repeating stanzas in folk song recordings, we modified our previous algorithm (Bohak & Marolt (2012)). In the original approach we used dynamic time warping in combination with shifted chroma vectors to find similarities of short excerpts of audio starting at vocal pauses with the entire song. We assumed that vocal pauses will occur at beginnings of segments where the signal will either have low magnitude or no detectable pitch. Detection of vocal pauses turned out to be unreliable, and was also problematic for choir and instrumental recordings, so with our new approach, we decided to omit this step. Our new approach is presented in Figure 1. Before applying the method we average the audio channels into a single channel and normalize it.

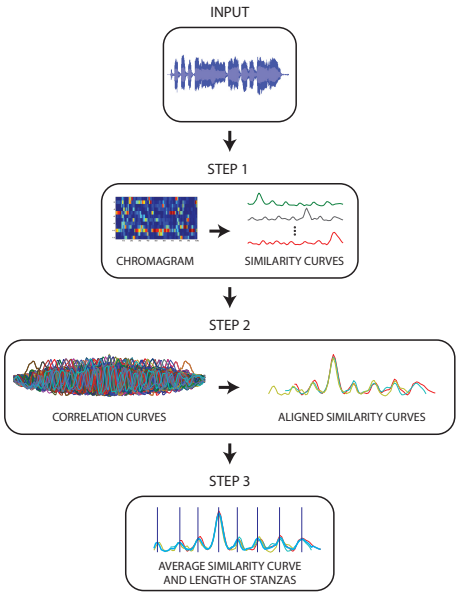


Figure 1: Outline of dual domain method for finding repeating stanzas.

2.1 Step 1 - Extracting chomagram and calculating the similarity curves for randomly selected parts

In the first step we calculate the CENS chromagram (Müller (2007)) for the whole audio track. Next we randomly select n locations in the audio track that are approximately equally distributed throughout the whole track. We use m seconds long chunks of the chromagram at selected locations and calculate similarity between the selected chunks and the entire song with Dynamic time warping and shifting chroma vectors. This results in n similarity curves that represent the similarity between selected chunks and the whole track. An example is given in Figure 2. Peaks in these curves represent repetitions of chunks in the track.

2.2 Step 2 - Aligning the similarity curves

In the second step we calculate cross-covariances between each pair of similarity curves to find those that are most similar. We select the similarity curves with above mean maximum cross-covariance value and smooth them with a low pass filter. Amongst these, we select the curve with the highest similarity to all other curves as the most representative one. As the curves are shifted in time, due to the randomly selected audio chunks, we then calculate time shifts between the selected most representative curve and

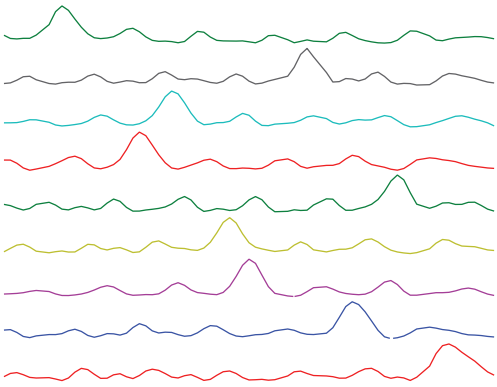


Figure 2: Similarity curves for random locations in a track.

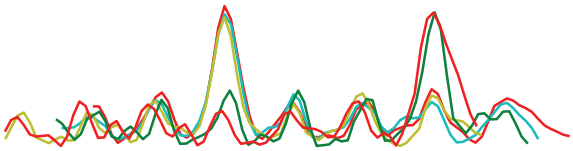


Figure 3: Curves aligned according to the calculated time shifts. Not all the curves were selected for alignment due to thresholding by mean covariance.

all the others. We use the calculated time shifts to align the curves as shown in Figure 3.

2.3 Step 3 - Calculating the average similarity curve and length of repeating stanzas

In the last step we calculate the average similarity curve that represents repetitions in the audio track. The average curve is calculated as the average of all aligned similarity curves as shown in Figure 4. By calculating the distances between peaks in the obtained average similarity curve we can calculate lengths of individual stanzas. We determine the beginning of the first stanza by cutting the silence from beginning of the track and use the calculated distances to find beginnings of all other stanzas.

3. CONCLUSIONS AND FUTURE WORK

In this paper we presented our current work in progress on a method for finding repeated stanzas in folk song recordings. In the future we are planning on extending our method to a double domain approach, in which we will augment results from the presented approach with an algorithm that works on the symbolic domain. Symbolic data will be obtained with polyphonic transcription. We plan to use approximate string matching approaches on the obtained data to find the repeating parts and merge both approaches into a robust segmentation algorithm.

Figure 4: The average curve (left) and calculated lengths of stanzas (right)

4. REFERENCES

Bartsch, M. A. & Wakefield, G. H. (2001). To catch a chorus: Using chroma-based representations for audio thumbnailing. In *Applications of Signal Processing to Audio and Acoustics*, (pp. 15–19)., New Platz, NY , USA.

Bohak, C. & Marolt, M. (2012). Finding repeating stanzas in folk songs. In *Proceedings of the 13th International Conference on Music Information Retrieval (ISMIR)*, (pp. 451–456).

Cooper, M. L. & Foote, J. (2002). Automatic music summarization via similarity analysis. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, (pp. 81–85)., Paris, France.

Foote, J. T. (1999). Visualizing music and audio using self-similarity. In *Proceedings of the 7th ACM international conference on Multimedia*, (pp. 77–80).

Foote, J. T. & Cooper, M. L. (2003). Media segmentation using self-similarity decomposition. In *Proceedings of SPIE Storage and Retrieval for Multimedia Databases*, (pp. 167–175).

Goto, M. (2006). A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech & Language Processing*, *14*(5), 1783–94.

Müller, M. (2007). *Information Retrieval for Music and Motion*. Springer Verlag.

Müller, M., Grosche, P., & Jiang, N. (2011). A segment-based fitness measure for capturing repetitive structures of music recordings. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*.

Müller, M., Grosche, P., & Wiering, F. (2009). Robust segmentation and annotation of folk song recordings. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, (pp. 735–740)., Kobe, Japan.

Peeters, G. (2002). Toward automatic music audio summary generation from signal analysis. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, (pp. 94–100)., Paris, France.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

A COMPUTATIONAL STUDY OF CHORUSES IN EARLY DUTCH POPULAR MUSIC

Jan Van Balen, Frans Wiering, Remco Veltkamp
Dept. of Information and Computing Sciences, Utrecht University, the Netherlands
J.M.H.VanBalen, F.Wiering, R.C.Veltkamp@uu.nl

1. INTRODUCTION

Recorded popular music in the Netherlands first appeared in the beginning of the 20th Century. This research studies a particular element of musical form that emerged as popular music came about: the chorus. A case-study dataset of Dutch music from before the Second World War has been assembled, annotated and analyzed using melody extraction and comparison of pitch characteristics between different segment [Bartsch & Wakefield, 2001].

2. CHORUS ANALYSIS

Choruses in Western popular music have been referred to as the 'most prominent, 'most catchy, 'most memorable and even 'most musical parts of songs, [Bartsch & Wakefield, 2001, Eronen & Tampere, 2007, Middleton, 2003]. While agreement on which section in a song constitutes the chorus generally exists, the above attributes are far from understood in music cognition and (cognitive) musicology [Honing, 2010]. On the other hand, as a frequent subject of study in the domain of Music Information Retrieval, the notion of chorus has been shown to correlate with a number of computable descriptors. Yet when studied more closely, the chorus detection systems that locate choruses most successfully turn out to rely on the amount of repetition and energy levels in the signal [Eronen & Tampere, 2007], with more sophisticated systems also taking section length and position within the song into account [Goto, 2003].

The term chorus originates in a denomination for the parts of a musical piece that feature a choir or some other form of group performance, as seen in many folk music traditions. With the early popular song and the development of Tin Pan Alley, Broadway, solo performance became the norm and the chorus became a structural unit of musical form while establishing itself as the site of the more musically distinctive and emotionally affecting material. The same evolution was observed for the analogs in European entertainment [Middleton, 2003]. The motivations to study the particularities of early Western choruses are two-fold: on the one hand, the concept is rather specific to popular music, and may tell us something about where to look for the historical shifts and evolutions that have resulted in the emergence of a new musical style. On the other hand, as choruses can be related to a catchy and/or memorable quality, to the notion of hooks, and perhaps to a general cognitive salience underlying these aspects, the nature of choruses may indicate some of the musical properties that constitute this salience. The choice to focus on early Dutch choruses stems from the conviction that the choice of a regional case-study allows to sample from a more consistent tradition, and because the data were at hand.

The following central question is formulated: are choruses observably different from other song sections, and specifically regarding melody, do choruses feature differences in their pitch structures when compared to other song sections? This question is now studied more closely for the case of early Dutch popular music.

3. DATASET

A dataset has been created as a diverse sample of the Netherlands popular music as it sounded before the 1950s. This Dutch50 dataset contains 50 songs by 50 different artists, all dated between 1905 and 1950. Recurring styles include cabaret, colonial history-related songs, advertisement tunes released on record and early examples of the *levenslied* musical style [Klöters, 1991]. An expert was consulted to judge the representativeness of the selected artists, and approved. Structural annotations were made by the author, indicating beginning and end of sections and labeling each with a section type chosen from a list of seven (intro, verse, chorus, bridge, outro, speech and applause). For all songs, the melody was then extracted using the *Melodia* Vamp plug-in [Salamon & Gomez, 2010]. This algorithm works best when applied to uncompressed audio with a prominent melody, as in our dataset. The resulting pitch contours and pitch salience were segmented along the annotated boundaries. For each section, statistics on the contour could then be computed and compared.

