Proceedings of the Fourth International Workshop on Folk Music Analysis (FMA2014)

12 and 13 June, 2014
Istanbul, Turkey

Editor:
Andre Holzapfel

Istanbul: Boğaziçi University

Title Proceedings of the Fourth International Workshop on Folk Music Analysis (FMA2014)

Editor Andre Holzapfel

Publishers Computer Engineering Department, Boğaziçi University

Program Committee

Chairs

- Peter Van Kranenburg (Meertens Institute)
- Matija Marolt (University of Ljubljana)

Members

- Christina Anagnostopoulou (University of Athens)
- Emmanouil Benetos (City University, London)
- Dániel P. Bíro (School of Music, University of Victoria)
- J. A. Burgoyne (ILLC, Universiteit van Amsterdam)
- Emilios Cambouropoulos (Aristotle University of Thessaloniki)
- Darrell Conklin (Universidad del País Vasco)
- Ewa Dahlig (The Institute of Art of the Polish Academy of Sciences)
- Bas De Haas (Utrecht University)
- Münevver Köküer Jancovic (Birmingham City University)
- Berit Janssen (Meertens Institute)
- Ali Cenk Gedik (Dokuz Eylül University)
- Emilia Gomez (Universitat Pomeu Fabra)
- Paco Gomez (Technical University of Madrid)
- Fabien Gouyon (INESC Porto)
- Olivier Lartillot (University of Jyväskylä)
- Aggelos Pikrakis (University of Piraeus)
- Robert Reigle (Istanbul Technical University)
- George Tzanetakis (University of Victoria)
- Anja Volk (Utrecht University)
- Tillman Weyde (City University, London)
- Frans Wiering (Utrecht University)

Organizing Committee

- Andre Holzapfel (Bogazici University)
- Taylan Cemgil (Bogazici University)
- Esra Mungan (Bogazici University)
- Baris Bozkurt (Bahcesehir University)

Preface

The present volume contains the proceedings of the Fourth International Workshop on Folk Music Analysis (FMA). The FMA workshops have a central goal to foster the application of computational analysis in (ethno)musicology. The two main reasons that motivate us to organize this series of workshops are, firstly, to explore systematic approaches to (ethno)musicology that have the potential to enrich the research in this field. And, secondly, we want to draw the attention of researchers in Music Information Retrieval (MIR) towards the analysis of musical styles that so far were not in the focus of MIR research. During the workshop, we provided an opportunity for debate regarding future directions in the field of computational (ethno)musicology, and provided an interdisciplinary platform to present our work to interested researchers and musicians.

We believe that the discussions we had during this workshop indicated ways to apply quantitative methods in (ethno)musicology, and we hope that we will see more collaborations between fields of engineering and humanities in general. Since music has been approached by researchers from various fields, such as (ethno)musicology, music cognition, anthropology, and engineering, it is about time to bridge the gaps between the scientific discourses in the individual fields. A difficult path to take, but the importance of music for all of us makes worth all efforts. We will continue our small contribution to this end with the series of FMA workshops.

We would like to thank the local team in Istanbul, all students and administrative staff who helped us. We would like to thank Martin Clayton for the keynote talk and for taking part in our discussions. We would like to thank the musicians Tolgahan Çoğulu and Erdem Şimşek for the inspiring presentation of Turkish folk music, and Peter van Kranenburg and Matija Marolt for their assistance in scientific aspects. And, last but not least, all participants and reviewers who support the FMA.

Istanbul, July 2014

The organizers

Contents

Chapter 1

Full papers

A STATISTICAL ANALYSIS OF THE ABC MUSIC NOTATION CORPUS: EXPLORING DUPLICATION

Chris Walshaw
Department of Computing & Information Systems,
University of Greenwich, London SE10 9LS, UK
c.walshaw@gre.ac.uk

ABSTRACT

This paper presents a statistical analysis of the abc music notation corpus. The corpus contains around 435,000 transcriptions of which just over 400,000 are folk and traditional music. There is significant duplication within the corpus and so a large part of the paper discusses methods to assess the level of duplication and the analysis then indicates a headline figure of over 165,000 distinct folk and traditional melodies. The paper also describes TuneGraph, an online, interactive user interface for exploring tune variants, based on visualising the proximity graph of the underlying melodies.

1. INTRODUCTION

1.1 Background

Abc notation is a text-based music notation system popular for transcribing, publishing and sharing folk music, particularly online. Similar systems have been around for a long time but abc notation was formalised (and named) by the author in 1993 (Walshaw, 1993). Since its inception he has maintained a website, now at abcnotation.com, with links to resources such as tutorials, software and tune collections.

1.1.1 Tune search engine

In 2009 the functionality of the site was significantly enhanced with an online tune search engine, the basis of which is a robot which regularly crawls known sites for abc files and then downloads them. The downloaded abc code is cleaned and indexed and then stored in a database which backs the search engine front end. Users of the tune search are able to view and/or download the staff notation, midi representation and abc code for each tune, and the site currently attracts around ½ million visitors a year.

1.1.2 Breadth

The aim of the tune search is to index all abc notated transcriptions from across the web. However there are a number of reasons why it is unable to do this completely:

- Unknown / new abc sites: the robot indexer is seeded from around 350 known URLs (some of which are no longer active), but it does not search the entire web.
- HTML based transcriptions: in the main, the indexer searches for downloadable abc file types (.abc, or sometimes .txt). However, there are a number of sites where the abc code is embedded directly into a webpage. Mostly these tend to be small collections (especially if the abc code has to be manually inserted into the HTML code) and so these are omitted from the search. However, there are 3 larger collections which are included (by parsing the HTML and looking for identifiable start and end tags).
- JavaScript links: for a small number of sites the file download is enacted via JavaScript, making the link to the .abc file difficult to harvest.

1.1.3 Growth

Starting with an initial database of 36,000 tunes in 2009 the index has expanded to over 435,000 abc transcriptions at the time of writing (May 2014). Most of these are folk tunes and songs from Western Europe and North America, although two massive multiplayer online role-playing games, Lord of the Rings Online and Starbound, have adopted abc for their in-game music system resulting in a number of dedicated websites with mixed collections of rock, pop, jazz and, sometimes, folk melodies. The ~35,000 transcriptions from these sites are ignored for the purposes of this paper, leaving just over 400,000 to be analysed (though this number changes every time the robot runs).

Importantly, although each of the transcriptions comes from a distinct URL, over half are duplicates and these are a major focus of this study.

1.2 Aims

The original intention for this paper was to present a statistical survey of the abc music notation corpus in its current state (i.e. mid-2014) including analyses of the corpus segmented by key, meter and tune type. The purpose was threefold:

- To provide a historical marker of the notation system in its 20th year (abc2mtex v1.0, a transcription package which contained the first description of the abc syntax, was released in December 1993).
- To discuss the composition of this large online resource and give some insights into the issues of curating and managing it.
- To invite other academics to explore the corpus in detail: the author is willing to grant exceptional access to the database for academic study

and interested in collaborating with projects that wish to make use of it.

For the most part this paper still has these aims. However, in investigating the data, a fundamental question arose: how many distinct tunes are there in the corpus? That, in addition to the supplementary question: what is meant by "distinct" in the context of aural traditions (with all the variation that implies) which are transcribed electronically (sometimes in sketch form), published online (sometimes temporarily), and subsequently copied and republished freely by other web users (often with no modifications, but sometimes with additional notes and corrections)?
The remainder of this paper is organised as follows:

- Section 2 discusses duplication and attempts to answer the question of how many distinct tunes there are in the corpus. It also presents on-going development of a user interface to allow the exploration of tune variants.
- Having decided on a methodology for discounting duplicates, section 3 presents a (straightforward) statistical analysis of the corpus together with a number of observations and comments on the data.
- Finally, section 4 presents some conclusions and ideas for further work.

2. DUPLICATION

Duplication occurs widely within the abc corpus for a number of observable reasons:

- **Compilations**: particularly in the past, certain enthusiasts have published compilations of all the abc tunes they could find, gathered from across the web.
- **Selections**: some sites, usually those containing repertoires (perhaps that of a band or an open session), publish a selection of tunes gathered from other sites.
- **Ease-of-access**: a number of sites publish collections or sub-collections both as one-tune-per-file together with a single file containing all of the tunes.

With respect to the tune search engine, there is little point in presenting users with dozens of identical results and so an important part of the pre-indexing clean-up involves identifying and, where appropriate, removing duplicates from the index. However, it is not necessarily clear which level of duplication to remove.
Furthermore, in the context of this paper, the elimination of duplicates is a fundamental process in determining how many distinct tunes there are in the corpus and the subsequent statistical analysis.

2.1 Eliminating duplicates

2.1.1 Classification

To discuss this topic further it is helpful to consider the structure of an abc tune transcription (see example in Figure 1).

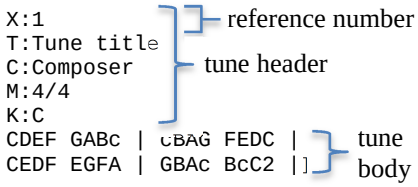


Figure 1. An example abc transcription.

Each tune consists of a **tune header** (including a **reference number**) and the **tune body**.
The header contains descriptive meta-data mostly, though not exclusively, with no musical information. Typically this includes the title and composer (where known), but amongst other data may also include information about where the tune was sourced (book, recording, etc.), who transcribed it, historical notes and anecdotes and instrumentation details (particularly for multi-voice music).
The tune body contains the music, and may also contain song lyrics.
With this structure in mind, duplication can be classified into 4 increasingly broad categories:

- **Electronic**: the duplicates are electronically identical (the exact same string of characters) – i.e. the tune headers and bodies are identical (although in practice this is relaxed somewhat by ignoring the reference number and any whitespace).
- **Musical**: the duplicates are musically identical (including song lyrics) although they may contain different meta-data in the tune header – i.e. the tune bodies are identical.
- **Melodic**: neglecting any song lyrics, grace notes, decorations and chord symbols, the first voice of each duplicate is identical – i.e. the primary melodies are identical.
- **Incipit**: when transposed to the same key, the duplicates are melodically identical over the first few bars of the tune.

2.1.2 Implementation

Code which analyses and counts the size of each category has been developed. In the first three categories this is done without actually parsing the abc music notation in the tune body: for the most part it involves stripping the transcriptions of data, for example by extracting parts and removing decorations, lyrics, grace notes, etc.
For each duplication class, the code derives a comparison string from each abc transcription which is then compared with all other comparison strings in that class: identical strings indicate duplicates.

As a small percentage of transcriptions contain errors and / or extraneous text, part of the parsing task involves exception handling. These can arise for a number of reasons including: misplaced characters which do not fit with agreed abc syntax and transcription errors such as unmatched start / end tags (for example, in abc syntax grace notes are delimited with curly braces, { … }, – an exception is thrown if one of the braces is missing).
Transcriptions which cannot be parsed, or are empty, fall back on the previous classification. In other words if an exception is thrown when a transcription is being parsed for *incipit* comparison, the comparison string reverts to a *melodic* comparison string. Likewise, if a transcription contains an empty tune body (as can often happen when abc headers are used as placeholders or for indexing purposes) then the *melodic* and *musical* comparison strings would revert to *electronic*.

2.1.3 Results

Table 1 shows the duplication results for the 4 duplicate classes. Here a **duplicate cluster** refers to a group of identical transcriptions. A cluster of size n has 1 primary transcription and $n - 1$ duplicates, so the number of duplicates (column 4) refers to the total number of duplicated transcriptions with a contribution of $n - 1$ duplicates from each cluster.

Class	#duplicate clusters	max. duplicate cluster size	#duplicates
Electronic	71,156	39	171,203
Musical	75,752	132	222,241
Melodic	73,199	132	232,528
Incipit	58,090	207	281,552

Table 1. The different levels of duplication.

As one might expect, the number of duplicates (and the maximum duplicate cluster size) increases with each successive class, since the duplication refers to a diminishing portion of each transcription. The increase and subsequent decrease of duplicate clusters is less intuitive, but is easily explained: for example, if there are two duplicate clusters of sizes n_1 and n_2 which differ from each other only after the 4th bar, then under melodic duplication this would result in two clusters whereas under incipit duplication it would result in a single cluster of size $n_1 + n_2$.
To interpret the figures further, consider melodic duplicates: of the 400,160 transcriptions, 232,528 (58.1%) are duplicates and can be excluded from the statistical analysis. Of the remaining 167,632 transcriptions, 73,199 (18.3%) have a duplicate in the excluded set and therefore 94,433 (23.6%) are not duplicated anywhere in the corpus. The maximum duplicate cluster size is 132 (in other words there is 1 tune with 131 excluded duplicates) and the average cluster size is 4.18, i.e. (232,528 + 73,199) / 73,199.
Whilst this indicates a very substantial amount of duplication within the corpus, this gives a headline figure of 167,632 distinct melodies, even when all of the metadata, decoration and lyrics are stripped away. Doubtless that some of these are very minor variants or corrections, but nonetheless it indicates that the abc music notation corpus represents a substantial online resource.

2.2 Exploring variants

The algorithm that is used for identifying incipit duplicates is actually based on a difference metric which numerically quantifies the difference between each pair of incipits. Pairs of melodies with a difference of 0 are duplicates (at least for the length of the incipit), but those with small difference values are very likely to be tune variants.
Tune variants are an important part of folk music's aural tradition and so near duplicates which appear only in the incipit category are of interest to researchers and musicians alike. However they are not always easy to identify by eye from a large number of search results.

2.2.1 TuneGraph

To facilitate user exploration of such variants the author is developing TuneGraph (Walshaw, 2014), an online tool for the visual exploration of melodic similarity, outlined below.
Given a corpus of melodies, the idea behind TuneGraph is to calculate the difference between each pair of melodies numerically with a difference metric or similarity measure (e.g. Kelly, 2012; Stober, 2011; Typke, Wiering, & Veltkamp, 2005). Next a proximity graph is formed by representing every tune with a vertex and including (weighted) edges for every pair of vertices which are "similar". Finally, the resulting graph can be visualised using standard graph layout techniques such as force-directed placement, (e.g. Walshaw, 2003), either applied to the entire graph or just to a vertex and its neighbours (i.e. a tune and similar melodies).
The concept is not dissimilar to a number of other software systems which give a visual display of relationships between tunes, often based on a graph (e.g. Langer, 2010; Orio & Roda, 2009; Stober, 2011).
TuneGraph consists of two parts – TuneGraph Builder, which analyses the corpus and constructs the required graphs, and TuneGraph Viewer, which provides the online and interactive visualisation.

2.2.2 The difference metric

In the current implementation, each melody is represented by quantising the first 4 bars (the incipit) into 1/64th notes and then constructing a pitch vector (or pitch contour) where each vector element stores the interval, in semitones, between the corresponding note and the first note of the melody (neglecting any anacrusis). Since everything is calculated as an interval it is invariant under transposition.
The difference metric then calculates the difference between two pitch vectors either using the 1-norm (i.e. the sum of the absolute values of the differences between each pair of vector elements) or the 2-norm (i.e. the square root of the sum of squared differences between each pair of vector elements). The 1-norm has long been available as part of the abc2mtex indexing facilities (Walshaw, 1994), but experimentation suggests that the 2-norm gives marginally better results (Walshaw, 2014).

If the pitch vectors have different lengths then the sum is over the length of the shorter vector (although see below – section 2.2.4).
Similarity measures of this kind are well explored in the field of music information retrieval, (e.g. Kelly, 2012; Typke et al., 2005), and there may be other, more advanced similarity measures that would work even better. However, in principle any suitable metric can be used to build the proximity graph, provided that it expresses the difference between pairs of melodies with a single numerical value. Indeed, even combinations of similarity measures could be used by forming a weighted linear combination of their values.

2.2.3 Building the proximity graph

The proximity graph is formed by representing every tune with a vertex and including (weighted) edges for every pair of vertices which are "similar" (i.e. every pair where the numerical difference is below some threshold value). However the question arises: what is a suitable threshold and how should it be chosen?
Perhaps the simplest choice, and one which is well-known for geometric proximity graphs, is to find the smallest threshold value which results in connected graph, i.e. a graph in which a path exists between every pair of vertices. Although computationally expensive, this can be done relatively straightforwardly starting with an initial guess at a suitable threshold and then either doubling or halving it until a pair of bounding values are found, one of which is too small (and does not result in a connected graph) and one of which is large enough (and does give a connected graph). Finally the minimal connecting threshold (minimal so as to exclude unnecessary edges) can be found with a bisection algorithm, bisecting the interval between upper and lower bounds each iteration.
This was the first approach tried but it resulted in graphs with an enormous number of edges; the test code ran out of memory as the number of edges approached 200,000,000 and the threshold under test had not, at that point, yielded a connected graph.
Further investigation revealed the basic problem: the graph is potentially very dense in some regions, with many similar melodies clustered together, whereas elsewhere there are outlying melodies which are not similar to any others. This means that in order to connect the outliers, and hence the entire graph, the threshold has to be so large that in the denser regions huge cliques are generated.

2.2.4 Segmentation by meter

In order to reduce the density of the graph, one successful approach tested was to segment the graph by meter – i.e. so that tunes with different meters are never connected. In fact a simple way to implement this is to avoid connecting pitch vectors with different lengths. This has the added benefit that some meters can be connected (i.e. those with the same bar length such as 2/2 and 4/4) meaning that the strategy is blind to certain variations in transcription preferences (although not universally as it will fail to connect related melodies, such as Irish single jigs, which are variously transcribed in 6/8 and 12/8, and French 3-time bourrées, which can be either 3/4 or 3/8).
Each pitch vector length results in a subset of graph vertices: in all there were 314 subsets, ranging in size from 63,581 vertices (for length 256 – e.g. 2/2 and 4/4 tunes), down to 115 subsets containing just one vertex. However, 98.7% of vertices are in a subset of size 100 or more and 99.7% are in a subset of size 10 or more.
The small subsets generally result from unusual vector lengths, usually because of errors in the transcriptions (i.e. extra notes or incorrect note lengths) and there was often no close relation between the melodies, meaning that a very high threshold would have to be used to connect that subset. To avoid connecting very different transcriptions, for each segment the edge threshold was somewhat arbitrarily limited to the length of the pitch vector for that segment. In most cases, this upper limit was never needed, but for very small subsets it sometimes meant that no edges were generated at all.

2.2.5 Average degree

Even with segmentation by meter in place the method can still generate huge graphs. However, there is no particular reason that the graph needs to be connected so the idea of trying to build a connected graph (or connected subgraphs, one for each pitch vector length) was abandoned as unpractical. Nevertheless, it is attractive as essentially parameter-free and it does work for small collections of relatively closely related tunes (for example, English morris tunes, where there are many similar variants of the same melody).
For the purposes of representing the entire corpus as a (disconnected) proximity graph, this still leaves the choice of a suitable edge threshold open, but rather than picking a value out of the air, instead a *target average degree* is chosen for the resulting graph. With this average degree as a user-selected parameter the same bounding and bisection method as above can be used to find the smallest threshold that yields this average degree.
An important observation was that the small number of vertices which have very many similar neighbours generate a relatively large number of edges in the graph. For example a cluster of, say, 100 very similar melodies will form a (near) clique with up to 4,950 edges. This significantly skews the average if it is expressed as the mean degree. However, using the median degree ignores these outlying values and gave much more useful results empirically and so the current implementation uses this measure to calculate the average.
Considerable experimentation has been carried out with a number of average degree values (see Walshaw, 2014, for a full discussion) and the best – i.e. the one which yields local graphs (see below) that are small enough to be useful in search but which are sufficiently rich enough to express similarities visually – seems to be an average (median) degree of 3.

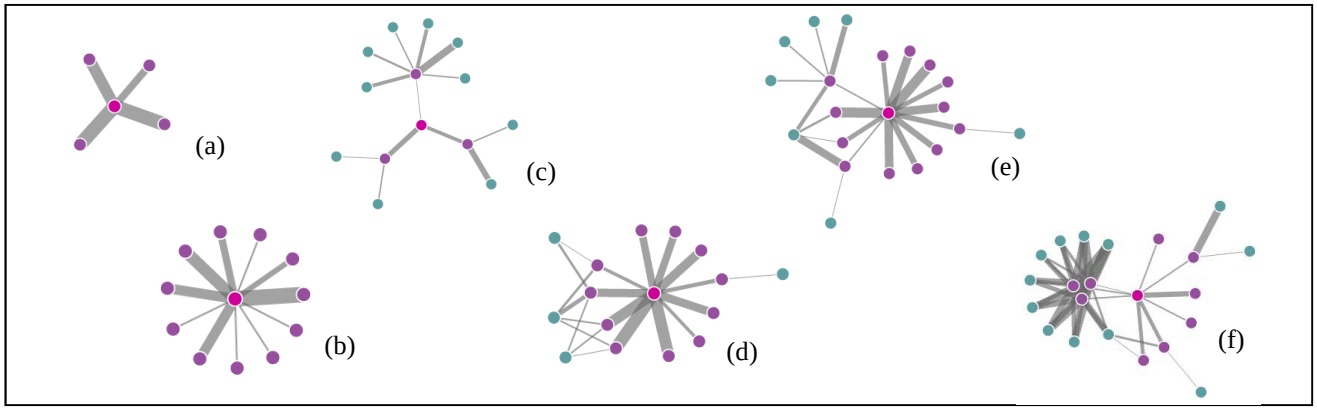


Figure 2. Some sample local graphs.

2.2.6 Extracting local graphs

Once suitable parameters have been chosen the graph is built as a series of proximity (sub-)graphs (one for each one for each pitch vector length). Each proximity subgraph is unlikely to be connected and as a result the graph as a whole can be highly disconnected.

One option is to use multilevel force directed graph placement (Walshaw, 2003), to find a layout for the entire graph. This has been tried and yields an interesting, but not necessarily very useful, representation of the corpus.

Instead, to allow exploration of similarities in an interactive online setting, the TuneGraph Builder code extracts a **local graph** for each non-isolated vertex. One way to do this is simply to extract the vertex, plus all its neighbours plus any edges between them. However, this can lead to clique-like local graphs where edges are hard to discern.

Instead, the local graph is built in layers: the seed (layer 0) is the original vertex for which the local graph is being built, layer 1 is any vertices neighbouring layer 0 and layer 2 is any vertices (not already included) neighbouring layer 1, etc. In order to maximise the clarity of the local graph, it only includes edges between layers and excludes edges between vertices in the same layer.

If the local graphs are just built from layers 0 and 1, each will be star-like, as in Figure 2(a) and Figure 2(b), yielding limited immediate visual information to the user (other than the number of neighbours and the strength of the relationships). Instead the builder code uses layers 0, 1 and 2, e.g. Figure 2(c) to Figure 2(f), to show some of the richness of certain neighbourhoods. Here colours indicate the layers, with layer 0 shown in crimson, layer 2 in light blue, and layer 1 interpolated between the two of them.

Finally, the graph edges are all weighted in inverse proportion to the difference between the two transcriptions that they connect (Walshaw, 2014). Since graph edge weights are indicated in the online tool by their thickness this conveys helpful visual information to the user by showing the more closely related tunes with thicker lines between them (and also affects how the graph is laid out by force directed placement).

2.2.7 Results

It is difficult to say exactly what features are desirable in the final graph, but experience with the local graphs suggests that they should be small enough not to overwhelm the user, but rich enough to convey some useful information. In particular the aim was to limit the maximum local graph size but maximise the average size. Experimentation was carried out with a number of different parameter settings (Walshaw, 2014) and often a small change can make a huge difference – for example, changing the target median degree from 3 to 4 increases the maximum local graph size from 121 to 724. However, the best parameters found were:

- Difference norm: $\|.\|_2$ – see section 2.2.2
- Segmentation by meter: true – see section 2.2.4
- Edge threshold limit: pitch vector length – see section 2.2.4
- Target average degree: median of 3 – see section 2.2.5

Using these settings results in a large number of isolated vertices, usually because there are no closely related melodies in the corpus or, less commonly, because there are no other transcriptions with the same pitch length. Eliminating these isolated vertices gave a final graph of 111,230 vertices in 31,784 connected subsets (many with as few as 2 vertices). The graph contains 250,182 edges, with a maximum degree of 68 and a minimum degree of 1, but is very sparse since the average degree is only 4.5.

From this 111,230 local graphs were produced with an average size of 6.1 vertices. The maximum size was 121 vertices and 468 edges. Whilst the largest local graphs can be difficult to visualise well, a random sample of the rest are of a size and complexity which both helps explore similarities without overwhelming the user.

Figure 2 shows some interesting examples: Here (a) and (b) come from local clique-like graphs with no immediate neighbours (recall that edges between vertices in the same layer are not included in the local graph so not all edges of the clique are shown). The tree shown in (c) indicates a number of tunes which are related but probably not immediate relations of each other. The graphs in (d) and (e) are similar to (b) only with some outlying tunes related to those in the clique. Finally the graph in (f) shows a tune on the edge of a tightly coupled clique.

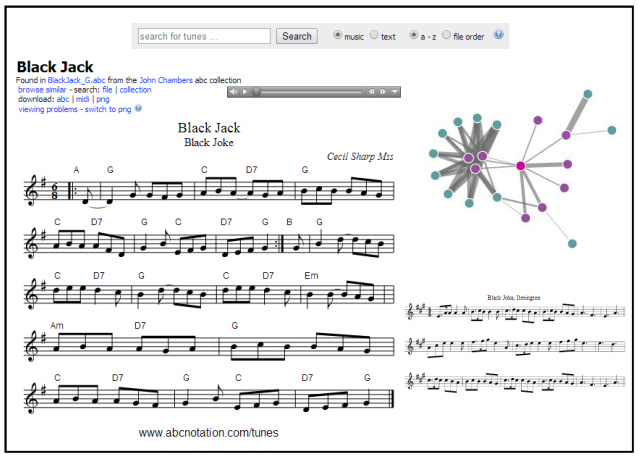


Figure 3. An example webpage.

2.2.8 TuneGraph Viewer

Although only in prototype version, TuneGraph Viewer contains a number of interactive features. The local graph is displayed on a webpage alongside the tune it corresponds to. It is visualised as a dynamic layout using D3.js (Bostock, 2012), a JavaScript library for manipulating documents based on data, and employing the inbuilt force-directed placement features.

It provides the following user interface:

- The graph vertices find their own natural position dynamically via force directed placement and vertices can be dragged to rearrange the layout (other vertices then relocate accordingly).
- Vertex colour indicates the relationship to the root vertex.
- Edge thickness indicates visually how closely related two vertices are (i.e. how similar their corresponding tunes are).
- Moving the mouse over a vertex reveals its name and displays the associated melody.
- Double clicking on a vertex (other than the root vertex) takes the user to the corresponding page (with its own tune graph).

Figure 3 shows an example webpage corresponding to the tune Black Jack (a well-known English tune). The tune is displayed on the left (the abc notation would appear underneath) and the local tune graph is shown on the right. If the user moves their mouse over one of the graph vertices, the tune associated with that vertex appears below.

3. STATISTICAL ANALYSIS

This section presents a brief and straightforward statistical analysis of the current abc music corpus (May 2014) based on those tunes found online by the abc search engine. It does not, of course, cover unpublished collections and so there are no real means to estimate what proportion of the abc corpus it represents.

Broadly speaking the analysis is qualitatively similar regardless of which method is used for eliminating duplicates. As a single example, neglecting the 171,203 *electronic* duplicates, 29.8% of the remaining melodies are transcribed in 4/4. With 222,241 *musical* duplicates removed this figure is 30.3% and respectively comes out at 30.7% and 32.4% when the 232,528 *melodic* or 281,552 *incipit* duplicates are removed.

To avoid filling the paper with statistics the rest of this section therefore concentrates on just one category of duplicates. In fact, *incipit* duplicates may not be duplicates at all – they may just have the same first four bars, so all of the following figures analyse the 167,632 distinct melodies remaining when the 232,528 *melodic* duplicates are removed from the corpus.

First note that, although abc is primarily used for monophonic tunes, of these 167,632 melodies, 6,480 (3.9%) are polyphonic and 12,574 (7.5%) are songs (i.e. with lyrics included in the abc transcription).

The tables below show an analysis of the corpus segmented by meter, rhythm (i.e. tune type), and in particular the key (a very expressive field in abc which allows the specification by mode).

It was also intended to include a table showing the corpus segmented by origin. However, this proved problematic for a number of reasons, specifically:

- The abc header field to specify origin (`O:`) allows free text and hence a wide variation in attribution and even spelling.
- The origin header field is not widely used and only 26.2% of tunes in the corpus make use of it.
- One particularly large collection (a compilation of other collections) has the default origin set to "England", when many of the tunes are clearly identifiable as Irish or Scottish – this significantly distorts the results.

Nevertheless, the origin analysis does indicate significant diversity, with substantial contributions (i.e. more than 1,000 transcriptions) from, in alphabetical order, China, England, France, Germany, Ireland, Scotland, Sweden & Turkey.

For each of the three tables that are included, key signature, meter and rhythm, the table shows all values with a count of 100 or more; any values with fewer than 100 instances are aggregated at the bottom.

3.1 Key signature

Table 2 shows the corpus segmented by key signature.

In abc, the key field is very expressive and allows the use of modes and even arbitrary accidentals, i.e. specified in the key signature and applied to all notes in the tune (unless overridden by another accidental applied to the individual note or notes in that bar).

There is even an option for the Great Highland Bagpipe (written `K:HP` in abc notation) where, by convention, tunes are usually played in Bb mixolydian but written in A mixolydian with no key signature (i.e. the C# and F#

are assumed but not written on the score). This is a throwback to the early days of abc and might now be better handled with an "omit-key-signature" output flag. Nonetheless, there are 2,326 transcriptions of this type.

Of more interest is the use of modes, the most common being A dorian with 3,638 transcriptions. In fact a survey of the entire range of key signatures (including aggregated values at the bottom of the table) shows that dorian is used for 9,008 transcriptions (5.4% of the corpus), mixolydian for 4,772 (2.9%), phrygian for 418 (0.3%), lydian for 85 (0.1%), aeolian for 84 (0.1%), ionian for 6 (0.0%) and locrian for 4 (0.0%). In addition, 19,596 of transcripitions (11.69%) are specified as being in a minor key.

Key signature	Count	Percentage	Cumulative
G	45,561	27.18%	27.18%
D	37,834	22.57%	49.75%
C	14,583	8.70%	58.45%
A	12,132	7.24%	65.69%
F	9,784	5.84%	71.52%
E minor	5,017	2.99%	74.52%
A minor	5,005	2.99%	77.50%
Bb	4,613	2.75%	80.25%
D minor	3,708	2.21%	82.46%
A dorian	3,638	2.17%	84.63%
G minor	2,995	1.79%	86.42%
E dorian	2,478	1.48%	87.90%
Great Highland Bagpipe	2,326	1.39%	89.29%
A mixolydian	1,957	1.17%	90.45%
B minor	1,945	1.16%	91.61%
none	1,798	1.07%	92.69%
D mixolydian	1,768	1.05%	93.74%
Eb	1,461	0.87%	94.61%
D dorian	1,172	0.70%	95.31%
E	1,115	0.67%	95.98%
G dorian	1,067	0.64%	96.61%
other	997	0.59%	97.21%
G mixolydian	593	0.35%	97.56%
C minor	561	0.33%	97.90%
Ab	286	0.17%	98.07%
C dorian	269	0.16%	98.23%
C mixolydian	240	0.14%	98.37%
B dorian	177	0.11%	98.48%
F minor	127	0.08%	98.55%
F# minor	127	0.08%	98.63%
E phrygian	117	0.07%	98.70%
other keys]	2,181	1.30%	100.00%

Table 2. A breakdown of the corpus by key.

3.2 Meter

Table 3 shows the corpus segmented by meter.

It is noticeable that much of the corpus is represented by meters common in Western European / North American folk music but there are significantly fewer of the more complex meters such as 7/8, 11/8, 15/8, etc., often found in Eastern Europe (9/8 is well represented but also includes slip jigs, commonly found in the British Isles).

Meter	Count	Percentage	Cumulative
4/4	51,493	30.72%	30.72%
6/8	34,840	20.78%	51.50%
2/4	22,378	13.35%	64.85%
2/2	19,764	11.79%	76.64%
3/4	19,614	11.70%	88.34%
free	6,006	3.58%	91.92%
9/8	3,679	2.19%	94.12%
3/8	2,166	1.29%	95.41%
6/4	1,868	1.11%	96.53%
12/8	1,425	0.85%	97.38%
3/2	1,411	0.84%	98.22%
4/2	409	0.24%	98.46%
7/8	371	0.22%	98.68%
8/8	317	0.19%	98.87%
9/4	302	0.18%	99.05%
10/8	293	0.17%	99.23%
5/4	122	0.07%	99.30%
5/8	113	0.07%	99.37%
[other meters]	1,061	0.63%	100.00%

Table 3. A breakdown of the corpus by meter.

3.3 Rhythm

Table 4 shows the corpus segmented by rhythm (tune type).

Unlike key signature and meter this is not a compulsory or assumed field (i.e. if no meter is specified, common time is assumed) and as result not all transcriptions have a rhythm indicated; nonetheless, 104,792 (62.5%) of them do.

Of interest in this table are the rhythms that indicate a specific origin. Reels, jigs and hornpipes are found widely in music from the British Isles and North America and the waltz, polka and schottische even more widely in Western European music. However, the strathspey indicates a Scottish origin – anecdotally there may be so many because of the large number of 19th Century tunebooks being transcribed into abc.

The polska and slängpolska indicate a Nordic origin, mostly likely Swedish, but found in other countries too and many come from a thriving wiki-based website, www.folkmusic.se.

Rhythm	Count	Percentage	Cumulative
no rhythm specified	62,840	37.49%	37.49%
reel	27,881	16.63%	54.12%
jig	20,353	12.14%	66.26%
hornpipe	6,943	4.14%	70.40%
waltz	4,636	2.77%	73.17%
strathspey	4,227	2.52%	75.69%
air	3,923	2.34%	78.03%
polka	3,863	2.30%	80.33%
march	2,343	1.40%	81.73%
slip jig	2,085	1.24%	82.98%
song	1,878	1.12%	84.10%
polska	1,663	0.99%	85.09%
barndance	1,181	0.70%	85.79%
country dance	1,126	0.67%	86.46%
slide	1,110	0.66%	87.13%
slängpolska	787	0.47%	87.60%
double jig	772	0.46%	88.06%
mazurka	581	0.35%	88.40%
dance	498	0.30%	88.70%
schottische	433	0.26%	88.96%
bourrée	386	0.23%	89.19%
triple hornpipe	379	0.23%	89.41%
quadrille	359	0.21%	89.63%
xiraldilla	247	0.15%	89.78%
minuet	221	0.13%	89.91%
miscellaneous	200	0.12%	90.03%
schottis	170	0.10%	90.13%
zwiefacher	135	0.08%	90.21%
single jig	123	0.07%	90.28%
other	108	0.06%	90.35%
set dance	106	0.06%	90.41%
[other rhythms]	16,075	9.59%	100.00%

Table 4. A breakdown of the corpus by rhythm.

4. CONCLUSION

This paper has presented a straightforward statistical analysis of the abc music notation corpus. The corpus contains around 435,000 transcriptions of which just over 400,000 are folk and traditional music.

There is significant duplication within the corpus and so a large part of the paper has discussed methods to assess the level of duplication. This has indicated a headline figure of over 165,000 distinct folk and traditional melodies. Much of the corpus seems to come from Western European and North American traditions, but there is a wide diversity included.

The paper has also described TuneGraph, an online interactive user interface for exploring tune variants, based on building a proximity graph of the underlying melodies. Although currently only in prototype form the intention is to deploy it on two sites with which the author is involved, abcnotation.com and the Full English Digital Archive at the Vaughan Williams Memorial Library (EDFSS, 2013).

4.1 Future work

The main focus for future work is to enhance the capabilities of TuneGraph. In particular it is intended to explore some of the wide range of similarity measures that are available as a means to build the proximity graph. As was indicated in section 2.2.2 there may be other, more advanced similarity measures, or combinations of similarity measures, that would work better than the 2-norm of the difference between pitch vectors.

5. REFERENCES

Bostock, M. 2012. Data-Driven Documents (d3.js), a visualization framework for internet browsers running JavaScript. http://d3js.org/

EDFSS. 2013. The Full English Digital Archive. *The Vaughan Williams Memorial Library*. http://www.efdss.org/efdss-the-full-english

Kelly, M. B. 2012. *Evaluation of Melody Similarity Measures*. Queen's University, Kingston, Ontario.

Langer, T. 2010. Music Information Retrieval & Visualization. In *Trends in Information Visualization*, pp. 15–22.

Orio, N., & Roda, A. 2009. A Measure of Melodic Similarity based on a Graph Representation of the Music Structure. *ISMIR*, pp 543–548.

Stober, S. 2011. Adaptive Distance Measures for Exploration and Structuring of Music Collections, Section 2, 1–10.

Typke, R., Wiering, F., & Veltkamp, R. C. 2005. A survey of music information retrieval systems. In *Proc. ISMIR*, pp. 153–160.

Walshaw, C. 1993. *ABC2MTEX: An easy way of transcribing folk and traditional music, Version 1.0*. University of Greenwich, London.

Walshaw, C. 1994. *The ABC Indexing Guide Version 1.2*. University of Greenwich, London.

Walshaw, C. 2003. A Multilevel Algorithm for Force-Directed Graph-Drawing. *Journal of Graph Algorithms and Applications*, *7*3, 253–285.

Walshaw, C. 2014. TuneGraph: an online visual tool for exploring melodic similarity. In *Proc. Digital Research in the Humanities and Arts (submitted)*. London.

A SYMBOLIC DATASET OF TURKISH MAKAM MUSIC PHRASES

M. Kemal Karaosmanoğlu
Yıldız Technical University, Istanbul
kkara@yildiz.edu.tr

Barış Bozkurt
Bahçeşehir University, Istanbul
baris.bozkurt@bahcesehir.edu.tr

Andre Holzapfel
Boğaziçi University, Istanbul
andre.holzapfel@boun.edu.tr

Nilgün Doğrusöz Dişiaçık
Istanbul Technical University, Istanbul
dogrusozn@itu.edu.tr

1. ABSTRACT

One of the basic needs for computational studies of traditional music is the availability of free datasets. This study presents a large machine-readable dataset of Turkish makam music scores segmented into phrases by experts of this music. The segmentation facilitates computational research on melodic similarity between phrases, and relation between melodic phrasing and meter, rarely studied topics due to unavailability of data resources.

2. INTRODUCTION

One of the rarely studied topics for Turkish makam music (TMM) is the structure inherent in its melodies. Indeed, as shown by Öztürk (2011), a large part of makam literature emphasizes the importance of "melody". Yet, computational studies dedicated to analysis of melody for TMM are mainly limited to n-gram and pitch class distribution based studies such as (Yener, 2004, Ünal et al, 2014). Such studies are agnostic to the structures of melodies, and focus on purely on observed intervals.

To fill this gap, we decided to collect a database of manually segmented melodic phrases. In the literature on Turkish makam music (e.g. (Kılınçarslan, 2006, Eroy, 2010, Gönül, 2010)), melodic analysis of a piece mainly involves the melodic phrase segmentation on scores followed by the labeling of the phrases with (tri - tetra - penta) chords used (which does not need to be one-to-one) and a further study of the functions within the makam. Our database contains melodic phrase boundaries and for a limited portion also the chord labels.

While such melodic analysis practice is part of the conservatory education today, in the TMM musicology literature, we do not find any explicitly formalized methodology for it. Hence, for the segmentation process, we preferred to give no specific instructions to the experts and asked them to perform the task as they would do it for melodic analysis of a piece. Three experts who received their education from different institutes / conservatories conducted the segmentation into phrases.

In the following sections we explain the design and the content of the database. In addition, we discuss potential computational studies that can make use of this data.

3. DATABASE DESIGN

The dataset is designed to contain pieces of the Turkish makam repertoire:

i) Composed in the most commonly used makams (Çevikoğlu, 2007): Acemaşiran, Beyati, Buselik, Hicaz, Hicazkar, Hüseyni, Hüzzam, Kürdilihicazkar, Mahur, Muhayyer, Neva, Nihavent, Rast, Saba, Segah and Uşşak,

ii) With a uniform distribution in time, from 17th century to today, dividing four main periods,

iii) From well-known composers such as Itri, Dede Efendi, Hacı Arif Bey and Sadettin Kaynak.

Nr.	Period	Composer
I	(…- 1750]	Itri
II	(1750-1850]	Dede Efendi
III	(1850-1930]	Hacı Arif Bey
IV	(1930-…]	Sadettin Kaynak

Overall, a set of 480 pieces was collected consisting of 30 pieces for each of the 16 distinct makams by rewriting the pieces using Mus2 microtonal notation software (http://www.mus2.com.tr/) in the Arel notation (Arel, 1968) and further converting this data to the machine readable text format of SymbTr (Karaosmanoğlu, 2012). Three experts were asked to mark the phrase boundaries and çeşni/geçki modulations on printed scores, as they would do it for makam melodic analysis. One of the experts completed the task and the other two could label only half of the scores. So each piece was labeled by at least one expert. Their segmentations of phrase boundaries and çeşni/geçki modulations were manually exported to the machine-readable format using a custom-designed interface. Table I summarizes the size of the data.