4. RESULTS

The first property of the pitch contours to be considered was the pitch strength, also referred to as the salience function. This salience is a measure of the strength of the fundamental frequency of the melody and its harmonics. For each section, the Average Pitch Strength (hereafter APS) was computed and normalized by subtracting the average pitch strength for the complete song. Figure 1 (left) shows the estimate of the mean APS across all sections in the dataset, with confidence intervals for the mean indicated. Note that 8 songs were not considered as they contained only one type of section, in which case the labeling (verse or chorus or other) was found to be rather arbitrary. The

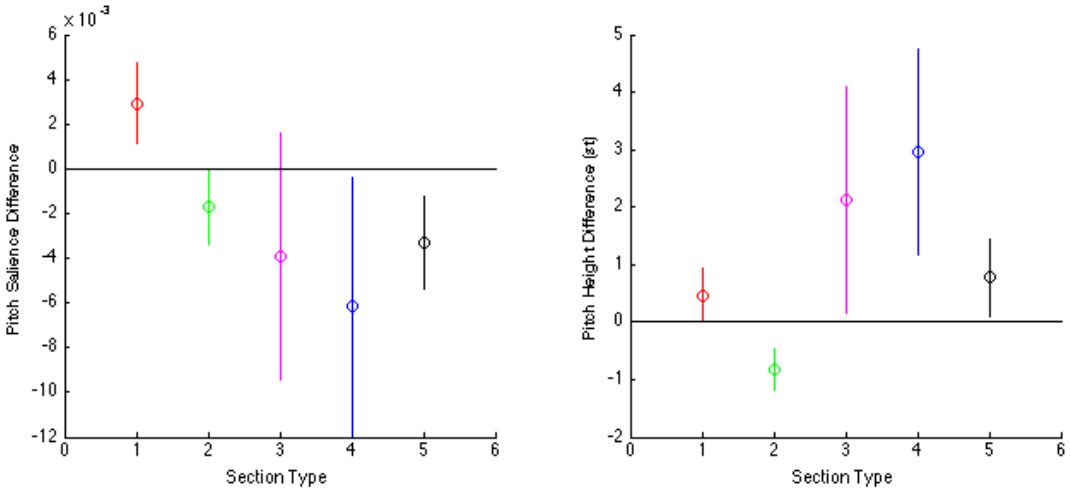


Figure 1: Mean APS and APH per section type. Section types from left to right: chorus (red), verse (green), bridge (pink), other (blue) and 'all non-chorus' (black).

figure illustrates how chorus APS significantly exceed the mean song APS (zero; dashed line) and are demonstrably higher than verse and other APS.

The next property considered is the pitch height. For each section, the Average Pitch Height (APH) was computed and again normalized. Figure 1 (right) shows the estimates of the mean APH. Interestingly, chorus APH are higher than the mean song APH as well as the verse APH (with around 20 semitone cents difference), though not compared to the bridge and other APH., which behave in rather extreme ways.

Another feature that was computed is the average pitch range (APR) With the pitch range as the standard deviation of the pitch height, chorus APR were, on average, higher than verse APR, suggesting a broader pitch range is used in choruses than is in the verse. The tendency does not generalize when chorus sections are compared to all non-chorus sections.

Finally, the average pitch direction (APD) is introduced. This measure aims to capture if the pitch contours in a section follow an up- or downward movement. It is currently computed simply as the difference between the pitch height of the sections end and its beginning. No movement can be shown for choruses, but the APD for verses is greater than zero with $p = 0.0134 < 0.05$, suggesting an upward tendency in pitch during the verse.

5. CONCLUSION

In the present research, a study of melodic pitch yields results that indicate a number of intrinsic musical differences between chorus and verse sections in early Dutch popular music. Given the widely spread discourse of choruses as the most catchy and memorable sections, these results present some reference points for a more elaborate study of cognitive salience in melodies. At the same time, they show the potential of a comparison of pitch structures between the considered genre and its precursors rooted in folk tradition. Future work will also include a similar analysis of post-1950 popular music (cfr. the Billboard dataset) and the design and testing of more detailed contour-based descriptors (cfr. [Salamon et al., 2012]).

6. REFERENCES

Bartsch, M. A. & Wakefield, G. H. (2001). To catch a chorus: Using chroma-based representations for audio thumbnailing. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

Eronen, A. & Tampere, F. (2007). Chorus detection with combined use of mfcc and chroma features and image processing filters. In *Proc. Int. Conf. Digital Audio Effects*.

Goto, M. (2003). A chorus-section detecting method for musical audio signals. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, (pp. 437–440).

Honing, H. (2010). Lure(d) into listening: The potential of cognition-based music information retrieval. *Empirical Musicology Review*, *5*(4).

Klöters, J. (1991). *J. Klöters, Omdat ik zoveel van je hou. Nederlandse chansons en cabaretliederen 1895-1958*. Amsterdam: Nijgh en Van Ditmar.

Middleton, R. (2003). Form. In D. L. J. Shepherd, D. Horn (Ed.), *Continuum Encyclopedia of Popular Music of the World*. Continuum International Publishing Group.

Salamon, J. & Gomez, E. (2010). Melody extraction from polyphonic music signals using pitch contour characteristics. In *IEEE Trans. on Audio, Speech and Language Processing*.

Salamon, J., Rocha, B., & Gomez, E. (2012). Musical genre classification using melody features extracted from polyphonic music signals. In *Int. Conf. Acoustics, Speech and Signal Processing*.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

Comparative description of pitch distribution in Cypriot melodies by analysing polyphonic music recordings

Maria Panteli
Department of Computer Science,
University of Cyprus
m.x.panteli@gmail.com

Hendrik Purwins
Neurotechnology Group, Berlin
Institute of Technology
hpurwins@gmail.com

1. INTRODUCTION

Folk music of Cyprus has a special character built upon a long-term history of cultural exchanges. Particularly emphasized are elements of the Greek and Turkish traditions that have been interacting mostly with the Cypriot culture due to geographical and historical reasons. In the music world, Greek and Turkish characteristics are reflected, amongst other, in Byzantine music (Levy & Troelsgård, 2013) and Turkish music (Feldman, 2013). Musicological studies of Cypriot music in mid 20th century, suggest strong influence from Byzantine music (Tombolis, 1966, Averof, 1978, Zarmas, 1993), whereas influence from Turkish music, is rarely mentioned or in case of (Zarmas, 1993) would only be admitted if scientific methods proved it existed.

The present research studies the characteristics of traditional music of Cyprus and tracks possible similarities with both Turkish and Byzantine music. Since Cypriot music belongs to the orally transmitted and mainly monophonic music traditions (Averof, 1989, Giorgoudes, 2013), the study is restricted to melodic and tonal features. The first objective addressed in this research is the description of pitch distributions in Cypriot folk tunes in order to understand the melodic particularities of this music tradition. This combines analysis of theoretical models of Cypriot music and empirical data extracted from audio recordings with computational methods. The second objective is a comparative study between pitch distributions of Cypriot music with Byzantine and Turkish music. Similarity is investigated in the tuning of the scales, the size of the underlying intervals, and the prominence of scale notes.

The modal system in Byzantine and Turkish music theory, namely *echos* and *makam* respectively, is revised in (Panteli, 2011). The theoretical models considered in this research is the *Chrisanthine* and *Arel* system, the widely recognised models of Byzantine and Turkish music respectively (Mavroeidis, 1999, Gedik & Bozkurt, 2009). Tonal features of Cypriot melodies are compared to echos and makam characteristics of Byzantine and Turkish recordings respectively.

Tools for computation of pitch histograms are revised and algorithms that respect the particularities of the analysed music traditions are proposed. Melodic characteristics of Cypriot music are summarized and similarity between Cypriot and Byzantine/Turkish pitch distributions is investigated in the use of scales, size of intervals and prominence of scale notes. Results contribute to tracking influence of Byzantine and Turkish music traditions to the music culture of Cyprus.

2. MUSIC MATERIAL

The music material gathered for the purpose of this research consists of a total of 210 vocal recordings (polyphonic and monophonic), of which, 74 recordings are religious and folk tunes of Cypriot music, 67 recordings are religious Byzantine music and, 69 recordings are religious Turkish music. Byzantine and Turkish music collections are further categorised by the Byzantine echos and Turkish makam respectively. In this study we consider in total 9 echos and 15 makams that share similarity

Byzantine echos								**Turkish makam**							
First/ First Plagal	10	8	12	12	10	8	12	Huseyni	8	5	9	9	8	5	9
								Ussak	8	5	9	9	4	9	9
								Buselik/ Nihavend	9	4	9	9	4	9	9
Second	8	14	8	12	8	14	8	Hicazkar/ Sedaraban	5	12	5	9	5	12	5
								Hicaz	5	12	5	9	4/8	9/ 5	9
Second Plagal	6	20	4	12	6	20	4	Suzidil	5	13	4	9	5	12	5
Third/ Grave	12	12	6	12	12	12	6	Mahur	9	9	4	9	9	9	5
								Suzinak	9	8	5	9	5	12	5
Grave Papadika	8	12	10	12	8	16	6	Segah	5	9	8	9	5	9/13	8/4
Fourth Heirmoi	8	12	12	10	8	12	10	Kurdilihicazkar	4	9	9	9	4	9	9
Fourth Plagal	12	10	8	12	12	10	8	Rast	9	8	5	9	9	8	5
								Saba	8	5	5	13	4	9	5
								Huzzam	5	9	5	9	5	13	4

Table 1. Echos and makam definition grouped by similarity of the scale step sequence.

in the relative size and sequence of the scale intervals (cf. Table 1). The Cypriot music collection includes, amongst other, folk tunes derived from the genre *Fones*. Fones describes a vocal folk music genre of Cyprus that uses specific melodic models on which many and different verses can be adapted (Giorgoudes, 2013). The use of Fones is often referred to as the most typical tradition in Cypriot music (Averof, 1989, Giorgoudes, 2013) and is essential for addressing the melodic characteristics of Cypriot music.