	Number of pieces	Number of phrases
Expert 1	488	20 293
Expert 2	200	4 312
Expert 3	201	6 757
Total	**889**	**31 362**

Table I. Size of the database

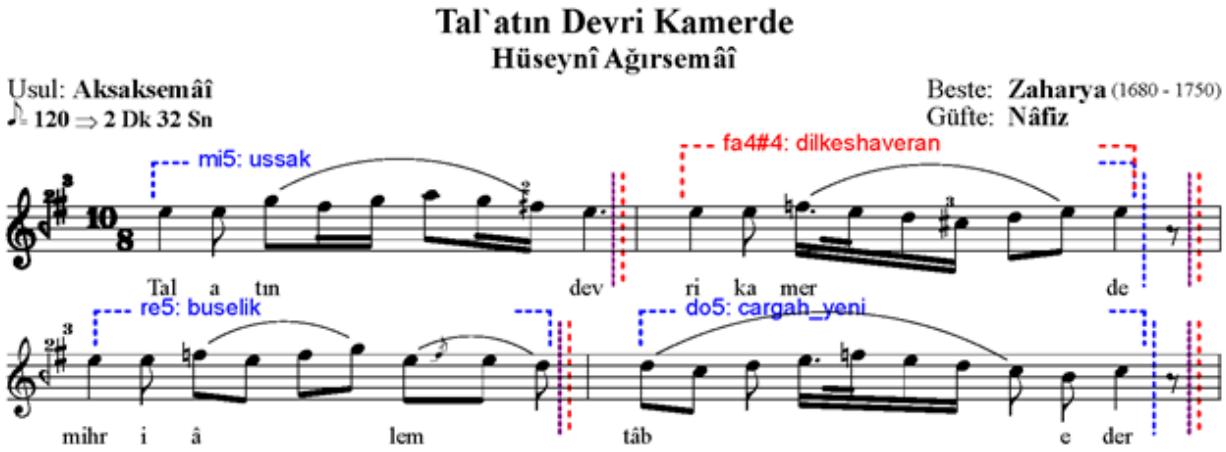


Figure 1. Example excerpt of a score

Code	NoteName	NoteName	Pitch	Pitch	Num.	Denum.	Ms	LNS	VelOn	Lyrics	Offset
9	Mi5	E5	336	336	1	4	1000	95	96	Tal	0.2
9	Mi5	E5	336	336	1	8	500	95	96	A	0.3
9	Sol5	G5	349	349	1	8	500	99	96	tın	0.4
…	…	…	…	…	…	…	…	…	…	…	…
9	Sol5	G5	349	349	1	16	250	99	96		0.65
9	Fa5#2	F5#1	342	341	1	16	250	95	96		0.7
9	Mi5	E5	336	336	3	8	1500	95	96	dev	1
53											1
54										dilkeshaveran@fa5#5	1
9	Mi5	E5	336	336	1	4	1000	95	96	ri	1.2
9	Mi5	E5	336	336	1	8	500	95	96	ka	1.3
…	…	…	…	…	…	…	…	…	…	…	…
9	Mi5	E5	336	336	1	8	500	95	96		1.7
9	Mi5	E5	336	336	1	4	1000	95	96	de	1.9
53											1.9
9			-1	-1	1	8	500	100			2

Table II. Machine-readable format of the dataset for the example in Figure 1 for one of the experts (red labels)

In Figure 1 and Table II we present a short example from the dataset. Machine readable files are in SymbTr format and basically contain *pitch* (columns 2 – 5), *duration* (columns 6 – 8, 12) and *lyric* (column 11) information for each note in a piece. It also includes microtonal pitch information which is an important feature for TMM. For this specific study, segmentation and çeşni/geçki information are added as separate lines (colored in Table II). So in the new format, each line contains information for a note or segmentation. This format is very similar to SymbTr and hence basics of the format is not explained in detail. New features are:

New *codes* (first column) are added to specify the type of segmentation: 53 for phrase boundary, 54 for the start of a *geçki*/*çeşni* label (the type of *çeşni* is specified in the *lyrics* column), 55 pointing the end of a *geçki*/*çeşni* label.

The dataset is shared along with Matlab code for importing the data into the standard format used by MIDI Toolbox (Eerola & Toiviainen, 2004) (while keeping the microtonal pitch information) so that melodic analysis tools available in the Toolbox can be easily used on the data. In addition to the symbolic data in text file format and computer codes, scanned images of the scores with manual segmentations are also shared.

The database can be accessed on:
http://akademik.bahcesehir.edu.tr/~bbozkurt/112E162_en.html
or http://www.rhythmos.org/shareddata/turkishphrases.html

In order to check for the degree of mutual agreement between the annotations, we conducted a comparison of the data obtained from the three experts. In Table III, we present F-measures obtained as a measure of match between each set of annotations. The first and the third expert agree on boundaries with an F-measure of 0.78. The second expert preferred comparatively longer phrases hence the corresponding F-measures are lower. As we discuss in the next section, there is a high agreement across the annotations conducted by the three experts considering phrase boundary choices with respect to makam pitches and usul beats.

	Precision	Recall	F-measure
Expert 1 vs 2	0.46	0.80	0.57
Expert 1 vs 3	0.74	0.85	0.78
Expert 2 vs 3	0.86	0.37	0.51

Table III. Measure of consistency between 3 experts

4. COMPUTATIONAL STUDIES

This dataset has been recently used in two computational studies: a study on automatic melodic segmentation (Bozkurt et al., 2014) and a study on characterization of segmented phrases (Bozkurt & Karaçalı, 2014). Below, we discuss parts of these studies and other observations on the data, in order to indicate potential research problems where the database can be of use.

4.1 Relation between phrase and meter

While the tonal content of Turkish makam music is guided by the modal framework of the *makam*, the metrical structure is implied by the *usul*. An usul is a rhythmic pattern that contains a series of strokes. While the length of such a pattern can be interpreted as defining the time signature of a piece, the locations of the strokes within the usul give further guidance for the rhythmic elaboration of a piece. As we showed by analyzing a large corpus of machine-readable notations of Turkish makam music (Holzapfel, 2014), note onsets and note durations tend to be related to the positions of the strokes in an usul. However, it is also evident from the results in (Holzapfel, 2014) that the metrical structure implied by onsets and durations is less clearly stratified than for Eurogenetic popular or classical music. This means, that given only the positions and durations of individual notes, we will necessarily do worse in tracking the temporal structure in Turkish music than in beat tracking of Eurogenetic popular music. In the oral tradition of Turkish makam music, the learning/memorization of pieces involves first learning and internalizing the usul pattern and then further memorizing the melodies in relation to the usul pattern (Behar, 1998). While we did not find any theoretical study discussing the link between the phrase boundaries and rhythmic cycles, our private communication with makam music experts revealed that melodic boundaries and usul cycles are strongly correlated.

A preliminary study on phrase and meter is carried out to compare phrase boundary positions with note onset positions for the Aksak usul. The Aksak usul is notated using a 9/8 time signature (Figure 2), and the stroke positions and accents of the usul are indicated by dotted lines in Figure 3a. The bold bars indicate the probability to encounter a phrase onset at the denoted location. It can be seen that the existence of a phrase boundary is a very strong indicator of the downbeat (*i.e.* the beginning of a measure), since it has a probability of almost 0.35, much higher than for the other locations. This result is consistent for all frequent short usul, with one remarkable exception. This exception is given by the *Ağıraksak*, an usul that is formally the same as the Aksak, but that is related to a slower tempo than Aksak and is usually notated as 9/4. The example in Figure 3b shows that the phrasing in Ağıraksak deviates widely from Aksak, even though the usul strokes are theoretically the same.

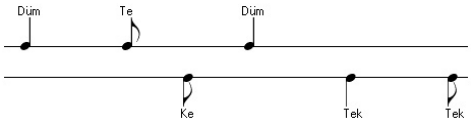


Figure 2. Symbolic description of the Aksak usul

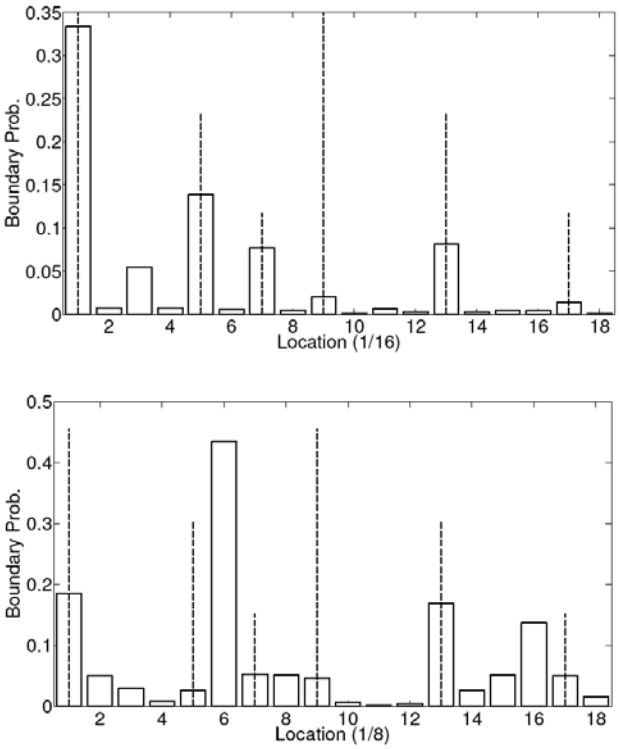


Figure 3. Location of annotated phrase onsets for Aksak and Ağıraksak.

Some of the research questions that raise immediately after these observations are: how can we explain the peculiarities of phrasing in the Aksak family? How can long usuls (of length 28 and 32) be subdivided into groups according to the annotated melodic phrasing? How this phrasing relates with the rhythmic structure encountered in the stroke sequences of these long usul? Is it possible to include phrasing in a model structure and phrase length as a constraint for tempo and beat tracking? Using the data presented in this paper, we can investigate in how far phrase information has the potential to improve automatic tempo/beat tracking on Turkish music. In our current research we develop generative models for this task, which combine assumptions about tempo development and rhythmic structure into a high-dimensional state-space. We aim at enhancing this model by including phrase information as a further high-level layer.

4.2 Phrase boundary distributions with respect to makam pitches

In (Bozkurt et al., 2014), a data-driven approach for automatic melodic phrase segmentation is proposed where the problem of segmentation is considered as a classification problem; each note is classified to be at a phrase boundary or not based on features computed from

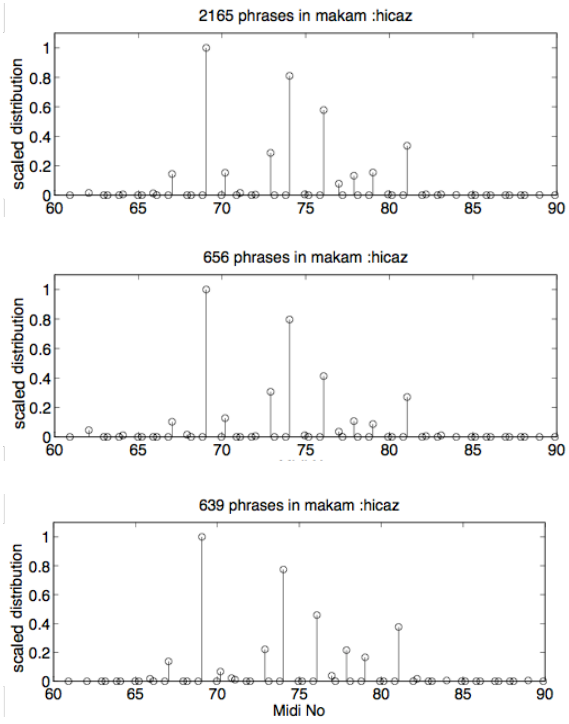


Figure 4. Feature values (scaled distributions) computed with respect to pitches of makam Hicaz from the three datasets.

the data. The feature vector computed for each note in the melody contains values of decision functions used in four state-of-the-art melodic segmentation techniques together with two novel features. The novel features are derived from probabilities of a note to appear at the boundary based on the makam and the usul of the piece. The probabilities are obtained from phrase boundary distributions computed for makam pitches and usul beats. The tests show that these new features carry complementary information to features from the literature and their inclusion in the feature vector results in a statistically significant improvement for the automatic segmentation task.

Our proposal of such features is motivated by the observation on our phrase annotated dataset that phrase boundaries are strongly correlated with the hierarchy of makam pitches and usul beats. In Figure 4, we present the phrase boundary features (computed as the scaled version of makam distributions) computed with respect to Hicaz makam pitches for the three annotators.

The first observation is that the distributions obtained from different annotators are very similar, indicating that all experts made similar choices with respect to these parameters. It is interesting to observe that this consistency holds to a very large extent for the three annotators for all makams and usuls. These distributions (both plots and data as Matlab file) are shared together with the database.

We observe that for makam Hicaz, the *karar* (tonic) pitch (A4, MIDI No: 69.06) appears as the most frequent phrase boundary (ending note). The second most frequent phrase boundary note is the *güçlü* (D5, MIDI No: 74.04). Similar observations hold for the rest of the makams in terms of hierarchies of scale degrees and frequent phrase ending notes. This supports the claims of Ayari (2005) that the hierarchy of makam pitches and phrase boundaries are related.

The interrelations between the metrical structure and grouping have been previously discussed by Lerdahl & Jackendoff (1983) and taken into consideration by Temperley (2001) for automatic melodic segmentation. While the link between phrase boundaries and hierarchies of makam pitch and usul beats is used in a data-driven approach effectively for an automatic task, the underlying phenomenon between phrasing and rhythmic cycle and pitch hierarchies is yet to be explored with a musicological perspective. Our database includes content in 16 makams and 10 usuls, which provides a large diversity for a systematic study of the interrelations between meter and melodic phrasing in Turkish music.

4.3 Phrase boundaries and personal style

We have carried out preliminary studies of relations between phrase boundaries and makam pitches on subsets of the database. Our observations suggest that the phrase boundary distributions, grouped with respect to composer and makam, provide important clues to understanding the makam concept and compositional choices of phrasing. Here, we present one simple example on makam Segah.

Comparing phrase distributions with respect to

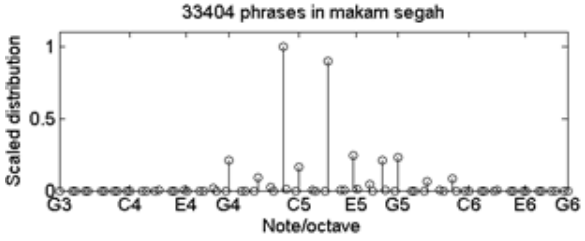


Figure 5. Feature values (scaled phrase boundary distributions) for all Segah pieces from the first dataset.

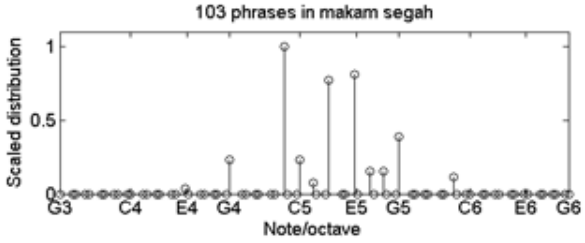


Figure 6. Feature values (scaled phrase boundary distributions) computed from Sadettin Kaynak's Segah pieces from the same dataset.

makam pitches for several subsets containing only compositions from a specific composer, we observe that distributions for compositions of Sadettin Kaynak stand out to be specifically different. In Figures 5 and 6, we present distributions for all Segah pieces and those of Sadetting Kaynak in the first data set.

The peculiarity of the subset including only Sadettin Kaynak's Segah pieces is the existence of many phrase boundaries on E5b pitch (in addition to the expected frequent boundaries on segah pitch B4 (the tonic) and on neva pitch D5 (the dominant)), which is not observed for other composers' data. The hierarchy of pitches can theoretically identify a makam but then these characteristic elements can be deliberately ignored by some composers like Sadettin Kaynak, a reformist composer.

Acknowledgement

This work is supported by the Scientific and Technological Research Council of Turkey, TUBITAK, Grant [112E162]. Andre Holzapfel is supported by a Marie Curie Intra European Fellowship (PIEF-GA-2012-328379).

1. REFERENCES

Arel, H.S. (1968). Türk Musikisi Nazariyatı (The Theory of Turkish Music). ITMKD yayınları, no: 2, Istanbul: Hüsnütabiat matbaası.

Ayari, M. (2005). *De la theorie musicale a l'art de l'improvisation: Analyse des performances et modelisation musicale.* Sampzon: Delatour France.

Behar, C. (1998). Aşk Olmayınca Meşk Olmaz. Istanbul: YKY Yayınları.

Bozkurt, B., Karaosmanoğlu, M.K., Karaçalı, B., Ünal, E. (2014), Usul and Makam driven automatic melodic segmentation for Turkish music. Accepted for publication in *Journal of New Music Research*.

Bozkurt, B., Karaçalı, B. (2014). A computational analysis of Turkish *makam* music based on a probabilistic characterization of segmented phrases. Accepted for publication in *Journal of Mathematics and Music*.

Çevikoğlu, T. (2007). Klasik Türk Müziğinin bugünkü sorunları. In *Proceedings of International Congress of Asian and North African Studies (Icanas 38')*. Ankara, Turkey.

Eerola, T., Toiviainen, P. (2004). MIDI Toolbox: MATLAB Tools for Music Research: http://www.jyu.fi/music/coe/materials/miditoolbox.

Eroy, O. 2010. "Tekirdağ Bölgesi Çingene Müziklerinde Kullanılan Ezgi Yapılarının İncelenmesi". Kırıkkale University.

Gönül, M. 2010. "Nevres Bey'in Ud Taksimleri Analizi ve Ud Eğitimine Yönelik Alıştırmaların Oluşturulması". PhD diss. Selçuk University.

Holzapfel, A. (2014) Relation between surface rhythm and rhythmic modes in Turkish makam music. Submitted for publication in *Journal of New Music Research*.

Karaosmanoğlu, M.K. (2012). A Turkish makam music symbolic database for music information retrieval: SymbTr. In *Proceedings of 13th Int. Society for Music Information Retrieval (ISMIR) conference, 223-228*. Porto, Portugal.

Kılınçarslan, H. 2006. "Dede Efendi'nin Hüzzam Mevlevi Ayininin Makam, Usul ve Ezgisel Yönden İncelenmesi". MA diss. Selçuk University.

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

Öztürk, O.M. (2011). Turkish Modernization and Makam Concept: Some determinations on two musical systems, *ICTM Yearbook*.

Ünal, E., Bozkurt, B., Karaosmanoğlu, M.K. (2014). A hierarchical approach to makam classification of Turkish makam music, using symbolic data. *Journal of New Music Research*, 43:1, 132-146.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.

Yener, S. (2004). Bilgisayar Destekli Analiz Yoluyla Geleneksel Türk Sanat Müziği Hicaz Taksimlerinde Kalıplaşmış Ezgilerin Araştırılması. MA diss. Gazi University.

INCORPORATING PITCH CLASS PROFILES FOR IMPROVING AUTOMATIC TRANSCRIPTION OF TURKISH MAKAM MUSIC

Emmanouil Benetos
City University London
emmanouil.benetos.1@city.ac.uk

Andre Holzapfel
Boğaziçi University, Istanbul
andre@rhythmos.org

ABSTRACT

In this paper we evaluate the impact of including knowledge about scale material into a system for the transcription of Turkish makam music. To this end, we extend our previously presented appoach by a refinement iteration that gives preference to note values present in the scale of the mode (*i.e.* makam). The information about the scalar material is provided in form of pitch class profiles, and they are imposed in form of a Dirichlet prior to our expanded probabilistic latent component analysis (PLCA) transcription system. While the inclusion of such a prior was supposed to focus the transcription system on musically meaningful areas, the obtained results are significantly improved only for recordings of certain instruments. In our discussion we demonstrate the quality of the obtained transcriptions, and discuss the difficulties caused for evaluation in the context of microtonal music.

1. INTRODUCTION

The derivation of a suitable notation for a music performance is an topic that has been discussed with many different goals and forms of music in mind. In ethnomusicology, the process of this derivation is referred to as transcription. Throughout the decades of the twentieth century, it was attempted to find extensions to the normal staff notation in order to capture micro-tonal information as well as other aspects related to rhythm and timbre that are hard to capture in usual staff notation (Abraham & von Hornbostel, 1994). A prominent point in the discussion about transcription was a symposium on transcription and analysis in 1963 (England, 1964), in which several experts conducted transcriptions of a Bushman song, with the output giving an interesting example of how transcriptions differ depending on personal interpretation and focus of analysis.

While at least Anglo-American ethnomusicology since then has taken a turn towards anthropology, the importance of deriving an adequate graphical representation for music remains an important task in many contexts. For instance, in many cultures notations are used in order to support the memorization of music, such as for Turkish makam music or for Eurogenetic classical music. In these contexts, a notation is a culturally conventional tool in order to compose and teach music and to facilitate performances. Apart from the immediate practical value, notations can also enable for a more focused analysis of musical aspects. Typically, when using a form of notation that is similar to Eurogenetic staff notation, the aspect that is most adequately represented is melody. Therefore, a notation is a useful tool to arrive, for instance, at an understanding of melodic phrases present in a repertoire, of the scale material that they use, and typical modulations between tonal modes.

Automatic music transcription, then, is the process of automatically converting some music signal into a form of notation. It is one of the pivotal areas of research in the field of Music Information Retrieval (MIR), where the difficulty of the task prevails after decades of research especially for the transcription of ensemble recordings (Klapuri & Davy, 2006). The results of an automatic transcription system can be improved by including the knowledge of the tonal material present in the recording, as it was demonstrated by Benetos et al. (2014) for Eurogenetic music. In the present paper, we will transfer these results to the context of Turkish makam music. This is facilitated by formulating the transcription as a probabilistic model, in which pitch class profiles can be included to enforce tonal structure assumed to be present in the piece which is to be transcribed. We apply a large set of pitch class profiles derived by Bozkurt (2008) from a large set of instrumental recordings. This way, we will continue our research on transcription of Makam music, which will hopefully help to shed light on the challenges of the AMT task for musics not part of Eurogenetic classical or popular music, which were at the focus of most MIR research so far.

In our paper we will start with a summary of the notation conventionally used in Turkish makam music practice, and the specific challenges for an AMT system when targeting such a transcription. We will then give a detailed description of our transcription system in Section 3. In Section 4, we describe the music collection which we aim to transcribe using our system, and we will conduct a quantitative evaluation of our system. We will then give some qualitative examples for the transcriptions, clarifying the shortcomings and the potentials of the method in Section 5. Finally, we will summarize our findings and give some overview of potential future research directions in Section 6.

2. CHALLENGES AND MOTIVATIONS

In our previous work we gave a detailed outline of the challenges related to the computational analysis of Turkish makam music (Bozkurt et al., 2014), and addressed challenges specific to automatic music transcription in Benetos & Holzapfel (2013). We will shortly summarize the most important musical aspects, that make a transcription system for Turkish makam music a challenging target for research.

First, Turkish music performance could be described as having a concept of relative pitch, which means that the frequency value of the tonic (*karar*) of a piece can vary depending on the choice of the performer(s). That means for a given performance, the pitch value of the tonic has to be determined either manually or using computational analysis in order to arrive at a valid notational representation of the piece.

Furthermore, once performers agree on a certain pitch for the tonic, certain instruments typically play the main melody in the distance of an octave due to their pitch range. This is the case for instance for the *tanbur* and *ney* instruments, which represent an often encountered combination in this music practice. Typically, melodies allow for a certain degree of freedom, resulting in a music practice that demands for the application of certain ornamentations that are supposed to give life and color to an interpretation. Especially in ensemble performances such ornamentations lead to deviations between the melodies played by the individual instruments, which results in an increased difficulty for a transcription.

Finally, the conventional notation system for Turkish makam music applies Eurogenetic staff notation with additional accidentals that signify certain micro-tonal intervals that deviate from well-tempered tuning. However, as discussed by Bozkurt (2008), these notated accidentals do often not match with the intervals encountered in performance practice. This further complicates AMT, since the set of note intervals to be expected in a piece strongly varies depending on tonal mode, performer, instrument, and possibly other parameters.

Our initial transcription system (Benetos & Holzapfel, 2013) for Turkish music targeted a transcription that included micro-tonal intervals. In the present publication we will evaluate, if the accuracy of our system can be further improved by including knowledge of the note intervals typically encountered in the performances of a certain makam. The inclusion of scale information was shown to improve AMT performance for Eurogenetic music Benetos et al. (2014), but it is an open question if this holds for Turkish makam music; There is a significant dissent between theory and practice, and the ongoing discussions among musicians indicate that the choice of certain intervals depends on personal choice, instrument, and historical period to some extent.

3. PROPOSED SYSTEM

The proposed transcription system takes as input a recording and information about the makam. Multi-pitch detection is performed using the efficient transcription system that was proposed in Benetos et al. (2013), modified for using ney and tanbur templates. Note tracking is performed as a post-processing step, followed by tonic detection. Given the tonic, the piece is then re-transcribed, using information from makam pitch profiles and the detected tonic. The final transcription output is a list of note events in cent scale centered around the tonic. A diagram of the proposed transcription system can be seen in Fig. 1.

3.1 Pitch Template Extraction

We use pitch templates extracted from 3 solo ney and 4 solo tanbur recordings, originally created in Benetos & Holzapfel (2013). The templates were extracted using probabilistic latent component analysis (PLCA) Smaragdis et al. (2006) with one component. The time/frequency representation used is the constant-Q transform (CQT) with a spectral resolution of 60 bins/octave, with 27.5Hz as the lowest bin (Schörkhuber & Klapuri, 2010). The range (in MIDI scale) for ney is 60-88 and the range for tanbur is 39-72.

3.2 Transcription Model

The proposed transcription model expands the probabilistic latent component analysis (PLCA) method, by supporting the use of multiple pre-extracted spectral templates per pitch and instrument, that are also pre-shifted across log-frequency for supporting frequency and tuning deviations; the latter is particularly useful for performing transcribing micro-tonal music. The model takes as input a log-frequency spectrogram $V_{\omega,t}$ (ω is the log-frequency index and t the time index) and approximates it as a bivariate distribution $P(\omega, t)$, which in turn is decomposed as:

$$P(\omega, t) = P(t) \sum_{p,f,s} P(\omega|s, p, f) P_t(f|p) P_t(s|p) P_t(p) \quad (1)$$

where $P(\omega|s, p, f)$ are the pre-extracted spectral templates for pitch p, instrument s, which are also pre-shifted across log-frequency according to parameter f. $P_t(f|p)$ is the time-varying log-frequency shift per pitch, $P_t(s|p)$ is the time-varying instrument contribution per pitch, and $P_t(p)$ the pitch activation over time (i.e. the transcription).

Since the log-frequency representation has a resolution of 60 bins/octave, and f is constrained to one semitone range, f has a length of 5. The unknown parameters of the model, $P_t(f|p)$, $P_t(s|p)$, and $P_t(p)$, can be iteratively estimated using the Expectation-Maximization algorithm of Dempster et al. (1977), with 30 iterations being sufficient for convergence. The spectral templates $P(\omega|s, p, f)$ are kept fixed using the pre-extracted pitch templates from Section 3.1 and are not updated.

The output of the transcription model is a MIDI-scale pitch activation matrix given by:

$$P(p, t) = P(t) P_t(p) \quad (2)$$

as well as a high pitch resolution time-pitch representation, given by:

$$P(f', t) = [P(f, p_{low}, t) \cdots P(f, p_{high}, t)] \quad (3)$$

where $P(f, p, t) = P(t) P_t(p) P_t(f|p)$. In (3), f' denotes pitch in 20 cent resolution. As an example of a time-pitch representation, Fig. 2 displays $P(f', t)$ for a ney recording.

3.3 Post-processing

The transcription output is a non-binary representation that needs to be converted into a list of of note events, listing

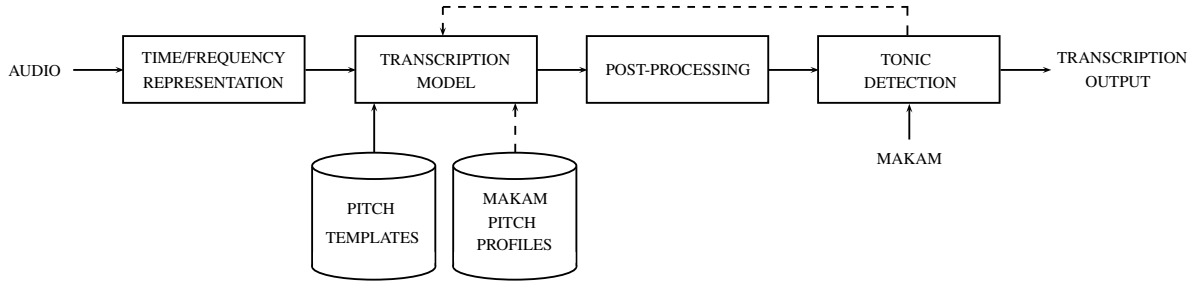


Figure 1: Proposed transcription system diagram. Dashed lines indicate operations taking place at re-transcription.

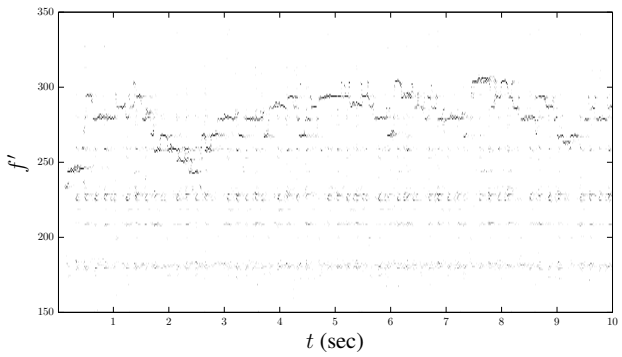


Figure 2: The time-pitch representation $P(f', t)$ for the 'Huseyni Peşrev' piece performed by ney. Note the percussive elements in the lower pitch range.

onset, offset, and pitch. Thus, we perform median filtering and thresholding on $P(p, t)$ (for converting it into a binary representation), followed by minimum duration pruning (with a minimum note event duration of 130ms).

As in Benetos & Holzapfel (2013), a simple 'ensemble detector' is used in order to detect heterophonic recordings. Subsequently, if a piece is detected as such, each octave interval is processed by merging the note event of the higher note with that of the lowest one. Then, using information from $P(f', t)$, each detected note is converted into the cent scale.

In order to detect the tonic frequency of the recording we apply the procedure described in Bozkurt (2008), which computes a histogram of the detected pitch values, and aligns it with a template histogram for a given makam using the cross-correlation function. A final post-processing step is made after centering the detected note events by the tonic, where note events that occur more than 1700 cents or less than -500 cents apart from the tonic are eliminated.

3.4 Pitch class profiles

For incorporating information on the pitch structure of the recording given a makam, we re-transcribe the recording having as additional information its detected tonic, and we impose a prior on the pitch activation $P_t(p)$. For enforcing a structure on pitch distributions, we employ the 44 makam pitch class profiles that were computed from large sets of instrumental recordings in Bozkurt (2008). An example of a pitch class profile is given in Fig. 3, for the Beyati makam.

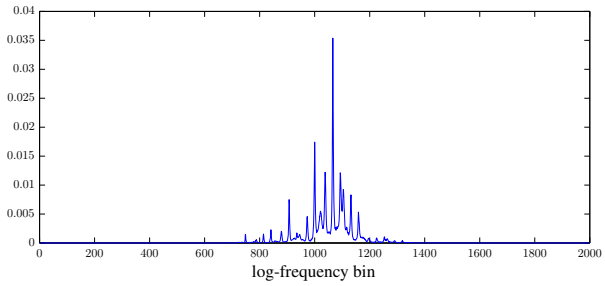


Figure 3: The pitch class profile for the Beyati makam, computed in Bozkurt (2008). The tonic is aligned with the bin 1000, and the resolution is 159 bins/octave (i.e. 1/3Hc).

As was shown in Smaragdis & Mysore (2009), PLCA-based models can use Dirichlet priors for enforcing structure on their distributions. We thus define the Dirichlet hyper-parameter for the pitch structure given a makam m and tonic τ as:

$$\alpha(p|t)_{m,\tau} = K_{m,\tau} P_t(p) \tag{4}$$

where $K_{m,\tau}$ is a pitch profile for makam m centered around tonic τ detected from 3.3. Essentially, $\alpha(p|t)_{m,\tau}$ represents a modified transcription, giving higher probability to pitches which are more frequently encountered in the specific makam than to pitches which are not.

Thus, a modified update rule is created for $P_t(p)$, which is as follows:

$$P_t(p) = \frac{\sum_{\omega,f,s} P_t(p, f, s|\omega) V_{\omega,t} + \kappa \alpha(p|t)_{m,\tau}}{\sum_{p,\omega,f,s} P_t(p, f, s|\omega) V_{\omega,t} + \kappa \alpha(p|t)_{m,\tau}} \tag{5}$$

where $P_t(p, f, s|\omega)$ is the posterior of the model and κ is a weight parameter expressing how much the prior should be imposed, which as in Smaragdis & Mysore (2009) gradually decreases from 1 to 0 throughout iterations (initialising the model but letting it converge in the end). After the modified transcription step, the output is then post-processed using the steps from Section 3.3.

4. EVALUATION

4.1 Music Collection

The music collection which is used is identical with the collection described in Benetos & Holzapfel (2013). It consists of 16 recordings of metered instrumental pieces of Turkish makam music, 10 of them solo performances, and

the remaining 6 recordings ensemble performances. The performances were note-to-note aligned with the microtonal notation available from the collection published in Karaosmanoğlu (2012). This results in a ground-truth notation that consists of a list of time-instances and note-values in cent assigned to each time instance, with the tonic of the piece at 0 cent. It is important to point out that the available ground-truth does not represent a descriptive transcription of the performance, but rather a summary of the basic melody played in the piece without ornamentations, as it is typical for notations in Turkish makam music.

4.2 Metrics

For the proposed evaluations, we use the note-based onset-only metrics that were defined in Benetos & Holzapfel (2013), and are based on the metrics used for the MIREX Note Tracking tasks for Music Information Retrieval: SymbTr (MIREX). Specifically, we consider a note to be correct if its F0 is within a +/-20 cent tolerance around the ground-truth pitch and its onset is within a 100ms tolerance, and use the proposed Precision, Recall, and F-measure metrics from Benetos & Holzapfel (2013), which are defined as follows:

$$\mathcal{P} = \frac{N_{tp}}{N_{sys}}, \quad \mathcal{R} = \frac{N_{tp}}{N_{ref}}, \quad \mathcal{F} = \frac{2\mathcal{R}\mathcal{P}}{\mathcal{R}+\mathcal{P}} \qquad (6)$$

where N_{tp} is the number of correctly detected notes, N_{sys} the number of notes detected by the transcription system, and N_{ref} the number of reference notes. Duplicate notes are considered as false alarms.

4.3 Results

We perform two types of evaluations, as in Benetos & Holzapfel (2013); The first uses the tonic automatically detected by the system of Bozkurt (2008), which attempts to automatically determine the frequency value in Hz of the tonic in a recording by comparing the pitch profile of the recording with a reference pitch profile for the makam. In the second evaluation, we use a manually annotated tonic. The proposed method is able to transcribe the 75min dataset in less than one hour, i.e. less than real time. Results comparing the proposed system with the system of Benetos & Holzapfel (2013) using the manually and automatically detected tonic can be seen in Tables 1 and 2, respectively. It can be seen that the F-measure increases by about 1.2% in both cases, indicating that the incorporation of prior information on pitch structure can improve transcription performance. This improvement is consistent for almost all 16 recordings, and is mostly evident for the subset of tanbur recordings (about 3% improvement). As in Benetos & Holzapfel (2013), there is a significant performance drop when comparing results with automatically detected tonic over the manually supplied one. This is attributed to the fact that the F0 tolerance for the evaluation is 20 cent, so even a slight tonic miscalculation might lead to a performance decrease.

In Tables 3 and 4, detailed results for ney, tanbur, and ensemble recordings can be seen, using manually aligned and automatically detected tonic, respectively. As in Benetos & Holzapfel (2013), the performance drops when the automatic tonic detection method is used. Likewise, the best results are reported for the subset of tanbur recordings, followed by the subset of ney recordings. The ensemble recordings, which additionally contain percussion along with heterophonic music, provide a greater challenge, with the F-measure reaching 47% for the case of manually annotated tonic.

	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
Benetos & Holzapfel (2013)	51.58%	52.85%	51.24%
Proposed system	52.72%	54.10%	52.42%

Table 1: Transcription results using manually annotated tonic.

	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
Benetos & Holzapfel (2013)	41.23%	42.07%	40.89%
Proposed system	44.44%	41.79%	42.09%

Table 2: Transcription onset-based results using automatically detected tonic.

5. DISCUSSION

The results we obtained when including pitch class profile information into our transcription system draw a slightly ambiguous picture. While we can observe a significant increase over the previous system for tanbur examples, for both ney and ensemble recordings no such conclusion can be drawn. This is astonishing since the pitch class profiles were derived by (Bozkurt, 2008) using a wide variety of different solo instrument recordings. Apparently, the fretted tanbur seems to be characterized by a more stable interval structure that fits well to the used profiles, while the pitch class profiles seem not to match with the ney recordings. It is an open question if this is due to the playing style or the variation between instruments. We will depict two examples in this section: One tanbur example, in which the inclusion of pitch class profiles proved to be of clear advantage (sample 8 in Table 1 of Benetos & Holzapfel (2013)), and a negative outlier from the set of ney recordings, which resulted in F-measures below 30%, independent from the usage of pitch class profiles (piece 10 on Table 1 of Benetos & Holzapfel (2013)). Audio of the depicted examples along with reference scores can be obtained from the second author's web-page [1] .

In Figures 4a and 4b the tonic (and its octave), and the dominant of the Rast makam are marked with dashed, and dotted lines, respectively. Black rectangles indicate the onsets of annotated notes, with the size of the rectangles determined by the given tolerances for pitch and timing inaccuracy. The red crosses designate the obtained annotations. It can be seen that the inclusion of pitch class profiles reduces the number of spurious notes and leads to a slightly

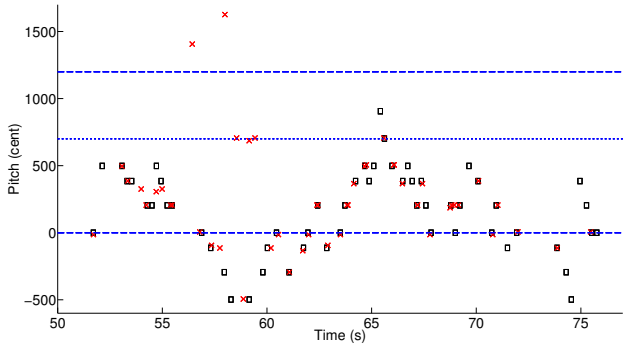
[1] www.rhythmos.org/shareddata/FMA2014

	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
Ney recordings	52.01%	49.25%	50.41%
Tanbur recordings	64.67%	51.67%	57.30%
Ensemble recordings	41.07%	58.37%	47.90%

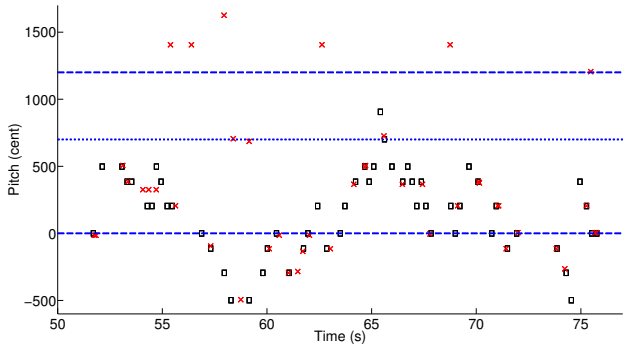
Table 3: Transcription onset-based results for each group of recordings, using manually annotated tonic.

	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
Ney recordings	47.16%	48.81%	46.62%
Tanbur recordings	48.42%	33.02%	39.14%
Ensemble recordings	35.20%	44.78%	38.95%

Table 4: Transcription onset-based results for each group of recordings, using automatically detected tonic.



(a) Transcription using pitch class profile



(b) Transcription without pitch class profile

Figure 4: Example for the transcription of the Teslim section of a Rast Peşrev.

more focused transcription. It is apparent, however, that the obtained result can only be considered a rough approximation of a precise transcription.

In Figure 5 we depict the allegedly negative outlier in our data, a Segah Peşrev played by a single ney. The astonishing insight provided by the figure is that the much lower F-measure for the Segah example (25.6% compared to 70.0% for the Tanbur example) is not related to a clearly worse automatic transcription. The detected notes are strongly related to the annotations, with one important exception. The fourth note of the Segah makam would be at an interval of approximately 500 cent above the tonic, which is a perfect fourth, and this interval was applied in the ground truth. However, the detected notes imply that this interval on the applied ney is sharper, with a size of about 550 cent, a fact not unusual for this interval in this specific makam in practice. Hence, it is apparent that in this case the dissent between theory and practice lead to an artificial underestimation of the actual quality of the transcription. In addition, the values of the metrics might have been influenced by the recording quality, since this recording is the only recording in the collection with a large reverberation, which further impedes a precise detection of note onsets in time.

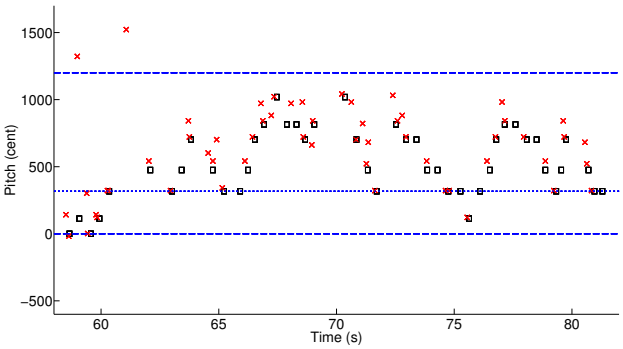


Figure 5: Transcription using pitch class profiles of the teslim section in a Segah Peşrev.

6. CONCLUSIONS

In this paper, we evaluated if the inclusion of pitch class profiles can improve the performance of a transcription system for Turkish makam music. The results imply that for solo instrument recordings of tanbur a consistent improvement can be achieved, when including this information about the interval structure of the makam. For the ney and for ensemble recordings, no increase in performance can be observed. Due to the limited number of included makams, it cannot be determined if this decrease is caused by the different type of instrument, or if makam with more variable interval structure have lead to this result. However, our discussion demonstrates that the values of the evaluation metrics might stand in no direct relation to the quality of the actual transcription, mainly due to the variable interval structure of this music. While the applied metrics represent a convention in the evaluation of AMT, the examples clarify that apparently objective measures should not be trusted blindly.

The quality of the achieved automatic transcriptions seems adequate as a starting point for a more precise manual transcription, or for a qualitative summary of the melodic content of the pieces, and can therefore serve as a method of practical value for the analysis of Turkish makam music. The achieved quality of the transcriptions is comparable to values achieved for Eurogenetic music. However, the demand of detecting notes apart from well-tempered tuning necessarily increases the search space for the notes to be detected, and the need to detect the frequency of the tonic represents another aspect that adds to the complexity. We

plan to improve our system by including pitch class profiles that are adapted to the piece at hand, and to adjust the ground truth annotations to the intervals encountered in a performance, in order to avoid the distortion of evaluation results that was observed in this paper.

7. ACKNOWLEDGEMENTS

Emmanouil Benetos is supported by a City University London Research Fellowship. Andre Holzapfel is supported by a Marie Curie Intra European Fellowship (PIEF-GA-2012-328379).

8. REFERENCES

Abraham, O. & von Hornbostel, E. M. (1994). Suggested methods for the transcription of exotic music. *Ethnomusicology*, *38*(3), 425–456. Originally published in German in 1909: "Vorschläge für die Transkription exotischer Melodien".

Benetos, E., Cherla, S., & Weyde, T. (2013). An efficient shift-invariant model for polyphonic music transcription. In *6th International Workshop on Machine Learning and Music*, Prague, Czech Republic.

Benetos, E. & Holzapfel, A. (2013). Automatic transcription of Turkish makam music. In *14th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 355–360)., Curitiba, Brazil.

Benetos, E., Jansson, A., & Weyde, T. (2014). Improving automatic music transcription through key detection. In *AES 53rd Conference on Semantic Audio*, London, UK.

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

Bozkurt, B., Ayangil, R., & Holzapfel, A. (2014). Computational analysis of makam music in Turkey: review of state-of-the-art and challenges. *Journal for New Music Research*, *43*(1), 3–23.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, *39*(1), 1–38.

England, N. M. (1964). Symposium on transcription and analysis: A Hukwe song with musical bow. *Ethnomusicology*, *8*(3), 223–277.

Karaosmanoğlu, K. (2012). A turkish makam music symbolic database for music information retrieval: Symbtr. In *ISMIR*.

Klapuri, A. & Davy, M. (Eds.). (2006). *Signal Processing Methods for Music Transcription*. New York.

MIREX. Music Information Retrieval Evaluation eXchange (MIREX). `http://music-ir.org/mirexwiki/`.

Schörkhuber, C. & Klapuri, A. (2010). Constant-Q transform toolbox for music processing. In *7th Sound and Music Computing Conference*, Barcelona, Spain.

Smaragdis, P. & Mysore, G. (2009). Separation by "humming": user-guided sound extraction from monophonic mixtures. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (pp. 69–72)., New Paltz, USA.

Smaragdis, P., Raj, B., & Shashanka, M. (2006). A probabilistic latent variable model for acoustic modeling. In *Neural Information Processing Systems Workshop*, Whistler, Canada.

HARMONY IN THE POLYPHONIC SONGS OF EPIRUS: REPRESENTATION, STATISTICAL ANALYSIS AND GENERATION

Maximos Kaliakatsos-Papakostas, Andreas Katsiavalos, Costas Tsougras, Emilios Cambouropoulos
School of Music Studies, Aristotle University of Thessaloniki, Greece
fmaxk, akatsiav, tsougras, emilios@mus.auth.gr

ABSTRACT

This paper examines a previously unstudied musical corpus derived from the polyphonic singing tradition of Epirus employing statistical methods. This analysis will mainly focus on unique harmonic aspects of these songs, which feature, for instance, unresolved dissonances (major second and minor seventh intervals) at structurally stable positions of the pieces (e.g. cadences). Traditional triadic tonal chord types are inadequate for this corpus' unconventional harmonic language; pc-set theoretic tools are too general/abstract. A novel chord representation has been devised that adapts to different non-standard tonal harmonic spaces. In the *General Chord Type (GCT)* representation, the notes of a harmonic simultaneity are re-arranged, depending on a given classification of intervals (that reflects culturally-dependent notions of consonance/dissonance), such that abstract idiom-specific types of chords may be derived. Based on these harmonic representations, statistical analyses are performed that provide insights regarding underlying harmony and, especially, on the idiosyncratic use of consonance/dissonance within this idiom. Then, characteristics of harmonic successions are examined via statistical analysis of the common chord transitions in the idiom. Finally, the learned statistical features are used to generate new harmonisations in the Epirus-song style for unseen Epirus-song melodies or for melodies from other distant idioms.

1. INTRODUCTION

The paper at hand paper examines a previously unstudied musical corpus derived from the polyphonic singing tradition of Epirus employing computational statistical methods. This analysis will mainly focus on unique harmonic aspects of these songs, which feature, for instance, unresolved dissonances (M2 and m7 intervals) at structurally stable positions of the pieces (e.g. cadences).

The statistical analysis over a corpus (Gjerdingen, 2014), mainly involves the definition of elements that are believed to occur within the corpus and the subsequent calculation of statistics on these elements, the norms of which indicate frequent relations and patterns within the corpus. Several studies with diverse orientations have been based on analysing the statistics on corpora. These studies can be divided in two categories: the "analytical" and the "compositional" approaches. Roughly, the analytical approaches aim to elucidate some facts about a studied corpus that remain ill-defined or unexplored in the domain of musical analysis. The compositional approaches on the other hand, are targeted towards utilizing the statistical values yielded by the corpus analysis and combine them with machine learning or other generative techniques, to compose novel music that complies with the yielded statistical properties of the corpus.

Among many other studies, the statistical analysis of music corpora has been performed for tracing the patterns of occurrence of specific musical phenomena, like the "cadenza doppia" (Gjerdingen, 2014) and Koch's metrical theory (Ito, 2014), or for providing additional insights about well–studied and popular idioms like rock music (de Clercq & Temperley, 2011; Temperley & de Clercq, 2013). Furthermore, through similar analytical techniques more detailed investigations on the functionality of perceptual mechanisms are allowed, like tracking the evolution of the major and minor keys in tonal music (Albrecht & Huron, 2014), measuring the historical development of musical style in relation to the cognition of key (White, 2014), analysing the perceptual functionality of tonality and meter in music (Prince & Schmuckler, 2014), or examining the insertion of preformed structures in improvisation (Norgaard, 2014). Statistical analysis on a corpus of a capella flamenco songs (Cabrera et al, 2008) has also been utilized to provide important tonal insights about this previously unstudied folk musical idiom, allowing thus a first ethnomusicological exploration of this idiom. Regarding the analytical part, the paper at hand contributes to the field of ethnomusicology by providing initial insights about statistical facts that concern the harmony in a corpus comprising several polyphonic songs of Epirus.

The polyphonic singing of Epirus, a region of northwestern Greece/southern Albania, is an intrinsically polyphonic Greek-language Balkan folk idiom, sharing features with similar Albanian, Aromanian and Vlachian idioms, and differing with them in language, structure and other vernacular aspects (Samson, 2013, p. 47-48; Ahmedaja, 2008; Kokkonis, 2008, p. 42-44). It is based on anhemitonic pentatonic modes and performed by 2-voice to 4-voice groups, with each voice having a specific musical and narrative role (Lolis, 2006; Kokkonis, 2008): leading solo voice (partēs [taker]), subsidiary solo voice (gyristēs [turner] or klōstēs [spinner]), drone group (isokratēs) and optionally, in 4-part polyphonic songs, the drone elaborator (richtēs [dropper]).

The present analysis focuses on the unique harmonic aspects of this polyphonic idiom, namely the structure and statistical distribution of the non-triadic sonorities and the use of unresolved dissonances (major second and minor

seventh intervals), especially at structurally stable positions of the pieces (e.g. cadences). The pentatonic pitch collection - described as pc set (0,2,4,7,9) - functions as source for both the melodic and harmonic content of the music. This analysis allows further musical hypotheses to be formed and conclusions to be drawn regarding harmonic mechanisms characteristic of the idiom.

Statistical analysis may provide confirmation of musicological hypotheses, comparing results with established musicological ground truth. For instance, in a statistical study on the harmony of the Bach Chorales (Prince & Schmuckler, 2014), the statistical analysis revealed chords with specific harmonic functionality; these results were aligned with the analytic musicological analysis encoded with roman numerals. However, for representing the harmonic elements of the non-standard examined idiom, the roman numerals or standard chord symbols cannot be employed, since chord simultaneities rarely comply with the standard major–minor categories. One can always resort to pc-set theory, however, in this study, the General Chord Type (GCT) representation Cambouropoulos et al, 2014) is utilized to represent this idiom's chord simultaneities that is more expressive than pc-sets – see brief description in next section.

There is extensive literature over the topic of automatic music composition by utilizing statistical quantities of a given corpus. Regarding these approaches, the statistical values are used either to straightforwardly provide estimations about the most probable future events (e.g. with variable length Markov models (Dubnov et al., 2003) or multiple view–points (Whorley et al., 2013) among others), or they are used as targeted fitness values to drive evolutionary processes (e.g. with genetic programming (Manaris et al., 2007) among others). The compositional part of the presented study discusses automatic melodic harmonisation through a machine learning technique which is based on the Hidden Markov Models (HMMs), namely the constrained HMMs (cHMMs) (Kaliakatsos-Papakostas et al., 2014). The presented compositional approach, utilizes in a straight forward manner the yielded statistical values to compose novel harmonies on given melodies, even if these melodies are not extracted from the studied idiom. Specifically, the harmonisations provided on melodies from the Bach chorales indicate that the "blended" musical result retains some important harmonic characteristics.

The present study examines only the vertical sonorities incorporated in the idiom, attempting to address the following questions: How can an unconventional musical style be encoded/represented so that meaningful computational analysis may be conducted? What are the most pronounced features of the harmonic language of polyphonic songs of Epirus? Which are the dominant pitches of the scale? Which is the frequency distribution of the sonorities? What constitutes a cadence phenomenon in this idiom? What is the dissonance's role, how freely is it employed and why/where? Finally, how might the statistical data obtained from the analysis be used creatively to generate epirus-style harmonisations of new melodies?

In the next section, the set of polyphonic songs used in this study is described and a novel chord representation, namely the *General Chord Type* representation, is employed to encode the song's verticalities. Then, the statistical methodology and results are presented along with musicological interpretations that unveil characteristic harmonic facts about the polyphonic style of Epirus. Finally, a constrained-HMM is used to generate novel melodic harmonisations in the style of the analysed songs. The paper concludes with remarks about the particular musical idiom and future directions of research.

2. DATASET AND *GCT* REPRESENTATION

2.1 The Epirus polyphonic song dataset

A collection of songs from the discussed idiom, transcribed from field recordings and notated in standard staff notation, is available in (Lolis 2006). From this collection, a dataset consisting of 22 songs (10 with 3 voices, 12 with 4) in minor pentatonic scale were selected; the minor pentatonic scale consists of pitch classes: [0,3,5,7,10] (for instance {A,C,D,E,G}). The 22 songs were manually segmented into 102 polyphonic phrases, while 37 monophonic phrases where excluded since the study revolves around the harmonic aspects of the idiom.

The songs were encoded at two reductional levels, corresponding to two adjacent levels of the metrical/time-span hierarchy: level *ms0* closely describes the musical surface by including embellishing figures, neighbour notes, etc. and corresponds to the metrical level of sixteenth or eighth notes (depending on the metrical tactus and beat level) of the transcription, while level *ms1* describes a deeper structure by omitting most elaborations and corresponds to the eighth or fourth notes of the transcription – see example in Figure 1. The lowest level encoding *ms0* consists of 1467 chords whereas the next reduction level *ms1* of 945 pitch class sets (chords); the pcs reduction ratio (ms1/ms0) of the polyphonic phrases is 0.64 (945 /1467 chord states). The reduction was deemed analytically necessary in order to disclose the idiom's harmonic functions and cadence patterns.

The songs were notated into music xml files; in addition to the original content, each xml file is annotated[1] with structural information using music notation that follows a specific formalism. More specifically, songs in the dataset consist of six staffs: two for the original song/content preserving voicing information (ms0), one for tonality annotations where the scale is written as a

[1] All annotations have been prepared by two of the authors.

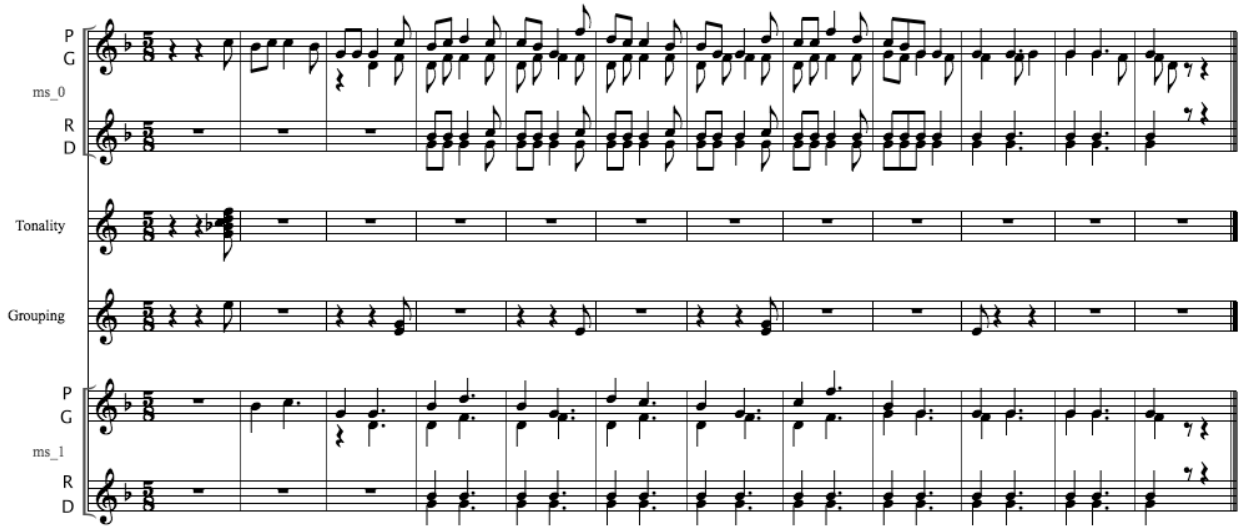


Figure 1. *Ammo ammo pigena,* polyphonic song from Epirus: Annotated xml file containing original song transcription (ms0), time-span reduction of the original content (ms1) as well as tonality and grouping information (see text). Symbols P, G, R and D denote the different voice groups (partēs, gyristēs, richtēs, drone).

note cluster, one for grouping boundaries where the number of notes indicates grouping level and two parts that contain an annotated time-span reduction of the original content (ms1) [see Figure 1].

The representations of the polyphonic songs were carried out without any transposition with the use of the GCT representation model. Afterwards, traditional note names or pc sets or other descriptions were employed when necessary. For ease of comprehension, the results of the statistical analysis and relevant figures/examples assume that the A minor pentatonic is used, i.e. A,C,D,E,G.

2.2 The GCT representation

The *General Chord Type (GCT)* representation (Cambouropoulos et al, 2014), allows the re-arrangement of the notes of a harmonic simultaneity such that abstract idiom-specific types of chords may be derived; this encoding is inspired by the standard roman numeral chord type labeling, but is more general and flexible.

Given a classification of intervals into consonant/dissonant (binary values) and an appropriate scale background (i.e. scale with tonic), the *GCT algorithm* computes, for a given multi-tone simultaneity, the 'optimal' ordering of pitches such that a maximal subset of consonant intervals appears at the 'base' of the ordering (left-hand side) in the most compact form. Since a tonal centre (key) is given, the position within the given scale is automatically calculated. For instance, the note simultaneity [C, D, F#,A] or [0,2,6,9] in a G major key is interpreted as [7,[0,4,7,10]] (appearing also in the figures for matter of space as 7.04710), i.e. as a dominant seventh chord.

The proposed representation is ideal for hierarchic harmonic systems such as the tonal system and its many variations, but adjusts to any other harmonic system such as the Epirus polyphonic system. In terms of the current paper, we have applied the GCT encoding for two different consonance vectors. The first is the standard vector for tonal music, i.e. consonant thirds/sixths and perfect fourths/fifths. The second vector, which is more adequate for 'atonal' music, includes additionally major seconds and minor sevenths (i.e., major seconds and minor sevenths are considered 'consonant' following the fact that in the Epirus songs these intervals are stable and require no resolution). In the example in Figure 2, we see the GCT encodings arising for the two different consonance vectors.

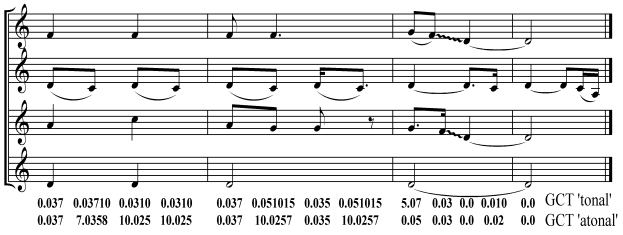


Figure 2. Excerpt from a traditional polyphonic song from Epirus. Top row: GCT encoding for standard common-practice consonance vector; bottom row: GCT encoding for atonal harmony – all intervals 'consonant' (this amounts to pc-set 'normal orders')

As can be seen in Figure 2, the two encodings are often different. Using the first consonance vector, the tonic is the root for almost all the GCTs (for A minor pentatonic we have tonic A: 85.39%). Using the second consonance vector, the chord's 'root' is more diverse (A: 61.58% , D: 12.27%, G: 22.85%) and additionally, since the consonance vector is less strict, the GCT chord encodings are more compact. Perhaps the most striking difference between the two encodings is this: the tonal-consonance-based labelling determines as the 'root' of almost all the chord types the *tonic* of the minor pentatonic scale which is actually the drone of each piece (notice that almost all

chord types on the upper row of Figure 2 have a 0 on the right before the period); the atonal-consonance based labelling gives 'roots' of chords that shift between the tonic (0) and the subtonic (10) (this is interesting as the dissonances that occur between the tonic and subtonic have a special role as will be discussed in the next section).

3. STATISTICAL ANALYSIS AND RESULTS

In this section a number of interesting statistical observations on the polyphonic Epirus dataset are presented. First, plain distributions of pitches and chord types are given and, then, conditional probability results that reflect transition regularities between chords are discussed. For ease of comprehension, whenever the GCT representation is not depicted, we assume that the A minor pentatonic is used, i.e. A,C,D,E,G (for readers not acquainted with the GCT, the traditional note names are easier to follow).

3.1 Pitch and chord distributions

The power set of the minor pentatonic scale has 30 different valid combinations (the empty set and the whole pentatonic set are excluded). The frequency of appearance of chords according to cardinality (number of pcs) is depicted in Figure 3.

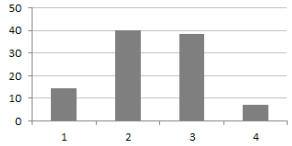


Figure 3. Frequency of chords according to cardinality (14.60% unison, 39.89% 2-note chords, 38.41% 3- note chords and 7.09% 4-note chords)

When single notes (i.e. unisons) appear in the polyphonic sections, 97.10% of these are the tonic (A). Unisons on other scale degrees are rare (<1%). From all 10 combinations with cardinality 2, dyads: AC, AD, AG are the most frequent (>25% each), AE is not so common, CE, GD, CG are rare (<1%) and DE, DG, EG are absent. For cardinality 3 (10 combinations), ACG is the most frequent (36.91%), ACE, ACD are common, ADG is less common, EAD and EGA are rare and CDE, CDG, CEG, DEG are absent. For cardinality 4 (5 combinations), ADGC is the most frequent of the cardinality class (67.16%) but rare in general (4.76%), ACEG and ACDE are rare and ADEG and CDEG are absent. Considering that from all the polyphonic phrases, almost half are taken from songs with 3 voices and also that it is common for one voice to double the drone tone, the cardinality percentages seem reasonable. Perhaps more 4-voice songs need to be added in the dataset to increase the frequency of higher cardinalities.

Since the drone tone (i.e. the tonic) is found in every chord, the absence of all the chords that don't contain it is also justified. However, from the 15 combinations of the power set of the scale that contain the tonic, only 4 from 6 chords with cardinality 3 are found (ADE and GAE don't exist) and only 1 from 4 with cardinality 4 (ACDE, ACEG, AEDG not found).

Apart from the quantitative difference in the frequency of appearance, the chord types found in the original surface (ms0) and the reduced musical surface (ms1) are identical (except chord AEDG that is found only 3 times in ms0 – not found in ms1). The statistical distribution of the dissonant sonorities (containing a major second) is almost identical at reductional levels ms0 and ms1, a feature indicating that dissonance is an integral/structural part of the harmonic idiom, and it is not reduced out as an elaborative event at the deeper level. On the contrary, this occurs in the tonal idiom, where most dissonant chords are the outcome of unstable non-chord tones and they are present mostly at the musical surface, not in reductions.

The appearance of the note E is rather rare. Considering that from the general percentage of all the existing chords that contain this note (19.68%), 8.99% corresponds to the only minor chord of the scale (ACE) and that another 5.18% corresponds to the perfect fifth interval (AE) from the tonal centre, it is clear that this note has relatively low usage.

Considering the pitch content of the chords, we can assume chord subsets and categories depending on various groupings relations. For example, from all the possible intervals that may appear in a chord in this pentatonic dataset (2,3,4,5,7,9,10 semitones) the most psychoacoustically dissonant is that of 2 and its inversion 10. Using this information we can create two groups of the most frequent chords depending on the presence of a 2-semitone interval: *Consonant* (A, AC, AD, AE, ACE) and *Dissonant* (AG, ACG, ADG, ADGC, ACD). Furthermore, we can split the *Dissonant* group according to the quantity of the dissonant intervals present in the chord and the quality of the pitch classes that comprise it. The *Dissonant* set may be split in three subsets: {ACD}, that contains the CD dissonant interval not involving the tonic A, {ACG, ADG, AG} that contain dissonant interval GA involving the tonic and {ADGC} that contains both 2-semitone intervals GA & CD.

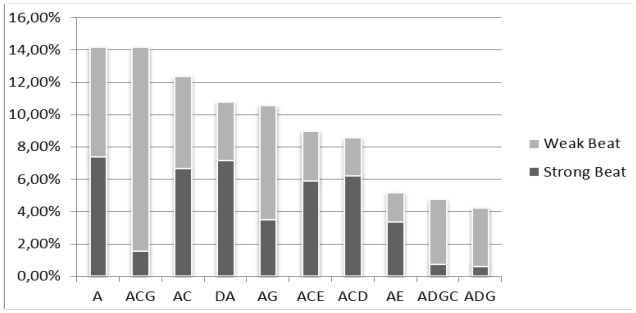


Figure 4. Frequency of appearance of the chords in the dataset (applying a threshold of >2%, 887/945 states, 6.14% reduction) appearing on strong beats (first beat of measure) and in weak beats (all other measure positions).

In Figure 4, the most common simultaneities are depicted according to frequency of occurrence on strong (first beat in measure) or weak beats (N.B. most of the Epirus polyphonic songs have a metre of 2/4, 3/8 or 5/8). Perhaps the most interesting finding in this figure is the fact that dissonant chords containing the interval AG (dissonance

with tonic) appear mostly on weak beats (not the first beat in the measure), whereas dissonant chord ADC containing interval CD (dissonance not with tonic) appears mostly on the strong beat. This fact shows a structural difference in the use of the two types of dissonant chords. The dissonant interval appearing between the tonic A and subtonic G is very common but seems to be a kind of 'colouring' of standard consonances that appear on strong beats (see below dissonance used as cadence extension). On the contrary, the CD dissonance is considered probably milder as it does not clash with the tonic and appears more often on strong beats.

Concerning the phrase positioning of chords, the most common final chords of final phrases are unison A, AG, AC, ACG, and AE (GCT: 0.0, 0.010, 0.03, 0.0310, 0.07) – see Figure 5. In the next section the use of dissonance at cadences is discussed more extensively. There is no significant preference for beginning chords.

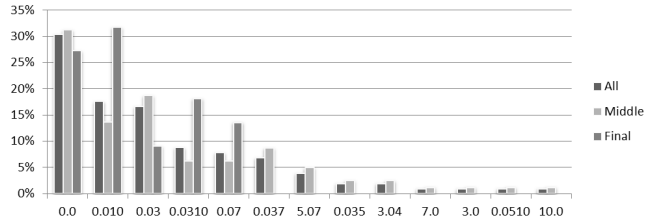


Figure 5. Frequency of appearance of the last chords of all (102), middle (non-final) & final phrases (22).

3.2 Chord transition statistics

A first order transition table reveals the general and immediate chord transition preferences of the chords. Utilizing this information we can track previous and next chords for single chords and for groups as well. The generated first-order transition matrix is presented in Appendix I. In Figure 6, the transition matrix of the Appendix I is converted in a different form that is simpler to read. The current chord is placed in the middle column; on the left we see the previous chords and on the right the following chord; distance from the middle column represents strength of transition (i.e. higher probability).

Variable-length higher-order transition probabilities are additionally being calculated using Prediction Suffix Trees (Ron et al., 1996). Such trees reveal common patterns occurring in the dataset. We have used such a PST to examine cadences in more detail. In Table 1, the most common chords appearing on phrase endings are depicted. Even though dissonances GA and GAC appear as last chords of phrases, we consider them as chord extensions assuming that the actual cadence end point is the unison on the tonic (A) or tonic plus third (AC); this assumption is supported by the fact that the last word of the songs' lyrics usually ends on a strong metric position consonance whereas the dissonance is introduced immediately after on a weaker beat as a kind of 'tension colouring' that requires no preparation or resolution. Approximately 30% of cadences ending in unison A or AC are extended with the addition of G creating a dissonance. More than half of the chords preceding the final chord in the cadence embody dissonances.

Common cadence chord transitions		
Semi-final chord	*Final chord*	*Extension*
GAC (12+2)	A (33+14+1)	GA (14) GAC (1)
GA (8+3)		
AC (3+6)		
AD (5+3)		
AE (2+1)		
GAC (6)	AC (17+2+4)	GAC (4+2)
AD (3+1)		
ACD (3+1)		
ACDG (3)		

Table 1. Most common chords appearing on phrase endings (chords appearing less than three times have been omitted). See text for details.

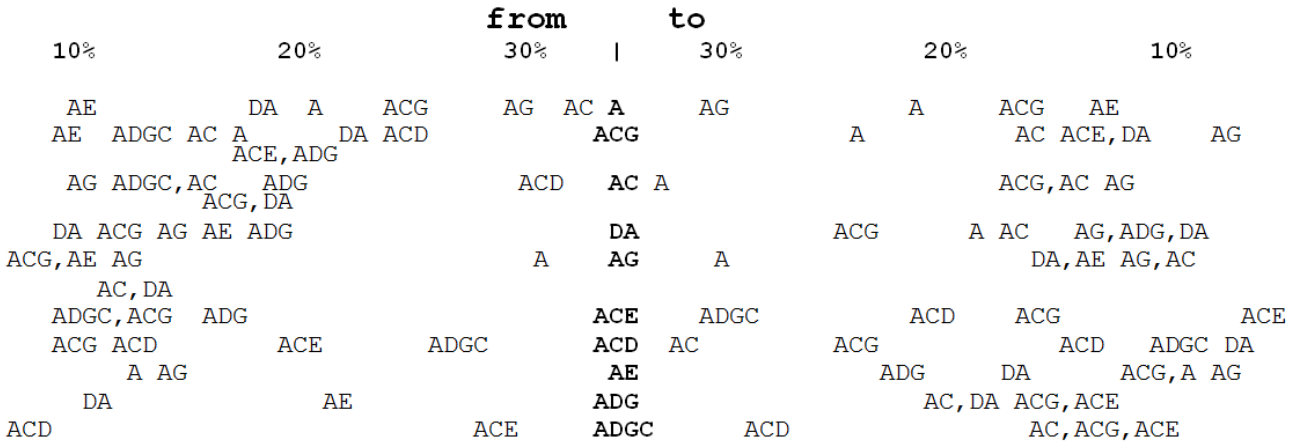


Figure 6. The transition map of the most frequent chords in the dataset in a graphical view.

In Figure 7 two prototypical cadential patterns appearing in this idiom are illustrated. Dissonance usually appears

just after the consonant downbeat final chord as a kind of cadence extension that adds a final brushstroke of tension.

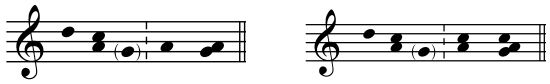


Figure 7. Prototypical cadences in polyphonic songs from Epirus.

Also, the high occurrence rate of the cadential pattern in this idiom and the fact that pitch class E is mostly absent in it possibly explains this pc's relatively low statistical appearance (see 3.1).

4. MELODIC HARMONISATION

The statistical results hitherto discussed provide some indications about harmonic regularities that are encountered in the examined idiom. The utilization of these statistics for melodic harmonisation are expected to produce harmonisations – in a statistical manner – that preserve these characteristics, even if the melody to be harmonised is not characteristic of this idiom. The statistical information that is utilized in the short examples presented in this section concern the chord–to–chord transitions, as reflected by the first order transition maps discussed in Section 3. To this end, a modification of the first order hidden Markov model (HMM) is utilized, namely the constrained-HMM (cHMM), which was presented in (Kaliakatsos-Papakostas et al., 2014). The cHMM methodology follows the typical HMM methodology, with the difference that specific chord constraints can be set by the user or by another algorithmic process at any point of the melody to be harmonised. From a user perspective, the cHMM methodology allows the user to choose specific beginning, intermediate or ending chords or fixed chord progressions (e.g. cadences), while the remaining harmonic content is filled with a probabilistic framework based on the HMM methodology.

The cHMM methodology requires statistical information from the corpus and deterministic information from the melody to be harmonised. The statistical information from the corpus concerns chord (or states in the HMM nomenclature) transitions, specifically the first order transitions under the Markov assumption (that the appearance of a chord depends only on its previous chord). The deterministic information, on one hand, involves chord-to-note (observation) rules[1] and, on the other, the fixed-chord constraints. Regarding the chord-to-note rules, a chord is considered "probable" to harmonise a given note of the melody if the melodic note's pitch class is a member of the chord's pitch class. In the presented application of the cHMM, the fixed-chord constraints are provided by the user (one of the authors), according to the harmonic checkpoints that are desired – primarily regarding the preservation of cadence characteristics. For a more detailed overview of the cHMM specifications, the interested reader is referred to (Kaliakatsos-Papakostas et al., 2014).

The results presented below discuss the harmonisation of some melodies, utilizing the corpus statistics for the first order Markov transitions between chords. Specifically, the transition tables are extracted from 30 random phrases of the pieces in the Epirus song corpus, in order to be aligned with scalability issues that demand efficient training with limited numbers of training data. Automated melodic harmonisation is performed on melodies of Epirus songs as well as on melodies from completely different idioms. It has to be noted that the Epirus songs that are re-harmonised, are not included in the 30 random phrases that comprised the training set. The aim of the results is to demonstrate the adequacy of these statistics regarding melodic harmonisation, not only in the context of melodies that are typical of the idiom, but also for less typical melodies.

The harmonisation in the polyphonic style of Epirus of three melodies is illustrated in Figures 8, 9 & 10. The first melody is the melody of the Epirus polyphonic song depicted in Figure 1, whereas the remaining two from different idioms, namely, J.S.Bach's *Chorale nr. 110* ("Vater unser im Himmelreich") melody and the second part of G. Gershwin's *Summertime* melody; all three are built on the minor pentatonic scale. In each of these figures, the melody is illustrated on the top, accompanied by the chords (in GCT encoding) generated by the aforementioned cHMM model that was trained on the polyphonic songs of Epirus. Below each melody, a full harmonisation following the idiom's voice leading is presented (created manually by one of the authors).

The harmonisation of the Epirus melody is very close to the original transcription depicted in Figure 1; this shows that the cHMM model is quite good at generating chords in the learned style (NB., this melody was not included in the training dataset). The generated harmonisations of the other two melodies, preserve characteristics of the polyphonic style of Epirus and have a unique harmonic flavour which shows that blending between diverse musical materials may give interesting original outcomes.

The cHHM constraints imposed to all three pieces related to chords of cadences; namely, for all pieces the final two chords were respectively: [0,[0,3]] and [0,[0,3],[10]] in GCT notation. When the cadence constraints were omitted, the generated cadence in most cases included two repetitions of the tonic unison which is also very common in the idiom (expressed with the GCT [0,[0]]). As can be seen, many characteristics of the idiom are preserved such as the used chord types, the progression of chords,

[1] The typical HMM methodology incorporate probabilistic rules that relate observations (melody notes) with states (chords).

the interchange of consonance-dissonance and the drone tone (common root in almost all GCTs); there are however, characteristics that are missed out in the current model such as the appearance of specific types of chords (e.g. chord with dissonances) on particular metric positions. The generation model requires embedding in a more general compositional framework that takes into account voice-leading, metrical structure, bass-line motion and so on. It is encouraging that such a simple learning model (cHMM) coupled with the GCT representation captures significant components of the harmonic language of the idiom.

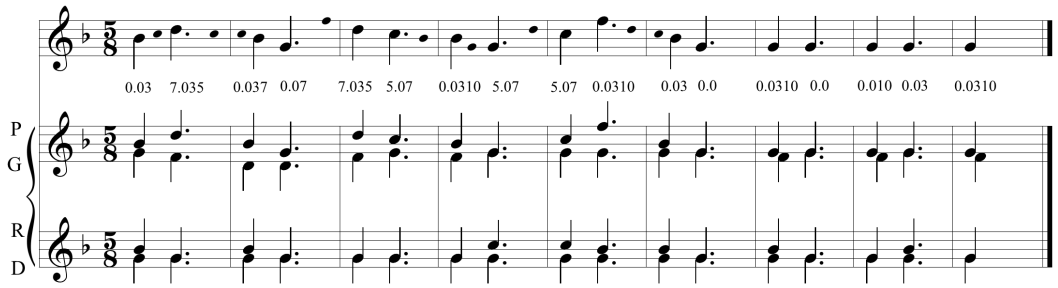


Figure 8. Melody of *Ammo ammo pigena* polyphonic song from Epirus harmonised by the cHMM model (GCT labels under the melody)

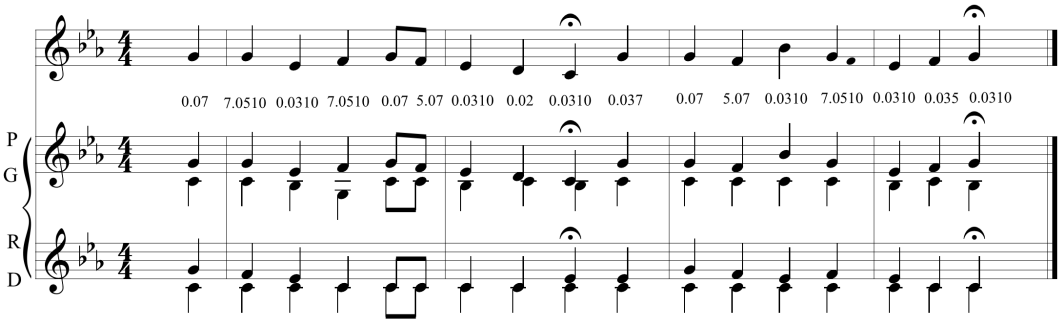


Figure 9. First two melodic phrases from J.S.Bach's *Chorale nr. 110* ("Vater unser im Himmelreich"), harmonised by the cHMM model (GCT labels under the melody)

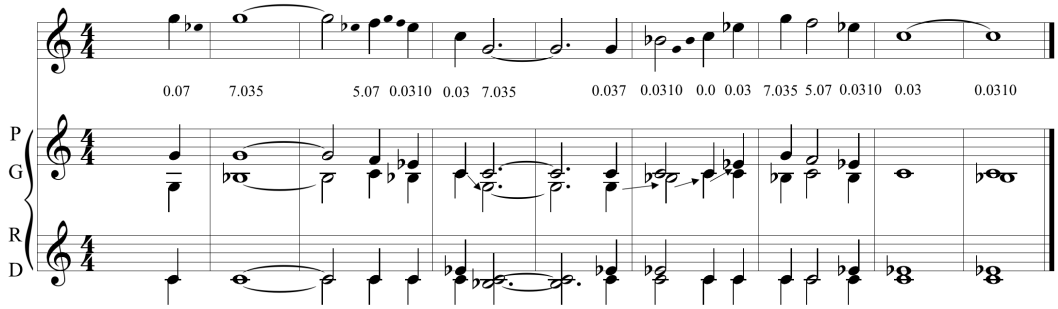


Figure 10. Second phrase from G. Gershwin's *Summertime*, harmonised by the cHMM model (GCT labels under the melody). The arrows indicate that the main melody was temporarily transferred to the *gyristēs* voice.

5. CONCLUSIONS

In this paper a study has been presented of a musical corpus derived from the polyphonic singing tradition of Epirus that employs computational statistical methods. The analysis focused primarily on harmonic aspects of these songs. Representing a non-tonal idiom in terms of harmonic content is non-trivial. A novel chord representation, namely the *General Chord Type (GCT)* representation, has been presented that adapts to different non-standard tonal harmonic spaces depending on a given classification of intervals that reflects culturally-dependent notions of consonance/dissonance. It has been shown that the GCT is appropriate for encoding note simultaneities in the given idiom and that different consonance vectors give rise to different labelings of chords that embody alternative facets of core harmonic concepts in the idiom (e.g. drone or tonic-subtonic relation).

Using the GCT as basis, statistical analytic tools have been employed that highlight characteristic distributional and transitional properties of chords in the idiom. It has been shown that some note simultaneities are more common than others, that 'dissonances' (major 2nds and minor 7ths) are structurally relatively 'stable' as they appear

in higher reductional levels and also on relatively strong metrical positions, and that certain cadential patterns seem to emerge that describe harmonic content at phrase endings. The transitional patterns learned from a section of the dataset (using the GCT representation and the cHHM methodology) are used to generate new harmonisations for melodies in the style of the studied idiom but also on melodies drawn from other foreign styles (e.g. Bach chorale melody and Gershwin's *Summertime* melody); the new harmonisations preserve qualities of the analysed idiom creating novel musical creations.

6. ACKNOWLEDGMENTS

The project COINVENT acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 611553.

7. REFERENCES

Albrecht, J. D. & Huron, D. (2014). A statistical approach to tracing the historical development of major and minor pitch distributions, 1400–1750. *Music Perception*, 31(3), 223–243.

Ahmedaja, Ardian (2008). "Changes within Tradition: Parts and their Number in Multi-part Songs among Albanians". In *European Voices I - Multipart Singing in the Balkans and the Mediterranean*, ed. A. Ahmedaja, G. Haid. Vienna: Böhlau Verlag.

Cabrera, J. J., Díaz-Báñez, J. M., Escobar-Borrego, F. J., Gómez, E., & Mora, J. (2008). Comparative melodic analysis of a cappella flamenco cantes. In *Fourth Conference on Interdisciplinary Musicology (CIM08),* Thessaloniki, Greece.

Cambouropoulos, E., Kaliakatsos-Papakostas, M., & Tsougras, C. (2014). An Idiom–independent Representation of Chords for Computational Music Analysis and Generation. In Proceeding of the joint 11th Sound and Music Computing Conference (SMC) and 40th International Computer Music Conference (ICMC), (Submitted), Athens, Greece.

de Clercq, T. & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music*, 30, 47–70.

Dubnov, S., Assayag, G., Lartillot, O., &Bejerano, G. (2003). Using machine-learning methods for musical style modeling. *Computer*, 36(10), 73–80.

Gjerdingen, R. O. (2014). "historically" informed corpus studies. *Music Perception*, 31(3), 192–204.

Ito, J. P. (2014). Kochs metrical theory and mozarts music: A corpus study. *Music Perception*, 31(3), 205–222.

Kaliakatsos-Papakostas, M. & Cambouropoulos, E. (2014). Probabilistic harmonisation with fixed intermediate chord constraints. In Proceeding of the Joint 11th Sound and Music Computing Conference (SMC) and 40th International Computer Music Conference (ICMC), (Submitted), Athens, Greece.

Kokkonis, Yorgos (2008). *Mousikē apo tēn Hēpeiro* [Music from Epirus]. Athens: Greek Parliament Foundation.

Lolis, Kostas (2006). *To Hēpeirōtiko Polyphōniko Tragoudi* [Epirus Polyphonic Song]. Ioannina.

Manaris, B., Roos, P., Machado, P., Krehbiel, D., Pellicoro, L., & Romero, J. (2007). A corpus-based hybrid approach to music analysis and composition. In *Proceedings of the 22nd national conference on Artificial intelligence* - Volume 1, (pp. 839–845). AAAI Press.

Norgaard, M. (2014). How jazz musicians improvise: The central role of auditory and motor patterns. *Music Perception,* 31(3), 271–287.

Prince, J. B. & Schmuckler, M. A. (2014). The tonal-metric hierarchy: A corpus analysis. *Music Perception*, 31(3), 254–270.

Rohrmeier, M., & Cross, I. (2008). Statistical properties of tonal harmony in Bach's chorales. In Proceedings of the 10th international conference on music perception and cognition (pp. 619-627).

Ron, D., Singer, Y., Tishby, N. (1996) The Power of Amnesia: Learning Probabilistic Automata with Variable Memory Length. *Machine Learning*, Vol. 25, pp. 117–149.

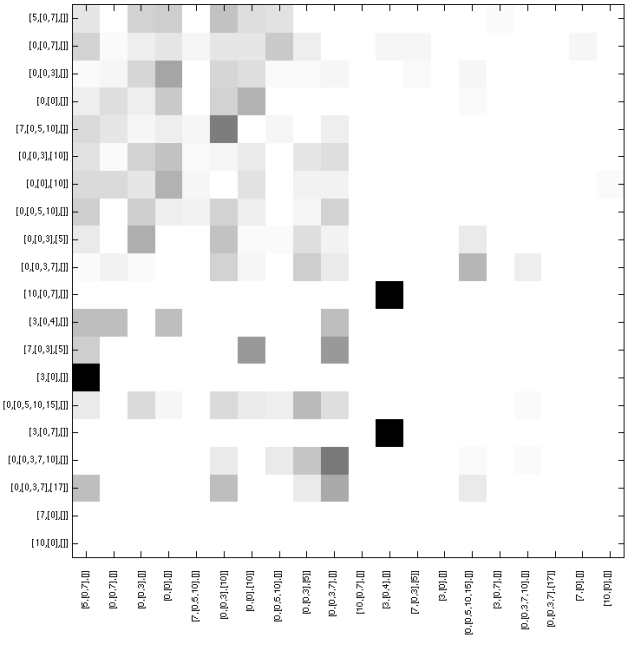
Samson, Jim (2013). *Music in the Balkans*. Leiden: Brill.

Temperley, D. & de Clercq, T. (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research*, 42(3), 187–204.

White, C. W. (2014). Changing styles, changing corpora, changing tonal models. *Music Perception*, 31(3), 244–253.

Whorley, R. P.,Wiggins, G. A., Rhodes, C., & Pearce, M. T. (2013). Multiple viewpoint systems: Time complexity and the construction of domains for complex musical viewpoints in the harmonization problem. *Journal of New Music Research*, 42(3), 237–266.