3. METHODOLOGY

We rely our investigation on pitch class profiles (also called chroma features or pitch histograms) derived from the leading melody of polyphonic music recordings. In order to compute them, we use a state-of-the-art source separation technique (FASST, Flexible Audio Source Separation) (Ozerov et al., 2010), as a pre-processing step to extract the leading melody from the input audio. We then employ the well-known Yin algorithm (de Cheveigné et al., 2002) to estimate the fundamental frequency envelop. The error rate of Yin estimates is reduced via incorporation of post processing filters to eliminate noisy or silent parts of the audio signal and correct octave and fifth errors. For each recording, a pitch histogram is computed from the frequency estimates reduced to the range of an octave. A Gaussian kernel function is employed in histogram computation that better overcomes the artificial discontinuities at the boundaries of the bins (Bishop, 2006). Histogram smoothness is further adjusted in order to eliminate spurious peaks and raise those peaks more relevant to the scale notes of the melody. The bin resolution is set according to the theory of the analysed music tradition; 72-bin resolution is used for Byzantine histograms according to Chrisanthine theory (Mavroeidis, 1999), and 53-bin resolution is used for Turkish histograms according to Arel theory (Bozkurt, 2008). For Cypriot music recordings two types of histograms are computed; 1) 72-bin resolution histograms are computed for comparison with Byzantine histograms, and 2) 53-bin resolution histograms are computed for comparison with Turkish histograms.

Byzantine and Turkish pitch histograms are further aligned to the tonic of the scale. The tonic is computed from the melody contour since as music theory suggests, the tonic of the Byzantine echos and Turkish makam considered in this study, is stated at the end of the musical phrase. The algorithm for detecting the tonic from the last phrase note integrates appropriate assumptions and limitations in order to overcome drawbacks of the method reported in (Bozkurt, 2008). The Cypriot pitch histograms are aligned to the bin of highest correlation with Byzantine and Turkish histograms, estimated by the correlation coefficients method. Note that, for each Cypriot recording, two comparisons are considered; 1) the Cypriot-Byzantine comparison with the correlation coefficients method applied to 72-bin resolution histograms, and 2) the Cypriot-Turkish comparison with 53-bin resolution histograms. The comparison with the highest correlation value also reveals the Byzantine echos or Turkish makam that best describes the scale of the corresponding Cypriot recording. Once histograms are aligned, histogram peaks are assigned a scale degree. Peak locations are used to evaluate the empirical value of scale notes and peak amplitudes are used to evaluate the prominence of scale notes. The distances of consecutive peaks are used to evaluate the size of empirical intervals employed in each music tradition.

4. RESULTS

Regarding the characteristics of Cypriot music, analysis of Cypriot pitch histograms, and specifically of Fones genre, revealed, amongst other, that: a) the pitch range of the melody is usually limited to a perfect fifth or sometimes a major sixth interval, b) Successive melodic steps usually do not exceed a major 3rd, with semitones being used often, and c) at the beginning and the end of the phrase the melody ascends or descends in usually four or five consecutive steps, (a tetra/penta-chordal movement that shares similarity with the construction of Byzantine and Turkish modes). Comparing the use of scales of Cypriot pitch distributions with respect to Byzantine echos and Turkish makam revealed that approximately 70% of Cypriot recordings are more similar to Byzantine echos and 30% to Turkish makam. The size of intervals used in Cypriot music revealed equal influence by both traditions whereas the prominence of scale notes showed more similarity with Byzantine pitch distributions rather than Turkish.

Our observations focus on similarity between folk music of Cyprus and religious Byzantine and Turkish music. Limitations of the proposed method are considered regarding the size of the music corpus, the choice of the Byzantine and Turkish scales, and the different bin resolutions employed for pitch histogram similarity. Improvement of the approach is considered in future work.

5. CONCLUSION

The research addresses a computational comparative analysis of recorded Cypriot music. It proposes algorithms to capture and compare tonal characteristics of Cypriot, Byzantine, and Turkish music. It is left to future work to extend the study to other fundamental music features such as rhythm and timbre, to investigate more similarity measures and to consider the mutual influence between Byzantine and Turkish music tradition in their relation to Cypriot music.

6. ACKNOWLEDGEMENTS

H. P. was supported in part by the German “Bundesministerium für Bildung und Forschung” (BMBF), Grant

BFNT, No. 01GQ0850. M. P. was supported by the grant "Ανθρωπιστικες/Ανθρω/0311(BE)/19" of the Cyprus Research Promotion Foundation. We are grateful to the Music Technology Group, Universitat Pompeu Fabra, Barcleona and in particular to Emilia Gomez and Xavier Serra for supporting this work.

7. REFERENCES

Averof, G. (1989). Ta dimotika tragoudia kai oi laikoi xoroi tis Kyprou [Popular songs and dances of Cyprus]. Nicosia: Bank of Cyprus Cultural Foundation.

Bozkurt, B. (2008). An automatic pitch analysis method for Turkish maqam music. *Journal of New Music Research*, *37(1)*, 1-13.

de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111(4)*, 1917.

Feldman, W. Z. (2013). Ottoman music. *Grove Music Online.* Oxford Music Online. Retrieved May 7, 2013, from http://www.oxfordmusiconline.com/subscriber/article/grove/music/52169

Gedik, A., & Bozkurt, Barıs. (2009). Evaluation of the makam scale theory of Arel for music information retrieval on traditional Turkish art music. *Journal of New Music Research*, *38(2)*, 103-116.

Giorgoudes, P. (2013). Cyprus. *Grove Music Online*. Oxford Music Online. Retrieved May 7, 2013, from http://www.oxfordmusiconline.com/subscriber/article/grove/music/40747.

Levy, K., & Troelsgård, C. (2013). Byzantine chant. *Grove Music Online*. Oxford Music Online. Retrieved May 7, 2013, from http://www.oxfordmusiconline.com/subscriber/article/grove/music/04494.

Mavroeidis, M. (1999). Oi mousikoi tropoi stin Anatoliki Mesogeio [The musical modes in East Mediterranean]. Athens: Fagotto.

Ozerov, A., Vincent, E., & Bimbot, F. (2010). A general flexible framework for the handling of prior information in audio source separation.

Panteli, M. (2011). Pitch patterns of Cypriot folk music between Byzantine and Ottoman influence (Master's Thesis). Pompeu Fabra University, Barcelona, Spain.

Tombolis, S. (2002). Kypriaki paradosiaki mousiki: syllogi, katagrafi, analysi (Tomos I) [Traditional music of Cyprus: collection, annotation, analysis (Part I)]. Nicosia: Bank of Cyprus Cultural Foundation.

Zarmas, P. (1993). Piges tis kypriakis dimotikis mousikis [Sources of folk music of Cyprus]. Nicosia: Kentro Meleton Ieras Monis Kykkou.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

MIR MODEL OF VOCAL TIMBRE IN WORLD'S CULTURES - WHERE DO WE START

Polina Proutskova
Goldsmiths, University of London
proutskova@googlemail.com

The talk will discuss the work in progress on our current study on expressive means of vocal tension in world's singing cultures. It will focus on the methodology that would enable us to explicitly define vocal timbres across research fields: from ethnomusicology via physiology and acoustics to MIR. Delineating vocal timbres by means of objective and measurable categories will allow to express complex relationships between them and to interpret the meaning of positive and negative MIR classification results.

Our current study is concerned with the vocal quality of tension, used by an ethnomusicologist Alan Lomax in his Cantometrics experiment. While vocal tension seems to have been somewhat subjective and not consistently defined, Lomax found that it possessed interesting qualities: in societies where "tense", "narrow" singing was the norm, the subordination of women tended to be higher than in societies where wide, open singing was preferred (Lomax, 1968, 1977).

Our aim in this study is to redefine the Cantometrics' descriptor called vocal tension in more objective, better measurable categories of vocal production that at the same time retain the correlation behaviour discovered in the Cantometrics experiment. In future studies the mapping of the categories onto the low-level signal features will be attempted (for a possible methodology see our preliminary experiment on phonation modes in singing (Proutskova et al., 2013)).

The prevalent tasks concerning vocal timbre that have been addressed by MIR include recognition and separation of timbres, classification or clustering (Berenzweig et al., 2002; Smit, 2011; Fujihara et al., 2010). The most common methodology employs similarity scoring based on MFCCs (Pachet & Aucouturier, 2004; Mesaros & Astola, 2005). This method functions largely like a black box, not providing any means to investigate for whom the results are meaningful and why (Allan et al., 2007).

A very large majority of previous research in vocal timbre has been concerned with Western music, leaving out the numerous other singing traditions, all of which pose specific timbral challenges (Födermayr, 1971; Lomax, 1977).