Appendix I. First-order transition matrix of GCT simultaneities for the Epirus polyphonic song dataset



USING POINT-SET COMPRESSION TO CLASSIFY FOLK SONGS

David Meredith
Aalborg University
dave@create.aau.dk

ABSTRACT

Thirteen different compression algorithms were used to calculate the normalized compression distances (NCDs) between pairs of tunes in the Annotated Corpus of 360 Dutch folk songs from the collection *Onder de groene linde*. These NCDs were then used in conjunction with the 1-nearest-neighbour algorithm and leave-one-out cross-validation to classify the 360 melodies into tune families. The classifications produced by the algorithms were compared with a ground-truth classification prepared by expert musicologists. Twelve of the thirteen compressors used in the experiment were based on the discovery of translational equivalence classes (TECs) of maximal translatable patterns (MTPs) in point-set representations of the melodies. The twelve algorithms consisted of four variants of each of three basic algorithms, COSIATEC, SIATECCOMPRESS and Forth's algorithm. The main difference between these algorithms is that COSIATEC strictly partitions the input point set into TEC covered sets, whereas the TEC covered sets in the output of SIATECCOMPRESS and Forth's algorithm may share points. The general-purpose compressor, bzip2, was used as a baseline against which the point-set compression algorithms were compared. The highest classification success rate of 77–84% was achieved by COSIATEC, followed by 60–64% for Forth's algorithm and then 52–58% for SIATECCOMPRESS. When the NCDs were calculated using bzip2, the success rate was only 12.5%. The results demonstrate that the effectiveness of NCD for measuring similarity between folk-songs for classification purposes is highly dependent upon the actual compressor chosen. Furthermore, it seems that compressors based on finding maximal repeated patterns in point-set representations of music show more promise for NCD-based music classification than general-purpose compressors designed for compressing text strings.

1. INTRODUCTION

For over a century, musicologists have been interested in measuring similarity between folk song melodies (Scheurleer, 1900; van Kranenburg et al., 2013), primarily with the purpose of classifying such melodies into *families* (Bayard, 1950) of tunes that have a common ancestor in the tree of oral transmission. Researchers have used a plethora of different features and methods in their attempts to automate (or at least formalize) this process of folk-song classification (see van Kranenburg et al., 2013, for an overview). In some cases, such methods have led to almost perfect models of the classifications produced by expert musicologists. For example, van Kranenburg et al. (2013) report a 99% success rate for classifying a set of 360 Dutch folk songs with a method based on local-features and string alignment. In contrast, in the study reported here, a universal, generally-applicable similarity metric, *normalized compression distance* (NCD, Li et al., 2004), is used to classify folk-song melodies based on compressing the melodies by discovering maximal repeated patterns within them.

Normalized compression distance has been used in several music classification studies in the past (Cilibrasi et al., 2004; Li & Sleep, 2004, 2005; Hillewaere et al., 2012). However, in these studies, only general-purpose compressors such as those based on the Lempel-Ziv algorithm (Ziv & Lempel, 1977, 1978) and bzip2 (Seward, 2010) have been used. In the study reported here, NCD was used to classify folk songs using a number of different compression algorithms specifically designed for producing compact structural analyses of pieces of music from symbolic encodings in the form of point sets (Meredith et al., 2003; Meredith, 2006a; Forth & Wiggins, 2009; Forth, 2012; Meredith, 2013). The results suggest that the choice of compressor has a very marked effect on the classification success rate.

2. NORMALIZED COMPRESSION DISTANCE

Li et al. (2004) introduced the *normalized information distance* (NID), a universal similarity metric based on *Kolmogorov complexity* (Li & Vitányi, 2008). The Kolmogorov complexity of any object is the length in bits of the shortest program that generates the object as its only output. The NID defines the distance between any two objects, x and y, as

$$d(x,y) = \frac{\max\{K(x \mid y^*), K(y \mid x^*)\}}{\max\{K(x), K(y)\}}$$

where $K(x)$ is the Kolmogorov complexity of x and $K(x \mid y^*)$ is the conditional complexity of x given a description of y whose length is equal to the Kolmogorov complexity of y. The Kolmogorov complexity of an object, however, is not computable. Therefore, $K(x)$ has to be substituted in practice by the length of a compressed encoding of x generated using a real-world compressor. Li et al. (2004) therefore propose the *normalized compression distance* (NCD) as a practical alternative to the NID and define it as follows:

$$\mathrm{NCD}(x,y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}$$

where $C(x)$ is the length of a compressed encoding of x and $C(xy)$ is the length of a compressed encoding of a concatenation of x and y.

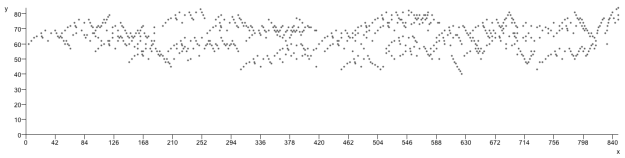


Figure 1: An example of a dataset. A two-dimensional point-set representing the fugue from J. S. Bach's Prelude and Fugue in C minor, BWV 846. The horizontal axis represents onset time in tatums; the vertical axis represents morphetic pitch. Each point represents a note or a sequence of tied notes.

3. REPRESENTING MUSIC WITH POINT SETS

In the algorithms considered in this paper, it is assumed that the music to be analysed is represented in the form of a multi-dimensional point set called a *dataset*, as described by Meredith et al. (2002). All the algorithms described below work with datasets of any dimensionality. However, it will be assumed here that each dataset is a set of two-dimensional points, $\langle t, p \rangle$, on an integer lattice, where t and p are, respectively, the onset time in tatums and the chromatic or morphetic pitch (Meredith, 2006b, 2007; Meredith et al., 2002) of a note or sequence of tied notes in a score. Figure 1 shows an example of such a dataset. When the music to be analysed is modal or uses the major-minor tonal system, the output of the algorithms described below is typically better when morphetic pitch is used. If morphetic pitch information is not available (e.g., because the data is only available in MIDI format), then, for modal or tonal music, it can be reliably computed from a representation that provides the chromatic pitch (i.e., MIDI note number) of each note, by using an algorithm such as PS13s1 (Meredith, 2006b, 2007). For pieces of music not based on the modal or major-minor tonal system, using chromatic pitch may give better results than using morphetic pitch.

4. MAXIMAL TRANSLATABLE PATTERNS

If D is a dataset, then any subset of D may be called a *pattern*. If $P_1, P_2 \subseteq D$, then P_1, P_2, are said to be *translationally equivalent*, denoted by $P_1 \equiv_{\mathrm{T}} P_2$, if and only if there exists a vector v, such that P_1 translated by v is equal to P_2. That is,

$$P_1 \equiv_{\mathrm{T}} P_2 \iff (\exists v \mid P_2 = P_1 + v)\,, \qquad (1)$$

where $P_1 + v$ denotes the pattern that results when P_1 is translated by the vector v. For example, in each of the graphs in Figure 2, the pattern of circles is translationally equivalent to the pattern of crosses. A pattern, $P \subseteq D$, is said to be *translatable* within a dataset, D, if and only if there exists a vector, v, such that $P + v \subseteq D$. Given a vector, v, then the *maximal translatable pattern* (MTP) for v in the dataset, D, is defined and denoted as follows:

$$\mathrm{MTP}(v, D) = \{p \mid p \in D \wedge p + v \in D\} \qquad (2)$$

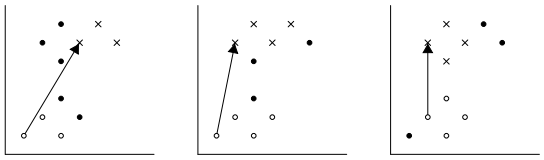


Figure 2: Examples of maximal translatable patterns (MTPs). In each graph, the pattern of circles is the maximal translatable pattern (MTP) for the vector indicated by the arrow. The pattern of crosses in each graph is the pattern onto which the pattern of circles is mapped by the vector indicated by the arrow.

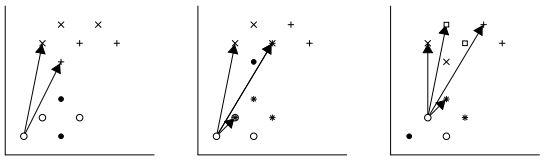


Figure 3: Examples of translational equivalence classes (TECs). In each graph, the pattern of circles is translatable by the vectors indicated by the arrows. The TEC of each pattern of circles is the set of patterns containing the circle pattern itself along with the other patterns generated by translating the circle pattern by the vectors indicated. The covered set of each TEC is the set of points denoted by icons other than filled black dots.

where $p + v$ is the point that results when p is translated by the vector v. Figure 2 shows some examples of maximal translatable patterns.

5. TRANSLATIONAL EQUIVALENCE CLASSES

When analysing a piece of music, we typically want to find *all the occurrences* of an interesting pattern, not just one occurrence. Thus, if we believe that MTPs are related in some way to the patterns that listeners and analysts find interesting, then we want to be able to find all the occurrences of each MTP. Given a pattern, P, in a dataset, D, the *translational equivalence class* (TEC) of P in D is defined and denoted as follows:

$$\mathrm{TEC}(P, D) = \{Q \mid Q \equiv_{\mathrm{T}} P \wedge Q \subseteq D\}\,. \qquad (3)$$

That is, the TEC of a pattern, P, in a dataset contains all and only those patterns in the dataset that are translationally equivalent to P. Figure 3 shows some examples of TECs.

We define the *covered set* of a TEC, T, denoted by $\mathrm{COV}(T)$, to be the union of the patterns in the TEC, T. That is,

$$\mathrm{COV}(T) = \bigcup_{P \in T} P\,. \qquad (4)$$

Here, we will be particularly concerned with *MTP TECs*—that is, the translational equivalence classes of the maximal

translatable patterns in a dataset.

A TEC, $T = \text{TEC}(P, D)$, contains all the patterns in the dataset, D, that are translationally equivalent to the pattern, P. Suppose T contains n translationally equivalent occurrences of the pattern, P, and that P contains m points. There are at least two ways in which one can specify T. First, one can explicitly specify each of the n patterns in T by listing all of the m points in each pattern. This requires one to write down mn, k-dimensional points or kmn numbers. Alternatively, one can explicitly list the m points in just one of the patterns in T (e.g., P) and then give the $n-1$ vectors required to translate this pattern onto its other occurrences in the dataset. This requires one to write down m, k-dimensional points and $n-1$, k-dimensional vectors—that is, $k(m+n-1)$ integers. If n and m are both greater than one, then $k(m+n-1)$ is less than kmn, implying that the second method of specifying a TEC gives us a *compressed* encoding of the TEC. Thus, if a dataset contains at least two occurrences of a pattern containing at least two points, it will be possible to encode the dataset in a compact manner by representing it as the union of the covered sets of a set of TECs, where each TEC, T, is encoded as an ordered pair, $\langle P, V\rangle$, where P is a pattern in the dataset, and V is the set of vectors that translate P onto its other occurrences in the dataset. When a TEC, $T = \langle P, V\rangle$, is represented in this way, we call V the *set of translators* for the TEC and P the TEC's *pattern*. We also denote and define the *compression ratio* of a TEC, $T = \langle P, V\rangle$ as follows:

$$\text{CR}(T) = \frac{|\text{COV}(T)|}{|P| + |V|}\,. \tag{5}$$

In this paper, the pattern, P, of a TEC used to encode it as a $\langle P, V\rangle$ pair will be assumed to be the lexicographically earliest occurring member of the TEC (i.e., the one that contains the lexicographically least point).

6. THE ALGORITHMS

6.1 SIA

All of the compression algorithms considered in this paper are based on Meredith, Lemström and Wiggins' SIA algorithm (Meredith et al., 2001, 2002, 2003; Meredith, 2006a). [1] SIA finds all the maximal translatable patterns in a set of n, k-dimensional points in $\Theta(kn^2 \lg n)$ time and $\Theta(kn^2)$ space. Figure 4 describes how the algorithm works with a simple example and Figure 5 gives pseudocode for a straight-forward implementation of SIA. In the pseudocode used in this paper, unordered sets are denoted by italic upper-case letters (e.g., D in Figure 5). Ordered sets are denoted by boldface upper-case letters (e.g., $\mathbf{V}$, $\mathbf{D}$ and $\mathbf{M}$ in Figure 5). When written out in full, ordered sets are denoted by angle brackets, "$\langle\cdot\rangle$". Concatenation is denoted by "$\oplus$" and the assignment operator is "$\leftarrow$". $\mathbf{A}[i]$ denotes the $(i+1)$th element of the ordered set (or one-dimensional array), $\mathbf{A}$, (i.e., zero-based indexing is used). If $\mathbf{B}$ is an ordered set of ordered sets (or a two-dimensional array), then

[1] SIA stands for "Structure Induction Algorithm".

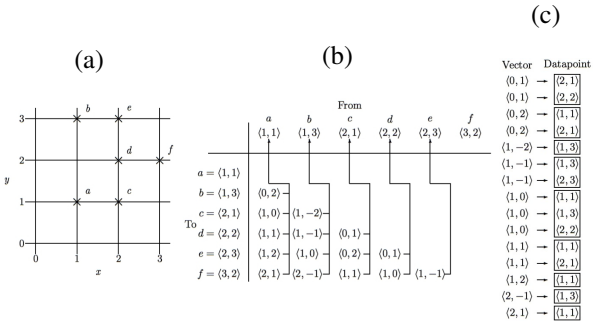


Figure 4: The SIA algorithm. (a) A small dataset that could be provided as input to SIA. (b) The vector table computed by SIA for the dataset in (a). Each entry in the table gives the vector from a point to a lexicographically later point. Each entry has a pointer back to the origin point used to compute the vector. (c) The list of $\langle$vector, point$\rangle$ pairs that results when the entries in the vector table in (b) are sorted into lexicographical order. If this list is segmented at points at which the vector changes, then the set of points in the entries within a segment form the MTP for the vector for that segment.

```
SIA(D)
1   D ← SORT_Lex(D)
2   V ← ⟨⟩
3   for i ← 0 to |D| − 2
4     for j ← i + 1 to |D| − 1
5       V ← V ⊕ ⟨⟨D[j] − D[i], i⟩⟩
6   V′ ← SORT_Lex(V)
7   M ← ⟨⟩
8   v ← V′[0][0]
9   P ← ⟨D[V′[0][1]]⟩
10  for i ← 1 to |V′| − 1
11    if V′[i][0] = v
12      P ← P ⊕ ⟨D[V′[i][1]]⟩
13    else
14      M ← M ⊕ ⟨⟨P, v⟩⟩
15      v ← V′[i][0]
16      P ← ⟨D[V′[i][1]]⟩
17  M ← M ⊕ ⟨⟨P, v⟩⟩
18  return M
```

Figure 5: Pseudocode for a straight-forward implementation of SIA.

$\mathbf{B}[i][j]$ denotes the $(j+1)$th element in the $(i+1)$th element of $\mathbf{B}$. Elements in arrays of higher dimension are indexed analogously. Block structure is indicated by indentation alone.

The algorithm can easily be modified so that it only generates MTPs whose sizes lie within a particular user-specified range. It is also possible for the same pattern to be the MTP for more than one vector. If this is the case, there will be two or more $\langle$pattern, vector$\rangle$ pairs in the output of SIA that have the same pattern. This can be avoided and the output can be made more compact by generating instead a list of $\langle$pattern, vector set$\rangle$ pairs, such that the vector set in each pair contains all the vectors for which the pattern is an MTP. In order to accomplish this, we merge the vectors for which a given pattern is the MTP into a single vector set which is then paired with the pattern in the output.

		From					
		$a = \langle 1,1 \rangle$	$b = \langle 1,3 \rangle$	$c = \langle 2,1 \rangle$	$d = \langle 2,2 \rangle$	$e = \langle 2,3 \rangle$	$f = \langle 3,2 \rangle$
	$a = \langle 1,1 \rangle$	$\langle 0,0 \rangle$	$\langle 0,-2 \rangle$	$\langle -1,0 \rangle$	$\langle -1,-1 \rangle$	$\langle -1,-2 \rangle$	$\langle -2,-1 \rangle$
	$b = \langle 1,3 \rangle$	$\langle 0,2 \rangle$	$\langle 0,0 \rangle$	$\langle -1,2 \rangle$	$\langle -1,1 \rangle$	$\langle -1,0 \rangle$	$\langle -2,1 \rangle$
	$c = \langle 2,1 \rangle$	$\langle 1,0 \rangle$	$\langle 1,-2 \rangle$	$\langle 0,0 \rangle$	$\langle 0,-1 \rangle$	$\langle 0,-2 \rangle$	$\langle -1,-1 \rangle$
To	$d = \langle 2,2 \rangle$	$\langle 1,1 \rangle$	$\langle 1,-1 \rangle$	$\langle 0,1 \rangle$	$\langle 0,0 \rangle$	$\langle 0,-1 \rangle$	$\langle -1,0 \rangle$
	$e = \langle 2,3 \rangle$	$\langle 1,2 \rangle$	$\langle 1,0 \rangle$	$\langle 0,2 \rangle$	$\langle 0,1 \rangle$	$\langle 0,0 \rangle$	$\langle -1,1 \rangle$
	$f = \langle 3,2 \rangle$	$\langle 2,1 \rangle$	$\langle 2,-1 \rangle$	$\langle 1,1 \rangle$	$\langle 1,0 \rangle$	$\langle 1,-1 \rangle$	$\langle 0,0 \rangle$

Figure 6: The vector table computed by SIATEC for the dataset shown in Figure 4 (a).

```
COSIATEC(D)
1  D′ ← COPY(D)
2  T ← ⟨⟩
3  while D′ ≠ ∅
4     T ← GETBESTTEC(D′, D)
5     T ← T ⊕ ⟨T⟩
6     D′ ← D′ \ COV(T)
7  return T
```

Figure 7: The COSIATEC algorithm.

6.2 SIATEC

SIATEC (Meredith et al., 2001, 2002, 2003; Meredith, 2006a) computes all the MTP TECs in a k-dimensional dataset of size n in $O(kn^3)$ time and $O(kn^2)$ space. In order to find the MTPs, the SIA algorithm only needs to compute the vectors from each point in a dataset to each lexicographically later point. However, to compute *all occurrences* of the MTPs, it turns out to be beneficial in the SIATEC algorithm to compute the vectors between *all* pairs of points, resulting in a vector table like the one shown in Figure 6. The SIATEC algorithm first finds all the MTPs using SIA. It then uses the vector table to find all the vectors by which each MTP is translatable within the dataset. The set of vectors by which a given pattern is translatable is equal to the intersection of the columns in the vector table headed by the points in the pattern (see Figure 6). In a vector table computed by SIATEC, each row descends lexicographically from left to right and each column increases lexicographically from top to bottom. SIATEC exploits these properties of the vector table to more efficiently find all the occurrences of each MTP (Meredith et al., 2002, pp. 335–338).

6.3 COSIATEC

COSIATEC (Meredith et al., 2003; Meredith, 2006a, 2013) is a greedy point-set compression algorithm, based on SIATEC. COSIATEC takes a dataset, D, as input and computes a compressed encoding of D in the form of an ordered set of MTP TECs, $\mathbf{T}$, such that $D = \bigcup_{T \in \mathbf{T}} \text{COV}(T)$ and $\text{COV}(T_1) \cap \text{COV}(T_2) = \emptyset$ for all $T_1, T_2 \in \mathbf{T}$ where $T_1 \neq T_2$. In other words, COSIATEC strictly partitions a dataset, D, into the covered sets of a set of MTP TECs. If each of these MTP TECs is represented as a $\langle \text{pattern}, \text{translator set} \rangle$ pair, then this description of the dataset as a set of TECs is typically shorter than an *in extenso* description in which the points in the dataset are listed explicitly.

Figure 7 shows pseudocode for COSIATEC. The first step in the algorithm is to make a copy of the input dataset which is stored in the variable D' (line 1). Then, on each iteration of the **while** loop (lines 3–6), the algorithm finds the "best" MTP TEC in D', stores this in T and adds T to $\mathbf{T}$. It then removes the set of points covered by T from D' (line 6). When D' is empty, the algorithm terminates, returning the list of MTP TECs, $\mathbf{T}$. The sum of the number of translators and the number of points in this output encoding is never more than the number of points in the input dataset and can be much less than this, if there are many repeated patterns in the input dataset.

The GETBESTTEC function, called in line 4 of COSIATEC, computes the "best" TEC in D' by first finding all the MTPs using SIA, then iterating over these MTPs, finding the TEC for each MTP, and storing it if it is the best TEC so far. In this process, a TEC is considered "better" than another if it has a higher compression ratio, as defined in Eq. 5. If two TECs have the same compression ratio, then the better TEC is considered to be the one that has the higher *bounding-box compactness* (Meredith et al., 2002), defined as the ratio of the number of points in the TEC's pattern to the number of dataset points in the bounding box of this pattern. Collins et al. (2011) have provided empirical evidence that the compression ratio and compactness of a TEC are important factors in determining its perceived "importance" or "noticeability". If two distinct TECs have the same compression ratio and compactness, then, in COSIATEC, the TEC with the larger covered set is considered superior.

6.4 Forth's algorithm

Forth (Forth & Wiggins, 2009; Forth, 2012) presented an algorithm, inspired by COSIATEC, that resembles the SIATECCOMPRESS algorithm to be described below. The first step in Forth's algorithm is to run SIATEC on the input dataset to generate a sequence of MTP TECs, $\mathbf{T} = \langle T_1, T_2, \ldots T_n \rangle$. The algorithm then post-processes the output of SIATEC to compute a cover for the input dataset. A weight, W_i, is assigned to each TEC, T_i, to produce a corresponding sequence of weights, $\mathbf{W} = \langle W_1, W_2, \ldots W_n \rangle$. W_i is intended to be a measure of the "structural salience" (Forth, 2012, p. 41) of the patterns in the TEC, T_i, and it is defined as $W_i = w'_{\text{cr},i} \cdot w'_{\text{compV},i}$ where $w'_{\text{cr},i}$ and $w'_{\text{compV},i}$ are normalized values representing the compression ratio and compactness of T_i. Having computed the sequence of weights, $\mathbf{W}$, Forth's algorithm then attempts to select a subset of the covered sets of the TECs in $\mathbf{T}$ that covers the input dataset and maximises the product of the coverage and weight of each TEC used in the encoding generated.

6.5 SIACT

Collins et al. (2010) claim that all the algorithms described above can be affected by what they call the 'problem of isolated membership'. This problem is defined to occur when "a musically important pattern is contained *within* an MTP, along with other temporally isolated members that may or may not be musically important" (Collins et al., 2010, p. 6).

```
SIATECCompress(D)
1   T ← SIATEC(D)
2   T ← SortTECsByQuality(T)
3   D′ ← ∅
4   E ← ⟨⟩
5   for i ← 0 to |T| − 1
6     T ← T[i]
7     S ← COV(T)
        ▶ Recall that each TEC, T, is an ordered pair, ⟨P, Θ⟩
8     if |S \ D′| > |T[0]| + |T[1]| − 1
9       E ← E ⊕ ⟨T⟩
10      D′ ← D′ ∪ S
11      if |D′| = |D|
12        break
13  R ← D \ D′
14  if |R| > 0
15    E ← E ⊕ ⟨AsTEC(R)⟩
16  return E
```

Figure 8: A straight-forward implementation of SIATECCompress.

Collins et al. (2010, p. 6) claim that "the larger the dataset, the more likely it is that the problem will occur" and that it could prevent the SIA-based algorithms from "discovering some translational patterns that a music analyst considers noticeable or important". Collins et al. propose that this problem can be solved by taking each MTP computed by SIA (sorted into lexicographical order) and 'trawling' inside this MTP "from beginning to end, returning subsets that have a compactness greater than some threshold a and that contain at least b points" (Collins et al., 2010, p. 6). This method is implemented in an algorithm that they call SIACT, which first runs SIA on the dataset and then carries out 'compactness trawling' (hence "SIA*CT*") on each of the MTPs found by SIA.

6.6 SIAR

In an attempt to improve on the precision and running time of SIA, Collins (2011, pp. 282–283) defines an SIA-based algorithm called SIAR. Instead of computing the whole region below the leading diagonal in the vector table for a dataset (as in Figure 4(b)), SIAR only computes the first r subdiagonals of this table. This is approximately equivalent to running SIA with a sliding window of size r (Collins et al., 2010; Collins, 2011).

6.7 SIATECCompress

COSIATEC uses SIATEC on each iteration of its **while** loop to compute the best TEC to add to the output encoding. Since SIATEC has worst case running time $O(n^3)$ where n is the number of points in the input dataset, running COSIATEC on large datasets can be time-consuming. On the other hand, because COSIATEC strictly partitions the dataset into non-overlapping MTP TEC covered sets, it tends to achieve high compression ratios for many point-set representations of musical pieces (typically between 2 and 4 for a piece of classical or baroque music).

Like COSIATEC, the SIATECCompress algorithm shown in Figure 8 is a greedy compression algorithm based on SIATEC that computes an encoding of a dataset in the form of a union of TEC covered sets. Like Forth's algorithm (but unlike COSIATEC), SIATECCompress runs SIATEC only *once* (line 1) to get a list of TECs. This list is then sorted into decreasing order of quality (line 2), where the decision as to which of any two TECs is superior is made in the same way as in COSIATEC (described above). The algorithm then finds a compact encoding, $\mathbf{E}$, of the dataset in the form of a set of TECs. It does this by iterating over the sorted list of TECs (lines 5–12), adding a new TEC, T, to $\mathbf{E}$ if the number of new points covered by T is greater than the size of its $\langle$pattern, translator set$\rangle$ representation (lines 8–12). Each time a TEC, T, is added to $\mathbf{E}$, its covered set is added to the set D', which therefore maintains the set of points covered so far after each iteration. When D' is equal to D or all the TECs have been scanned, the **for** loop terminates. Any remaining uncovered points are aggregated into a *residual point set*, R, (line 13) which is re-expressed as a TEC with an empty translator set (line 15) that is added to the encoding. SIATECCompress does not generally produce as compact an encoding as COSIATEC, since the TECs in its output may share points. However, it is faster than COSIATEC and can therefore be used practically on much larger datasets. Unlike Forth's algorithm, SIATECCompress always produces a complete cover of the input dataset.

7. USING THE ALGORITHMS TO CLASSIFY FOLK SONGS

COSIATEC, Forth's algorithm and SIATECCompress were used to classify the melodies in the *Annotated Corpus* (van Kranenburg et al., 2013; Volk & van Kranenburg, 2012) of 360 Dutch folk songs from the collection, *Onder de groene linde* (Grijp, 2008), hosted by the Meertens Institute and accessible through the website of the Dutch Song Database (`http://www.liederenbank.nl`). The algorithms were used as compressors to calculate the normalized compression distance between each pair of melodies in the collection. Each melody was then classified using the 1-nearest-neighbour algorithm with leave-one-out cross-validation. The classifications obtained were compared with a ground-truth classification of the melodies carried out by expert musicologists.

Four versions of each of the three algorithms were tested:

- the basic algorithm as described above,
- a version incorporating the compactness trawler from Collins et al.'s SIACT algorithm,
- a version using SIAR instead of SIA and
- a version using both SIAR and the compactness trawler.

As a baseline, one of the best general-purpose compression algorithms, bzip2 (Seward, 2010), was also used to calculate NCDs between the melodies.

Table 1 shows the results obtained in this task. In this table, algorithms with names containing "R" employed the SIAR algorithm with $r = 3$ in place of SIA. Algorithms

Table 1: Results of using different compressors to classify the *Annotated Corpus* of Dutch folk songs using NCD, 1-nn and leave-one-out-cross-validation. *SR* is the classification success rate, CR_{AC} is the average compression ratio over the melodies in the *Annotated Corpus*. CR_{pairs} is the average compression ratio over the pairs of files used to obtain the NCD values.

Algorithm	SR	CR_{AC}	CR_{pairs}
COSIATEC	0.8389	1.5791	1.6670
COSIARTEC	0.8361	1.5726	1.6569
COSIARCTTEC	0.7917	1.4547	1.5135
COSIACTTEC	0.7694	1.4556	1.5138
ForthCT	0.6417	1.1861	1.2428
ForthRCT	0.6417	1.1861	1.2428
Forth	0.6111	1.2643	1.2663
ForthR	0.6028	1.2555	1.2655
SIARCTTECCompress	0.5750	1.3213	1.3389
SIATECCompress	0.5694	1.3360	1.3256
SIACTTECCompress	0.5250	1.3197	1.3381
SIARTECCompress	0.5222	1.3283	1.3216
bzip2	0.1250	2.7678	3.5061

with names containing "CT" used Collins et al.'s (2010) compactness trawler, with parameters $a = 0.66$ and $b = 3$. The column headed "SR" gives the classification success rate—i.e., the proportion of songs in the corpus correctly classified. The third and fourth columns give the mean compression ratio achieved by each algorithm over, respectively, the corpus and the file-pairs used to compute the compression distances.

The highest success rate of 84% was obtained using COSIATEC. Table 1 suggests that algorithms based on COSIATEC performed markedly better on this song classification task than those based on SIATECCOMPRESS or Forth's algorithm. All three algorithms use compression ratio and compactness to select the TECs used in their output encodings. The main difference between COSIATEC and the other two algorithms is that COSIATEC removes the points covered by each selected TEC and reruns SIATEC on the remaining points to select the next TEC. This produces a strict partition of the input dataset into TEC covered sets that are *collectively exhaustive* in that they collectively cover the input dataset and *mutually exclusive* (i.e., they do not intersect). On the other hand, the covered sets of the TECs computed by Forth's algorithm and SIATECCOMPRESS may share points—i.e., they may not be mutually exclusive. Moreover, the set of TEC covered sets generated by Forth's algorithm may not be collectively exhaustive. The results in Table 1 suggest that the strategy adopted in COSIATEC may better model the cognitive processes used by the musicologists who created the ground-truth classification.

Using SIAR instead of SIA and/or incorporating compactness trawling reduced the performance of COSIATEC. However, using both together, slightly improved the performance of SIATECCOMPRESS. Forth's algorithm performed slightly better than SIATECCOMPRESS. The performance of Forth's algorithm on this task was improved by incorporating compactness trawling; using SIAR instead of SIA in Forth's algorithm slightly reduced the performance of the basic algorithm and had no effect when compactness trawling was used. The results obtained using bzip2 were much poorer than those obtained using the SIA-based algorithms, suggesting that general-purpose compressors may fail to capture certain musical structure that is important for this task—at least when run on point-set representations of the type used in this study. Of the SIA-based algorithms, COSIATEC achieved the best compression on average, followed by SIATECCOMPRESS and then Forth's algorithm. COSIATEC also achieved the best success rate. However, since Forth's algorithm performed slightly better than SIATECCOMPRESS, it seems that degree of compression *alone* was not a reliable indicator of classification accuracy on this task—indeed, the best compressor, bzip2, produced the worst classifier. None of the algorithms achieved a success rate as high as the 99% obtained by van Kranenburg et al. (2013) on this corpus using several local features and an alignment-based approach. The success rate achieved by COSIATEC is within the 83–86% accuracy range obtained by Velarde et al. (2013, p. 336) on this database using a wavelet-based representation, with similarity measured using Euclidean or city-block distance.

8. CONCLUSIONS

The results in Table 1 suggest that the implicit and explicit knowledge and cognitive processes used by the musicologists who developed the ground-truth classification for the *Annotated Corpus* of the Dutch folk-song database can be modelled reasonably well by using normalized compression distance (NCD) as a measure of similarity. However, the results also show that the success of such an NCD-based model depends critically on which compressor one uses to produce NCDs and how encoding length is measured. In particular, in this study, compressors based on point-set pattern discovery and TEC compression-ratio performed much better than a baseline general-purpose, string-based compressor of the type used in previous studies that have used NCD for music classification.

9. ACKNOWLEDGEMENTS

The author is grateful to Peter van Kranenburg for providing the data files for the *Annotated Corpus* from the collection *Onder de groene linde* (Grijp, 2008) currently hosted by the Meertens Institute and accessible through the website of the Dutch Song Database (Nederlandse Liederenbank, `www.liederenbank.nl`). The work was carried out within the EU project, "Learning to Create" (Lrn2Cre8). The project Lrn2Cre8 acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859.

10. REFERENCES

Bayard, S. (1950). Prolegomena to a study of the principal melodic families of British-American folk song. *Journal of American Folklore*, *63*(247), 1–44.

Cilibrasi, R., Vitányi, P. M. B., & de Wolf, R. (2004). Algorithmic clustering of music based on string compression. *Computer Music Journal*, *28*(4), 49–67.

Collins, T. (2011). *Improved methods for pattern discovery in music, with applications in automated stylistic composition*. PhD thesis, Faculty of Mathematics, Computing and Technology, The Open University, Milton Keynes.

Collins, T., Laney, R., Willis, A., & Garthwaite, P. H. (2011). Modeling pattern importance in Chopin's Mazurkas. *Music Perception*, *28*(4), 387–414.

Collins, T., Thurlow, J., Laney, R., Willis, A., & Garthwaite, P. H. (2010). A comparative evaluation of algorithms for discovering translational patterns in baroque keyboard works. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010), Utrecht, The Netherlands, 9–13 August 2010*, (pp. 3–8).

Forth, J. & Wiggins, G. A. (2009). An approach for identifying salient repetition in multidimensional representations of polyphonic music. In J. Chan, J. W. Daykin, & M. S. Rahman (Eds.), *London Algorithmics 2008: Theory and Practice* (pp. 44–58). London: College Publications.

Forth, J. C. (2012). *Cognitively-Motivated Geometric Methods of Pattern Discovery and Models of Similarity in Music*. PhD thesis, Department of Computing, Goldsmiths, University of London.

Grijp, L. P. (2008). Introduction. In L. P. Grijp & I. van Beersum (Eds.), *Under the Green Linden—163 Dutch Ballads from the Oral Tradition* (pp. 18–27). Meertens Institute/Music & Words.

Hillewaere, R., Manderick, B., & Conklin, D. (2012). String methods for folk tune genre classification. In *International Society for Music Information Retrieval Conference (ISMIR 2012)*, (pp. 217–222).

Li, M., Chen, X., Li, X., Ma, B., & Vitányi, P. M. B. (2004). The similarity metric. *IEEE Transactions on Information Theory*, *50*(12), 3250–3264.

Li, M. & Sleep, R. (2004). Melody classification using a similarity metric based on Kolmogorov complexity. In *Sound and Music Computing Conference (SMC)*.

Li, M. & Sleep, R. (2005). Genre classification via an LZ78-based string kernel. In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005)*, (pp. 252–259).

Li, M. & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications* (Third ed.). Berlin: Springer.

Meredith, D. (2006a). Point-set algorithms for pattern discovery and pattern matching in music. In *Proceedings of the Dagstuhl Seminar on Content-based Retrieval (No. 06171, 23–28 April, 2006)*, Schloss Dagstuhl, Germany. Available online at <http://drops.dagstuhl.de/opus/volltexte/2006/652>.

Meredith, D. (2006b). The *ps13* pitch spelling algorithm. *Journal of New Music Research*, *35*(2), 121–159.

Meredith, D. (2007). *Computing Pitch Names in Tonal Music: A Comparative Analysis of Pitch Spelling Algorithms*. PhD thesis, Faculty of Music, University of Oxford.

Meredith, D. (2013). COSIATEC and SIATECCompress: Pattern discovery by geometric compression. In *MIREX 2013 (Competition on Discovery of Repeated Themes & Sections)*. Available online at `http://www.titanmusic.com/papers/public/MeredithMIREX2013.pdf`.

Meredith, D., Lemström, K., & Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, *31*(4), 321–345.

Meredith, D., Lemström, K., & Wiggins, G. A. (2003). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. In *Cambridge Music Processing Colloquium*.

Meredith, D., Wiggins, G. A., & Lemström, K. (2001). Pattern induction and matching in polyphonic music and other multi-dimensional datasets. In Callaos, N., Zong, X., Verges, C., & Pelaez, J. R. (Eds.), *Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI2001)*, volume X, (pp. 61–66).

Scheurleer, D. (1900). Preisfrage. *Zeitschrift der Internationalen Musikgesellschaft*, *1*(7), 219–220.

Seward, J. (2010). bzip2 version 1.0.6, released 20 September 2010. `http://www.bzip.org`. Accessed 19 April 2014.

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., & Meredith, D. (2013). An approach to melodic segmentation and classification based on filtering with the haar-wavelet. *Journal of New Music Research*, *42*(4), 325–345.

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.

Ziv, J. & Lempel, A. (1977). A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, *23*(3), 337–343.

Ziv, J. & Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, *24*(5), 530–536.

AN OPEN WEB AUDIO PLATFORM FOR ETHNOMUSICOLOGICAL SOUND ARCHIVES MANAGEMENT AND AUTOMATIC ANALYSIS

Thomas Fillon
PARISSON / LAM,
Institut Jean Le Rond d'Alembert,
UPMC Univ. Paris 06,
UMR CNRS 7190
thomas.fillon@parisson.com

Guillaume Pellerin, Paul Brossier
PARISSON
16 rue Jacques Louvel-Tessier
Paris, France
guillaume.pellerin@parisson.com

Joséphine Simonnot
CREM, LESC, UMR CNRS 7186
MAE, Université Paris Ouest
Nanterre La Défense

ABSTRACT

Since 2007, ethnomusicologist and engineers have joint their effort to develop a scalable and collaborative web platform for management of and access to digital sound archives. This platform has been deployed since 2011 and hold the archives of the *Center for Research in Ethnomusicology*, which is the most important collection in Europe. This web platform is based on *Telemeta*, an open-source web audio framework dedicated to digital sound archives secure storing, indexing and publishing. It focuses on the enhanced and collaborative user-experience in accessing audio items and their associated metadata and on the possibility for the expert users to further enrich those metadata.

Telemeta architecture relies on *TimeSide*, an open audio processing framework written in Python which provides decoding, encoding and streaming methods for various formats together with a smart embeddable HTML audio player. TimeSide also includes a set of audio analysis plugins and additionally wraps several audio features extraction libraries to provide automatic annotation, segmentation and musicological analysis.

1. INTRODUCTION

In social sciences like anthropology and linguistics, researchers have to work on multiple types of multimedia documents such as photos, videos, sound recordings or databases. The need to easily access, visualize and annotate such materials can be problematic given their diverse formats, sources and given their chronological nature.

In the context of ethnomusicological research, the Research Center on Ethnomusicology (CREM) and Parisson, a company specialized in big music data projects, have been developing an innovative, collaborative and interdisciplinary open-source web-based multimedia platform since 2007.

This platform, *Telemeta* is designed to fit the professional requirements from both sound archivists, researchers and musicians to work together on big music data. The first prototype of this platform has been online since 2010 and is now fully operational and used on a daily basis for ethnomusicological studies since 2011. A description of theses archives and some use cases are given in Section 4.

The benefit of this collaborative platform for ethnomusicological research has been described in several publications (Simonnot, 2011; Julien Da Cruz Lima, 2011; Simonnot et al., 2014).

This work is partly supported by a grant from the french National Research Agency (ANR) with reference ANR-12-CORD-0022.

Figure 1: Screenshot excerpt of the Telemeta web interface

Recently, an open-source audio analysis framework, TimeSide, has been developed to bring automatic music analysis capabilities to the web platform and thus have turned Telemeta into a complete resource for *Computational Ethnomusicology* (Tzanetakis et al., 2007; Gómez et al., 2013).

2. THE TELEMETA PLATFORM

2.1 Web audio content management features and architecture

The primary purpose of the project is to provide the ethnomusicological communities with a scalable system to access, preserve and share audio research materials together with their associated metadata, as these data provide key information on the context and significance of the recording. Telemeta [1], as a free and open source [2], is a unique scalable web audio platform for backuping, indexing, transcoding, analyzing, sharing and visualizing any digital audio or video file in accordance with open web standards.

The time-based nature of such audio-visual materials and some associated metadata as annotations raises issues of access and visualization at a large scales. Easy and on demand access to these data, as you listen to the recording, represents a significant improvement.

An overview of the Telemeta's web interface is illustrated in Figure 1. Its flexible and streaming safe architec-

[1] http://telemeta.org

[2] Telemeta code is available under the CeCILL Free Software License Agreement

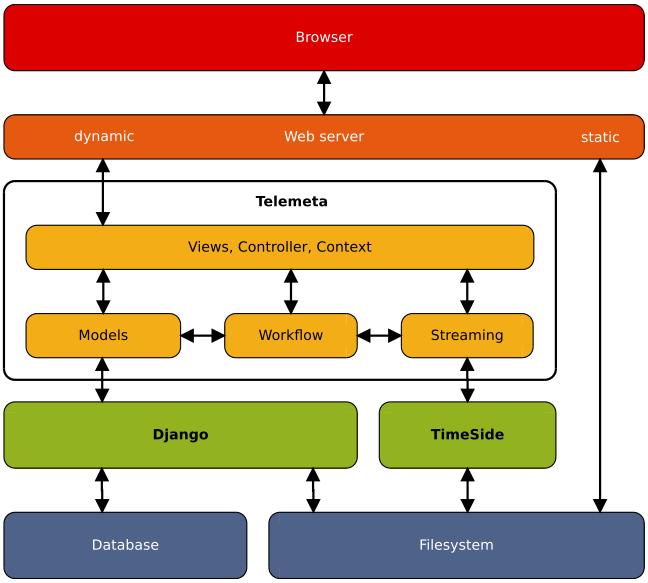


Figure 2: Telemeta architecture

ture is represented in Figure 2.

The main features of *Telemeta* are:

- Pure HTML5 web user interface including dynamical forms
- On the fly audio analyzing, transcoding and metadata embedding in various formats
- Social editing with semantic ontologies, smart workflows, realtime tools, human or automatic annotations and segmentations
- User management with individual desk, playlists, profiles and group access rights
- High level search engine (geolocation, instruments, ethnic groups, etc...)
- Data providers : DublinCore, OAI-PMH, RSS, XML, JSON and other
- Multi-language support (now english and french)

Beside database management, the audio support is mainly provided through an external component, TimeSide, which is described in Section 3.

2.2 Metadata

In addition to the audio data, an efficient and dynamic management of the associated metadata is also required. Consulting metadata provide both an exhaustive access to valuable information about the source of the data and to the related work of peer researchers. Dynamically handling metadata in a collaborative manner optimises the continuous process of knowledge gathering and enrichment of the materials in the database. One of the major challenge is thus the standardization of audio and metadata formats with the aim of long-term preservation and usage of the different materials. The compatibility with other systems is facilitated by the integration of the metadata standards protocols *Dublin Core* [3] and *OAI-PMH* (Open Archives Initiative Protocol for Metadata Harvesting) [4] .

Metadata provide two different kinds of information about the audio item: contextual information and annotations.

2.2.1 Contextual Information

In ethnomusicology, contextual information could be geographic, cultural and musical. It could also store archive related information and include related materials in any multimedia format.

2.2.2 Annotations and segmentation

Metadata also consist in temporally-indexed information such as a list of time-coded markers associated with annotations and a list of of time-segments associated with labels. The ontology for those labels is relevant for ethnomusicology (e.g. speech versus singing voice segment, chorus, ...).

Ethnomusicological researchers and archivists can produce their own annotations and share them with colleagues. These annotations are accessible from the sound archive item web page and are indexed through the database.

It should be noted that annotations and segmentation can also be produce by some automatic signal processing analysis (see Section 3).

3. TIMESIDE, AN AUDIO ANALYSIS FRAMEWORK

One specificity of the Telemeta architecture is to rely on an external component, TimeSide [5] , that offers audio player

[3] Dublin Core Metadata Initiative, `http://dublincore.org/`
[4] `http://www.openarchives.org/pmh/`
[5] `https://github.com/yomguy/TimeSide`

web integration together with audio signal processing analysis capabilities.

TimeSide is an audio analysis and visualization framework based on both python and javascript languages to provide state-of-the-art signal processing and machine learning algorithms together with web audio capabilities for display and streaming. Figure 3 illustrates the overall architecture of TimeSide together with the data flow between TimeSide and the Telemeta web-server.

3.1 Audio management

TimeSide provides the following main features:

- Secure archiving, editing and publishing of audio files over internet.
- Smart audio player with enhanced visualisation (waveform, spectrogram)
- Multi-format support: reads all available audio and video formats through Gstreamer, transcoding with smart streaming and caching methods
- "On the fly" audio analyzing, transcoding and metadata embedding based on an easy plugin architecture

3.2 Audio features extraction

In order to provide Music Information Retrieval analysis methods to be implemented over a large corpus for ethnomusicological studies, TimeSide incorporates some state-of-the-art audio feature extraction libraries such as Aubio [6] (Brossier, 2006), Yaafe [7] (Mathieu et al., 2010) and Vamp plugins [8] .

As a open-source framework and given its architecture and the flexibility provided by Python, the implementation of any audio and music analysis algorithm can be consider. Thus, it makes it a very convenient framework for researchers in computational ethnomusicology to develop and evaluate their algorithms.

Given the extracted features, every sound item in a given collection can be automatically analyze. The results of this analysis can be stored in a scientific file format like Numpy and HDF5 and serialized to the web browser through common markup languages: XML, JSON and YAML.

3.3 Automatic Analysis of ethnomusicological sound archives

Ongoing works lead by the DIADEMS project consist in implementing advanced classification, indexation, segmentation and similarity analysis methods dedicated to ethnomusicological sound archives.

Besides music analysis, such automatic tools also deal with speech and noises classification and segmentation to enable a full annotation of the audio materials.

In the context of this project, both researchers from Ethnomusicological, Speech and Music Information Retrieval communities are working together to specified the tasks to be addressed by automatic analysis tools.

[6] http://aubio.org/

[7] https://github.com/Yaafe/Yaafe

[8] http://www.vamp-plugins.org

4. SOUND ARCHIVES OF THE CNRS - MUSÉE DE L'HOMME

Since June 2011, the Telemeta platform has been deployed to hold the *Sound archives of the CNRS - Musée de l'Homme* [9] and is managed by the CREM (Center for Research in Ethnomusicology). The platform aims to make these archives available to researchers and to the extent possible, the public, in compliance with the intellectual and moral rights of musicians and collectors.

4.1 Archiving research materials

The archives of CREM, the most important in Europe, are distinguished by their wealth:

- Nearly 3,500 hours of recordings of unpublished field.
- Approximately 3700 hours of material published (more than 5000 discs, many of which are very rare).

The collection is sustained by the field missions of researchers on all continents.

Through this platform, archivists can properly ensure the long-term preservation of data and continuously maintain and enrich the associated metadata.

Accessing the collections aid laboratory research, diachronic and synchronic comparisons, the preparation of new fieldwork and the training of PhD students.

Publishing collections also helps researchers making their work more visible. Besides making available and listenable the related corpora, researchers can also append related academic publications and provide temporal annotations to further illustrate their work.

4.2 A collaborative platform

Given the collaborative nature of the platform, both research and archivist can cooperate with colleagues to continuously enrich metadata associated to a sound item or a collection.

Collaborative tools like markers and comments enable researchers from different institutions to work together on some common audio materials. It also allows researchers to exchange data online with communities producing their music in their home countries.

5. CONCLUSION

The Telemeta open-source framework provides the researchers in musicology with a new platform to efficiently distribute, share and work on their research materials. The platform has been deployed since 2011 to manage the *Sound archives of the CNRS - Musée de l'Homme* which is the most important european collection of ethnomusicocological resources.

[9] http://archives.crem-cnrs.fr

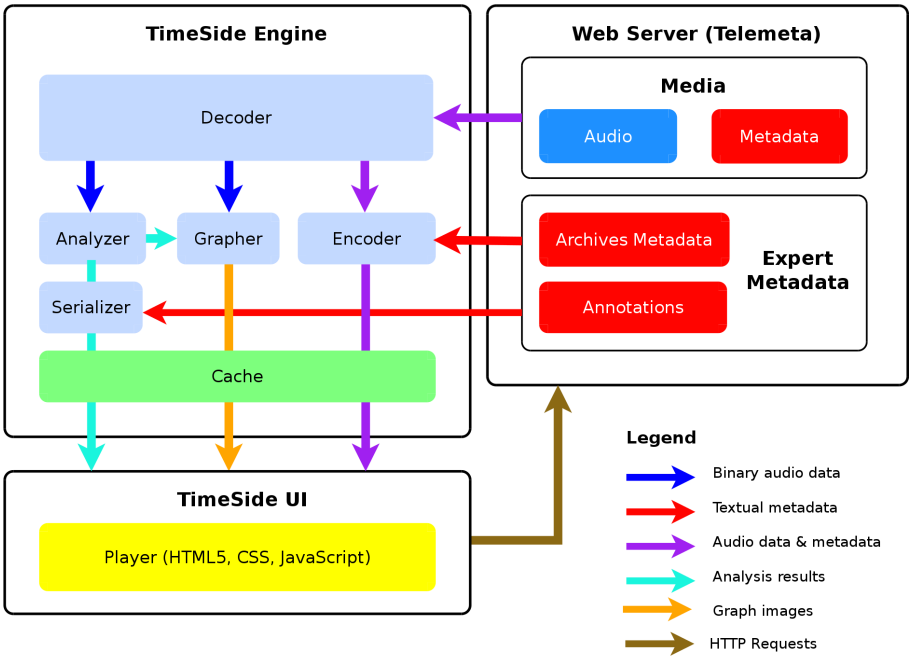


Figure 3: TimeSide engine architecture and data flow with Telemeta web-server

Furthermore, this platform is offered automatic music analysis capabilities through an external component, TimeSide that provides a flexible computational analysis engine together with web serialization and visualization capabilities. As an open-source framework TimeSide could be an appropriate platform for researchers in computational ethnomusicology to develop and evaluate their algorithms

Further works on the user interface will enhance the visualization experience with time and frequency zooming capabilities and will thus improve the accuracy and the quality of time-segment base annotations.

Acknowledgments

The authors would like to thank all the people that have been involved in Telemeta specification and development or have provide useful input and feedback. The project has been partially funded by the French National Centre for Scientific Research (CNRS), the French Ministry of Culture and Communication, the TGE Adonis Consortium, and the Centre of Research in Ethnomusicology (CREM).

6. REFERENCES

Brossier, P. (2006). *Automatic annotation of musical audio for interactive systems*. PhD thesis, Centre for Digital music, Queen Mary University of London, UK.

Gómez, E., Herrera, P., & Gómez-Martin, F. (2013). Computational ethnomusicology: perspectives and challenges. *Journal of New Music Research*, *42*(2), 111–112.

Julien Da Cruz Lima, A. (2011). The CNRS — Musée de l'Homme audio archives: a short introduction. *International Association of Sound and Audiovisual Archives journal*, *36*.

Mathieu, B., Essid, S., Fillon, T., Prado, J., & Richard, G. (2010). Yaafe, an easy to use and efficient audio feature extraction software. In *Proc. of ISMIR 2010, Utrecht, Netherlands*, (pp. 441–446). International Society for Music Information Retrieval.

Simonnot, J. (2011). TELEMETA: an audio content management system for the web. *International Association of Sound and Audiovisual Archives journal*, *36*.

Simonnot, J., Mifune, M.-F., & Lambert, J. (2014). Telemeta: Resources of an online archive of ethnomusicological recordings. Panel accepted at ICTM Study Group on Historical Sources of Traditional Music, Aveiro, Portugal, May 12-17 2014.

Tzanetakis, G., Kapur, A., Schloss, W. A., & Wright, M. (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, *1*(2), 1–24.

Uncovering Semantic Structures within Folk Song Lyrics

Gregor Strle
Institute of Ethnomusicology
Research Centre of the Slovene Academy of Sciences and Arts
gregor.strle@zrc-sazu.si

Matija Marolt
University of Ljubljana
Faculty of Computer and Information Science
matija.marolt@fri.uni-lj.si

ABSTRACT

In this paper, we focus on computational methods for natural language processing (NLP) and evaluate some possibilities that NLP methods offer to folkloristics. Due to inherent dialectical diversity and strong intertextuality, folkloristic materials have generally proven to be very challenging for NLP. Our goal was therefore to evaluate different NLP methods and study the semantics generated by respective approaches and their practical implica-tions. Three experiments analyzing a collection of Slovenian folk narrative poems are presented and results discussed.

1. INTRODUCTION

Growth of digital collections in recent years has motivated interdisciplinary research that connects various fields of computer science and humanities. For example, machine learning techniques can today complement researcher's analysis to uncover latent semantic structures in music, visual materials and text. They go beyond limitations of human (manual) analysis and are especially useful in processing and classification of materials, as they can inspect large amounts of data in relatively short time.

In this article, we focus on computational methods for natural language processing (NLP). NLP is used to solve a wide variety of tasks: machine translation, optical character recognition (OCR), parsing (grammatical analysis), speech recognition and semantic analysis (word-sense disambiguation, topic recognition). We are interested in the latter.

We present three experiments of using NLP to analyze a collection of Slovenian folk poems. In the first, the goal was to get an insight into the conceptual structure of the materials and at the same time discover advantages and disadvantages of two main NLP approaches. In the second experiment, we wanted to assess whether the automatically obtained topics correspond in any way to annotated song families. In the third experiment, we wanted to assess whether topic distributions in any way correspond to the major themes of individual variant types.

2. CORPUS

For our experiments, we selected 1,965 variants of Slovenian folk narrative poems (Golež Kaučič, Kumer, Terseglav, & Vrčon, 1998; Golež Kaučič, Kumer, Šivic, Terseglav, & Vrčon, 2007), part of our multimedia digital library EthnoMuse (Strle & Marolt, 2012). The selection includes narrative poems about love and fate conflicts and about family fates and conflicts. The songs date back to 18th and 19th century, with some variant types represented by only one variant, whereas others have up to 180 versions and are still sung today. Thematically, the variants are closely related, as they share similar stories about death, murder, suicide, infidelity, punishment, etc. Moreover, strong intertextuality is present through the whole corpus, which reflects a characteristic folk song phenomenon: traveling of verses, motifs, and thematic patterns from one song to the other. This has strongly affected the results, as most occurring themes and motifs dominated over rarer variant types.

3. EXPERIMENT 1

Our first experiment has focused on the general characteristics of Slovenian folk song lyrics at the level of poetic (variant) types and topics that intertwine within them. We wanted to get an insight into the conceptual structure of the materials and at the same time discover advantages and disadvantages of two main NLP approaches: a statistical associative approach using Latent Semantic Analysis (LSA (Landauer & Dumais, 1997)) and a probabilistic topic-modeling approach using Latent Dirichlet Allocation (LDA (Blei, Ng, & Jordan, 2003)). There are significant differences between the two in terms of semantics and context they generate.

The synthetic nature of Slovenian language with many morphological rules, as well as strong dialects in singing, which are reflected in transcriptions, made it necessary to lemmatize song lyrics before analysis. We first replaced special characters used for encoding characteristics of dialect groups (such as semivowels, diphthongization, pitch accents etc.) by their grammatical equivalents. A dialect dictionary was then used to translate the words into literary language and finally a statistical morphosyntactic tagger for the Slovenian language (Grčar, Krek, & Dobrovoljc, 2012) to lemmatize the text.

3.1 Results

LSA and LDA were both performed on lemmatized documents that were converted into word-document matrices with word counts weighted according to the tf-idf statistics. For both types of analyses, we limited projections of the word-document space to 10 dimensions/topics, as the size of the corpus is relatively small. Table 1 shows the most salient words and documents for each method according to variant types they mostly represent. While

LSA variant types and dimensions	*LDA variant types and topics*
DEATH OF A BRIDE BEFORE WEDDING	DEATH AT A REUNION
d1: mother child young baby shepherd wreath blood	t1: heart boy Breda head sad hunter Danube
d4: Ljubljana linden lover boy seduce chamber Tonček	MURDER OUT OF JEALOUSY
d5: Breda Ljubljana groom mother-in-law linden baby Turk	t2: love sword kneel sharp neighbor boyfriend blame
d6: Breda accident evil house mother-in-law sister groom	BRIDE INFANTICIDE
d8: Ljubljana brother linden sea shirt prefer wash lover	t3: home shepherd Mary uncle birth shred rockcradle
NUN'S SUICIDE FOR LOVE	UNFAITHFUL STUDENT/NEW PRIEST
d2: convent Ursula nun baptism godmother ring blood	t4: undertaker love priest parish love promise letter
d3: convent Ursula nun baptism godmother shepherd wreath	NUN'S SUICIDE FOR LOVE
HUNTER SHOOTS HIS LOVER AND HIMSELF	t5: love Uršika convent boy Jesus farewell sword
d7: newpriest grave bury church rifle hunter student	REJECTED LOVER
d9: Ljubljana linden rifle grave hunter shaking leaves	t6: seduce blood house Vida linden Ljubljanians death
d10: rifle hunter shaking Tonček leaves face pale	WIDOWER ON BRIDE'S GRAVE
	t7: tender abandon blood bread jesus rockcradle married
	ABANDONED ORPHANS
	t8: bury window chamber wound grow crying dead
	PUNISHMENT FOR THE WICKED SONS AND DAUGHTERS-IN-LAW
	t9: gold sea mountain rooster fear crying darling son
	MISTRESS' LOYALTY REPAID
	t10: boy fenced heart nosegay dead grieve loyal

Table 1. Top words and the corresponding variant types for LSA dimensions and LDA topics.

similarities between words in relation to variant types are relatively similar for both analyses - compare for example LSA and LDA analysis of variant type 'Nun's suicide for love' in dimensions 2-3 and topic 5 - there is a significant difference in the detection of variant types across the corpus. As shown in the table, LSA could detect only three variant types that dominate across semantic space, whereas LDA successfully detected heterogeneity of the corpus and associated each topic with a different variant type. As LSA cannot account for topical distribution, and hence lacks additional hierarchy, it has difficulty detecting heterogeneity and the resulting semantic space repeatedly generalizes towards the most salient aspects of the corpora. LDA's projections, on the other hand, are more balanced: less overlapping and better at detecting heterogeneity.

4. EXPERIMENT 2

The goal of our next experiment was to assess whether LDA topics correspond in any way to song families. As described previously, our corpus consists of two main song families: poems about love and fate conflicts and poems about family fate conflicts. Although the themes in both are similar, we wanted to assess whether the topics discovered by LDA have any correspondence to the two families. We first calculated distributions over LDA topics for individual variant types within both families by averaging topic distributions of all variants of each type. The purpose of averaging was to reduce the imbalance of the number of variants of each poem type in our corpus, as some types have many (over 50) variants, while others have just a few. This resulted in a set of 90 topic distributions for all 90 variant types in our corpus.

We then used the cosine similarity measure, often used in text retrieval, to cluster the variant types based on similarities of their topic distributions by using the agglomerative hierarchical cluster tree method. We examined the obtained clusters to find out whether they relate to the division of poems into families. The ratio between love and family ballads taken from 4 major clusters is shown in Table 2. As shown in the table, topics about family relations (e.g. 'son', 'mother', 'brother', 'father', 'wife', 'mother-in-law' etc.) correspond more to clusters 1 and 4, which have a higher ratio of family ballads, whereas clusters 2 and 3 include more love oriented topics. Thus even though the entire corpus contains strong intertextuality and themes in both families are very related, the obtained semantic space does include some notion of song families and enables us to place individual (also new or unknown) songs into this space and study their relations to existing materials.

family clusters 1 (2:6) and 4 (13:31)	*love clusters 2 (17:11) and 3 (6:4)*
hunter earth unfortunately rifle son mother remember	field three maid sun golden mara ark sea lover
noble castle son stand cry dress letter dress give	things husband voice eat say young white know sin school
mom wife children find gold adultery measure colorful stick boy	mistress unlock boy saint window pot die lie
mountain will water mom hero angry dam girlfriend mother-in-law	stepmother run home getup graveyard rough get out go home
brother father house dear ours sister see	
tender live leave quickly name call barely crown world beg	

Table 2. Distribution of songs about love and family - clusters (love:family) and topic descriptions (top words) are shown.

5. EXPERIMENT 3

The third experiment focused on topic distributions among variants in order to assess whether LDA can detect major themes characteristic for individual variant types. We used a supervised learning method, Labeled LDA (LLDA). The procedure is similar to basic LDA analysis, with few minor differences. LLDA is a supervised topic model that uses predefined labels for calculating the topical distributions of the corpus. Thus, we first manually annotated selected variants with labels corresponding to the major theme (or themes) of particular variant (many of the labels represent the most occurring themes in the corpus). Next, we used this annotated dataset (around 18% of the whole corpus) to train our model. In the final phase, the whole corpus was used for inference to find variants best associated with each label globally. Results are shown in Figure 1, which shows thematic structure for selected variant types.

Most variant types share multiple topics, with the main topic for each type shown as most salient. For example, variants of type '*Mother prevents her son's marriage*' cover a wide range of topics, predominantly 'forced marriage', 'family', 'rejection' and 'tragic fate' of course, which in some variants result in 'murder', 'suicide' or even 'infanticide' due to illegitimately born child. In '*Punishment of a broken vow*', a young bride-to-be is torn between her (or her mother's) vow to Jesus (for her to live in a monastery as a nun) and a vow given to her lover, consequently being punished for her weakness; hence topics 'fidelity'/'infidelity', 'rejection', 'murder', 'death' and 'suicide', with a pinch of 'priesthood' in-between. LLDA can disambiguate different senses of unhappy love. For example, thematically similar to the previously mentioned '*Punishment of a broken vow*' are two variant types '*Death of a girl married far away*' and '*Death of a bride before marriage*', but here the emphasis is on 'forced marriage' and consequently 'death'.

Some variant types have a single dominant topic. For example, type '*Homicide because of incest*' has appropriate topic distribution with 'incest' prevailing over other context-related topics, such as 'family' and 'tragic fate', which are activated to a lesser extent. The extreme cases of single topic dominance are for example the three variant types '*The condemned infanticide*', '*Stepmother and her stepchild*' and '*Deceitful abduction of a young mother*', with the former two having at least 'tragic fate' in common.

An interesting example is variant type '*Poisoning of own sister*'. Here, the predominant topic given by LLDA is 'kidnapping', even though in the second part of the ballad the story slowly gravitates towards enmity between the two sisters, Zarika (Dawn) and Sončica (Sun), and ends with the former murdering the latter. But it is important to point out that 'murder' is only vaguely expressed in the ballad, with no hint before the metaphorical last verse "*The sister does not recognize her / And gives snake's poison for her to drink*", whereas 'kidnapping', and later ransom and rescue from "strong and evil" Turks, is made explicit.

As a rule, topic distributions for variant types with strongly expressed or dominant topics (such as 'infanticide' and 'kidnapping') will often show single topic domination, whereas types with less dominant or more obscure themes (e.g. 'tragic fate' or 'rejection') will commonly share multiple topics. Moreover, due to strong in-

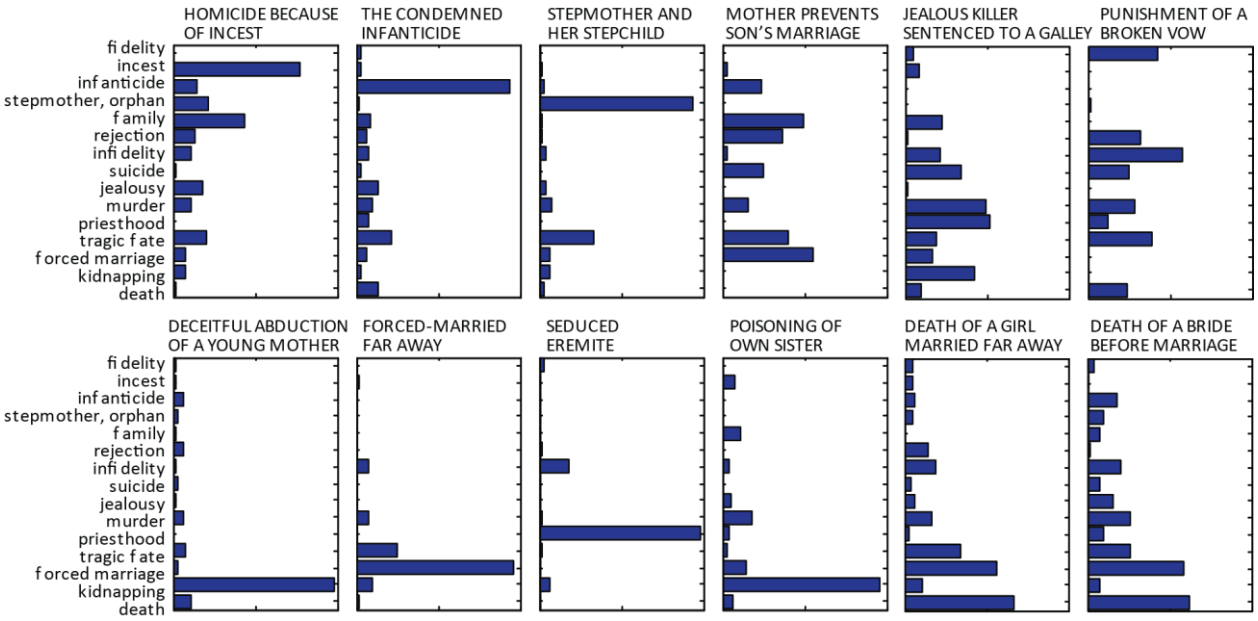


Figure 1. Topic distribution for selected variant types. 15 annotated topics from top down: fidelity, incest, infanticide, stepmother and orphan, family, rejection, infidelity, suicide, jealousy, murder, priesthood, tragic fate, forced marriage, kidnapping and death.

tertextuality, stories about murder, for example, prevail throughout the whole corpus, whereas stories about kidnapping, which typically involves "evil" Turks, are being represented by only a few variants. The fact that in our choice of tf-idf weighting for topic modeling, the significance of a word is inversely related to its frequency, is another factor weighing in on the topic distribution in favor of dominant, but not globally represented words. Hence, when considering 'murder' and 'kidnapping' in the same context, 'kidnapping' will often be taken as more salient. Nevertheless, LLDA has correctly assigned appropriate main topics for all selected variant types in Figure 1.

6. CONCLUSION

Our aim in the above experiments was to replicate some of the real world scenarios, similar to folklorist's approach to analysis and classification of topics and uncovering of latent semantic structure of a folk song. Experiment 2 has shown LDA could be used in classification of large folkloristic corpora, as it is able to disambiguate between major family types despite strong intertextuality of the corpus. Experiment 3, on the other hand, has shown that based on a set of few annotated examples we are able to infer general thematic structure and prevalent topics for different variant types in the corpus. Of course, results of these preliminary analyses should be further examined, but they nevertheless show LDA can uncover typical characteristics of individual variant type (e.g., compare variant type names and corresponding topics detected by LLDA in Figure 1). And, the ability of the LDA to detect multiple topics can further help us discover general relationships (similarities and differences) in the corpus, as shown by the three experiments.

NLP methods should be chosen with care. Our comparative analysis has shown that there are different representational structures generated by statistical and probabilistic models (see Experiment 1). These representations significantly differ in their composition, with LSA generated space having in general more 'unbalanced' distribution compared to LDA. This difference is especially evident in the Voronoi tessellation of the semantic space, where salience of individual regions is highly disproportional (e.g. dimensions overlap) given the topical distributions of the corpus. This is in line with the results from previous studies, (Blei et al., 2003; Steyvers & Griffiths, 2007) that show that probabilistic models generally output more discriminative and hence interpretable semantic structures compared to statistical similarity-space models.

In our future work we plan to enrich the studied materials with other song families and use the described techniques to visualize and explore the obtained semantic spaces. Also, we plan to study interrelations of lyric spaces to melodic spaces obtained by analyzing relationships between song melodies.

7. REFERENCES

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res., 3*, 993-1022.

Golež Kaučič, M., Kumer, Z., Terseglav, M., & Vrčon, R. (1998). *Slovenske ljudske pesmi IV: Ljubezenske pripovedne pesmi*. Ljubljana: Slovenska matica.

Golež Kaučič, M., Kumer, Z., Šivic, U., Terseglav, M., & Vrčon, R. (2007). *Slovenske ljudske pesmi V: Družinske pripovednepesmi*. Ljubljana: Založba ZRC, ZRC SAZU.

Grčar, M., Krek, S., & Dobrovoljc, K. (2012). *Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik*. Paper presented at the Osma konferenca jezikovnih tehnologij, Ljubljana.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *104*(2), 211-240.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. S. McNamara, S. Dennis & K. W. (Eds.), *Handbook of Latent Semantic Analysis* (pp. 424–440). Hillsdale, NJ: Erlbaum.

Strle, G., & Marolt, M. (2012). The EthnoMuse digital library: conceptual representation and annotation of ethnomusicological materials. *International Journal on Digital Libraries, 12*(2-3), 105-119.

PERFORMER PROFILING AS A METHOD OF EXAMINING THE TRANSMISSION OF SCOTTISH TRADITIONAL MUSIC

Scott Beveridge
Glasgow Caledonian University
sbe4@gcu.ac.uk

Ronnie Gibson
University of Aberdeen
r02rg11@abdn.ac.uk

Estefanía Cano
Fraunhofer IDMT
cano@idmt.fhg.de

1. INTRODUCTION

This paper presents work on profiling the playing styles of individual performers of Shetland fiddle music as a first step towards modelling the transmission of fiddle music throughout Scotland. The Shetland Isles, an archipelago situated 100 miles north of the Scottish mainland, consist of over 100 islands spread over approximately 551 square miles (Figure 1). Being equidistant from Scotland and Norway, the Isles are home to an autonomous heritage rich in social customs which, due to geographic isolation and the relatively low level of infrastructure, are further marked by a high degree of regional variation. As such, the Islands' fiddle tradition(s) represent an extremely interesting case study for modelling the transmission and evolution of Scottish fiddle music, encompassing local, regional, and national perspectives in different capacities.

This work represents a small scale investigation of computational music analysis and machine learning techniques suitable for modelling fiddle performance style. We examine a core set of fiddle performers from across Shetland with the aim of identifying distinguishing characteristics and developing models which can help study the transmission of this musical tradition.

2. AIM

There are two distinct aims of this study. First we seek to identify and extract salient musical features which can discriminate performance styles across fiddlers in the Shetland Isles. Second we want to test the efficacy of these features in simple modelling procedures aimed at grouping these musicians based on performance style. This work presents a nascent framework for the musicological investigation of the transmission of traditional fiddle music in this area.

3. METHOD

3.1 Corpus

As a basis for this work we collected a corpus of 52 music tracks from 5 performers across the Shetland Isles. These tracks were collected from the Tobar an Dualchais website [1] . Tobar an Dualchais is a project led by Sabhal Mòr Ostaig, University of the Highlands and Islands, and the School of Scottish Studies at the University of Edinburgh.

[1] http://www.tobarandualchais.co.uk/

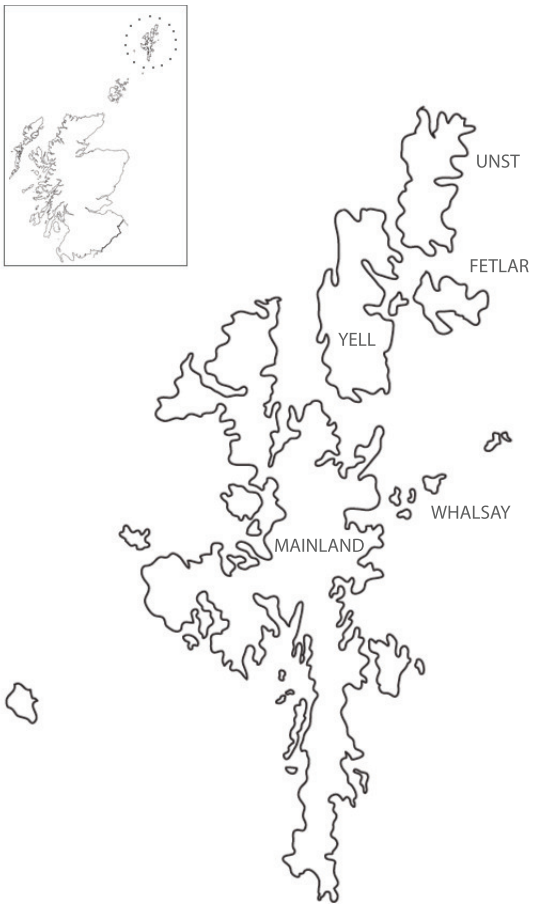


Figure 1: The Shetland Isles.

The project website provides material consisting of folklore, songs, and music. The original sound recordings are a combination of field recordings made on a Nagra III reel-to-reel tape recorder with Sennheiser microphones and copies of tapes of unknown origin. These were captured as monophonic wave files at 22.1KHz, 16 bit resolution.

3.2 Music descriptors

Many factors contribute to the unique sound of Shetland fiddle music, including an immediately distinguishable rhythmic quality determined by bowing styles intimately related to indigenous dance choreographies, and a degree of microtonality arising from the position in which the instrument is held. Figure 2 illustrates the 'old' style of holding the violin against the player's chest rather than under the

Figure 2: Andrew Poleson.

chin, which afforded his left hand less flexibility in stopping the strings compared to the modern classical hold (Cooke, 1986). For the purposes of this study we conducted an initial experiment seeking to characterise these stylistic differences by examining spectral features. We also present some early work examining the use of pitch contours to capture objective measures relating to microtonality.

We extracted the standard set of spectral and timbral features provided by the MIR Toolbox for Matlab (Lartillot & Toiviainen, 2007). For a representation of pitch contour, we implemented an algorithm by Cano et al. (2014), this is fully described in Section 5.

3.3 Modelling

We implement a simple K-Nearest Neighbour (K-NN) algorithm using the Euclidean distance measure and 1 nearest neighbour. Our objective is to produce a simple, informative model based an interpretable set of features. In order to reduce the size of the feature space we use the ReliefF feature selection algorithm (Frank & Witten, 2005). Using this method we identified the 3 most informative features. The model was evaluated using 5 fold cross-validation.

4. RESULTS

In the initial experiment we implemented the ReliefF feature selection process to identify a subset of 3 salient features. Our motivation for this feature space reduction is the relatively small corpus size, and amount of performer classes. The features selected by the algorithm largely describe statistics of the spectral frame-based representation. These include:

- Spectral brightness (frame-wise mean)
- Spectral centroid (frame-wise mean)
- Spectral kurtosis (frame-wise std)

Training the K-NN classifier with these features, the algorithm reported a cross validation accuracy of 73.1%. The confusion matrix in Table 1 shows performer class-by-class accuracy (See Table 2).

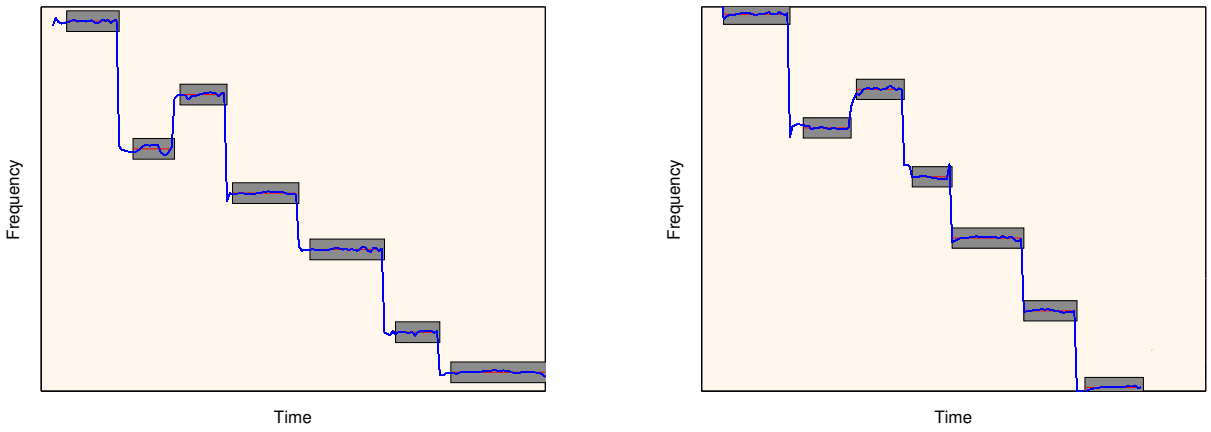
		Predicted class				
		rb	bp	wh	ji	ap
Actual class	rb	1	2	1	0	1
	bp	3	10	0	1	0
	wh	1	0	5	0	1
	ji	0	0	0	11	0
	ap	1	1	2	0	11

Table 1: Confusion matrix.

Both spectral centroid and brightness are measures of high frequency content in an audio signal. These initial results indicates the importance of high frequency content in traditional fiddle playing style. The third feature kurtosis, which describes the extent to which the peak of a distribution differs from that of the normal distribution, is perhaps less intuitive. As these are preliminary results, we plan to fully investigate the meaning of this feature in future work.

5. CURRENT WORK

To further characterize each performance, the use of pitch contours in conjunction with a transcription algorithm is proposed. We use the method presented in (Cano et al., 2014) further improved to better handle glissandos. This method extracts frame-wise instantaneous frequency sequences using the method proposed in (Salamon & Gómez, 2012). Each pitch sequence is then processed to obtain the sequence of tones that represents each performance. While having tone information can evidence important performance characteristics related to timing and phrasing, having frame-wise instantaneous frequency sequences allows the study of micro-variations in pitch and intonation that could be representative features of the different performance styles. In Figure 3, an example plot is shown where the pitch sequence and transcription results of a segment of a fiddle performance are presented. This excerpt is from the album *Bold* by contemporary Shetland fiddler,



(a) Traditional

(b) Contemporary

Figure 3: Performance style comparison of 'Three Drunken Fiddlers'.

Catriona MacDonald (MacDonald, 2000). The piece begins with an archive recording of the tune 'Three Drunken Fiddlers' by Gibbie Gray, a fiddle player from Shetland who lived between 1909 and 1989. The tune is then picked up by MacDonald and highlights the stark stylistic differences between traditional and contemporary performance styles. In both plots, the blue lines show the pitch contours, the different tones are shown as grey bars, and the mean frequency of each tone is indicated with a red line. It can be seen that the traditional performance in Fig. 3a shows more micro-variations in pitch throughout the duration of each tone. In contrast, the contemporary performance in Fig. 3b shows more stable pitch contours with less variations and more precise transitions between tones. These are only general observations on this particular example; however, they are good indicators of the analysis possibilities available when pitch contour studies and automatic music transcription are conducted.

The development of pitch contour features also helps to overcome the phenomenon known as the *album effect*, a common problem in the field of Music Information Retrieval (Kim et al., 2006). In our corpus, many of the recordings were captured with tape-based equipment which may be subject to idiosyncrasies. In order to mitigate the effects of these extra-musical characteristics, and to avoid biasing our classifier, we turned to pitch contour features. As these focus on tonal representations they are essentially independent of the recording process.

6. CONCLUSION

In this paper we presented work towards performer profiling of traditional Shetland fiddle musicians. In our preliminary experiment, the modelling process shows good results using musical descriptors based on spectral features. With this approach we reached 73.1% accuracy using K-NN classifier combined with ReliefF feature selection algorithm.

As part of ongoing work we also presented features based on pitch contour. The aim was to capture microtonal differences in performance, and mitigate idiosyncrasies relating to the original recording method which could bias our classification results.

Although in very early stages, both of these methods show considerable promise for the task of performer classification in this context. In future work we hope to expand the corpus, investigate new features, and move towards unsupervised feature identification. The overall driving force of this work will be interpretable features and models which support the musicological study of the transmission of traditional music across Scotland.

Performer	ID	No Tracks	Region
Robert Bairnson	rb	5	Mainland
Bobby Peterson	bp	14	Mainland
Willie Hunter	wh	7	Mainland
John Jamieson Irvine	ji	11	Whalsay
Andrew Poleson	ap	15	Whalsay

Table 2: Performer information.

7. REFERENCES

Cano, E., Schuller, G., & Dittmar, C. (2014). Pitch-informed solo and accompaniment separation towards its use in music education applications. *To appear in EURASIP Journal on Advances in Signal Processing*.

Cooke, P. (1986). *The fiddle tradition of the Shetland Isles*. CUP.

Frank, E. & Witten, I. H. (2005). *Data Mining: Practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann.

Kim, Y. E., Williamson, D. S., & Pilli, S. (2006). Towards quantifying the album effect in artist identification. Poster at the International Conference on Music Information Retrieval (ISMIR 2006), Canada.

Lartillot, O. & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects*, (pp. 237–244).

MacDonald, C. (2000). Bold, Peerie Angel. Audio CD.

Salamon, J. & Gómez, E. (2012). Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics. *IEEE Transactions on Audio, Speech and Language Processing*, *20*(6), 1759–1770.

EXPLORING PHRASE FORM STRUCTURES. PART II: MONOPHONIC JAZZ SOLOS.

Klaus Frieler, Wolf-Georg Zaddach
Liszt School of Music Weimar
{klaus.frieler, wolf-georg.zaddach@hfm-weimar.de

Jakob Abeßer
Fraunhofer IDMT Ilmenau
{jakob.abesser@idmt.fraunhofer.de

ABSTRACT

In this explorative study, we investigate the phrasal structure of a set of 100 monophonic jazz solo taken from the WEIMAR JAZZ DATABASE. The main purpose was to see whether phrase form structure might lead to useful features for computational jazz solo analysis. To this end, we extracted basic statistical descriptors for phrases such as the number of notes, event density, total duration etc. Furthermore, we analysed the self-similarity of phrase sequences with regard to semitone intervals and duration classes and in combination. Phrase form structure can be characterised by coherence values and runlengths. As expected, form coherence values are generally very low with duration-based form coherence being higher than interval-based or combined form structure. John Coltrane was found to be an exceptional case with very high duration-based coherences. Furthermore, a global tendency to increase in event density (i. e., intensity) at the beginning of solos was observed.

1. INTRODUCTION

In the second part of our explorations into phrase form structure (PFS), we examine monophonic jazz solos. Compared to folk songs, these solos represent quite a different set of data: solos are improvised not composed (or fixed by tradition), they are much longer, do not contain lyrics, and employ much more advanced rhythmic and tonal devices. Nevertheless, they are also structured in phrases, on a generative as well as on a perceptional level. Moreover, jazz musicians and their solos are also rooted in Western music traditions, which might be reflected in the PFS. From the outset, we do not except very large inner coherence in the PFS, since jazz solos are often thought to be structured like ad-lib speeches (Johnson-Laird, 2002). But motivic improvisation might lead to phrase coherence. Moreover, one could suppose that form coherence is used as a dramaturgical device. For example, a series of phrases with high similarity in the rhythmic or tonal domain can be used by the soloist to increase intensity or just to convey coherence. Likewise, phrase lengths and event densities can be employed to increase or lower the "power" level. The present studies sets out to explore such relationships, to check some of the stated hypotheses and to search for useful features.

To our knowledge, this is the first study which explores phrase form structure in jazz solos using a large set of data. Jazz researchers examined phrase structures in the past (e. g., Downs, 2001; Love, 2012), but rarely with regard to phrase similarities, and usually not using statistical methods but focusing on selected examples.

Form in music is generally a hierarchical and multi-layered phenomenon. Phrase form structure as considered here is not to be confused with the form of the underlying composition. Those structural levels might coincide, but often do not (Love, 2012). In the following, "form" is nearly always to be understood as "phrase form structure".

2. DATA

The analysis was carried with help of the WEIMAR JAZZ DATABASE included in the MeloSpySuite software toolkit [1] (Frieler et al., 2013), which at the time of the study (February 2014) contained 106 annotated monophonic jazz solos covering a wide range of soloist and styles. For technical reasons, only a subset of 100 solos from 38 soloist with a total of 368 choruses, 2643 phrases and 42015 tones was included in the study. Phrase annotation for jazz solos suffers from the same issues as folk songs, even though in the case of wind instruments phrase boundaries often coincide with breathing rests. The phrase annotation for this study was done by the transcribers of the solos which are musicology and jazz students. The data are in a pre-final status, which means they were cross-checked by an independent transcriber for basic correctness, which, however, did not include a revisal of phrase boundaries.

3. METHOD

To extract the form information, similarity values between each phrase of a song were calculated using edit distance (Levenshtein, 1965) on a interval-based or duration-class-based representation of the melody. Form strings (such as `AAAA`, `ABCD` etc.) were extracted from the resulting self-similarity matrices using fixed numerical thresholds. Two phrases p_i, p_j were deemed similar if $\sigma_I(p_i, p_j) \geq 0.6$ for intervals and $\sigma_D(p_i, p_j \geq 0.7$ for duration classes, where $\sigma(p_i, p_j)$ is the normed edit similarity taking values $\in [0, 1]$ (Müllensiefen & Frieler, 2004). For a more in-depth discussion on the choice of these values see the accompanying paper Frieler (2014), this volume. Similar phrases were denotated using the same symbol, i. e., no distinction between identical and similar phrases was made for sake of simplictity. The analysis was carried out on the level of whole solos and of single choruses. Additionally, number of notes, total duration (in seconds and in bar units), and event densities (per seconds or per bar)

[1] Available at `http://jazzomat.hfm-weimar.de`

were calculated for each phrase. Duration of a phrase is defined here as the time-interval between the onset of the first and the offset of the last tone. Bar units are based on a fractional representation of metrical positions, where each measure equals one numerical unit.