We would like to approach vocal timbre from a different perspective, which is ethnomusicologically motivated and employs interdisciplinary methods. The direction we take is to explicitly define vocal production in objective, measurable categories and to map these categories onto audio features by means of signal processing and statistical classification. Achieving this would allow us to acquire a much more detailed understanding of vocal timbre, to define more complex relationships than just 'similar - different'. Employing distinct vocal production categories as opposed to MFCCs would allow for precise cognitive investigations of timbre based on responses to each of the categories.

Our approach belongs to the larger class of ideas that address the 'semantic gap' in MIR, in which a middle layer of more objective and better measurable categories is designed (such as scales or musical instrument classes) in order to provide a link between cognitive or social descriptors (such as mood or similarity) and low-level signal features (Wiggins, 2009; Lew et al., 2006).

Timbre in MIR is related to the spectrum of a sound. For a musical instrument its timbre is determined mainly by its physical qualities such as size, form, material, and in performance only one or a small number of parameters are usually varied. Not so with singers: a vocal apparatus is a highly complex system, the form of the vocal tract, the adduction of the vocal folds and the air pressure of the breath are varied constantly (Sundberg, 1987). At the same time we humans are very good at recognising voices by their timbres. Previous research in speaker and singer recognition clearly demonstrates that timbres can be distinguished and detected by MIR methods (Fujihara et al., 2010; Mesaros et al., 2007; Nwe & Li, 2007; Mesaros & Astola, 2005).

There are several expert groups who deal with vocal timbre (Kreiman & Van Lancker Sidtis, 2011). Singing teachers have the deepest, most nuanced knowledge of vocal timbre and the mechanisms to produce its various colours and shades (Miller, 2011; Soto-Morettini, 2006). Health professionals like phoniatricians, speech therapists and vocologists understand the practice of voice production based on physiology (Titze, 1996; Benninger, 2011). Voice scientists make new advances in investigating vocal physiology and acoustics, which still remains a sparsely researched field (Sundberg, 1987; Titze, 2000).

Unfortunately, none of these fields has developed an ontology of timbre that allows for its explicit categorisation. At the same time the experts in the named groups have an extensive tacit knowledge of voice quality (Sofranko & Prosek, 2012). To elicit this knowledge scientifically sound tools are required.

We suggest to employ a qualitative approach which is widely used in anthropology, psychology, sociology, mar-

ket research and other areas (Creswell, 2006; Bryman, 2006). In contrast to quantitative methods it is based on a detailed study of a relatively small group of people followed by a rigorous analysis and reduction of collected data. Our study follows in its methodology knowledge elicitation techniques from artificial intelligence (Cooke, 1999; Ford & Sterman, 1998; Patton, 2005). It will involve semi-structured interviews with vocal timbre experts from various fields and various cultures. The goal of the study is the verification and the adjustment of a preliminary ontology of vocal tension we have developed through previous research.

The talk will include

- The relationship between music and culture - the methodology to revise the Cantometrics experiment based on modern MIR methods
- Our preliminary ontology of vocal tension in world's cultures
- The data collection and analysis guidelines for the qualitative study of vocal tension
- Musical examples that will be used in the study
- Expected outcomes
- Preliminary results

1. REFERENCES

Allan, H., Müllensiefen, D., & Wiggins, G. (2007). Methodological considerations in studies of musical similarity. *Austrian Computer Society*.

Benninger, M. S. (2011). Quality of the voice literature: What is there and what is missing. *Journal of Voice*, *25*(6), 647 – 652.

Berenzweig, A., Ellis, D. P. W., & Lawrence, S. (2002). Using voice segments to improve artist classification of music. *AES 22nd International Conference*.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative research*, *6*(1), 97–113.

Cooke, N. J. (1999). Knowledge elicitation. *Handbook of applied cognition*, 479–510.

Creswell, J. W. (2006). *Qualitative Inquiry And Research Design: Choosing Among Five Approaches Author: John W. Creswell, Publisher: Sage Publica*. Sage Publications, Inc.

Födermayr, F. (1971). *Zu gesanglichen Stimmgenbung in der außereuropäischen Musik. Ein Beitrag zur Methodik der vergleichenden Musikwissenschaft*. Wien: Stieglmayr.

Ford, D. N. & Sterman, J. D. (1998). Expert knowledge elicitation to improve formal and mental models. *System Dynamics Review*, *14*(4), 309–340.

Fujihara, H., Goto, M., Kitahara, T., & Okuno, H. G. (2010). A modeling of singing voice robust to accompaniment sounds and its application to singer identification and vocal-timbre-similarity-based music information retrieval. *Audio, Speech, and Language Processing, IEEE Transactions on*, *18*(3), 638–648.

Kreiman, J. & Van Lancker Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Wiley.

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, *2*(1), 1–19.

Lomax, A. (1968). *Folk Song Style and Culture*. New Brunswick, New Jersey: Transaction Books.

Lomax, A. (1977). *Cantometrics: A Method of Musical Anthropology (audio-cassettes and handbook)*. Berkeley: University of California Media Extension Center.

Mesaros, A. & Astola, J. (2005). The mel-frequency cepstral coefficients in the context of singer identification. In *International Conference on Music Information Retrieval*, (pp. 610–613). Citeseer.

Mesaros, A., Virtanen, T., & Klapuri, A. (2007). Singer identification in polyphonic music using vocal separation and pattern recognition methods. In *Proc. ISMIR*, (pp. 375–378).

Miller, R. (2011). *On the art of singing*. Oxford University Press, USA.

Nwe, T. L. & Li, H. (2007). Exploring vibrato-motivated acoustic features for singer identification. *Audio, Speech, and Language Processing, IEEE Transactions on*, *15*(2), 519–530.

Pachet, F. & Aucouturier, J.-J. (2004). Improving timbre similarity: How high is the sky? *Journal of negative results in speech and audio sciences*, *1*(1), 1–13.

Patton, M. Q. (2005). *Qualitative research*. Wiley Online Library.

Proutskova, P., Rhodes, C., Crawford, T., & Wiggins, G. (2013). Breathy, resonant, pressed - automatic detection of phonation mode from audio recordings of singing. *Journal of New Music Research*.

Smit, C. (2011). *Characterization of the singing voice from polyphonic recordings*. PhD thesis, Columbia University.

Sofranko, J. L. & Prosek, R. A. (2012). The effect of experience on classification of voice quality. *Journal of Voice*, *26*(3), 299 – 303.

Soto-Morettini, D. (2006). *Popular Singing: A Practical Guide To: Pop, Jazz, Blues, Rock, Country and Gospel*. A&C Black.

Sundberg, J. (1987). *The science of the singing voice*. Illinois University Press.

Titze, I. R. (1996). What is vocology? *Logopedics Phonatrics Vocology*, *21*(1), 5–6.

Titze, I. R. (2000). *Principles of voice production*. National Center for Voice and Speech.

Wiggins, G. (2009). Semantic gap?? schemantic schmap!! methodological considerations in the scientific study of music. *Proceedings of IEEE AdMIRe*.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

TIMBRE AND TONAL SIMILARITIES BETWEEN THE TURKISH, WESTERN AND CYPRIOT MONOPHONIC SONGS USING MACHINE LEARNING TECHNIQUES

Andreas Neocleous
University of Groningen
and University of Cyprus
neocleous.andreas@gmail.com

Maria Panteli
University of Cyprus
m.x.panteli@gmail.com

Nicolai Petkov
University of Groningen
n.petkov@rug.nl

Christos N. Schizas
University of Cyprus
schizas@ucy.ac.cy

1. INTRODUCTION

The aim of this work is mainly to explore timbre and tonal similarities between Cypriot folk songs in comparison with Turkish and Western monophonic songs using a computational approach. Several studies exist that identify similarities between the Cypriot folk music and Byzantine music [Kallinikos, T. (1951), Tobolis S. (1980)]. On the other hand, very little information exists on the possible influence of eastern music such as Arabic and Turkish on Cypriot folk music. The main musical instruments that are used in the Cypriot folk music are the lute, violin, a kind of wooden Cypriot flute called "pithkiavli" and a Cypriot percussion instrument called "tamboucha". The existence of the lute and violin in the Cypriot folk music creates a complex combination between Greek and Turkish traditional rhythms. In a previous study [Neocleous et.al (2012)], non equal temperament between notes was identified in some of the Cypriot folk songs, using pitch histograms. The present work is focused on the Cypriot flute "pithkiavli" [Michael, E. (2008).]. This is compared to the Turkish "ney" [http://compmusic.upf.edu/node/55] as well as to western wood instruments such as flute, saxophone, bassoon and oboe.

2. DATA

A database of 127 monophonic songs was used. The 37 songs were Cypriot folk songs performed by Cypriot folk musician Andreas Gristakkos and Giannis Zavros. The 34 Cypriot songs were recorded specifically for the purposes of this research using professional audio equipment. The remaining of the 90 songs were consisted of 47 western songs and solo improvisations, as well as 43 Turkish Makams. The 47 western songs were consisted by songs and solo improvisations with flute, bassoon, clarinet, oboe and saxophone. The songs that were played with flute were 4 movements from partita for solo flute by Bach, 12 fantasias for solo flute by Telemann, the song "Syrinx" of Debussy, the song "Soliloquy for Solo Flute Op. 44" by Lowell Liebermann, the song "Image for solo flute" by Bozza, the song "Danse de la chevre" by Arthur Honegger, the song "Tango Etude" by Piazzolla, the song Daphnis et Chloe by Ravel and nine solo improvisations played with flute. In our library we used also 3 monophonic bassoon solos, 3 pieces for solo clarinet from Stravinsky, two monophonic oboe solos, and eight monophonic saxophone solos. The 43 makam songs were consisted by six makams with 7 Hicaz, 7 Huseiny, 7 Huzzam, 7 Nihavend, 8 Saba, and 7 Ussak.
All the Turkish music and all the western songs were extracted from original audio cd's, while the western monophonic solos were downloaded from youtube.