This analysis was carried out using the `melfeature` commandline tool from the MeloSpySuite (Frieler et al., 2013). The resulting data were imported into R (R Core Team, 2013) for further analysis.

4. RESULTS

A majority of 75% of the solos comprise 4 or less choruses, the median is 2 and the mean is 3.5 choruses. 30 solos consist of only 1 chorus and 28 of 2 choruses. A small peak can be found at 8 choruses (6 instances), possibly due to a previous arrangement of the musicians. Descriptive statistics of all used variables can be found in Tab. 1. Phrases have an average duration of about 3 sec, which coincides with estimates for the subjective presence time (Fraisse, 1982). This is also in good agreement with previous studies on ideational flow in jazz piano solos (Schütz, 2011; Lothwesen & Frieler, 2013). The mean length in bar units is about 2 bars.

Curiously, the mode of the distribition lies at 19 phrases with 8 instances, most likely just a chance result. For phrase lengths less or equal 16 phrases, there is a preference for an even number of phrases, which is the case for 77% solos. This is in agreement with our observations for folk songs (Frieler, 2014), and might be a reflection of the form of the underlying compositions, which are rooted in Western music traditions and thus prefer even-numbered structures. Beyond a length of 16 phrases, the situation is rather opposite—about 60% have odd number of phrases. Interestingly, the number of phrases is strongly decreasing with chorus position. A Spearman rank test for phrase count and chorus position became highly significant ($p < 0.0001, \rho = -0.42$, cf. Tab. 2). While the first choruses have a mean of 10.6 phrases, this number drops down to 4.0 for the sixth chorus. This correlation is rather stable; disregarding very short (less than 3 choruses) and very long solos (more than 11 choruses it is still highly significant ($p < 0.0001, \rho = -0.46$).

In Tab. 2, Spearman rank correlations between chorus and phrase position and various descriptors can be found. Most of these correlation are either non-significant or they indicate uncorrelatedness. Only event density (tones per second) is increasing with chorus and phrase position. However, this correlation is only occasionally present on the level of single solos (9 solos had significant correlation of event density with chorus position, and 13 with phrase position, but not always in the same direction).

In total, 72 different interval phrase form strings (IF) and 97 duration-based phase form strings (DF) were found, which gave rise to 58 distinct combined phrase form classes (CF). Combining was done by enumerating each unique pair of IF and DF symbols of a song with a new form symbol. As expected, many form strings were basically sequences of different form parts with only occasional repetitions. The most frequent form was `ABCDEFGH` with 5 instances for IF and 3 instances for DF.

One can define the *coherence* of a form as the amount of contained repetition, i. e. the number of unique elements divided by the total length of the form (subtracted from 1 for better interpretation). A coherence of 0 means that no form part is repeated, whereas a coherence of 1 can only be reached in the limit of a single, infinitely repeated part. We found a median coherence for IFs and CFs of 0.0, and for DFs of 0.25 (c.f. Tab. 1). As for folk songs, DF coherence is much higher than IF coherence (Frieler, 2014). Generally, the very low IF coherence might indicate that motivic improvisation is not very common or, alternatively, that it is not captured by our (rather simplistic) method of similarity calculation. Moreover, motivic improvisations should result in runs of similar phrases. To check this, we calculated run lengths for all form strings of all types. The median IF runlength is 0 (AM=0.42), the median DF runlength is 1 (AM=1.52), and the median CF runlenght is 0 (AM=0.18). Three-quarter of all solos have not a single IF run, whereas 38 have no DF run, and 88 solo no CF run. The longest IF run with 5 repetitions, which was also the longest CF run, occured in Freddie Hubbards solo on "Society Red". The longest DF with 11 elements was performed by John Coltrane in his legendary solo on "Giant Steps". A analysis of variance on DF run lengths revealed that John Coltrane is the only soloist in the database with a tendency to use DF runs ($F(37, 62) = 1.704, p = 0.031, R^2_{adj} = 0.21$). With regard to IF and CF runs, no difference between performers could be found, the same holds for IF and CF coherences. Again, John Coltrane was the only soloist which differed in DF coherence ($F(37, 62) = 2.011, p = 0.007, R^2_{adj} = 0.27$). Coltrane's mean DF coherence is 0.22 [2], about twice as high as the overall mean of 0.13. Furthermore, a analysis of variances for mean run lengths of the three different form types was carried out. For IF, Curtis Fuller and Kenny Garrett showed a tendency to longer runlengths. For DF, this was the case for Coleman Hawkins, Bob Berg, and Clifford Brown. However, this analysis has to be taken with care, because in general runs occur rarely, and for the majority of soloist there are only 1 or 2 solos in the database.

5. CONCLUSION

In this explorative study, we reported on basic statistics of jazz solo phrases in a relatively large set of 100 solos. We found that event density correlates high with chorus and phrase position, which might result in an increase in intensity during the course of a solo. The fact, that this correlation is only found at the corpus level and only occasionally for individual solos, can be interpreted such, that solos in general show a tendency to increase in intensity at the beginning. The same effect is observed for phrase positions as well.

[2] The DF coherence values for all 7 solos of John Coltrane are "Central Park West": 0.00, "26-2": 0.13, "Mr. P.C.: 0.16, "Blue Train": 0.17, "Countdown": 0.26, "Giant Steps": 0.31, "So What": 0.54.

Variable	Median	AM	SD	Range
Choruses/solo	2.00	3.54	3.57	$(1, 19)$
Phrases/solo	23.00	25.94	15.15	$(3, 66)$
Phrases/chorus	18.00	20.97	13.26	$(1, 66)$
Tones/solo	370.50	411.90	261.87	$(50, 1172)$
Tones/chorus	92.00	113.90	80.71	$(10, 616)$
Tones/phrase	12.00	15.9	13.5	$(1, 129)$
Event density/phrase (sec)	5.38	5.76	2.43	$(0.26, 19.47)$
Event density/phrase (bars)	7.62	8.65	4.36	$(0.74, 38.79)$
Duration/phrase (sec)	2.37	2.95	2.28	$(0.03, 20.64)$
Duration/phrase (bars)	1.53	2.04	1.71	$(0.23, 12.62)$
IF coherence	0.00	0.04	0.06	$(0.00, 0.25)$
DF coherence	0.26	0.28	0.20	$(0.00, 0.83)$
CF coherence	0.00	0.02	0.04	$(0.00, 0.22)$

Table 1: Descriptive statistics of various indices. Range is indicated in the format (min, max). AM=arithmetic mean, SD=standard devisation. X/Y means "(Number of) X per Y". Densities and duration are measured absolutely in seconds or relatively in bar units. Very small event-densities are due to one-note phrases. IF, CF, and DF coherences measure the amount of repetition in a phrase form structure. For more details see text.

Position	Variable	ρ	p
Chorus position			
vs.	Number of phrases	-0.51	0.000^{***}
	Number of notes	$+0.07$	0.002^{***}
	Event density (sec)	$+0.23$	0.000^{***}
	Event density (bars)	-0.03	0.129
	Total duration (sec)	-0.04	0.024^{*}
	Total duration (bars)	$+0.09$	0.000^{***}
	IF coherence	-0.06	0.204
	DF coherence	-0.08	0.112
	CF coherence	-0.03	0.566
	Coherence difference	-0.06	0.253
Phrase position			
vs.	Number of notes	-0.02	0.248
	Event density (sec)	$+0.17$	0.000^{***}
	Event density (bars)	$+0.04$	0.029
	Total duration (sec)	-0.11	0.000^{***}
	Total duration (bars)	-0.03	0.083

Table 2: Spearman rank correlations ρ of phrase statistics vs. chorus and phrase position.

For the significant decrease of phrase count with chorus position, we have currently no explanation. Regarding coherence, a few individual differences could be found, especially for John Coltrane, who has a much higher DF coherence. As expected, all coherences are generally rather low, but DF coherence is always higher than IF or CF coherence, just as for folk songs (Frieler, 2014). This could hint to an universal phenomenon in music. However, it might primarily be the result of a much smaller space of possible durations classes (here: five) compared to a larger event space for intervals, even though both event spaces are constrained by further tonal or metrical conditions. This need further examination.

All in all, phrase form structure as such seems to be only weakly exploitable for useful features, though, as the solos of John Coltrane show, it might be possible to capture exceptional cases. Other phrase characteristics such as event densities are more likely to be useful. However, our findings nicely corroborate the above cited analogy of jazz solos to ad-lib speeches (Johnson-Laird, 2002). In contrast, folk songs might more comparable to poems, which classically show more inner coherence and regularities than speeches or other narrative prose.

6. REFERENCES

Downs, C. (2000-2001). Metric displacement in the improvisation of Charlie Christian. *Annual Review of Jazz Studies*, *11*, 39–68.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 149–180). New York: Academic Press.

Frieler, K. (2014). Exploring phrase form structure. Part I: European Folk Songs. In Holzapfel, A., Cemgil, T., Mungan,

E., & Bozkurt, B. (Eds.), *Proceedings of the Fourth International Workshop on Folk Music Analysis (FMA2014), Istanbul*. Bogazici University.

Frieler, K., Abeßer, J., Zaddach, W.-G., & Pfleiderer, M. (2013). Introducing the Jazzomat Project and the Melo(S)py Library. In van Kranenburg, P., C. Anagnostopoulou, C., & Volk, A. (Eds.), *Proceedings of the Third International Workshop on Folk Music Analysis, Meertens Institute and Utrecht University Department of Information and Computing Sciences*, (pp. 76–78).

Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception*, *10*, 415–442.

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, *163*(4), 845–848. Englische Übersetzung in: Soviet Physics Doklady, 10(8) S. 707-710, 1966.

Lothwesen, K. & Frieler, K. (2013). Einflussfaktoren und Gestaltungsprinzipien formelbasierter Jazzimprovisation. In Jeßulat, A., Lehmann, A., & Wünsch, C. (Eds.), *Kreativität - Struktur und Emotion*, (pp. 256–265)., Würzburg. Königshausen & Neumann.

Love, S. (2012). An approach to phrase rhythm in jazz. *Journal of Jazz Studies*, *8*(1), 4–32.

Müllensiefen, D. & Frieler, K. (2004). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology*, *13*, 147–176.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Schütz, M. (2011). Improvisation im Jazz: Ideenflussanalyse von Jazzpiano-Improvisationen. Master, University of Hamburg.

EXPLORING THE MUSIC OF TWO MASTERS OF THE TURKMEN DUTAR THROUGH TIMING ANALYSIS

David Fossum
Brown University
dcfossum@gmail.com

Andre Holzapfel
Boğaziçi University, Istanbul
andre@rhythmos.org

ABSTRACT

In this paper, we analyze onset characteristics to try to identify important differences between two famous Turkmen dutar performers in terms of patterns of timing. We first analyzed annotated onset data for equivalent excerpts from recordings by these two musicians. We then analyzed unannotated onset data for a larger set of entire recordings. These analyses showed several conclusions. First, during introductory strumming outside the context of a composed melody, the two have different timing habits. Mylly aga is more consistent and Pürli aga more varied in terms of recurring inter-onset-intervals (IOIs). Second, during through-composed melodies, the timing profiles of the two musicians are very similar. This perhaps reflects the traditional Turkmen emphasis on preserving the form of traditional compositions in great detail and the attention paid to strumming technique. Finally, we found that automatically derived representations of rhythmic patterns, referred to as pulsation matrices, could be useful for identifying departures from typical timing patterns, which we could then analyze in order to understand such variations and their possible significance.

1. INTRODUCTION

The Turkmen dutar is the most popular traditional instrument in Turkmenistan, a former Soviet republic in Central Asia. It is a two-stringed, fretted lute which is strummed without a pick, using a variety of right-hand techniques. Virtuosos in some parts of the country have developed a repertoire of complex, through-composed pieces which have traditionally been transmitted orally through master-disciple lineages. In Turkmenistan, musicians and listeners often focus on right hand strumming style and technique in their discussions of dutar performances. Rhythm and timing are thus emphasized in evaluations of playing style.

In this paper, we analyze onset characteristics to try to identify important differences between famous performers in terms of patterns of timing. We will focus specifically on perhaps the two most prominent dutar players of the 20th century: Mylly Tamyradow (1885-1960), and Pürli Saryÿew (1905-1970). According to local convention, we will refer to these two as Mylly aga and Pürli aga, respectively. Mylly aga enjoys a reputation as a heritage-bearer, a great teacher who preserved old versions of traditional pieces and passed them on unchanged to students who would be leaders of the next generation of dutar greats. He is also celebrated for his unsurpassed technical mastery of the instrument. Pürli aga, by contrast, is known as an innovator and stylist, and is often credited with improvising new passages in traditional compositions, constantly exploiting spaces for variation and development on the fly. For more details on the Turkmen dutar and its players, please refer to Fossum (2010).

We took two different approaches in order to identify differences and similarities between Mylly aga and Pürli aga in terms of their timing habits. The first applies a semi-automatic approach, in which onsets are first determined using a signal analysis software, and these detected onsets are then corrected manually. Using this first method, we are able to focus on short excerpts played by both musicians and compare the timing of onsets observed for the two musicians in a representation that summarizes the observed Inter-Onset-Intervals (IOI) in a histogram. The second approach uses a representation presented in Holzapfel (2013), referred to as pulsation matrix. A pulsation matrix can be derived from an audio signal without any annotation, and allows therefore for an inspection of rhythmic properties in larger sections of audio.

2. ANALYSIS METHODS AND RESULT SUMMARY

For the first semi-automatic analysis, we selected portions of these musicians recordings, annotating the onsets identified by Sonic Visualizers [1] automatic onset detection feature. For the automatic detection, we applied the note onset detector from the Queen Mary plug-in set, using Spectral Difference with maximum sensitivity as a parameter choice. Due to this choice of parameters, annotation is necessary as the automatic detection in this software sometimes identified onsets that did not correspond to the right hand strokes that we considered meaningful for our timing analysis. We used this annotated data to create summary inter-onset interval histograms (IOI-H) of the recorded segments in question and compare how the two performers had played a collection of sections of musically identical phrases.

This approach allowed us to draw several conclusions. First, when looking at introductory portions of a piece, when the musicians perform a kind of open strumming that is free of melodic content, the two musicians have statistically significantly differences in timing habits. But, second, when we compared the musicians playing through-composed melodic phrases, from a statistical standpoint, it appears that Turkmen musicians attend to timing accuracy to such a level of detail that identifying meaningful timing variations is not feasible using the chosen approach

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

for detecting and summarizing onset times in a recording. While the pulsation matrix for different phrases was quite distinct, the profiles for each musician playing the same phrases was similar, with a few exceptions which we will discuss.

For the automatically obtained IOI representations, the pulsation matrices, we analyzed a number of entire recordings of each performer. While we were able to focus on specific phrases using a limited sample size using our first approach, in the second fully automatic approach we were able to analyze a larger duration of recordings. By inspecting the obtained results, we hoped to spot locations in the composed pieces in which the two players differ regarding to their rhythmic style. The approach used the same method as was applied in Holzapfel (2013), where as a first step an onset function is computed that takes large magnitudes in the vicinity of strokes by the dutar players. Then autocorrelations of this onset function are computed in small shifting windows of 2s length with a hop size of 0.2s, which provides us with a pulsation matrix that has the time of the audio recording on its x-axis, and the estimated IOI on its y-axis. Since our manually annotated data showed that the maximum duration between two onsets is 0.5s, we restrict the visualization of the obtained pulsation matrix to this range. We give a simple artificial example in order to explain the visualization in form of the pulsation matrix: In Figure 1 we depict a pulsation matrix that was obtained from a simple sequence of noise bursts of 8s length. Until the middle of this artificial example, only bursts with IOI of 0.2s and 0.4s exist, and in the second half a strong fast component at 0.1 is added. The bright areas indicate these active IOI.

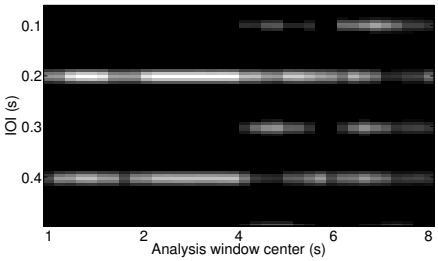
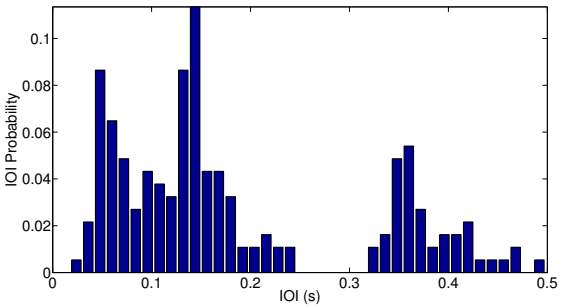
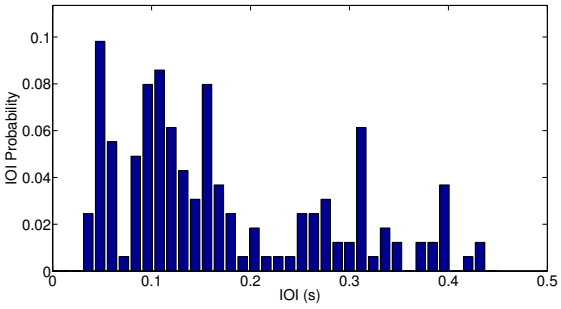


Figure 1: Artificial example of a pulsation matrix that was obtained from a simple sequence of noise bursts of 8s length.

This approach confirmed what we found in our analysis of the shorter melodic phrases: that this method of visualizing timing reveals how similarly, from a broader point of view, these musicians play the same piece, and how comparing pulsation matrices of different pieces shows the extent to which the pieces look rhythmically distinct from each other. However, this approach does allow us to scan a large number of pieces and visually identify occasional moments when one musician momentarily diverges meaningfully from the usual pulsation matrix.



(a) Mylly aga



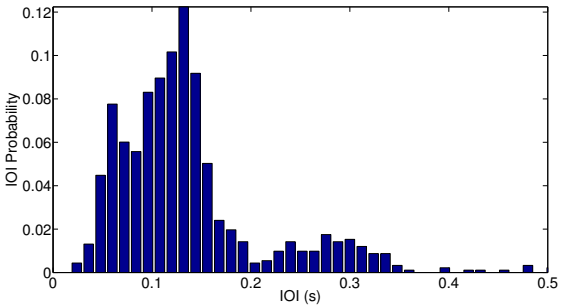
(b) Pürli aga

Figure 2: Summary IOI Histograms for introductory strumming (kakuw), derived from pieces in 2/4 time.

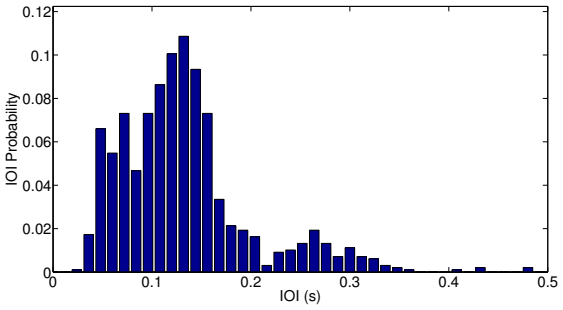
3. ANALYZING ANNOTATED TIMING DATA

Turkmen dutar players always begin a performance of a traditional piece by strumming the strings of the dutar open and then playing a short cadential formula. Such an introduction, called a *kakuw*, is not specific to the piece that is about to be played. Performers improvise the kakuw by stringing together stock rhythmic formulae. Kakuw affords an important opportunity for analysis in that it allows us to see the timing patterns of musicians outside the context of a melody. The musicians are just strumming the strings of the dutar to establish the metrical framework for the approaching composition. Musicians strum the kakuw in a meter that corresponds to that of the piece they are about to play.

For our first analysis, we visualized the timing profile for a number of introductory kakuw performances by each musician when they were playing in the same meter, 2/4. We analyzed 4 such introductions by Pürli aga and 4 by Mylly aga. What this analysis revealed was that Mylly aga was remarkably consistent in the typical inter-onset intervals that recur in his strumming of kakuw. Across different pieces, both tended to play at a consistent tempo, with the same few IOIs appearing over and over for Mylly aga, and a large variety of changing patterns appearing for Pürli aga. We conducted a statistical test to evaluate if the changes in IOI are likely to be generated by exponential distributions with parameters that are different for the two players. The result was a significantly higher variation in consecutive IOI values for Pürli aga. In Figure 2a and Figure 2b, respectively, we show the two summary IOI Histograms which depict the more stable IOI for Mylly aga, and clearly more multi-modal IOI Histogram for Pürli aga.

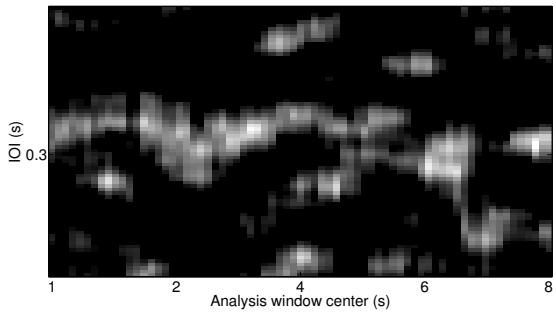


(a) Mylly aga

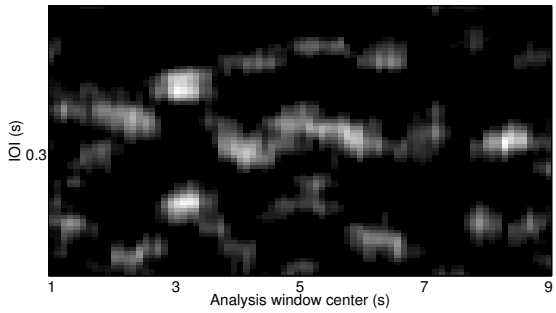


(b) Pürli aga

Figure 3: Summary IOI Histogram of a collection of the same melodic phrases played by both players.



(a) Mylly aga



(b) Pürli aga

Figure 4: Pulsation matrices for Mylly and Pürli aga playing the same excerpt from Ene.

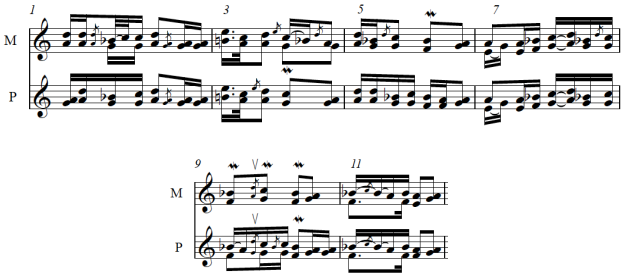


Figure 5: Transcription of a passage of Ene in which Pürli aga disrupts the timing profile. Beats are numbered for reference in analysis below.

Our next analysis visualizes timing profiles for the two musicians playing the same melodic phrases, the same excerpts from a traditional composition. This analysis revealed several things. First, when we compare the timing profile for the two musicians playing the same phrases, they look strikingly similar. This reflects the fact that the musicians are conservative in guarding the form and rhythmic characteristics particular to the traditional compositions they play. We illustrate this in Figures 3a and 3b, summary IOI Histograms of the musicians playing a set of the same melodic phrases.

Comparing the timing profile of the same musician playing phrases from two different pieces reveals that the timing profile changes according to the piece much more so than according to the musician playing it, at least as far as this method of analysis is able to illustrate.

4. ANALYZING NON-ANNOTATED TIMING DATA OF ENTIRE RECORDINGS

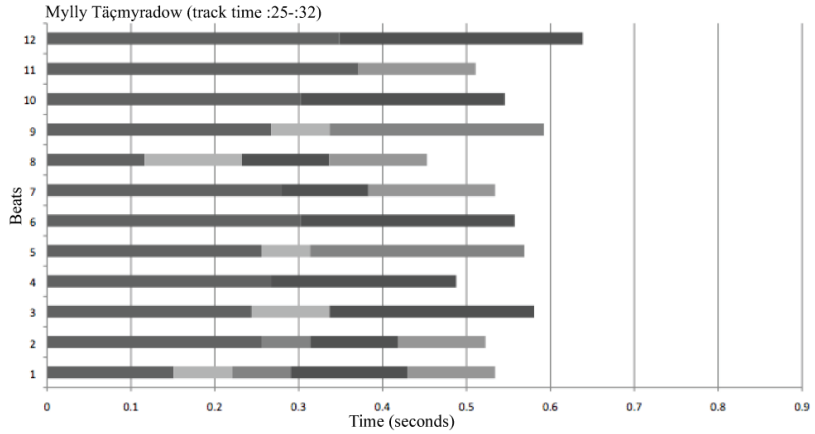
In addition to these analyses of select excerpts from these musicians recordings, we also created IOI representations in the form of pulsation matrices for entire recordings, without annotating the timing data. While the unannotated data was not as accurate in terms of consistently representing the timing of the right hand strokes that we wanted to isolate, it allowed us to scan a larger body of data, comparing the two musicians recordings of the same pieces to spot moments of divergence from each other.

We found that Pürli aga occasionally disrupted the usual timing profile by dramatically shortening or elongating a beat in the music. These moments were rare but may be nonetheless perceptually significant for listeners of Turkmen music who describe Pürli aga as a spontaneous performer who varies material more so than Mylly aga does. One such moment appears in the piece Ene. An IOI Histogram of Mylly agas performance of this excerpt is shown in Figure 4a.

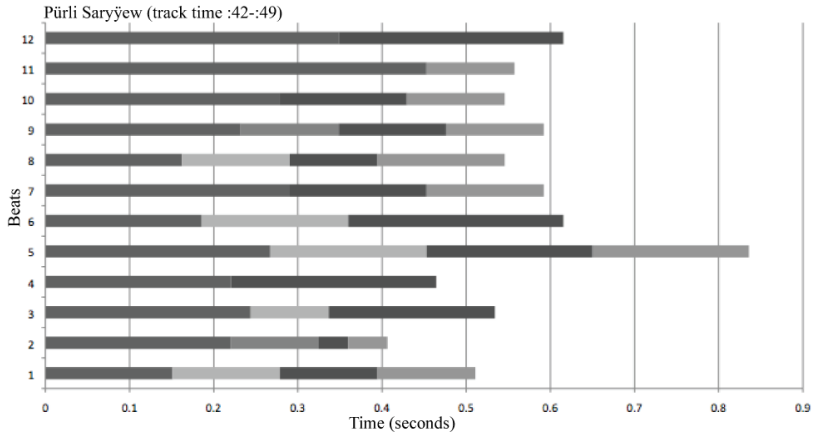
In Figure 4b, at the 3 second mark, the histogram of Purli agas performance reveals a disruption of the glowing white line that marks an otherwise consistently recurring IOI in this piece.

If we look at a transcription of this section of Ene, we can explain this disruption of the pulse as a microtiming device Pürli aga uses to draw out the phrasing of the melody. The following graphic (Figure 6) charts the durations, in seconds, of each stroke and beat in this passage.

Here Pürli agas beat 5 is more than double the length of his beat 2. What accounts for such dramatic shifts in tempo? The reason for this truncation of the beat is that it



(a) Mylly aga



(b) Pürli aga

Figure 6: Chart graphing the IOIs in a phrase of Ene as performed by Mylly aga (top) and Pürli aga (bottom). Lengths of each bar on the graph represent one beat (one quarter note in the transcription in Figure 5); shaded segments of each bar represent an IOI within the beat.

helps to stress the ensuing downbeat. In this case the emphasized downbeat, the chord b-e, constitutes a variation on the main tune. The earlier (standard) appearances of this melody are shown in Figure 6a.

In the passage we just heard, both performers have replaced the progression d-c-d at beat 3 with e-c-d, holding the e for a dotted 16th note to emphasize both the pitch climax and the fact that this represents a deviation from the norm for the tune. While both musicians hold this note, Pürli aga even further emphasizes it by rushing the previous beat to create an anticipatory flourish. We might call this anticipatory flourish a kind of microtiming device.

Pürli aga employs a contrasting microtiming device just three beats later, in the elongated beat 5 illustrated in the graph we just saw. Again, the melodic context reveals the apparent logic behind such timing: the melodic phrase is set to resolve to the tonic a during beats 6 and 7. Anticipating this, Pürli aga taps on the temporal brakes, so to speak. He does so not only by holding the first note in the beat, but also by playing four more or less even 16th notes at this reduced rate, ensuring that the listener hears this as a tempo shift rather than a mere fermata. The precadential positioning of the device is perfectly placed to help the melody shed some excess momentum before settling on the tonic. A pleasing byproduct of this precadential deceleration is its syncopated-sounding disruption of the expected pulse.

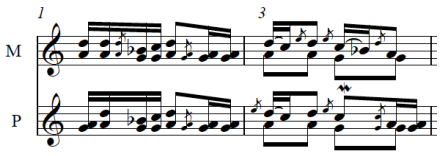


Figure 7: Transcription of an earlier permutation of the phrase depicted in Figures 5 and 6.

Such microtiming devices as precadential deceleration and anticipatory flourishes may offer another focal point of microtiming analysis. If we could identify a limited set of such devices and track their appearance across many performances, what patterns would emerge from this data? Do Mylly aga and Pürli aga have favorite microtiming devices or employ them characteristically at particular moments? How do listeners interpret such habits, if they notice them? Turning to the theme of transmission, we might ask which microtiming devices seem to get transmitted and survive

across time, what the particular microtiming habits of a musical lineage are, and whether musicians seem to perceive particular microtiming devices as an essential part of a traditional piece to be preserved.

One potential use for creating automatically derived pulsation matrices, then, would be to allow for scanning over a large number of pieces and identifying such moments when they appear. Then a closer analysis of the phrase in question could allow us to consider how the microtiming device compares to other examples we find and draw conclusions about any patterns in the recurrence of such devices.

5. CONCLUSIONS

Our analysis revealed several aspects of timing in Turkmen dutar performance, at least as regards two of the most famous musicians, Pürli aga and Mylly aga. First, during introductory strumming outside the context of a composed melody, the two have different timing habits. Mylly aga is more consistent and Pürli aga more varied in terms of recurring IOIs. Second, during through-composed melodies, the timing profiles of the two musicians are very similar. This perhaps reflects the traditional Turkmen emphasis on preserving the form of traditional compositions in great detail and the attention paid to strumming technique. Finally, we found that creating IOI Histograms could be useful for identifying departures from typical timing patterns, which we could then analyze in order to understand such variations and their possible significance.

6. ACKNOWLEDGEMENTS

Andre Holzapfel is supported by a Marie Curie Intra European Fellowship (PIEF-GA-2012- 328379).

7. REFERENCES

Fossum, D. (2010). Turkmen dutar and the individual. Master's thesis, Wesleyan University, Middletown, CT.

Holzapfel, A. (2013). Tempo in Turkish improvisation. In *Proceedings of the 3rd Workshop on Folk Music Analysis*, Amsterdam, Netherlands.

TOWARDS ALIGNMENT OF SCORE AND AUDIO RECORDINGS OF OTTOMAN-TURKISH MAKAM MUSIC

Sertan Şentürk, Sankalp Gulati, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{sertan.senturk,sankalp.gulati,xavier.serra}@upf.edu

1. ABSTRACT

Audio-score alignment is a multi-modal task, which facilitates many related tasks such as intonation analysis, structure analysis and automatic accompaniment. In this paper, we present a audio-score alignment methodology for the classical Ottoman-Turkish music tradition. Given a music score of a composition with structure (section) information and an audio performance of the same composition, our method first extracts a synthetic prominent pitch per section from the note values and durations in the score and a audio prominent pitch from the audio recording. Then it identifies the performed tonic frequency by using melodic information in the repetitive section in the score. Next it links each section with the time intervals where each section performed in the audio recording (i.e. structure level alignment) by comparing the extracted pitch features. Finally the score and the audio recordings are aligned in the note-level. For the initial experiments we chose DTW, a standard technique used in audio-score alignment, to show how well the state-of-the-art performs in makam musics. The results show that our method is able to handle the tonic transpositions and the structural differences with ease, however improvements, which address the characteristics of the music scores and the performances of makam musics, are needed in our note-level alignment methodology. To the best knowledge this paper presents the first audio-score alignment method proposed for makam musics.

2. INTRODUCTION

Audio recordings and scores are the two most relevant representations of music, both of which provide invaluable information about the melodic, metrical, expressive and cultural characteristics of the music. Audio-score alignment is a multimodal music information retrieval task, which aims to synchronise the musical events in an audio performance of a music piece with corresponding events in the score of the same piece. The aligned information from both sources may be used to enhance or facilitate tasks such as automatic accompaniment, audio-lyrics alignment, source separation, structure analysis, music discovery, tuning and intonation analysis, rhythmic analysis (Müller, 2007).

The current state-of-the-art focuses on aligning scores and audio of Eurogenetic musics. Nevertheless, incorporating knowledge specific to different music traditions in computational tasks might produce more accurate results (Şentürk et al., 2014). In this paper, we propose a audio-score alignment method which addresses some of specificities of Ottoman-Turkish *makam* music. To the best of our knowledge, this is the first audio-score alignment method applied to *makam* musics.

This remainder of the paper is structured as follows: Section 3 introduces the basic concepts of Ottoman-Turkish *makam* music, Section 4 explains the proposed methodology. Section 5 presents the data collection. Section 6 presents the experiments. Section 7 presents the results and the discussions and Section 8 give a brief conclusion.

3. MAKAM MUSIC

*Makam*s are modal structures, which are commonly used in Turkey, Middle East, North Africa, Greece and Balkans. Makams typically involve intervals smaller than a semitone. Moreover the notes are not equal tempered and the tuning/intonation of the intervals might differ with respect to the region, *makam* and performer. Each performance involves expressive usage of note repetitions, omissions, embellishments and timings (Ederer, 2011). Moreover, the musicians may decide to repeat, insert or omit musical phrases or even entire sections. It is also common to transpose the tonic of a performance due to instrument/vocal range or aesthetic concerns (Ederer, 2011). Makam music is also rich with heterophony, i.e. simultaneous variations of the same melody performed in the register of the instrument(s) or the voice(s) in the vocal pieces (Cooke, 2013). Typically the degree of heterophony increases with the number of instruments/voices, i.e. a solo ney recording is monophonic, while an ensemble/chorus recordings typically consist of complex heterophonic interactions.

In this paper, we focus on the classical Ottoman-Turkish tradition. Arel-Ezgi-Uzdilek (AEU) theory is the mainstream music theory used to explain the classical Ottoman-Turkish music. Note that the tuning and/or the intonation of a performed note might be substantially different from the theoretical interval (Bozkurt et al., 2009). We test our methodology on a collection of audio recordings and scores of *peşrev* form, which is one of the most common instrumental forms of classical tradition of the Ottoman-Turkish music. *Peşrev*s commonly consist of four distinct hanes and a teslim section, which typically follow a verse-refrain-like structure.

While *makam* musics are predominantly oral, in classical Ottoman-Turkish music a score representation extending the Western music notation is typically used to comple-

ment the practice (Popescu-Judetz, 1996). The scores do not show any embellishments, expressive decisions, heterophonic interactions, tuning or intonation information of the *makam* but only the basic melodic progression. The symbols in the music notation typically follow the rules of AEU music theory (Popescu-Judetz, 1996).

4. METHODOLOGY

We define *audio-score alignment* as synchronisation of the musical events in the score of a composition with the corresponding events in the audio recording of the same composition. In this paper we deal with two levels of granularity in the alignment: **1)** Section level, **2)** Note level. Our method addresses the some of main challenges of computational analysis of Ottoman-Turkish *makam* musics namely the transpositions, structural differences and the tuning.

Given a machine-readable score of a composition with the note symbols, the note durations and the annotated section information (i.e. the beginning and ending notes of each section) and an audio recording of the same performance, our method first extracts a prominent pitch of the audio recording [1] and synthesises a prominent pitch from the basic melody of each section indicated in the score such. In the synthesised prominent pitch of each section, the tonic symbol is assigned to 0 and the other note symbols are mapped to the melodic intervals defined by the AEU theory. The audio prominent pitch is also normalised such that the tonic frequency is assigned to 0. We then compute a distance matrix per section in the score between the synthesised prominent pitch of the section and the audio prominent pitch. We convert the distance matrix to a binary similarity matrix by assigning 1 to each point having a pitch distance less than 2.5 Holdrian commas ($\approx$ 56.6cents, a little bit higher than a quarter tone) and 0 otherwise. Binarisation takes care of the tuning and intonation differences between the synthesised and the performed pitches. In the binary matrices some pixels are distributed such that they form blobs similar to diagonal line segments. These line segments hint the locations of the sections in the audio. We use Hough transform (Duda & Hart, 1972), a common line detection algorithm to detect these lines and hence link the score sections with their corresponding time-intervals in the audio recording using the methodology. [2]

Note that for the above methodology to work the tonic frequency of the audio recording should be correctly identified. To handle this issue, we compute a pitch class distribution from the audio prominent pitch and extract the peaks in the distribution as tonic candidates. Assuming each peak as the tonic, attempts to linking the repetitive section using the methodology explained above. The tonic is identified as the pitch class which produces the most confident links. [3]

After we link the score sections with the audio recording, the time-interval of each section link is extended by 3 seconds to deal with the tempo differences. Then, we attempt to align the note events within each section with the corresponding time-intervals of the section links in the audio recording. For the preliminary experiments, we aim to present how well a standard alignment technique applied to other musics performs on makam musics.

We use dynamic time warping (DTW), which is a standard technique for audio-score alignment (Müller, 2007, Chapter 4). In order to avoid pathological warping during alignment, a known issue in DTW computation, we apply local constraints as discussed in (Sakoe & Chiba, 1978). We select the step size condition as $\{(2,1), (1,1), (1,2)\}$, which is a common local constraint in audio-score alignment (Müller, 2007, Chapter 4). In addition to the local constraint we also apply global constraint as discussed in (Sakoe & Chiba, 1978). The bandwidth of the global constraint is selected as 20% of the query length. Further, we also leverage the condition for the alignment path to be from the start of the strings to the end by implementing a subsequence version of the DTW as described in (Müller, 2007, Chapter 4).

For each section link, we apply subsequence DTW between the features audio prominent pitch extracted from each section link in the audio recording and the corresponding section in score. The extracted features are prominent pitch from the audio recording and synthetic prominent pitch from the section as explained above. We octave-wrap the features to deal with the octave errors and use City Block Distance (L1) for the distance computation at each step in DTW. Octave-wrapped City Block distance was previously shown as an effective and intuitive distance metric for comparing pitch values (Şentürk et al., 2014).

5. DATA COLLECTION

For the initial experiments, we collected 6 audio recordings of 4 *peşrev* compositions from the classical Ottoman-Turkish tradition. [4] The recordings are performed in a variety of transpositions. There are 51 sections in the audio recordings in total. The duration of the sections are 36.1 seconds on average with a standard deviation of 16.2 seconds.

The scores for each composition are obtained from the SymbTr collection (Karaosmanoğlu, 2012). The SymbTr-scores are machine-readable files, which contains note values on 53-TET (tone-equal tempered) resolution and note durations. These SymbTr-scores are divided into sections that represent structural elements in *makam* music. The beginning and ending notes of each section are indicated in the instrumental SymbTr-scores. These scores also follow the section sequence of the composition. [5]

[1] We use the Essentia implementation (Bogdanov et al., 2013) of the Melodia algorithm (Salamon & Gómez, 2012)

[2] For detailed information on the section linking methodology, we refer the reader to (Şentürk et al., 2014)

[3] For detailed information on the tonic identification methodology, we refer the reader to (Şentürk et al., 2014)

[4] In the text and in the supplementary results, we use MusicBrainz Identifier (MBID) as an unique identifier for the audio recordings and compositions. For more information on MBIDs please refer to *http://musicbrainz.org/doc/MusicBrainz_Identifier*.

[5] The SymbTr-scores, the audio and composition metadata, the annotations and the complete results are available at *http://compmusic.upf.edu/node/218*.

symbTr-score	Audio MBID	Instrumentation	#Anno	t_p	f_p	f_n	$F_1\%$
beyati–pesrev–hafif—-seyfettin_osmanoglu	70a235be-074d-4b9b-8f94-b1860d7be887	ensemble	906	790	116	116	87.2
huseyni–pesrev–muhammes—-lavtaci_andon	8b78115d-f7c1-4eb1-8da0-5edc564f1db3	ensemble	614	482	132	132	78.5
	9442e4cf-0cb3-4cb3-a060-77aa37392501	ney & percussion	302	260	45	42	85.7
rast–pesrev–devrikebir—-giriftzen_asim_bey	31bf3d56-03d8-484e-b63c-ae5ae9a6e733	tanbur	658	374	306	281	56.0
	5c14ad3d-a97a-4e04-99b6-bf27f842f909	ney	673	418	262	255	61.8
segah–pesrev–devrikebir—-yusuf_pasa	e49f33b8-cf8a-4ca9-88cf-9a994dbad1c0	ney & kanun	743	267	490	476	35.6

Table 1: Results of note-level alignment per experiment

6. EXPERIMENTS

Given a score of a composition and an audio recording of the same composition, we align the note onsets in the symbTr-score of a composition with the corresponding audio performance of the same composition using the methodology explained in Section 4 and obtain aligned note onsets in the audio recording.

As the ground truth, we use the manual note annotations collected for the evaluation of (Benetos & Holzapfel, 2013). For the data collection explained in Section 4, the total number of the note annotations in the audio recordings are 3896. These annotations typically follow the note sequence in the symbTr. Note that there are 3 inserted and 49 omitted notes in the annotations with respect to the symbTr-scores.

To evaluate the tonic identification, we compare the distance between the pitch class of the estimated tonic and the pitch class of the annotated tonic as explained in (Şentürk et al., 2013). If the distance is less than 1 Hc, the estimation is marked as correct.

To evaluate section linking, we check the time distance between the time interval of annotated sections and sections links as explained in (Şentürk et al., 2014). A section link is marked as a true positive, if an annotation in the audio recording and the link has the same section label, and the link is aligned with the annotation, allowing a tolerance of ± 3 seconds. All links that do not satisfy these two conditions are considered as false positives. If a section annotation does not have any links in the vicinity of ± 3 seconds, it is marked as false negative.

To evaluate the note-level alignment, we compare the aligned onset and the corresponding annotated onset. We consider the aligned onset as a true positive if the distance is less than $\pm$200ms. If the distance is higher than $\pm$200ms, the aligned onset and the annotated onset are labeled as false positive and false negative, respectively. The insertions are ignored in the evaluation. If the aligned note corresponding to an omitted annotation is not rejected (i.e. the duration is non-zero), it is deemed as a false positive.

From these quantities we compute the F_1-scores for section linking and note-level alignment separately as:

$$P = \frac{t_p}{t_p + f_p}, \quad R = \frac{t_p}{t_p + f_n}, \quad F_1 = 2\,\frac{P\,R}{P + R} \qquad (1)$$

t_p, f_n, f_p, P, R and F_1 stand for number of true positives, number of false negatives, number of false positives, precision, recall and F_1-score, respectively.

7. RESULTS AND DISCUSSION

Across all the experiments, the tonic is identified correctly (100% accuracy in tonic identification) and all the sections were linked perfectly ($F_1 = 100\%$ for section linking). In the note level our methodology is able to align 2591 notes out of 3896 notes correctly, yielding to an F_1-score of 66.1%. The mean, median and standard deviation of the time-distance between the aligned note and the corresponding annotation are 299, 93 and 498 milliseconds, respectively. Moreover, 89.2% of the notes are aligned with a margin of ± 1 second, implying that DTW does not lose track of the melody.

Previously in (Şentürk et al., 2013) and (Şentürk et al., 2014) we showed that our linking methodology is highly reliable for tonic identification and section linking. The results in this paper also comply with these previous findings.

To understand the common mistakes in the note-level, we examined the aligned notes against annotated notes. Table 1 shows the results per experiment. The expressive embellishments in the performance (*portamento*s, *legato*s, *trill*s etc.) are common reasons of misalignment. For example, DTW infers portamentos as an insertion and the note onsets are aligned around the time when the portamento reaches to the stable note pitch. Similarly when there is a melodic interval less than a whole tone, a trill might cause a note onset to be marked earlier. Since these embellishments are not shown in the score, standard (subsequence) DTW was expected to fail. While we can argue that the section-level alignment is accurateXX, the results in the note-level alignment show that there is still more room for improvement for note-level alignment.

From Table 1, it can be seen that the note-level alignment fails for most of the notes in the audio recording of *Segah Peşrev* within the $\pm$200ms tolerance. This is a recording with ney and kanun, which consists of heterophonic interactions such as embellishments played by a single musician and time differences in note onsets between the performers. Due to such cases, the time distance between aligned onset and the annotated is typically larger than 200ms. Note that 75% of the notes are still aligned correctly within a tolerance of ± 1 second.

8. CONCLUSION

In this paper, we propose a method to align scores of *makam* musics with their associated audio recordings. Our system is able to handle the transpositions and structural rep-

etitions and omissions in the audio recordings, which are common phenomenon in *makam* musics. The results obtained from the data collection present a proof-of-concept that a standard technique such as DTW can be effective for audio-score alignment for *makam* musics in the note level. Nevertheless, we need incorporate additional steps to handle non-notated embellishments and note omissions, insertions and repetitions.

Currently method relies on manual section segmentations in music scores. Manual segmentation of the score is not an difficult task compared to the note-level audio-score alignment itself. Nevertheless, it might be desirable to use other methodologies that do not require structural segmentations (e.g. (Gasser et al., 2013)), especially when we are working on large audio-score collections.

While we didn't have such an example in our data collection, there can be also omissions, insertions and repetition of phrases inside the sections. Currently, our methodology cannot handle such cases. In the future we want to use the JumpDTW proposed by (Fremerey et al., 2010) to handle phrase omissions, insertions and repetitions. Another approach might be segmentation of the symbolic score into melodic phrases and link extracted phrases from score with the corresponding audio recording. Recently, Bozkurt et al. (Bozkurt et al., pted) came with a method for segmenting music scores into melodic phrases according to the makam and usual information. Our initial experiments using the extracted phrases show that phrase linking is highly accurate. We observed that the erroneously linked phrases are almost identical to the true phrase, differing by very few pitches or durations, hence note-level alignment does not suffer a large number of errors.

We are extending the data collection to cover more examples from the CompMusic collection. In audio recordings with heterophonic interactions (such as the audio recording of *Segah Peşrev*) there is an ambiguity of the exact timings in the note onsets. To study the implications we plan to make several annotators, annotate the notes in the same set of scores and audio recordings. We will jointly compare the onset markings from each annotator with the aligned onsets produced by the future iterations of our automatic audio-score alignment method.

9. ACKNOWLEDGEMENTS

We would like to thank André Holzapfel for providing the note annotations. This work is partly supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

10. REFERENCES

Benetos, E. & Holzapfel, A. (2013). Automatic transcription of Turkish makam music. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*.

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., & Serra, X. (2013). Essentia: An audio analysis library for music information retrieval. In *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR)*.

Bozkurt, B., Karaosmanoğlu, M. K., Karaçalı, B., & Ünal, E. (accepted). Usul and makam driven automatic melodic segmentation for Turkish music. *Journal of New Music Research*.

Bozkurt, B., Yarman, O., Karaosmanoğlu, M. K., & Akkoç, C. (2009). Weighing diverse theoretical models on Turkish maqam music against pitch measurements: A comparison of peaks automatically derived from frequency histograms with proposed scale tones. *Journal of New Music Research*, *38*(1), 45–70.

Cooke, P. (accessed April 5, 2013). Heterophony. Grove Music Online. `http://www.oxfordmusiconline.com/subscriber/article/grove/music/12945`.

Duda, R. O. & Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, *15*(1), 11–15.

Ederer, E. B. (2011). *The Theory and Praxis of Makam in Classical Turkish Music 1910-2010*. PhD thesis, University of California, Santa Barbara.

Fremerey, C., Müller, M., & Clausen, M. (2010). Handling repeats and jumps in score-performance synchronization. In *Proceedings of 11th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 243–248).

Gasser, M., Grachten, M., Arzt, A., & Widmer, G. (2013). Automatic alignment of music performances with structural differences. In *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 607–612)., Curitiba, Brazil.

Karaosmanoğlu, K. (2012). A Turkish makam music symbolic database for music information retrieval: SymbTr. In *Proceedings of 13th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 223–228).

Müller, M. (2007). *Information retrieval for music and motion*, volume 6. Springer Heidelberg.

Popescu-Judetz, E. (1996). *Meanings in Turkish Musical Culture*. Istanbul: Pan Yayıncılık.

Sakoe, H. & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *26*(1), 43–49.

Salamon, J. & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(6), 1759–1770.

Şentürk, S., Gulati, S., & Serra, X. (2013). Score informed tonic identification for makam music of Turkey. In *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 175–180)., Curitiba, Brazil.

Şentürk, S., Holzapfel, A., & Serra, X. (2014). Linking scores and audio recordings in makam music of Turkey. *Journal of New Music Research*, *43*, 34–52.

AUTOMATIC LYRICS-TO-AUDIO ALIGNMENT IN CLASSICAL TURKISH MUSIC

Georgi Dzhambazov, Sertan Şentürk, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{georgi.dzhambazov,sertan.senturk,xavier.serra}@upf.edu

ABSTRACT

We apply a lyrics-to-audio alignment state-of-the-art approach to polyphonic pieces from classical Turkish repertoire. A phonetic recognizer is employed, whereby each phoneme is assigned a hidden Markov model (HMM). Initially trained on speech, the models are adapted on singing voice to match the acoustic characteristics of the test dataset. Being the first study on lyrics-to-audio alignment applied on Turkish music, it could serve as a baseline for singing material with similar musical characteristics. As part of this work a dataset of recordings from the classical music tradition is compiled. Experiments, conducted separately for male and female singers, show that female singing is aligned more accurately.

1. INTRODUCTION

Lyrics are one of the most important musical components of vocal music. When a performance is heard, most listeners will follow the lyrics of the main vocal melody. From this perspective, the automatic synchronization between lyrics and music poses a user-demanded research question.

By applying a lyrics-to-audio alignment state-of-the-art approach to classical Turkish songs, we aim to outline the research challenges, raised by the musical aspects peculiar to this music tradition. To this end we compile a dedicated evaluation corpus. This work is performed in the context of the CompMusic project [Serra, 2011], which aims to analyze non-western music traditions in a culture-specific manner. In this respect, the corpus is built as well with the intention to be useful for further music retrieval tasks for the Turkish tradition.

2. ELEMENTS OF CLASSICAL MUSIC OF TURKEY

Şarkı - the scope of this study - is a vocal form in the classical repertoire. Typical for it is that vocal and accompanying instruments follow the same melodic contour in their corresponding registers with slight melodic variations. However, the vocal line has usually melodic predominance. This musical interaction is termed heterophony.

Additionally, the *şarkı* form adheres to a well-defined verse-refrain-like structure: a *şarkı* contains *zemin* (verse), *nakarat* (refrain), *meyan* (second verse), *nakarat* (refrain) sections, which are preceded by *aranağme* (an instrumental interlude) [Ederer, 2011].

Concerning language, unlike modern Turkish, Ottoman Turkish is characterized by more loanwords from Arabic and Persian origin. The lyrics language for the *şarkı* compositions in our evaluation dataset spans both modern and Ottoman Turkish. The Turkish phonology comprises 38 distinctive phonetic sounds, 8 of which are vowels. There are no diphthongs, and when vowels come together, they retain their individual sounding. Lengthening of vowels is realized by a non-pronounced character *ğ*. However vowel length has a negligible importance in sung Turkish.

In classical Turkish singing an expressive effect occurs frequently: When performing vibrato, singers tend to alternate between the original vowel and another helper one, simultaneously to alternating the pitch.

3. RELATED WORK

To date most of the studies of singing voice in general and the automatic lyrics-to-audio alignment in particular are focused on western polyphonic popular music. Many approaches exploit phonetic acoustic features.

An example of such a system [Fujihara et al., 2011] relies on a forced alignment scheme and was tested on Japanese popular music. Since the forced alignment technique was originally developed to carry out the alignment between clean speech and text, accompanying instruments and non-vocal sections deteriorate the alignment accuracy. To address this issue, the authors perform automatic segregation of the vocal line and the alignment is run using phonetic features extracted from the vocal-only signal.

A diametrically different approach is to deploy external information sources. Müller et al. [2007] uses MIDI files, which are manually synchronized to lyrics. By performing mapping of timestamps between an audio recording and a MIDI version of the composition, lyrics are implicitly aligned to the audio.

4. METHOD

Combining aspects of these two methods, in this work we develop a system for the automatic synchronization between vocal *şarkı* recordings and their lyrics. Similar to the approach of Fujihara et al. [2011] we train a hidden Markov model (HMM) for each phoneme, present in Turkish language.

Furthermore, we exploit a lyrics representation, for which sections are labeled. Songs are segmented into structural

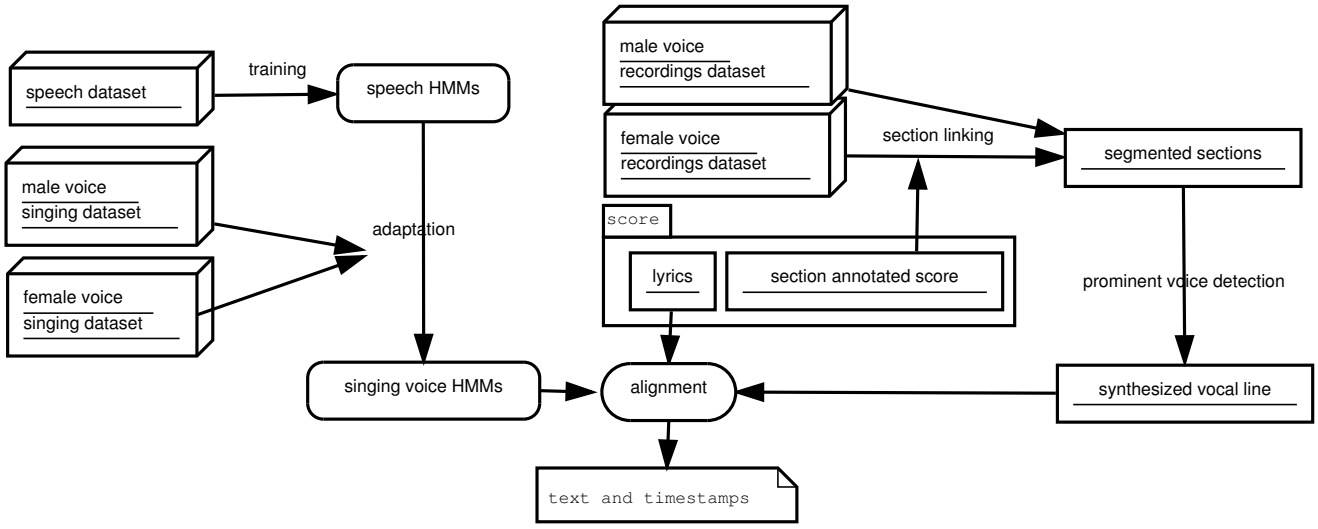


Figure 1: Training steps (on the left) and alignment process (on the right)

parts through an external module, which links sections from the musical score to temporal anchors in the audio [Şentürk et al., 2014]. Using a Viterbi forced alignment the system aligns in a non-linear way the extracted phonetic features to a network of the trained phoneme models.

Figure 1 gives an overview of the layout of the system.

4.1 Training

4.1.1 Training a speech model

In the absence of annotated data of singing phonemes, we train mono-phone models on a big corpus of annotated Turkish speech. Later, we adapt the speech models to match the acoustic characteristics of clean singing voice using a small singing dataset (see Figure 1).

The acoustic properties (most importantly the formant frequencies) of spoken phonemes can be induced by the spectral envelope of speech. To this end, we utilize the first 12 mel-frequency cepstral coefficients (MFCCs) and their difference to the previous time instant.

A 3-state HMM model for each of 38 Turkish phonemes is trained, plus a silent pause model. For each state a 10-mixture Gaussian distribution is fitted on the feature vector.

4.1.2 Adaptation

Mesaros & Virtanen [2008] have proposed to apply an adaptation technique for speaker-dependent speech modeling. The Maximum a posteriori (MAP) transform - applied as well in this work - shifts the mean and variance components of the Gaussians of the speech model towards the acoustic characteristics of the singing voice. An advantage of the MAP transform compared to other adaptation techniques is that it allows the manipulation of each phoneme model independently.

In singing the articulatory characteristics of unvoiced consonants do not vary significantly from these in speech. This is because unvoiced consonants do not bear any melodic line. For this reason we adapted only the vowels and voiced consonants.

4.2 Preprocessing steps

4.2.1 Section Linking

In the *şarkı* form each vocal segment is associated with a section (e.g. *nakarat*). Furthermore, a given performance of a *şarkı* composition typically contains section repetitions or omissions, which are not indicated in the score. Thus, for each audio recording, prior to alignment, we utilize a method for linking score sections to their beginning and ending timestamps [Şentürk et al., 2014] (see Figure 1). Non-vocal *aranağme* sections are discarded.

To each segmented vocal section we assign the corresponding lyrical strophe automatically, because lyrics syllables are manually anchored to musical notes in the score.

4.2.2 Predominant vocal detection

After section linking a melody extraction algorithm is applied [Salamon & Gómez, 2012]. It extracts the contour of the predominant melodic source and generates time series of pitch values. It performs in the same time a dominant source detection: it returns pitch values of zero for regions with no dominant melody.

Re-synthesis Using a harmonic model [Serra & D, 1989], based on the extracted pitch series, we re-synthesize the spectral components corresponding to the first 30 harmonic partials of the singing voice. This results ideally in a vocal line, with no audible instruments. However, some spectral artifacts from accompanying instruments are inevitable, because they follow in parallel the melodic line of the voice. A side effect of the synthesis is that non-voiced consonants are not re-synthesized, which leaves regions of silence (see Figure 2). To handle these special-case-consonants, they are replaced in the pronunciation dictionary by the HMM for silence. MFCCs are extracted from the vocal part, because the harmonic partials keep the information about articulation.

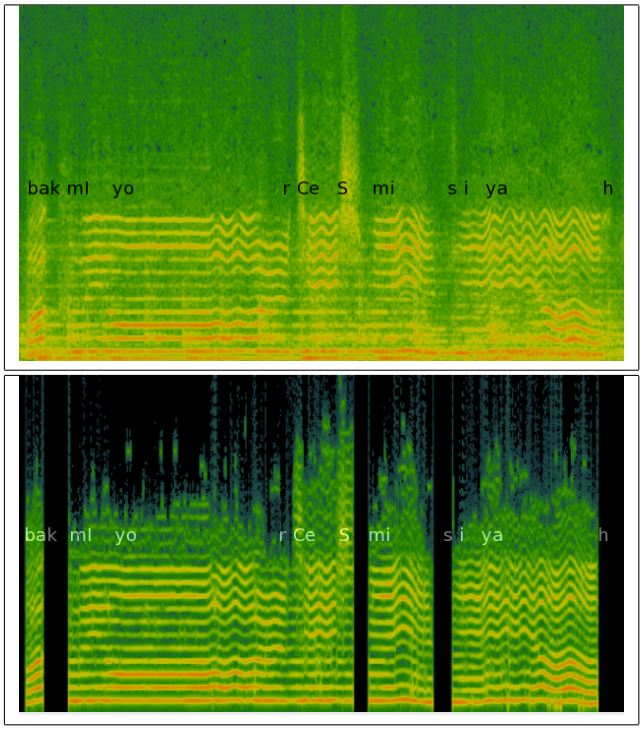


Figure 2: A snippet from the test dataset: spectrum of original audio (above) and after re-synthesis (below)

A challenge for lyrics-to-audio alignment is posed by regions, in which the predominant source is a solo instrument, because its pitch is detected instead of voice and respectively the synthesized audio does not represent voice. In all sections of the *şarkı* form (except *aranağme*), solo instruments can be present at interludes preceding the vocal phrases. To accommodate these instrumental regions, we train a single-state background noise HMM that captures the timbre of background instruments.

4.3 Alignment

The lyrics are expanded to phonemes based on grapheme-to-phoneme rules for Turkish [Özgül Salor et al., 2007, Table 1]. In this way, the HMMs are concatenated into a phoneme network. At the beginning and end of the network for each section, the background noise model (NOISE) is appended. Setting it as optional allows the recognizer to activate it or not, depending on whether sound from background instruments was re-synthesized.

Figure 3 presents an example of such a network.

```
{sil|NOISE} kUSade [sil] taliim {sil|NOISE}
```

Figure 3: Example of recognition grammar. Curly brackets denote one or more occurrence, square brackets denote zero or more occurrence, and vertical bars denote alternatives.

The phoneme network is then aligned to the extracted features by means of the Viterbi forced alignment. The alignment is run for each segmented section separately.

	abs. median	abs. mean	standard deviation
male	0.78	2.18	3.42
female	0.39	1.82	3.37
total	0.52	1.98	3.4

Table 1: Alignment error (in seconds) for the female and male subsets of the test dataset. Evaluation metrics are the median and mean of the absolute value of the misaligned time

5. EXPERIMENTAL SETUP

To train the speech model and adapt it to singing voice, as well as to run the forced alignment, the HMM Toolkit (HTK) [Young & Young, 1993] is employed. The alignment is run on the re-synthesized voice-only counterparts of the original.

5.1 Datasets

The speech dataset, used for training, encompasses clean speech totaling to approximately 500 minutes of speech [Özgül Salor et al., 2007].

The test dataset consists of 10 single-vocal *şarkı* performances (5 distinct male and 5 distinct female). The recordings are selected from a musicBrainz collection of Turkish music [1], whereas scores are provided in the machine-readable *symbTr* format [Karaosmanoğlu, 2012] [2]. The phrase boundaries of each song section were manually annotated, whereby a phrase corresponds roughly to a musical bar and contains 1 or 2 words. Each recording contains on average 15 phrases resulting in total to 150 phrases to be aligned. Phrase-level alignment accuracy of the detection results is reported (in terms of the absolute error in seconds).

5.2 Results

Statistics of the alignment errors are summarized in table 1.

The alignment accuracy for female singing voice is slightly better than for male. A reason for this is that for female voice the extracted melody line has less errors than for male. We observed that, when the extracted pitch is wrong (e.g by an octave), the vowels are not recognized correctly.

Another problem is that alignment performs poorly towards the end of longer sections, which results in outliers of huge magnitude (seen at the distribution of the alignment error figure 4). A glance at the distribution of the alignment error of a lyrics-to-audio system for western popular music [Mesaros & Virtanen, 2008] reveals that the frequency and magnitude of outliers are comparable to ours. Further, our mean error lies not far from theirs of 1.4 seconds.

[1] http://musicbrainz.org/collection/544f7aec-dba6-440c-943f-103cf344efbb

[2] The dataset will be made available on http://compmusic.upf.edu/datasets

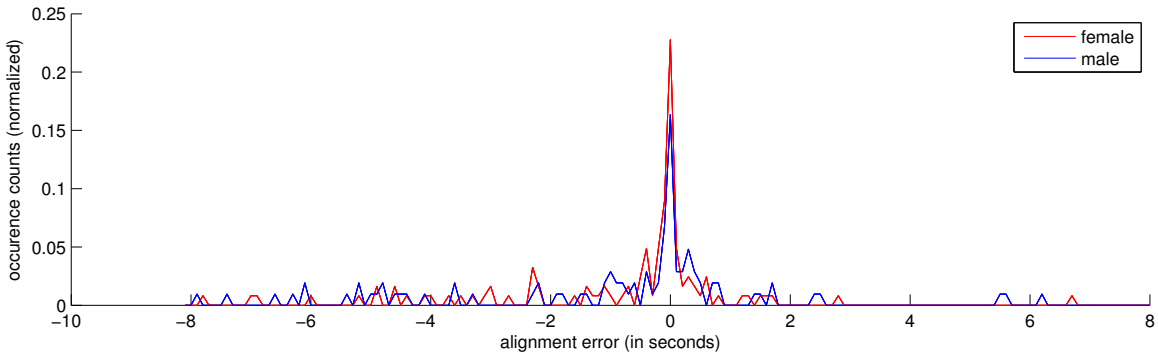


Figure 4: Distribution of the alignment error (in seconds) for male and female. Negative values mean that the beginning timestamp is detected as being earlier than it actually is

6. CONCLUSION

In this work was presented a method for the automatic alignment between lyrics and audio recordings of vocal compositions in the classical Turkish tradition. Performance was evaluated on a dataset, compiled and annotated by us especially for this task. The method showed better results for female singers, which is partly explained by the greater amount of erroneously recognized male pitch, which distorts the re-synthesized pitch as well.

We expect that the generated phrase-to-audio alignment may be a starting point for subsequent musicological analysis tasks.

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

7. REFERENCES

Ederer, E. B. (2011). *The Theory and Praxis of Makam in Classical Turkish Music 1910–2010*. University of California, Santa Barbara.

Fujihara, H., Goto, M., Ogata, J., & Okuno, H. G. (2011). Lyricsynchronizer: Automatic synchronization system between musical audio signals and lyrics. *Selected Topics in Signal Processing, IEEE Journal of*, *5*(6), 1252–1261.

Karaosmanoğlu, M. K. (2012). A turkish makam music symbolic database for music information retrieval: Symbtr. *Proc. Int. Society for Music Information Retrieval (ISMIR)*.

Mesaros, A. & Virtanen, T. (2008). Automatic alignment of music audio and lyrics. In *in Proceedings of the 11th Int. Conference on Digital Audio Effects (DAFx-08*.

Müller, M., Kurth, F., Damm, D., Fremerey, C., & Clausen, M. (2007). Lyrics-based audio retrieval and multimodal navigation in music collections. In *Research and Advanced Technology for Digital Libraries* (pp. 112–123). Springer.

Salamon, J. & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *Audio, Speech, and Language Processing, IEEE Transactions on*, *20*(6), 1759–1770.

Şentürk, S., Holzapfel, A., & Serra, X. (2014). Linking scores and audio recordings in makam music of turkey. *Journal of New Music Research*, *43*, 34–52.

Serra, X. (2011). A multicultural approach in music information research. In *Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, (pp. 151–156)., Miami, Florida (USA).

Serra, X. & D, X. S. P. (1989). A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition. Technical report.

Young, S. J. & Young, S. (1993). *The HTK hidden Markov model toolkit: Design and philosophy*. Citeseer.

Özgül Salor, Pellom, B. L., Ciloglu, T., & Demirekler, M. (2007). Turkish speech corpora and recognition tools developed by porting sonic: Towards multilingual speech recognition. *Computer Speech and Language*, *21*(4), 580 – 593.

AUTOMATED DETECTION OF SINGLE-NOTE ORNAMENTS IN IRISH TRADITIONAL FLUTE PLAYING

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

ABSTRACT

This paper presents an automatic system for the detection of single-note ornaments in Irish traditional flute playing. The presented ornament detection method is based on detecting onsets. We employ several methods for onset detection and explore customisation of their parameters to our task. This includes two methods based on signal amplitude, with processing performed in time domain and in spectral domain, and one method based on the fundamental frequency estimation. The discrimination between notes and single-note ornaments is based on assessing the duration of segments, formed by adjacent detected onsets. Experimental evaluations are performed on audio recordings from Grey Larsen's CD which accompanied his book. Manual annotation of ten recordings was performed. The onset and ornament detection performance is presented in terms of the precision, recall and F-measure. In terms of the F-measure, we achieved onset detection of 91% and ornament detection of 87% and 64% for 'cut' and 'strike', respectively.

1. INTRODUCTION

Ornamentation is used extensively by all melody instruments within Irish traditional music. Ornaments are central to the style of the music, adding to its liveliness and expression. The basic melody is only a point of reference for traditional players who use embellishment, and melodic and rhythmical variations to make each performance unique (Breathnach, 1996). Williams (2010) and Keegan (2010) discuss how each instrument must handle a basic melody differently based on instrumental advantages and limitations. In this work, we focus on the detection of ornamentation played on the wooden concert flute in Irish traditional music.

Single-note ornaments, such as 'cut' and 'strike', are the most common in Irish traditional music. They are pitch articulations of very short duration, created by quickly lifting a finger from a tonehole or closing an open hole (Larsen, 2003).

Methods for ornament detection are typically based on detection of note onsets. A variety of approaches have been proposed for the detection of note onsets in music recordings, e.g., (Scheirer, 1998; Klapuri, 1999; Bello et al., 2005; Collins, 2005; Dixon, 2006; Holzapfel et al., 2010). The proposed methods are usually based on assessing the change in the envelope of the signal amplitude, in fundamental frequency and in phase and their combinations. However, there have been only few studies on detection of ornaments. Automated detection of ornaments is a challenging problem. This is because ornaments are of very short durations, which may cause them being fused with neighbouring notes played. Transcription of baroque ornaments in two piano recordings by analysing rhythmic groupings and expressive timing was studied in Boenn (2007). They employed onset values from time-tagged audio data that was manually edited. An automatic location of ornaments for lute recordings based on MPEG-7 features was investigated in (Casey & Crawford, 2004). Another work in Puiggros et al. (2006) analysed ornamentation from Bassoon recordings. There is only one group which have investigated detection of ornaments in Irish traditional music (Kelleher, 2005; Gainza & Coyle, 2007). They employed spectral domain energy-based onset detectors and provided initial evaluations, which were on a smaller dataset than we used in this paper.

In this paper, we investigate automatic detection of single-note ornaments, namely cut and strike, in flute playing. We employ three different methods for onset detection and explore their customisation and suitability for detection of soft onsets in flute recordings. We perform statistical analysis of the duration of notes and ornaments. This is then employed to decide whether each detected segment, defined based on adjacent detected onsets, corresponds to a note or an ornament. Experimental evaluations are performed on recordings of ten Irish traditional tunes played by flute (Larsen, 2003). Results of onset and ornament detection are presented in terms of the precision, recall and F-measure. Considerably better detection performance is achieved for both onset and ornament detection than that previously published on similar dataset.

2. DETECTION OF ORNAMENTS IN IRISH FLUTE MUSIC

This section first introduces the ornamentation. It then describes the methods we employed for onset detection. It follows with statistical analysis of the duration of notes and ornaments and then presents how this is employed for the detection of single-note ornaments.

2.1 Single-note ornament definition

Ornaments are notes of a very short duration and are used as embellishments in Irish traditional music (Larsen, 2003).

Their places are mostly not marked in the score and the choice of their usage usually depends on the performer's style. Single-note ornaments, namely 'cut' and 'strike', are pitch articulations. They are created through the use of special fingered articulations. The 'cut' involves quickly lifting and replacing a finger from a tonehole, and corresponds to a higher note than the ornamented note. The 'strike' is performed by momentarily closing an open hole, and corresponds to a lower note than the ornamented note.

A schematic representation of single-note ornaments is given in Figure 1.

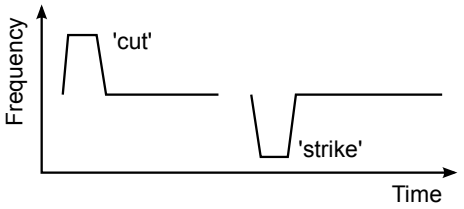


Figure 1: A schematic representation of single-note ornaments.

2.2 Methods for detection of onsets

We employed three onset detection methods, which are briefly described below. Two of the methods exploit the change of the signal amplitude over time, with processing being performed in temporal domain and in spectral domain (Bello et al., 2005; Dixon, 2006). The third method is based on the fundamental frequency estimation (Collins, 2005; Holzapfel et al., 2010). Each method requires several parameters to be set and their values are explored during experimental evaluations. The implementation of the temporal domain amplitude method used in parts some functions from the *MIRtoolbox* (Lartillot & Toiviainen, 2007).

2.2.1 Amplitude change: temporal domain

In the temporal domain amplitude-based method we employed, the signal is passed through a bank of fourteen band pass filters, each tuned to a specific note on the flute in the range from $D4$ to $B5$. The filters have non-overlapping bands, with the lower and the upper frequency being half way between the adjacent note frequencies. These fourteen notes are readily playable on an unkeyed concert flute. The same range of notes, but an octave higher, can be found on small D whistle, which was used in Gainza et al. (2004). The signal in each band is full-wave rectified and then smoothed, resulting in an amplitude envelope. The time derivative of the amplitude envelope is calculated for each band. This is further smoothed by convolving it with a half-Hanning window. The detection function is obtained by summing the smoothed amplitude derivative signals from all bands. The peaks in this detection function, whose values are above a given threshold, are used as the detected onsets. We explored the use of a fixed as well as adaptive threshold. If two consecutive peaks are found within a given time distance, only the first peak is used.

2.2.2 Amplitude change: spectral domain

In the spectral-domain amplitude-based method, also sometimes referred to as spectral-flux method, the signal is first split into frames, with a given overlap between adjacent frames. For each frame, weighted by the Hamming window, the short-time magnitude spectrum is calculated. The difference between the magnitude spectra of successive signal frames is computed for each frequency bin and this is then half-wave rectified. We used the L_2 norm of the resulting spectral vector to provide the value of the detection function at the current frame. The peaks of the detection function, whose amplitude is above a threshold, are used as detected onsets. We explored setting the threshold to a fixed value as well as adaptively changing the value throughout the signal. If two consecutive peaks are found within a given time distance, only the first peak is used.

2.2.3 Fundamental frequency

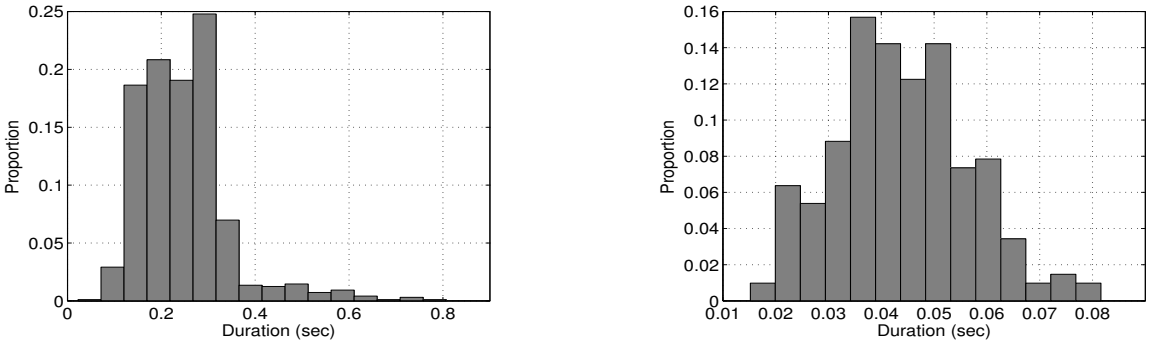
In addition to methods exploiting the signal amplitude, we also explore the use of the fundamental frequency (F_0). This has been reported to be beneficial for soft onset detection in Holzapfel et al. (2010). Among a large variety of existing F_0 estimation algorithms, we employed the YIN algorithm (de Cheveigne & Kawahara, 2002) in this work. The F_0 estimation may result in accidental errors, so called doubling / halving errors, at some frames. To help dealing with these errors, the F_0 estimates are postprocessed using a median filter. The length of this filter needs to be set sensitively in our application – a longer filter may be preferable to deal with the F_0 estimation errors but this may also cause filtering out ornaments, which are characterised by their short duration.

The value of the detection function at the frame time t is based on calculating the difference between the F_0 estimate at the frame $t + \Theta$ and $t - \Theta$, where Θ is a lag parameter to be set. The onset is detected as the first frame for which the absolute value of the detection function is above a given threshold. The value of this threshold, denoted as γ, relates to the minimum frequency difference between adjacent notes.

2.3 Ornament detection

The detected onsets, as obtained using the methods described in Section 2.2, provide a segmentation of the signal, where each segment is formed based on the adjacent detected onsets. The task of the ornament detection system is to determine for each segment whether it corresponds to a note or ornament, and if ornament then determine its type.

We explored whether the discrimination between notes and ornaments can be achieved based on assessing the duration of the detected segments. Using the manual annotation of our recordings, we performed statistical analysis of the duration of notes and ornaments. The obtained distributions of the durations are depicted in Figure 2. These indicate that the duration can provide a good discrimination between notes and ornaments. Based on these results, we consider that a segment is classified as an ornament



(a) (b)

Figure 2: The distribution of the duration of notes (a) and ornaments (b).

when its duration is below 90 ms, otherwise it is classified as note.

Finally, the decision whether a detected ornament is a 'cut' or 'strike' can be made based on comparing the values of the F_0 of the current and the following segment. This reflects the musical knowledge of ornamentation. Each segment is characterised by F_0^{seg}, calculated as the median value of the F_0s corresponding to all signal frames assigned to that segment. If F_0^{seg} of the segment detected as ornament is higher than F_0^{seg} of the following segment, the ornament is classified as 'cut' and as 'strike' otherwise.

3. EXPERIMENTAL RESULTS

3.1 Data description

Evaluations are performed using recordings of ten Irish traditional tunes and training exercises played by flute from Grey Larson's CD which accompanied his book "Essential Guide to Irish Flute and Tin Whistle" (Larsen, 2003). The tunes are between 20 sec and 1 min 11 sec long. All recordings are monophonic and are sampled at 44.1 kHz sampling frequency. We performed manual annotation of the recordings by an experienced listener to indicate the times of onsets and offsets and the identity of notes and ornaments. This is used as the ground truth in evaluations. The total number of onsets, including notes and single- and multi-note ornaments, is 1354. Out of these there are 204 single-note ornaments, consisting of 174 cuts and 30 strikes. A list of the recordings used in this paper, with the number of notes and single-note ornaments in each recording, is given in Table 1.

3.2 Evaluation measures

Performance of the onset and ornament detection is evaluated in terms of the precision (P), recall (R) and F-measure. The definition of these measures is the same as used in MIREX onset detection evaluations, specifically,

$$P = \frac{N_c}{N_c + N_{ins}}, R = \frac{N_c}{N_c + N_{del}}, F = \frac{2PR}{P + R}$$

where N_c is the number of correctly detected onsets / ornaments and N_{ins} and N_{del} is the number of insertions and deletions, respectively. The onset detection is considered as correct when it is within ± 50 ms around the onset annotation.

The single-note ornament is considered to be detected correctly when both onsets, corresponding to the start and to the end of the ornament, are within the ± 50 ms range. Note that when calculating the detection performance for each type of ornament, the deletions also include the cases when the ornament was detected correctly but the type of the ornament was incorrect, i.e., 'cut' was substituted by 'strike' or vice-versa.

3.3 Results of onset detection

Experimental results of onset detection achieved by each of the method are presented in Table 2. Note that these results include onsets corresponding to both notes and ornaments. It can be seen that all methods provide very good onset detection performance. The F-measure values are nearly the same for all the methods. These results were obtained after extensive evaluations with different parameter values – the following values were used for each of the method: i) amplitude-based method (temporal domain processing): half-Hanning window of length 35 ms, the threshold set to 20% of the maximum of the normalised detection function, and the minimum distance between peaks set to 18 ms; ii) amplitude-based method (spectral domain processing): frame length of 1024 samples (23.2 ms), frame shift of 896 samples (20.3 ms), both the fixed (3% of the maximum normalised detection function) and adaptive thresholds performed similarly, the minimum distance between peaks set to 10 ms; iii) fundamental frequency method: frame length of 1024 samples, frame shift of 128 samples (2.9 ms), F_0 median filter of length 9, parameter Θ set to 6 (corresponding to approx. 17 ms) and parameter γ set to 10 Hz. In the following, we use only the F_0-based method for evaluating the detection of ornaments.

An example of a signal extract from one of the tune and the corresponding F_0 estimate and the detection function, with indicated true label and detected onsets, are depicted in Figure 3.

Table 1: List of tunes used in experimental evaluations, indicating the number of notes and 'cut' and 'strike' ornaments and the duration of each tune.

Tune Index	Tune Title	Number of			Time (sec.)
		Notes	Single-note ornaments		
			Cut	Strike	
1	Study 5	55	16	0	25
2	Study 6	56	24	0	25
3	Study 11	76	20	0	30
4	Study 17	48	19	0	20
5	Study 22	127	0	28	51
6	Maids of Ardagh	98	23	0	37
7	Hardiman the Fiddler	112	12	0	32
8	Lady on the Island	118	18	1	26
9	The Lonesome Jig	153	27	0	71
10	The Whinny Hills of Leitrim	117	15	1	35
	Total	960	174	30	352

Table 2: Experimental results of onset detection obtained by each of the employed onset detection method.

Method	Onset detection performance (%)		
	Precision	Recall	F-measure
amplitude (temp)	91.8	90.2	91.0
amplitude (spect)	94.2	87.9	90.9
F_0	88.9	93.2	91.0

3.4 Results of ornament detection

The results of single-note ornament detection are presented separately for 'cut' and 'strike' in Table 3. The achieved detection performance is significantly higher than that presented in previous flute studies using similar data (Kelleher, 2005; Gainza & Coyle, 2007). The performance for 'cut' is only little lower than the overall onset detection performance as presented in Table 2. The performance for 'strike' is considerably lower than that for 'cut'. Such trend has been reported also in previous research. It may be due to the way the 'strike' is realised. We have observed that the errors happened mainly in tunes with indices 5, 6, 7, and 10.

Table 3: Experimental results of single-note ornament detection obtained by employing the F_0-based onset detection method.

	Ornament detection performance (%)		
	Precision	Recall	F-measure
Cut	91.3	90.8	87.0
Strike	62.5	66.7	64.5

4. CONCLUSION AND FUTURE WORK

In this paper, we presented work on detection of single-note ornaments in Irish traditional flute music. The presented method was based on first detecting note onsets and then deciding whether ornament is detected or not based on the duration of segments defined by the adjacent detected onsets. Three methods for onset detection were employed, specifically, two amplitude-based methods and one method based on the fundamental frequency. We explored on customisation of the parameters used within each of these methods for our task of detecting soft onsets, with possibly of short note durations. We have demonstrated that all the three methods performed similarly when suitable parameter setup was used. The onset detection performance of over 91% in terms of the F-measure was achieved. The detection performance for single-note ornaments was over 87% for 'cut' and over 64% for 'strike' in terms of the F-measure.

In our future work, we plan to analyse the errors made by each of the onset detection methods and accordingly explore whether a combination of the methods could lead to detection performance improvements. We also plan to explore on the use of the sinusoidal detection method we presented in (Jančovič & Köküer, 2011) which could be employed standalone or in combination with other methods and could also potentially help in dealing with the doubling / halving errors present in F_0 estimation. We are also working on expanding this work to the detection of multi-note ornaments.

5. REFERENCES

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *IEEE Trans. on Speech and Audio Processing*, 1–13.

Boenn, G. (2007). Automated quantisation and transcription of musical ornaments from audio recordings. In *Proc.*

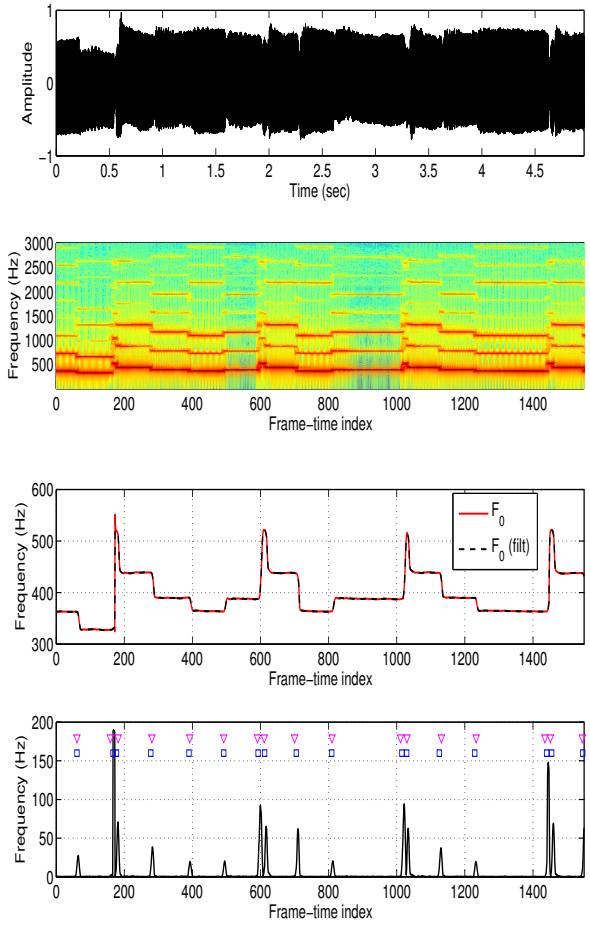


Figure 3: An extract from the tune 'Study 5', depicting (from top to bottom) the waveform, spectrogram, F_0 estimation (unfiltered (dashed black) and filtered (red)) and the detection function with indicated detected onsets (blue $\square$) and true label (magenta ∇).

of the Int. Computer Music Conf. (ICMC), Copenhagen, Denmark, (pp. 236–239).