3. METHODS

3.1 Timbre features

Each song was segmented into approximately 13000 frames of 1024 bins length. For each frame, 16 low-level features were extracted and the mean and the standard deviation of each feature were stored in another database, thus creating a vector of 34 features for each song. The features used were the 1)Zero crossing rate - mean, 2)Zero crossing rate - standard deviation, 3)Spectral centroid - mean, 4)Spectral centroid - standard deviation 5)Roll off - mean, 6)Roll off - standard deviation 7)Entropy - mean, 8)Entropy - standard deviation, 9-21)Mel frequency cepstrum coefficients (13 coefficients) - mean, 22-34)Mel frequency cepstrum coefficients (13 coefficients) - standard deviation.

3.2 Tonal features

For each song, the pitch track was extracted using the YIN algorithm [De Cheveigné, A., and Kawahara, H. (2002)] and pitch histograms with resolution 1200 bins per octave were created using the information from the pitch tracks. The histograms were shifted to the first bin according to the highest peak of the histogram. We assume that the highest peak of the histogram is the note that was most common played in a song and we consider this as the tonic. From the histograms we extracted the location and the amplitude of the 7 higher peaks and we created 14 mid-level tonal features.

3.3 Classification

In a following step, we separated the dataset in-to a "training set" and "validation set". The training set was consisted of 34 western songs and 32 Turkish songs. The validation set was consisted of 13 western, 11 Turkish and 37 Cypriot songs. We built models with supervised learning using artificial neural networks with one hidden layer, K-nearest neighbour with 1-nearest neighbour, and

support vector machines with kernels 1, 2 and 3. We made three experiments in order to understand the tonal and timbre similarities and differences between the Western/Cypriot and Turkish/Cypriot music. In the first experiment we created models using only the timbre features, in the second experiment we created models using only the tonal features and in the third experiment we created models by combining the tonal and timbre features together.

4. RESULTS

The first experiment was built with timbre features and K-nearest neighbours classified correctly 100% of the Western songs and 90% of the Turkish songs (one song misclassified). The classification performance of the rest of the models was around 90% for Western and 92% for Turkish music. Table 1 shows the tabular confusion matrix for the model built with K-nearest neighbours. In the confusion matrix, the validation of the Cypriot songs is included.

Table1: Confusion matrix of the models built with K-nearest neighbours.

confusion matrix		Predicted class	
		Western songs	Turkish songs
Actual class	Western songs	13	0
	Turkish songs	1	10
	Cypriot songs	8	29

The second experiment was built with tonal features. The support vector machines with linear kernel correctly classified 100% of the Turkish songs and 77% of the Western songs. Neural networks classified correctly 54% of the Western songs and 82% of the Turkish songs and K-nearest neighbours 85% and 90% of the Western and Turkish songs accordingly. Table 2 shows the tabular confusion matrix for the model built with support vector machines with linear kernel.

Table 2: Confusion matrix of the models built with support vector machines.

confusion matrix		Predicted class	
		Western songs	Turkish songs
Actual class	Western songs	10	3
	Turkish songs	0	11
	Cypriot songs	15	22

The third experiment consisted with both tonal and timbre features. Neural networks and Support vector machines with kernel 3 classified correctly 100% both the Western and Turkish songs while the K-nearest neighbours misclassified one Turkish song. Table 3 shows the tabular confusion matrix for the model built with artificial neural networks.

Table 3: Confusion matrix of the models built with artificial neural networks.

confusion matrix		Predicted class	
		Western songs	Turkish songs
Actual class	Western songs	13	0
	Turkish songs	0	11
	Cypriot songs	30	7

5. CONCLUSIONS

From the results, a straight forward observation regarding the discrimination of the Western music from the Turkish music is that only the models used both timbre and tonal features were able to absolutely discriminate them. The models built only with timbre features, showed that 78% of the Cypriot songs were classified as Turkish songs while the rest 22% of the Cypriot songs were classified as Western songs. The models built only with tonal features were not able to completely discriminate the Western music from the Turkish music and these models were classifying the 60% of the Cypriot music as Turkish songs and the other 40% as Western songs. The models built with tonal and timbre features were able to completely discriminate the Western songs from the Turkish songs and these models classified 81% of the Cypriot music as Western music and 19% of the Cypriot music as Turkish music.

6. ACKNOWLEDGMENTS

This work is supported by the research grant Ανθρωπιστικες/Ανθρω/0311(BE)/19 funder by the Cyprus Research Promotion Foundation.

7. REFERENCES

Kallinikos, T. (1951). Kypriaki laiki mousa.

Tobolis S. (1980). Traditional Cyprian songs and dances.

Neocleous, A., Panteli, M., Petkov, N., Schizas, C. (2012). Identification of similarities between the Turkish Makam scales and the Cypriot folk music. HELINA's 5th National Conference.

Michael, E. (2008). The pithkiavli flute in Cyprus. Master thesis, Free University of Berlin.

http://compmusic.upf.edu/node/55.

De Cheveigné, A., and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America, 111(4), 1917.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

TOWARDS A COMPREHENSIVE AND MODULAR FRAMEWORK FOR MUSIC TRANSCRIPTION AND ANALYSIS

Olivier Lartillot
Finnish Center of Excellence in
Interdisciplinary Music Research
olartillot@gmail.com

Mondher Ayari
University of Strasbourg
IRCAM/CNRS
ayari.mondher@gmail.com

ABSTRACT

Computational tools for musicology (and for folk music analysis in particular) usually restrict on particular aspects of music analysis. For instance, automated transcription does not generally deal with microtonality, or computational modal analysis rarely addresses the detection of modal transitions.

We introduce a computational framework for music transcription and analysis composed of a set of specialized modules that form complex interdependencies. Music recordings are progressively analyzed throughout the whole set of modules in a bottom-up fashion. One objective of the framework is that modules dedicated to low-level operations will take benefit from higher-level information already inferred from other modules while analyzing the previous instants, as well as predefined cultural knowledge.

A purely bottom-up analysis through the framework can be briefly described as follows:

1. PITCH CURVE

Pitch is extracted from audio through the computation of autocorrelation functions on a moving window. This transcription is performed using the default configuration of the *mirpitch* operator in *MIRtoolbox* (Lartillot & Toiviainen, 2007), which includes refinements such as filterbank decomposition and enhanced autocorrelation computation (Lartillot, 2012). The method theoretically allows the transcription of polyphonies, but in a first step, the study restricts on simple monodies.

2. PITCH-BASED NOTE SEGMENTATION

The pitch curve is segmented into a succession of stable parts (often corresponding to notes), separated by either silence, absence of pitch content, continuous transition between stable pitch levels, or longer unstable parts. In a purely bottom-up configuration of this module, segmentation is purely based on such local discontinuities, such that each stable pitch can be characterized by an average pitch level (for instance in Hz.) that is independent on any scale and in particular is not constrained by chromatic scale discretization.

3. PITCH SCALE DISCRETISATION

Pitch levels, initially expressed on a continuous axis, are discretized through the progressive bottom-up constitution of a pitch scale. Each new pitch level is successively integrated into the scale under constitution: either the new pitch level is clustered into an existing scale level, or a new pitch level is added to the scale. Mechanisms are also proposed to track the possible progressive drift in pitch of the whole pitch scale.

4. CURRENT PROJECT

4.1 Modal analysis

Various sets of pitch levels are formed and are associated with activation score that varies across time. These sets of pitches model different musicological concepts such as chords, modal scales (such as diatonic scale, as opposed to chromatic scale) and modal subscales (such as chord-degrees in western classical music or genres in Maqam music) (Lartillot & Ayari, 2011).

4.2 Pattern analysis

All configurations (chord classes, subscales, pitch level within modal scales, etc.) are stored in associative memory, implemented as reverse tables. In this way, any repetition of a same configuration is detected; consequently, the configuration enters a dictionary of repeated patterns, and all subsequent repetitions of the configuration will be simply identified as occurrences of that pattern.

Succession of configurations (notes, chords, scales, etc.) are also stored in associative memory. In particular, a succession of two notes is stored as an interval with underlying pitch interval and temporal (inter-onset) interval. This allows the detection of repetition of successions, leading to the inference of sequential patterns made of two notes, two chords, etc. This can be extended recursively to the detection of sequential patterns of any length. Successive occurrences of a same pattern form a cyclic pattern (Lartillot, 2005).

This formalization of cyclic pattern enables to model the emergence of rhythm and meter. The start of a regular pulsation is immediately detected as a cyclic pattern composed of a single temporal interval. Each subsequent pulse is simply identified as a new extension of the cyclic pattern. The multi-level metrical structure can also be progressively constructed as a set of cyclic patterns that are interconnected (Lartillot, 2010).

Other musicological concepts can be formalized based on patterns, such as motives, themes, forms, etc.

4.3 Top-down influence

The modularity of the model enables the formalization of top-down influences. Some examples are listed below:

- The detection of particular notes, and in particular of their temporal location, duration and pitch level, can be guided by expectation driven by knowledge related to the underlying pitch scale, modal analysis, metrical structure, motivic development, etc.
- Transformation of pitch scales and pitch sets can be anticipated and tracked based on an analogy driven by particular motivic pattern developments: If a transposition of a known motivic pattern is detected, any modification of the current modal scale implied by the transposition will be anticipated and thus more easily detected, and transposed modal subscales will be inferred as well.
- Modal analysis is influenced by metrical analysis and vice-versa: for instance, metrical pulse can guide the detection of modal transitions, and modal transitions outside the expected metrical pulse might indicate a modification of the metrical structure, etc.
- Configurations collected from analyses of other pieces can guide the analysis of a new piece. Cultural knowledge can also be predefined as well, in order to test its impact on the analysis.

5. DISCUSSION

In folk music analysis, the resulting computational tool, under development, allows the automation of transcription with or without guidance on cultural knowledge such as predefined pitch scale or rhythmic structure. Subtle modal and rhythmical structures can be therefore discovered that can question the musicologists initial hypotheses. For instance, a transcription can detect whether the musician uses subtle microtonalities, and whether there is any interdependency with the motivic logic of the improvisation.

The model will be made available in a new Matlab framework for audio and music analysis called *The MiningSuite* (Lartillot, 2011). The second step in the bottom-up description above (pitch curve segmentation) have also been integrated in *MIRtoolbox* 1.5.

6. REFERENCES

Lartillot, O. (2005). Multi-dimensional motivic pattern extraction founded on adaptive redundancy filtering. *Journal of New Music Research*, *34*(4), 375–393.

Lartillot, O. (2010). Reflexions towards a generative theory of musical parallelism. *Musicae Scientiae*, *Discussion Forum 5*, 195–229.

Lartillot, O. (2011). A comprehensive and modular framework for audio content extraction, aimed at research, pedagogy, and digital library management. In *130th Audio Engineering Society Convention*.

Lartillot, O. (2012). Computational analysis of maqam music: from audio transcription to structural and stylistic analyses, everything is tightly intertwined. In *Acoustics 2012 Hong Kong Conference and Exhibition*.

Lartillot, O. & Ayari, M. (2011). Cultural impact in listeners' structural understanding of a tunisian traditional modal improvisation, studied with the help of computational models. *Journal of Interdisciplinary Music Studies*, *5*(1), 85–100.

Lartillot, O. & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

FOLK TUNE CLASSIFICATION WITH MULTIPLE VIEWPOINTS

Darrell Conklin
Department of Computer Science and Artificial Intelligence
Universidad del País Vasco UPV/EHU, San Sebastián, Spain, and
IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Extended abstract

Folk tune classification is the inference of tune properties such as geographic location, tune family, tonality, meter, and genre directly from the musical content of folk songs. The standard way to approach this problem is by machine learning: by training and evaluating a classifier on a set of labelled pieces. Trained classifiers can subsequently be used for labelling of unlabelled pieces, for suggesting missing labels, or for identifying possible labelling errors in a large digital collection.

Machine learning approaches to symbolic folk song classification range widely based on the type of representation chosen for the pieces, and the specific machine learning method applied (Hillewaere et al., 2009). Recently Conklin (2013) presented a method for music classification which is based on the theory of *multiple viewpoint systems* (Conklin & Witten, 1995), which are ensembles of multiclass n-gram models, each one promoting a different abstraction of the music surface. Multiple viewpoint systems were shown to be highly effective for several folk tune classification tasks (Conklin, 2013).

In addition to the four datasets considered previously (Conklin, 2013: European national and Basque regional folk tunes, partitioned into genres, and regions) this presentation will illustrate the performance of multiple viewpoints on several further corpora including Swiss and Austrian folk tunes derived from the Essen folksong collection (Schaffrath, 1995) (this dataset was also used in an earlier classification study by Conklin, 2009); songs from the Finnish folk song database (Eerola & Toiviainen, 2004) partitioned in various ways; and Dutch folk songs grouped into tune families (Volk & van Kranenburg, 2012). The corpora and their partitioning were selected and designed to ensure a wide range of intra-group melodic divergence.

On each dataset a multiple viewpoint system will be compared with two other classifiers and representations: a 1-NN classifier using the Levenshtein distance of pairs of melodic interval strings; and a classifier based on a large set of global features (McKay & Fujinaga, 2006). Supporting the conclusions of Conklin (2013), multiple viewpoint systems perform well relative to other methods on a wide range of folk song classification tasks.

References

Conklin, D. (2009). Melody classification using patterns. In *MML 2009: International Workshop on Machine Learning and Music*, (pp. 37–41), Bled, Slovenia.

Conklin, D. (2013). Multiple viewpoint systems for music classification. *Journal of New Music Research*, *42*(1), 19–26.

Conklin, D. & Witten, I. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73.

Eerola, T. & Toiviainen, P. (2004). Suomen Kansan eSävelmät. Finnish Folk Song Database. [11.3.2004]. Available: http://www.jyu.fi/musica/sks/.

Hillewaere, R., Manderick, B., & Conklin, D. (2009). Global feature versus event models for folk song classification. In *ISMIR 2009: 10th International Society for Music Information Retrieval Conference*, (pp. 729–733), Kobe, Japan.

McKay, C. & Fujinaga, I. (2006). jSymbolic: A feature extractor for MIDI files. In *Proceedings of the International Computer Music Conference*, (pp. 302–305), New Orleans.

Schaffrath, H. (1995). The Essen folksong collection. In D. Huron (Ed.), *Database containing 6,255 folksong transcriptions in the Kern format and a 34-page research guide*. Menlo Park, CA: CCARH.

Volk, A. & van Kranenburg, P. (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, *16*(3), 317–339.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

SOME QUANTITATIVE INDEXES IN THE STUDY OF TRADITIONAL MUSICAL SCALES AND THEIR GENESIS

Rytis Ambrazevičius
Kaunas University of Technology, Kaunas, Lithuania
rytisam@delfi.lt

1. INTRODUCTION

1.1 On the Alexeyev's Theory

Rationale of the current research rests on the theories of historical development of musical scales. Here I will refer to the theory proposed by Russian ethnomusicologist Eduard Alexeyev (1986), although similar and well-known concepts are reflected by different authors (Sachs, 1962; Levman, 1992; Brown, 2000; etc.). Alexeyev considers, inter alia, several stages in the phylogenesis of musical scales. After some initial stages, he arrives at the so-called 'gamma-intonation' characterized by already crystalized pitch categories (≈ scale degrees), yet they are not subordinated but rather coordinated. It means that tonal hierarchies are not yet clearly fixed, and the pitches are 'wandering', i.e., intonation is far from constant across performance. Additionally, the constituent intervals (between the neighboring pitch categories) are very roughly equal. This stage is succeeded by the so-called 't-intonation' which corresponds roughly to the present type of modal thinking. It means, the tonal hierarchies are finally fixed, the functions and weights of scale degrees are clearly divergent, and the intonation is relatively stable across performance. The scale is no longer equidistant.

To summarize, according to the theory, the musical thinking experienced several interrelated phylogenetic changes: 1) the divergence of the weights of scale degrees, 2) the asymmetrization of the intervals (divergence of the sizes of constituent intervals), and 3) the acoustical stabilizing of intonation.

1.2 Aim of the Study

The aim of the current study is to test the interrelations just mentioned based on the measurements of a set of traditional vocal performances. To reach this aim, the corresponding three quantitative indexes are proposed and applied.

2. THE QUANTITATIVE INDEXES

2.1 Diatonic Contrast

Actually, the very fact of deviation of the scales from 12TET and the obvious traces of the 'equidistant' scales in the Lithuanian traditional music were already noted in a number of earlier studies (Ambrazevičius, 2005–2006 and later). For the generalization of interval asymmetries, the index of diatonic contrast (DC) was introduced (Ambrazevičius, 2006). The more clearly the constituent intervals of a musical scale cluster into two groups ('small' and 'large' intervals), the larger DC is. The expression for the index of DC is normalized so that exactly equidistant scale gives DC = 0 and diatonic versions of 12TET give DC = 1. For Pythagorean tuning, DC = 1.26, DC = 0.5 means exactly medial case between ideal equitonics and 12TET-diatonics, and so on.[1]

2.2 Modal Contrast

The index of modal contrast (MC) stands for the divergence in the weights of scale degrees (the modal weights). Different strategies could be chosen for the evaluation of modal weights. For instance, the tonal profiles resulting from the well-known probe-tone experiments (Krumhansl, 1990) could be applied. For the meantime, I used simple summation of several factors such as rhythm value of the pitch and its weight in the hierarchical tree structures of metrorhythm and cadences. The MC was then defined as the measure of contrast of the modal weights of individual scale degrees.[2]

2.3 Acoustical Stability of Intonation

The stability of intonation of a certain scale degree is defined as the reciprocal of the standard deviation of pitch across the occurrences in the performance. Then the individual stabilities can be averaged to get the general stability of intonation (SI). Or, instead, the averaged standard deviation ($\bar{s}$) can be employed.

3. SAMPLES AND PROCEDURE

Three samples of sound recordings of the Lithuanian traditional monophonic vocal performances were compiled. The first one (denoted as S) contains 25 songs recorded in 1930s, in Suvalkija (Southwestern Lithuania); from different male and female singers. It represents ethnomusical dialect of the region. The second and third samples (PZ and JJ) contain, respectively, 29 and 26 songs recorded in the end of 20th century (1970s–1990s), in Dzūkija (Southern Lithuania), from the prominent male singers Petras Zalanskas and Jonas Jakubauskas. Thus these samples represent two typical ethnomusical idiolects of the region.

[1] Concerning the mathematical procedure of DC calculation, see our other paper at FMA2013 (Ambrazevičius & Budrys, „Traces of Equidistant Scale in Lithuanian Traditional Songs").

[2] The mathematical procedure of MC calculation is detailed in Ambrazevičius, 2008: 165-166, and will be presented in more detail at FMA2013.

The structural pitches of the songs were measured. Praat software was applied. Then the scales of all songs were calculated based on the averages of occurrences of the scale degrees.

4. RESULTS

A positive correlation between DC and MC was found. The values of Pearson's *r* equal .42 (S), .43 (PZ), and .47 (JJ). The correlations between SI and DC, and between SI and MC (or between $\bar{s}$ and DC, $\bar{s}$ and MC), are not that distinct. More exactly, they differ significantly for the three samples. The two idiolects from Dzūkija show negative values of the correlation coefficients: $r(\bar{s}; \text{DC}) = -.57$; $r(\bar{s}; \text{MC}) = -.53$ (PZ); and $r(\bar{s}; \text{DC}) = -.39$; $r(\bar{s}; \text{MC}) = -.21$ (JJ). However, for the sample S, $r(\bar{s}; \text{DC}) = -.07$; $r(\bar{s}; \text{MC}) = +.09$ (JJ).

5. DISCUSSION

The positive correlation between the diatonic and modal contrasts supports the theoretical presumption that the strengthening of modal functions of scale degrees is in step with asymmetrization of equidistant scale, i.e., with the formation of diatonics. Importantly, the phenomenon is observed for geographically and historically not distant musical materials. For instance, here it was registered for individual vocal idiolects. Possibly, this manifests "multimusicality", in terms of modal thinking, and/or some relation of physiology/psychology of intonation to the modal features (in terms of MC) of the piece performed.

The general trends of the interrelations between the stability of intonation and the diatonic and modal contrasts (the prevailing negative correlations between $\bar{s}$ and DC, MC) also seemingly supports the theoretical presumption that the development of mode (in terms of DC and MC) is accompanied by the acoustical stabilizing of intonation. However, the significant scatter of *r* values including even some small and "illogical" positive numbers, shows that there are robust additional phenomena at work. For instance, naturally, the perceived stability could be considered as being not of purely acoustical origin, but also modified by certain perceptual phenomena. One can refer to the observations of intonation of the tonal center in the traditional vocal performance: quite surprisingly, the intonation of the perceptually "most stable" tonal center (or, more exactly, the lower tonic) is frequently quite loose compared to the intonation of other scale degrees (Alexeyev, 1986: 67, etc.).

6. REFERENCES

Alexeyev, E. (1986). *Rannefol'klornoe intonirovanie. Zvukovysotnyj aspekt*. Moscow: Sovetskij kompozitor.

Ambrazevičius, R. (2005–2006). Modelling of Scales in Traditional Solo Singing. *Musicae Scientiae*, Special Issue. Interdisciplinary musicology, 65–87.

Ambrazevičius, R. (2006). Pseudo-Greek Modes in Traditional Music as Result of Misperception. In *ICMPC9. Proceedings of the 9th International Conference on Music Perception and Cognition. 6th Triennial Conference of the European Society for the Cognitive Sciences of Music. Alma Mater Studiorum University of Bologna, Italy, August 22–26* (pp. 1817–1822). Bologna: Bononia University Press.

Ambrazevičius, R. (2008). *Psichologiniai muzikinės darnos aspektai*. Kaunas: Technologija.

Brown, S. (2000). The 'Musilanguage' Model of Music Evolution. In N. Wallin, B. Merker, and S. Brown, eds. *The Origins of Music* (pp. 271–300). Cambridge: MIT Press.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Levman, B. G. (1992). The Genesis of Music and Language. *Ethnomusicology*, *36*(2), 147–170.

Sachs, C. (1962). [J. Kunst, ed.] *The Wellsprings of Music*. The Hague: Marinus Nijhoff.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

ON COMPUTATIONAL MODELING IN ETHNOMUSICOLOGICAL RESEARCH: BEYOND THE TOOL

Peter van Kranenburg
Meertens Insitute, Amsterdam
peter.van.kranenburg@meertens.knaw.nl

1. INTRODUCTION

During the last decade, a rapidly increasing amount of work has been done in the area of *Digital Humanities*. Conferences were established. Research programs were launched. Departments and institutes were founded. Although there is still much confusion and debate about the essence of the field, some recurring themes can be observed. It seems that much of the current effort goes into creating research environments and infrastructures combining large interoperable, harmonized data sets and user friendly search and visualization tools that enable humanities researchers to search and explore the data and make new, previously unimaginable, discoveries.

Among these humanities data sets, there are musical data as well. Ethnomusicological archives have been digitized. Many scores of important composers are currently available in various digital formats. Massive amounts of user tags from services such as Last.fm are available. This enables data-rich research on music on a large scale.

The question is how to extract new knowledge from this data. One approach is to use generic search and visualization *tools*. Such tools enable discoveries in the data that also could have been done 'by hand', but would take a lot of time. For example, finding all occurrences of the name 'Joachim' in Brahms' letters, or finding all occurrences of the Landini cadence in the compositions of Gilles Binchois. This kind of automatic retrieval is of course very important, since it can save a tremendous amount of time. However, the resulting knowledge is typically not of computational nature. After being used, the computational tool is put away. One could call this computer-aided, or computer-assisted research.

Notwithstanding the importance of such tools, there is a next level of integration of computational methods and musicological research. It is on this level that the current contribution focuses. The core of this approach is to construct computational *models*, or rather to perform *computational modeling*. In contrary to the tool-scenario, the resulting knowledge is expressed in computational terms.

On a theoretical level, methodology for this approach has been proposed by Willard McCarty in his 2005 book *Humanities Computing*. There are, however, not much practitioners already following these research methods. In this contribution, I will take McCarty's rather abstract approach as point of departure and present some concretizations for doing computational ethnomusicology.

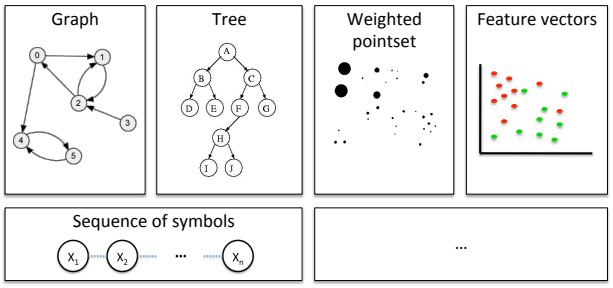


Figure 1: Examples of abstract data structures.

2. ABSTRACT DATA STRUCTURES AND ALGORITHMS

Computer Science provides numerous abstract data structures and algorithms that operate on these data structures. Examples of data structures are sets, trees, graphs, weighted point sets, sequences of symbols, vectors of feature values, etc. Some are depicted in Figure 1. Examples of algorithms are classification algorithms, Bayesian inference algorithms, alignment algorithms, sorting algorithms, etc., etc. A very important property of many of these data structures and algorithms is that they are abstract. E.g., symbols in a sequence could represent anything: characters in words, words in sentences, notes in melodies, chords in hymns, etc. Exactly this property enables the inclusion of domain knowledge.

3. COMPUTATIONAL MODELING

Computational modeling of musical knowledge involves expressing a musical problem in terms of abstract data structures and algorithms. It is the creativity of the researcher to find or design appropriate data structures and algorithms for the musical problem at hand. The better the musicological problem can be expressed in terms of data structures and algorithms, the more relevant the results are from a musical point of view. As foreseen by Leonard Meyer (1996): "I have no doubt about the value of employing computers in such studies, not merely because they can save enormous amounts of time but, equally important, because their use will force us to define terms and traits, classes and relationships with precision – something most of us seldom do."

The general research cycle is as follows (Van Kranenburg et al., 2011):

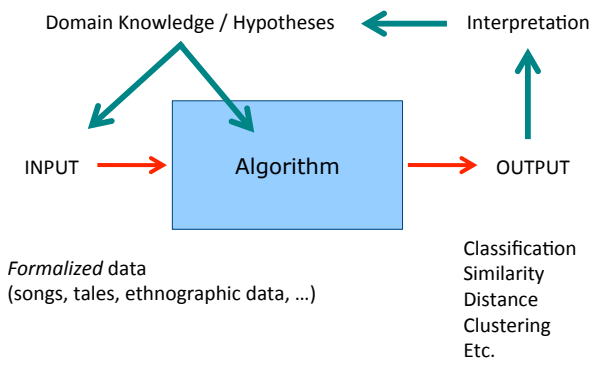


Figure 2: Research cycle for computational modeling.

1. Understanding the musical problem, which involves studying relevant musicological literature, especially the specific discourse concerning the research question at hand;
2. Designing musically meaningful data-structures and algorithms: the computational model as hypothesis;
3. Interpreting the algorithmic output;
4. Revising the model in case of failure;
5. Integrating the results in the musicological discourse.

Especially steps 1 and 5 are often absent in studies that can be found in the research area of Music Information Retrieval. Such studies are less relevant from the perspective of musicology.

The data structures and algorithms that are developed in the second step can be considered hypotheses. They reflect the current understanding of the musical phenomenon in a formalized way. The formalization of data is a research topic in itself. A computational data structure is a model of – in our case – musical data. In the third step, the hypotheses are 'tested' by interpreting the algorithmic output. The fourth step is a key idea of McCarty's approach: knowledge gain is possible in cases where the model fails. Therefore, these cases offer opportunities to improve the model in the next iteration of the modeling.

The third step poses us for serious problems. The question is: what to compare the output of an algorithm with? For relatively simple questions it might be possible to collect a set of examples that can be used as reference set or *ground-truth*. This is common practice in fields such as Music Information Retrieval, in which ground-truth data is collected by inquiring musicological experts or by crowdsourcing. These data are considered the intended output of the computational model and the model is evaluated in terms of its ability to *reproduce* the ground-truth. From a musicological point of view, a pitfall of this research method is to mainly focus on accuracy rates, resulting in algorithms that might be able to reproduce ground-truth, but do not reveal any knowledge about the involved musical phenomena.

In most cases, however, mere collecting of proper ground-truth data is already problematic. Reasons for this include the multi-dimensional character of music and musical phenomena, lack of knowledge about the subject of study, or differences in opinion in musicological discourse.

One approach to avoid these problems is to involve the construction of ground-truth data in the research cycle, and by making the assumptions behind the ground-truth explicit and questionable. Then, not only the algorithm will be subject to revision in each iteration of the modeling, but also the reference or ground-truth data. One step further would be to employ the algorithmic output to explore and explain the properties of empirical data.

4. TOOLS AND MODELS

A tool is supposed to function properly. Therefore, tool-building complements modeling, in which, on the contrary, failure is the main focus of interest. In each iteration of the process of modeling, the model reflects the current understanding of a musical phenomenon in computational terms. Construction of a tool based on that model necessarily inherits the limitations of the model. In many cases, it is important to understand the underlying model to make proper use of the tool.

5. MUSIC

Some examples of ethnomusicological research that can be addressed by taking a computational approach are: studying the geographical differentiation of recitation on a large scale, classification of folk song melodies, modeling oral variation, finding common rhythmic or pitch patterns in large bodies of music, and ultimately, testing hypotheses on human musicality as such.

6. FUTURE WORK

The full potential of computational modeling has by far not yet been realized. On the longer term, computational modeling could build a body of explicitly defined (ethno)musicological knowledge, which would enrich and complement traditional approaches.

7. REFERENCES

McCarty, W. (2005). *Humanities Computing*. Basingstoke: Palgrave Macmillan.

Meyer, L. (1996). *Style and Music—Theory, History, and Ideology*. Chicago: University of Chicago Press.

Van Kranenburg, P., Wiering, F., & Volk, A. (2011). On operationalizing the musicological concept of tune family for computational modeling. In Maegaard, B. (Ed.), *Proceedings of Supporting Digital Humanities (SDH2011)*, Copenhagen.

Proceedings of the Third International Workshop on Folk Music Analysis, June 6-7, 2013. Amsterdam, Netherlands. P. van Kranenburg, C. Anagnostopoulou, and A. Volk (Eds.).
Amsterdam: Meertens Institute, Utrecht: Department of Information and Computing Sciences, Utrecht University.

ANALYSIS OF "POLISH RHYTHMS"

Ewa Dahlig-Turek
Polish Academy of Sciences
Institute of Art, Warsaw
Ewa.Dahlig-Turek1@ispan.pl

1. PROBLEM

The paper presents a method of studying morphology of so-called "Polish rhythms". This term stands for triple-time rhythmic structures of decreasing rhythm condensation within a measure. They can be as simple as *iambus*  and *ionicus a minore*  (so-called "mazurka formula"), and as complex as in polonaise ().

Such rhythms may and do appear in many repertoires, but nowhere is their significance as striking as in Polish (not only music) culture in which they function as a crucial identity factor, a true symbol of Polishness. For example, Polish national anthem is built almost exclusively of such rhythms. However, they were also widely present in the European art music, from the "Polish dance" (*chorea polonica, Polnischer Tanz, baletto polacco, danza polacca*) of the 16th-17th century till the 19th century mazurka and polonaise.

The most spectacular export of "Polish rhythms" took place to Sweden where so-called "Polish dance" (*polska*) significantly contributed to the creation of the local musical idiom. "Polish dances" and "Polish rhythms" became an object of interest for Polish and Scandinavian (Swedish, Finnish, Danish) music historians focused on historical aspect – e.g., Norlind (1911), Hławiczka (1967) - as well as ethnomusicologists interested mostly in origins and morphology of these rhythmic structures – e.g. Ala-Könni (1956), Aksdal (2003). However, approaches applied in both subdisciplines of musicology were rather descriptive than strictly analytical. To reconstruct the historical development of the analyzed music phenomenon, a quantitative method seemed the best solution.

2. MATERIAL

The proposed method of analysis was developed to help reveal changes in morphology of "Polish Rhythms" throughout the centuries, in all kinds of music repertoire (Dahlig-Turek, 2006). It was applied to 791 musical pieces: from the 16th lute and organ tablatures till 19th century (Chopin's mazurkas and polonaises), with a few examples from the early 20th century, as well as selected folk dance-tunes notations from the 19th - 20th century.

All the material was taken from printed collections edited by specialists, thus freeing the author-ethnomusicologist from possible consequences of her incompetence in transcribing original sources.

3. METHOD

The applied method is a "do-it-yourself" solution which does not require any particular experience in computer-aided research, as it does not go beyond the functionalities offered by MSOffice (Word and Excel).

Rhythm of analyzed music pieces has been recorded in the form of a digital code in which rhythmic organisation of metric unit was the basic portion of information. The main body of the "alphabet" contains nine items:

00 
01 
02 
03 
04 
05 
06 
07 
08 

with possible extension of variants, e.g. a3 ()

Rhythmic groups extending over more than one beat were substituted by a combination of the above units or their variants.

Sequences of numbers replaced the original rhythm notation, thus changing notes into a quasi-text, e.g.:

Rhythmic formula	Traditional notation	Encoding
iambus		010100
ionicus a minore (mazurka formula)		020101
polonaise formula		060202

Encoding was the most time-consuming part of the procedure, as it seems hardly possible to automatize the process of conversion.

In the first stage of the analysis, pertinent calculations were made to define two main rhythm parameters necessary to characterize "Polish rhythms":

1. **average density**, i.e. the average number of impulses corresponding to a metric unit;
2. **descendentality**, i. e. the difference between rhythm density in the first and last beat of a measure.

For this purpose, the encoded text was converted into a sequence of numbers indicating the number of impulses on each beat, e.g.:

Encoded measure	Number of impulses per beat		
	1st beat	2nd beat	3rd beat
010100	1	1	0
020101	2	1	1
060202	3	2	2

Presented in diagrams for music from different periods, these two parameters differentiate very well the examined repertoire and perfectly co-harmonise with its chronology. In brief:

- parameter of descendentality is crucial to distinguish "Polish" rhythms from others (e.g. in old German tablatures in which so-called "Polnische Tänze" have a positive descendality while other music pieces – negative);
- the average density of rhythm differentiates historical strata of analyzed music: it clearly and consequently grows through centuries, as rhythms more complicated than *iambus* (first mazurka formula, then polonaise formula with their variants) emerged.

For the second stage of analysis, a classification of triple-time rhythms has been developed as a modified version of Bielawski's proposal (1970). It departs from the obvious premise that a rhythmic formula within one bar contain in each beat: no impulse (0), one impulse (1), two impulses (2), three or more impulses (R). Additional auxiliary symbols were introduced for syncope (S) and punctuated rhythm extending over two metric units (Pp). Combinations of these symbols allowed the reduction of all the possible bar-structures to fewer types - as in the following example in which six different rhythms represent only three types:

Encoding	Rhythm type
010100	110
020101	211
040101	211
060202	R22
070402	R22
080202	R22

The classification was then used in matrices presenting frequency of all the newly defined types of rhythmic figures in the whole analyzed repertoire.

Placing all the music pieces in a common matrix in chronological order, one can follow in details the changes in rhythmic figures piece by piece, period by period. Thus, rhythmic matrices give the precise view of the rhythmic properties of compositions over centuries.

For more spectacular results, the numbers indicating frequency of particular rhythms were substituted by graphic symbols of different intensity.

4. RESULTS

The applied method made it possible to base the discussion on "Polish rhythms" on quantitative ("objective") data instead of descriptive, often emotional ("subjective") statements typical of many previous studies. It helped verify some of the former hypotheses relating to the periodisation of old repertoire.

The method allows detailed examination of the changes in rhythm – either historically, between repertoires of different epochs, or geographically, between regions and countries. Applied to the music from different centuries, it revealed the beginnings, peaks and declines of specific rhythmic structures. Its application to the ethnomusicological sources helps understand relations between folk music and art/popular music of different periods.

Recently the method will be used to study the Scandinavian phenomenon of *polska* dances and their rhythmic relation to Polish folk music.

5. REFERENCES

Aksdal, B. (2003). Polish Dance with Walking and Jumping Dance. A Historical Perspective on the Pols Music in Norway up to the Late 1800s. In *The Polish Dance in Scandinavia and Poland*. *Skrifter utgivna av Senskt visarkiv 17*, 53-75.

Ala-Könni, E. (1956). Die Polska-Tänze in Finnland. Eine ethno-musikologische Untersuchung. Helsingfors.

Bielawski, L. (1970). Rytmika polskich pieśni ludowych. Kraków.

Dahlig-Turek, E. (2006). Rytmy polskie w muzyce XVI-XIX wieku. Studium morfologiczne. Warszawa: ISPAN.

Hławiczka, K. (1968). Grundriss einer Geschichte der Polonaise bis zum Anfang des 19. Jahrhunderts. „Svensk tidskrift för musikforsking" L, 51-124.

Norlind, T. (1911). Zur Geschichte der polnischen Tänze. *Sammelbände der Internationalen Musikgesellschaft* XII, 505-525.