Breathnach, B. (1996). *Folk music and dances of Ireland*. London: Ossian.

Casey, M. & Crawford, T. (2004). Automatic location and measurement of ornaments in audio recording. In *Proc. of the 5th Int. Conf. on Music Information Retrieval (ISMIR), Barcelona, Spain*, (pp. 311–317).

Collins, N. (2005). Using a pitch detector for onset detection. In *Proc. of the 5th Int. Conf. on Music Information Retrieval (ISMIR), Barcelona, Spain*, (pp. 100–106).

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.

Dixon, S. (2006). Onset detection revisited. In *Proc. of the 9th Int. Conf. on Digital Audio Effects (DAFx), Montreal, Canada*, (pp. 133–137).

Gainza, M. & Coyle, E. (2007). Automating ornamentation transcription. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii.

Gainza, M., Coyle, E., & Lawlor, B. (2004). Single-note ornaments transcription for the Irish tin whistle based on onset detection. *Proc. of the Digital Audio Effects (DAFX), Naples*.

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.

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

Keegan, N. (2010). The parameters of style in Irish traditional music. *Inbhear, Journal of Irish Music and Dance*, *1*(1), 63–96.

Kelleher, A. (2005). *Onset and ornament detection and music transcription for monophonic traditional Irish music*. MPhil thesis, Dublin Institute of Technology, Dublin.

Klapuri, A. (1999). Sound onset detection by applying psychoacoustic knowledge. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 6, (pp. 3089–3092).

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

Lartillot, O. & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects, Bordeaux*.

Puiggros, M., Gómez, E., Ramirez, R., Serra, X., & Bresin, R. (2006). Automatic characterization of ornamentation from bassoon recordings for expressive synthesis. In *9th Int. Conf. on Music Perception and Cognition*, Bologna, Italy.

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, *103*(1), 588–601.

Williams, S. (2010). *Irish Traditional Music*. Abingdon, Oxon: Routledge.

A COMPUTATIONAL RE–EXAMINATION OF BÉLA BARTÓK'S TRANSCRIPTION METHODS AS EXEMPLIFIED BY HIS *SIRATÓ* TRANSCRIPTIONS OF 1937/1938 AND THEIR RELEVANCE FOR CONTEMPORARY METHODS OF COMPUTATIONAL TRANSCRIPTION OF QUR'AN RECITATION

Dániel Péter Biró
School of Music, University of Victoria
birouvic@gmail.com

Peter van Kranenburg
Meertens Institute, Amsterdam
peter.van.kranenburg@meertens.knaw.nl

ABSTRACT

This is a study about furthering transcription methods via computational means. In particular we re-examine Bartók's methods of transcription to see how his project of transcription might be continued incorporating 21st century technology. We then go on to apply our established analytical and computational tools to examples of Qur'an recitation, in order to test hypotheses about connections between the rules of Qur'an recitation (*tajwīd*) and the establishment of salient tones within Qur'an recitation performance.

1. INTRODUCTION

The current study is an outgrowth of a talk given during the panel session on methods of folk music transcription at the third International Workshop on Folk Music Analysis, Amsterdam, Netherlands, June 7, 2013. The panel was moderated by John Ashley Burgoyne (University of Amsterdam) and included Kofi Agawu (Princeton University), Dániel P. Biró (University of Victoria, Canada), Olmo Cornelis (University College Ghent, Belgium), Emilia Gómez (Universitat Pompeu Fabra, Barcelona), and Barbara Titus (Utrecht University).

In transcribing indigenous and world music, ethnomusicologists have to deal not only with subjective hearing, imagination and technologies but also with the history or histories of transcription. The present study re-examines Bartók's methods of transcription by testing it with technology recently developed as part of a research project undertaken by researchers at the University of Victoria, Utrecht University and the Meertens Institute.[1]

In comparing Bartók's transcriptions of *siratók* (Hungarian improvised ritual laments) with those done with the help of computer technology, we are able to reassess Bartók's production process and methodology, as well as to see if and how Bartók's projects of transcription can be applied, reevaluated and continued. Taking this analysis further, we apply the resulting computational analysis procedures to examples of Qur'an recitation.

The *sirató* is a lament ritual from Hungary that goes back at least to the Middle Ages. This improvised song type is integral for our study, as it exemplifies inherent relationships between speech and singing while demonstrating stable melodic formulae within an oral/aural ritual context. While the performance practice of *siratók* (plural of *sirató*) had been determined by traditions of textual and melodic improvisation, the performance framework for Qur'an recitation are determined by rules of recitation that are primarily handed down orally.

2. METHODOLOGIES OF PITCH ALALYSIS

In this study we have taken Bartók's recordings of *siratók,* applying computational audio analysis of these recordings to find the most prevalent pitches. We have set out to find these pitches in order to reinterpret Bartók's transcription using a scale derived by automatically detecting the peaks in a density estimation of the distribution of pitches. These density-estimation based scales can be analyzed in terms of their functionality in forming melodic contour and pitch identity. Such analysis helps to demonstrate salient structural features of oral transmission.

In early 2014 we applied the same analysis to examples of Tunisian Qur'an recitation. In so doing, we have been able to investigate the practice of *maqamat* (a set of pitches for melodic performance found in Qur'an recitation) in terms of quasi–Schenkerian analysis showing the most prevalent pitches within a given performance in terms of foreground, middle-ground and background frequency analysis. These pitches relate roughly to *maqamat* traditions of instrumental music, although the melodic

[1]This research project, which involved scholars of musicology and computer science, similarity measures for melodies from oral tradition have been developed, especially designed for the monophonic chant repertoires in coordination with the Department of Computer Science, University of Victoria, School of Music, University of Victoria, Department for Information and Computer Science, University of Utrecht and the Meertens Institute, Amsterdam Netherlands. This research has resulted in a series joint journal papers including Steven R. Ness, Dániel P. Biró, and George Tzanetakis: "Computer-Assisted Cantillation and Chant Research Using Content-Aware Web Visualization Tools," in *Multimedia Tools and Applications* (2009), Van Kranenburg, P, D.P. Biró, S.R. Ness, and G. Tzanetakis (2011), "A Computational Investigation of Melodic Contour Stability in Jewish Torah Trope Performance Traditions". In: *Proceedings of the International Society on Music Information Retrieval (ISMIR2011) Conference*, pp. 163-168., and D.P. Biró, P. van Kranenburg, S.R. Ness, G. Tzanetakis, and A. Volk (2012) "Stability and Variation in Cadence Formulas in Oral and Semi-Oral Chant Traditions – a Computational Approach." *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*. Thessaloniki. 2012. pp. 98-105.

entities within Qur'an recitation do not adhere directly to *maqam* traditions within instrumental music.[2]

We set to compare such analysis to the cultural perception within traditional *maqamat* practice. Our method sets out to delineate the main pitches of a given scale or *maqam* from secondary "ornamental" pitches. By doing so, we present a hierarchy of scale degrees, thereby showing how surrounding "ornamental" pitches structurally interact with the main "skeletal" notes of the scale. In addition, we investigate how such scale structures function in the context of the rules of Qur'an recitation (*tajwīd* and *tartīl*). In this manner, we are able to show how the parameters of textual recitation, pronunciation and interpretation interact with melodic contour and scale structures within these two traditions.

With these case studies we show how to establish a direct interaction between automatically derived scales and traditional practices of transcription, therewith enriching the arsenal of methods ethnomusicologists have at their disposal.

3. DATA

We have employed recordings of *siratók* found in the sound recording *Hungarian Folk Music: Gramophone Records with Bartók's Transcriptions,* edited by László Somfai. *Magyar népzenei hanglemezek Bartók Béla lejegyzéseivel,* szerkesztette, Somfai László, (Budapest: Hungaroton, 1981).

In our study we employed recordings of Qur'an recitation from *Sura Al-Fatiha, Sura Ash Shams, Sura Al-Qaria* and *Sura Ghafir*. These have been extracted from Michael Sells' *Approaching the Qur'an: The Early Revelations* (Ashland: White Cloud Press, 2007) and field recordings conducted by the authors in Rotterdam (Netherlands).

Each recording has been segmented in terms of syntactical units (phrases) and, in the recordings of Qur'an recitation, analysis has also been based on audio segments corresponding to the individual words of the given *sura*.

Each recording has been converted to a sequence of frequency values using the YIN pitch extraction algorithm (De Cheveigne & Kawahara, 2002) by estimating the fundamental frequency in a series of overlapping time-windows of 40ms, with a hopsize of 10ms. The frequency sequences have been converted to sequences of real-valued MIDI pitches with a precision of approximately 1 cent (which is 1/100 of an equally tempered semitone, corresponding to a frequency difference of about 0.06%). A MIDI-value of 60.0 corresponds with the c', 61 with equal tempered c#', 62, with d', and so on. A value of e.g., 60.23 would correspond to a pitch that is 23 cents higher than c'.

We employ two post-processing filters to correct for possible errors of the pitch extractor. First, all samples with a signal energy of more than 40 dB below the maximum value are considered silence. Second, all pitches that are further than an octave away from the average pitch of the entire curve are considered silence as well.

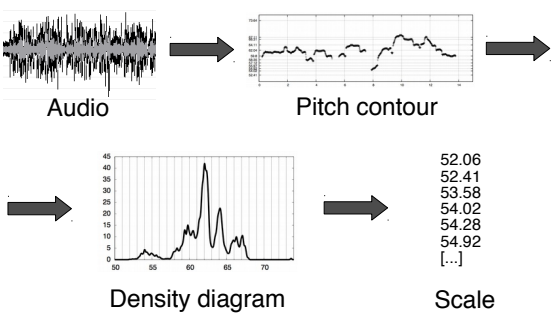


Figure 1. Schematic overview of the computational analysis method.

In the resulting pitch contour, we perform a non-parametric density estimation using a Gaussian kernel with σ=5 cents. The Gaussian kernel has a smoothing effect on the resulting density curve. By performing peak detection in the density curve, we obtain the pitches that recur most often during the recording. The height of a peak indicates the frequency of occurrence of the corresponding pitch. The set of pitches that correspond to the peaks in the density estimation can be considered the scale the singer, or reader, is adhering to. In the peak detection, we set the constraint that the peaks should be at least 10 cents apart to be considered separate scale tones. This leads to a fine-grained scale that reduces the pitch content of a recording to a relatively small set of pitches, while retaining enough detail to perform precise analyses. By sorting the scale tones according to their density values (the heights of the peaks), we get an ordering of pitches according to their prevalence in the recording.

By replacing each detected pitch with the pitch of its nearest scale tone, we obtain a reduced contour. It is in these contours that we perform an analysis of occurrences of pitches. For the Qur'an readings, we determine for each verse the durations of all pitches that occur in the verse as percentage of the duration of the entire verse. Thus, we get an indication of the importance of each pitch within the verse. We do the same for the pitches in each word.

[2] Lois Ibsen al Faruqi: "The Cantillation of the Qur'an" *Asian Music*, Vol. 19, No. 1 (Autumn - Winter, 1987), p. 9, "Although Qur'anic recitation does not adhere strictly to the modal (*maqām*) practice of the secular music, and although members of the culture maintain a notion of rigid boundaries separating Qur'ānic chant from all the other sound arts, Qur'ānic recitation does conform to many of the theoretical aspects of Arabian music. It employs many of the interval combinations (trichordal, tetrachordal or, pentachordal) that identify *ajnās* (s. jins) on which the secular music is based - e.g., bayyātī, ḥijāz, kurd, rāst, nahāwand, ṣabā, sīkāh, etc. It similarly evidences a predominance of serial treatment of individual ajnās rather than utilization of the whole modal scale in a single phrase. Cantillation of the scripture is punctuated, like secular improvisations, by returns to the tone (qarār) of resolution in the jins. It is marked by transpositions and modulations internal or external to the phrase, which are also characteristic of the secular genres. Some reciters are knowledgeable about maqām practice and the theory of music. Others have no formal exposure to music theory and only conform to its rules in such measure as their ears and listening experience have trained them."

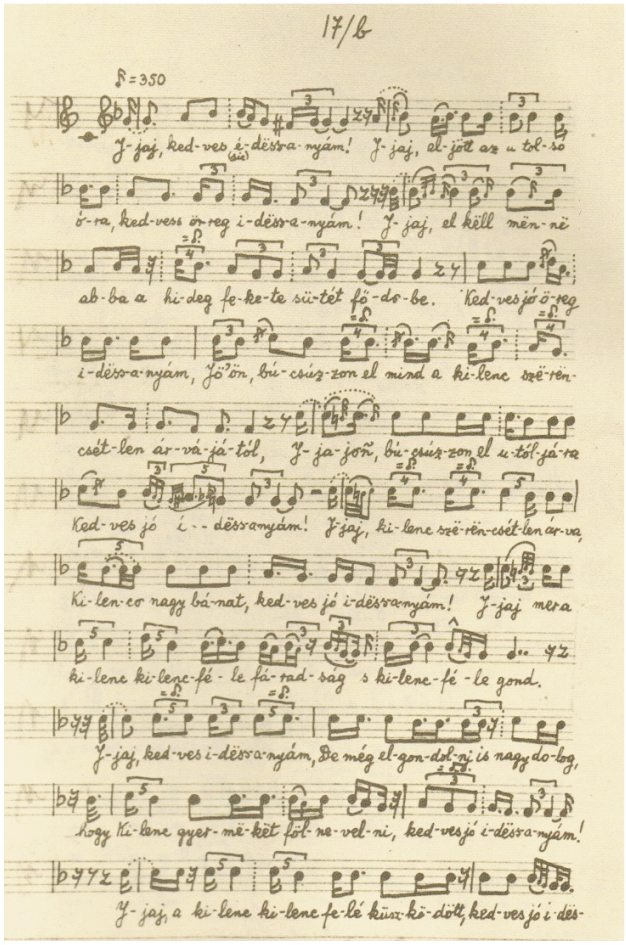


Figure 2. Béla Bartók's sirató Transcription. Recording of Mrs. János Péntek completed in Körösfő on December 14, 1937 and transcribed by Bartók in 1937–1938.

4. SIRATÓK

4.1 Cultural context of the sirató

The *sirató's* performance connects the life of the singer not only with her ancestral past but also with the larger community. It is an integral part of the performer's life: the enactment of the sirató most often has no clear beginning or ending. Although the *sirató* is improvised, each time exhibiting a personal melodic expression, the song-type is clearly discernible and exhibits a remarkable consistency of textual and musical form. Elements of both formal semblance and improvisational variability can be observed among the various examples of *sirató*, proving its mythical nature. The song is not determined individually but collectively, as the boundaries of its enactment are explicit enough to be reconstructed by the individual and recognized by the village collective.[3]

The *sirató* is most often improvised in a recitative manner. In its unfolding, melody and text function symbiotically, as gestures of improvised speaking are applied to the melodic and rhythmic domain.[4] Kodály described such a dichotomy between singing and speaking: "This is the only type of musical prose of this kind and can only be done spontaneously […]. Musical prose, on the border of music and speaking […]. The rhythm therein is no different from the rhythm of spoken speech […] the sections between the rests are not the same."[5] The *musical vocabulary* of the *sirató* is comprised of the same cadential formulas and modality found in other types of Hungarian folksong and chant.

The recording of Mrs. János Péntek was completed in Körösfő on December 14, 1937 and transcribed by Bartók in 1937–1938.[6] The *sirató* is a lament sung by women after a loved-one passes away. Sometimes a lament is sung by a so-called "professional" and it seems that here Ms. János Péntek is indeed a "professional" who would do this type of lament for a relative or someone else in her village if necessary. Here the lament is done for a deceased mother using the language common in *siratók*, which can be found in renditions done by a variety of performers across large territories.[7]

[3] Aleida and Jan Assmann, "Schrift, Tradition und Kultur," Zwischen Festtag und Alltag, Zehn Beiträge zum Thema 'Mündlichkeit und Schriftlichkeit, (Tübingen: Günter Narr Verlag, 1988), "In schriftloser Kommunikation vermag das, was beim Publikum nicht auf unmittelbare Akzeptanz stößt, schon den Augenblick der Darbietung nicht zu überdauern. Weltbildkonformität ist hier schon in dem Merkmal 'haltbarer' Geformtheit eingebaut. Sie reguliert durch "Präventivzensur," die gar nicht erst Form gewinnen läßt, was seines Ortes im kulturellen Gedächtnis sicher sein kann und daher riskiert, als a-topos ('albern,' eigentlich: ortlos) eingestuft zu werden." "In non-textual communication, what is not met with immediate acceptance by the general public, can not survive in the moment of its rendition. Conformity of world view is built in to the atribute of the preserved formal unit. This is regulated through 'preventive censorship.' Such censorship does not even start to allow for forms to take hold, which, risking to be labled a-topos (absurd; actually without place), do not have a secure function within the context of cultural memory." (English translation by Biró)

[4] Although the *sirató* is most often sung by a female relative of the deceased, a "professional," or a designated singer from the local vicinity, will often sing the sirató; this type of sirató is termed *parodia* or "parody." Even in this form the song type remains intact as the "professional" takes the place of the mourner and "improvises" the typical expressions of mourning like "My dearest mother, why have you left me?" "What will I do without you?" "You were so good to us" etc..

[5] Zoltán Kodály, *A Magyar népzene* (Budapest: EMB, 1952) 38-39 "Egyetlen példája a prózai recitáló énekeknek és szinte egyedüli tere a rögtönzésnek [. . .] zenei próza, a zene és beszéd határán [. . .] Ritmus nincs benne más mint a beszéd ritmusa [. . .] a nyugvópontok közti részek nem egyenlõk."

[6] Hungaroton

[7] The text reads as follows: Jaj, kedves idëssanyám!
Jaj, eljött az utolsó óra, kedves öreg idëssanyám!
Jaj, el këll mënë abba a hideg fekete sütét födbe.
Kedves jó öreg idëssanyám,
Jö'ön, búcsúzzon el mind a kilenc szërëncsétlen árvájától,
Jajon, búcsúzzon el utóljára
Kedves jó idëssanyám!
Jaj, kilenc szërëncsétlen árva,
Kilenc nagy bánat, kedves jó idëssanyám!
Jaj mer a kilenc kilencféle fáradság s kilencféle gond.
Jaj, kedves idëssanyám, De még elgondolni is nagy dolog, hogy
Kilenc gyermëkët fölnevelni, kedves jó idëssanyám!
Jaj, a kilenc kilenc felé küszködött, kedves jó idëssanyám!
Alas, my dear sweet mother!
Alas, the last hour has come, my dear, old sweet mother!
Alas, one has to descend into that cold, black, dark earth.
My dear, old sweet mother,
Come and say goodbye to all your nine orphans,
Alas, say goodbye for the last time,
My dear, good sweet mother!
Alas, nine wretched orphans,
Nine great sorrows, my dear, good sweet mother!
Alas, because nine is nine kinds of weariness and nine kinds of worry.

4.2 Computational transcription of the sirató: a re-evaluation of Bartók's transcriptions

One can imagine how Bartók the composer might have been intrigued by this text and by the performance, and the resulting, for him both archaic and modernist musical structure. Bartók often carried the very heavy cylinder recorder to record his subjects, as this was the most sophisticated way to transcribe folk music in his day. In transcribing the recordings to paper he would often slow down the recordings, allowing him to achieve detailed transcription of ornaments.

Bartók always transcribed the recordings to have g' be the *tonus finalis*. This was done in order to better compare the tonal language of large quantities of transcriptions. So in this way a tonal "reduction" or transposition served to allow for easier analysis within and across folk music traditions. This is also the case in his transcription of Mrs. János Péntek shown in **Figure 2.**

A product of his education and European musical culture, Bartók employs the five-line stave in his transcriptions. He was very aware of tuning and the differences in tuning within folk music. **Figure 3** shows the sequence of pitches as estimated by the YIN-algorithm. **Figure 4** shows the estimated pitch density. The peaks demonstrate the most prevalent pitches in the scale.

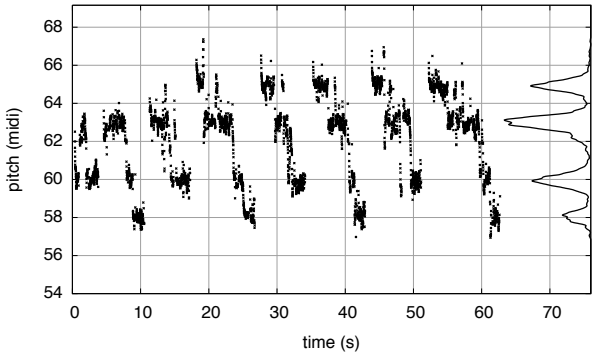


Figure 3. Pitch contours over time in the recording of the *sirató*, with the density curve shown at the right.

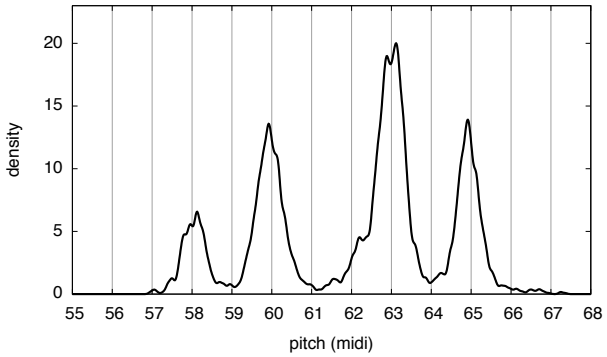


Figure 4. Density plot of frequencies occurring in recording of *sirató.*

Alas, my dear sweet mother, It is a great deed - to think that you raised your nine children, my dear, good sweet mother!
Alas, the nine, you suffered nine times, my dear good sweet mother!

We set out to find these pitches in order to reinterpret Bartók's transcription using a scale derived by a density estimation of the pitch-content of the recording. These density-estimation based scales can be compared in terms of their melodic contour and pitch identity and such comparison helps to demonstrate salient structural features of oral transmission.

Figure 5 shows the pitches corresponding to the peaks in the density curve as they are ordered in terms of their density-value.

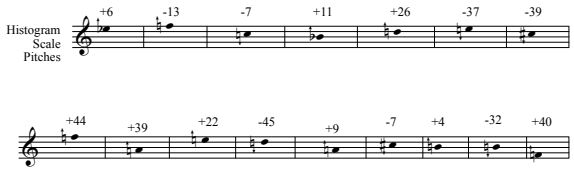


Figure 5. Pitches occurring in recording of *sirató* in order of density.

Figure 6 shows their ordering in terms of scale tone.

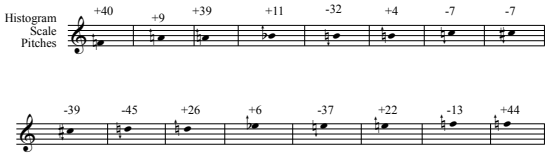


Figure 6. Pitches occurring in recording of *sirató* in order of scale-tone.

The density-estimation based scale presents a series of pitches determined by the frequency of use in a particular recording. **Figure 7** presents Bartók's transcription on top and the more-or-less same transcription on the bottom now with cent deviations according to the density-estimation based scale analysis. Here we can see how Bartók *perceived* certain microtonal deviations and *integrated* them into the conventional tonal framework he had knowledge of.

In employing density-estimation based scales it is possible to examine the levels of pitch hierarchy in the scales. **Figure 8** shows the most prevalent pitches and where they occur in the transcription. In this way we are able to see a kind of foreground, middle–ground and background of pitch hierarchies. We are also better able to appreciate the diversity of scale species present in these recordings.

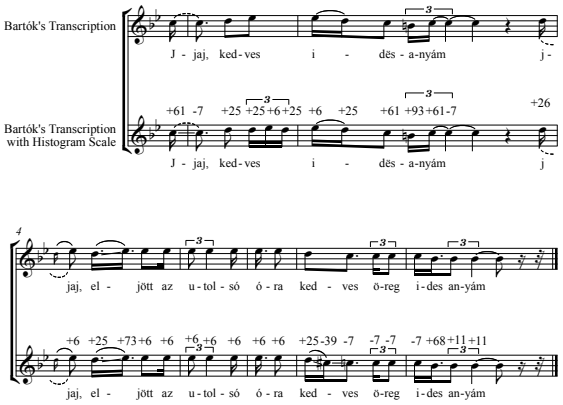


Figure 7. Comparison of Bartók's transcription with a transcription according to the density-estimation based scale: cent differences are indicated

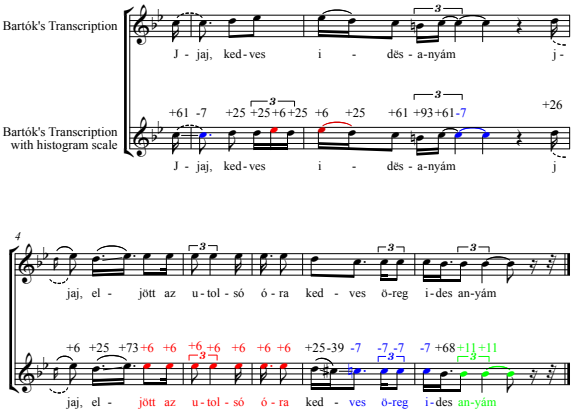


Figure 8. Bartók's transcription juxtaposed by transcription done with density-estimation based scale. Colors indicate primary, secondary and tertiary degrees of occurrence in recording: Primary pitches in red, secondary pitches in blue, and tertiary pitches in green.

5. QUR'AN RECITATION

5.1 Computational analysis of Qur'an recitation

We have applied this method of analysis of scale-tones to a repertoire of Qur'an recitation. We show how the durations of tones, determined by the rules of *tajwīd,* affects the density of structurally salient pitches within a given recording of recitation.

Figure 9. Qur'an text of *Sura al-Fatiha* with color indications for rules of *tajwīd.*

5.2 Cultural context of Qur'an recitation

The performance framework for Qur'an recitation is not determined by text or by notation but by rules of recitation that are primarily handed down orally (Zimmermann 2000, p. 128).[8] Here the hierarchy of spoken syntax, expression and pronunciation play a major role in determining the rules of *Tajwīd.*[9] The resulting melodic phrases, performed not as "song" but "recitation" are determined by both the religious and larger musical cultural contexts. In the context of "correct" recitation contexts, improvisation and repetition exist in conjunction (as is the case with the Hungarian *sirató).* In this way the traditions of *siratók* and Qur'an recitation are examples of "oral literature,"[10] as their given modes of production are collective-

[8] "Like the Hebrew *miqra'* the primary name 'Qur'an' derives from the root q-r, i.e., 'reading': the visual implication of text is not implied with this root. Rather the concepts 'pronounce, calling, reciting' are expressed with the word, so that an adequate translation of Qur'an (Qur'ān) could be 'the recited'" (Translation from the German by Dániel Péter Biró). Heidi Zimmermann, *Tora und Shira: Untersuchungen zur Musikauffassung des rabbinischen Judentums* (Bern: Peter Lang, 2000), 27.

[9] "*Tajwīd* [is] the system of rules regulating the correct oral rendition of the Qur'an. The importance of *Tajwīd* to any study of the Qur'an cannot be overestimated: *Tajwīd,* preserves the nature of a revelation whose meaning is expressed as much as by its sound as by it content and expression, and guards it from distortion by a comprehensive set of regulations which govern many of the parameters of the sound production, such as duration of syllable, vocal timbre and pronunciation." Kristina Nelson, *The Art of Reciting the Qur'an* (Austin: University of Texas Press, 1985), 21.

[10] Al Faruqi p. 21, "The term "literature" is derived from *littera*, a Latin word meaning "letter" or, in plural form, "writing." However, to limit literature to only that imaginative and artistic organization of words which is written would be to succumb to a type of cultural chauvinism dictated by a Western European emphasis on the written word. The

ly determined by set of rules which are transmitted orally from generation to generation.

5.3 Analysis Criteria for Qur'an recitation

Within a given recording we investigated the relative durations of scale-pitches within verses and within words. In particular, we looked at the sections of elongation within a given *sura* and tried to see how these sections of elongation affected the central tones within the *sura* in terms of a) scale degree and b) degree of density. We looked at the results in several recordings of a given *sura* and compared pattern relationships between scale degree and scale of density in these examples.

In our study we found relationships between scale degree and density of occurrence in final words of verses and especially in sections where syllables are elongated.[11] Locating these tones we are able to determine "stable" and "variable" structural tones within the recording. While certain scale tones show a clear stability in some examples, other tones display an amount of identity variability, and entail an "ornamental" functionality, as in the higher pitches in **Figure 10**. This may relate to the culture of *maqāmat* from instrumental music, which affects the tonal structure of Qur'an recitation.

5.4 Qur'an recitation: outcomes of analysis

In our examinations of recorded renditions of *sura al-Fatiha* we observe relationships between structural scale tones within verses (*ayat*) and scales tones used in sections of syllable elongation (*madd*). The rules of *tajwīd* specify that specific sections of words require either variable extensions of vowels from two to six beats (*ḥarakāt*) or an obligatory six beats (*ḥarakāt*).[12] While such elongations have been studied in terms of correct pronunciation based on the rules of *tajwīd,* there has not been considerable study of how such elongations might contribute to the establishment of structurally salient tones, as employed in a given recitation, and how these relate to *maqāmat*. We base our analysis on the frequency of occurrence of scale pitches in each verse (*ayah*) of the *sura* and in the final words of each verse, as displayed in **Figure 12.**

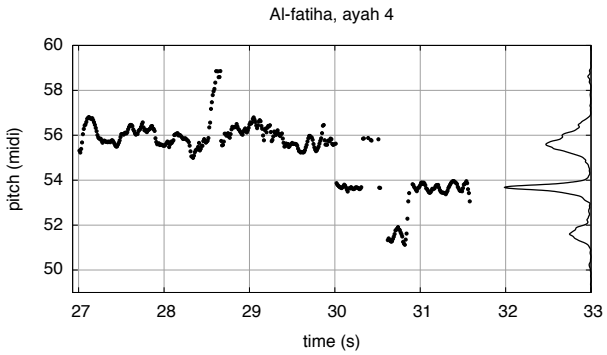


Figure 10. Pitch contour for fourth verse (*ayah*) of *sura al-Fatiha* in recording of al-Minshawi.

In our computational analysis we have looked for salient pitches within a recording, within each verse and within words with syllable elongations (*madd*). Comparing the pitches employed in sections of textual elongation with those employed throughout a given recitation, we have found that the rules of *tajwīd* display a profound influence on creating and stabilizing salient pitches. In addition, through comparative analysis of the same *sura* recited by the same person,[13] we are able to show scale relationships and recurring patterns of final selections of tones within these sections of elongation.

In our study, we have set out to test the performance of syllable elongation in Qur'an recitation via computational means. *Sura al-Fatiha* contains syllables to be specifically performed with types of elongations (*madd*)*;* either with a variable duration of two to six beats or with a obligatory six beats (*ḥarakāt*). These are displayed in **Figure 11.**

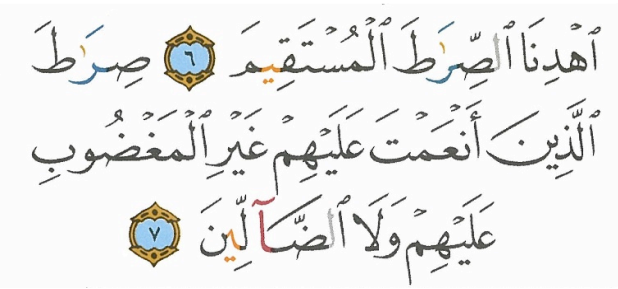


Figure 11. Last two *ayaat* (verses) of *Sura al-Fatiha* with indications of elongation (*madd*) in color; orange: variable elongation of 2-6 beats (*ḥarakāt*); red: obligatory elongation of 6 beats (*ḥarakāt*).

"literature" of many peoples in the world - and even of certain genres in Western culture - is not preserved in written form. Such genres are usually designated as "oral literature."

[11] Al Faruqi, p. 10, "Durations of tones and rhythmic motifs are strongly affected by the rules of pronunciation set down in the manuals on *tajwīd.* Those rules prescribe determined durational relationships between the short vowels or *ḥarakāt* (the *fatḥah*, *ḍammah*, and *kasrah*) and the long vowels (i.e., the letters *alif*, *waw* and *yā*). *Tajwīd* also determines the extension or *madd* of the long vowels according to their place in the word, their combination with certain other letters of the Arabic alphabet, and their use with unvowelled consonants (i.e., with *sukūn*) or doubled consonants (*tashdīd*). These rules insure that the difference between the short and long syllables does not exceed the ratio of 1 to 6 (e.g., the difference between a 16th note and a dotted quarter). Often the actual differences are much less. Many of the prohibited practices regarding Qur'ānic chant, which have been repeated and reemphasized in each successive century of Islamic history, have been restrictions against vocal practices that exaggerated durational contrast. Among these condemned practices are the exaggerated lengthening of short vowels (*tshbā' al ḥarakāt*), the lengthening of the long vowels (*ziyādah al, madd*), omission of short vowels (*taqtī' al ḥurūf*), and the improper addition of short vowels (*taḥrīk al ḥarf al sākin*) (al Sa'īd 1967:347)."

[12] Nelson, 24. "Arabic prosody classifies the syllable into long and short durations, one long being approximately equivalent to two short. The durations of syllables in Qur'anic range from one to six beats (*ḥarakāt*) or longer."

[13] Mesrur Coşkun performed multiple recitations of specific *surat* in Rotterdam in May 2014.

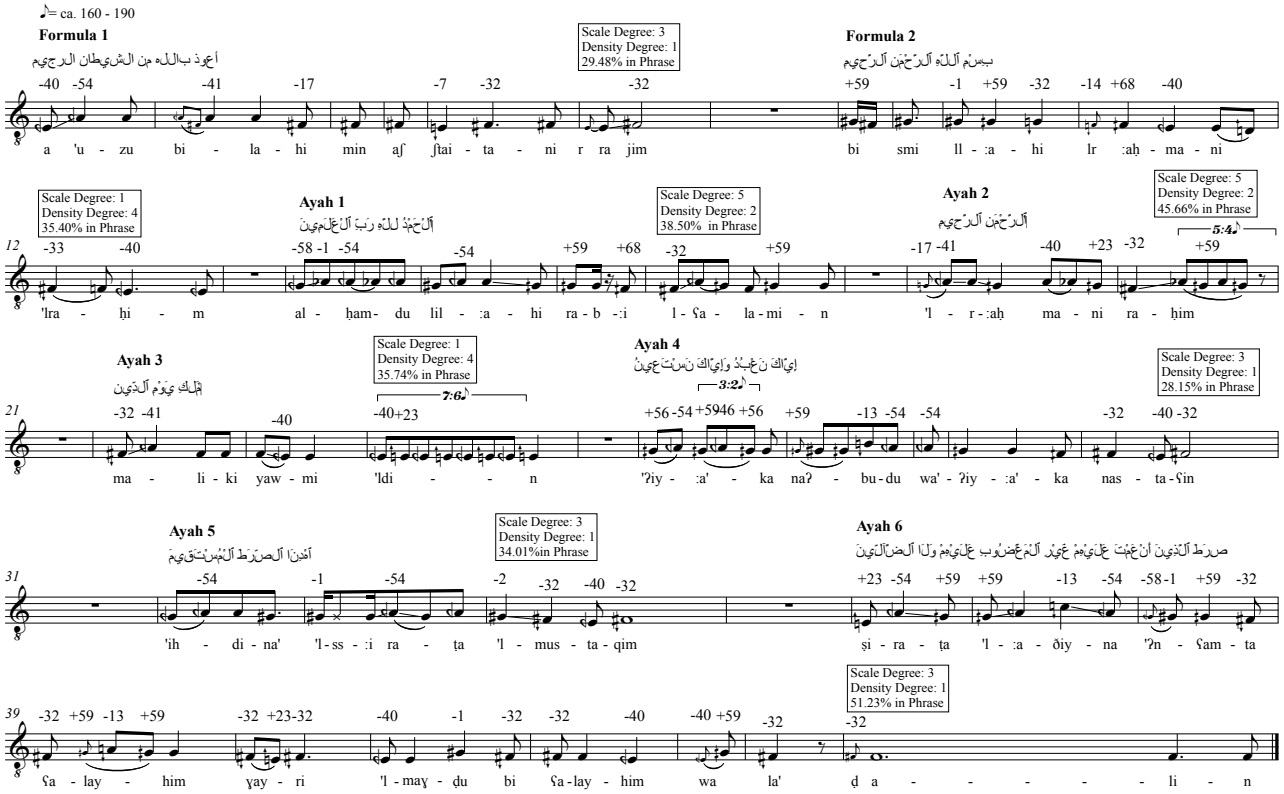


Figure 12. Transcription of *Sura Al-Fatiha* as performed by Siddiq al-Minshawi. Computational analysis shows scale degrees and density degrees of each word with elongation (*madd*) as well as overall percentage of pitch within a given verse.

The salient tones used for elongation of syllables within given words (*madd*) correspond to salient frequencies used within the given recitation. Comparing recorded examples of the *sura al-Fatiha,* we found paradigmatic connections between salient tones within given words and verses as displayed in **Figure 12**. While some tones used in association with variable elongation of durations (two to six beats) show more variability in terms of their relationship to the salient pitches within corresponding verses, the pitch associated with obligatory elongation of durations of six beats (*ḥarakāt*) show an exact correlation with the salient pitches of corresponding verses, as displayed in **Figures 13 and 14.**

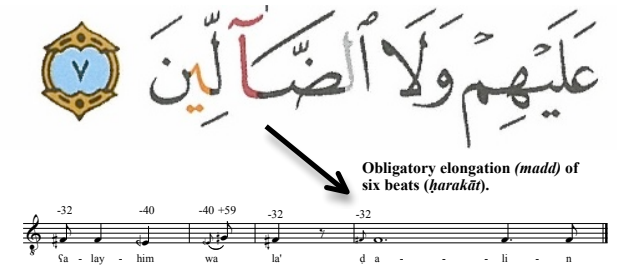


Figure 13. Last words of *Sura al-Fatiha* with indications of elongation (*madd*) in color; orange: variable elongation of 2-6 beats (*ḥarakāt*); red: obligatory elongation of 6 beats (*ḥarakāt*); comparison to transcription of performance by Muhammad Siddiq al-Minshawi.

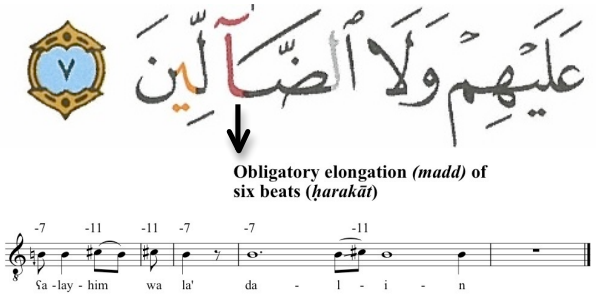


Figure 14. Last words of *Sura al-Fatiha* with indications of elongation (*madd*) in color; orange: variable elongation of 2-6 beats (*ḥarakāt*); red: obligatory elongation of 6 beats (*ḥarakāt*); comparison to transcription of performance by Muhammad Khalil al-Husari.

Figure 15 displays how, in five recorded recitations of *sura al-Fatiha*, central tones relate in terms of a) their use in elongation of syllables (*madd*), and b) their use in verses. In four of the five examples, the variable elongation of the fourth verse (*ayah*) displays variability to the predominant tone used in the verse. In all five examples, the obligatory elongation helps to define the predominant tones of the verses as well as the tonic of the *maqām* or set of tones employed for recitation.

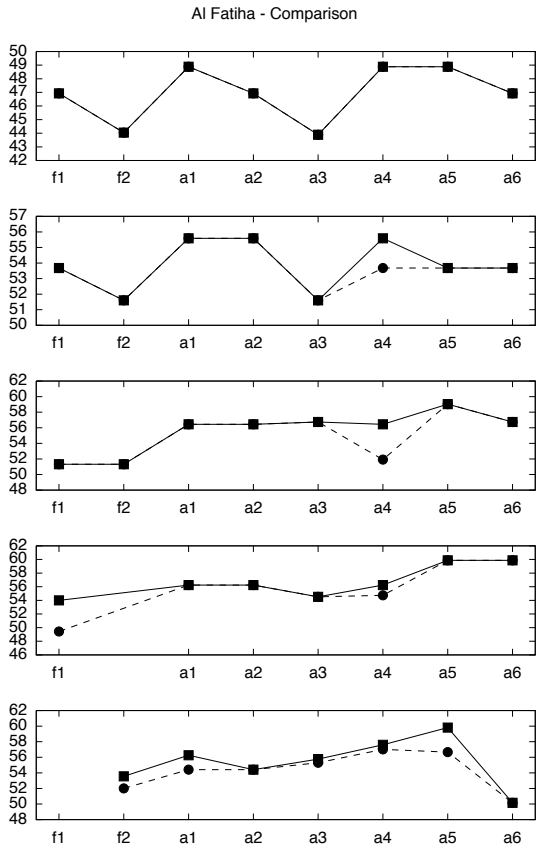


Figure 15. Each graph show the sequence of MIDI-values of the most frequently occurring pitches in the entire verses (squared), and in the last words of the verses (dotted) in various readings of *Sura Al Fatiha*. The readers are from top to bottom: Muhammad Khalil Al-Husari, Muhammad Sidduq Al-Minshawi, Mesrur Coşkun (first rendition), Mesrur Coşkun (second rendition), and Mesrur Coşkun (third rendition).

6. CONCLUSIONS

In our analysis of *siratók* and Qur'an recitation we have set out to discover how salient pitch structures are determined and how these relate to performance parameters of these traditions. The computational analysis of *siratók* transcribed by Béla Bartók in the 1930s, allows us to re-evaluate his work and to better comprehend how pitch hierarchies function within this type of chant ritual. Combining computational analysis of resultant pitch structures determined by specific rules of *tajwīd*, such as the use of elongation (*madd*) with analysis of tonal structures (*maqāmat)*, we better understand how the rules for correct enunciation help to form and interact with given tonal hierarchies found within performances of Qur'an recitation, as the pitches of the elongated syllables help to create the salient pitches within a given recitation.

7. FUTURE WORK

We are currently examining recordings of *sura al qadr, sura al-qaria, sura ghafir* and *sura ash-shams.* In addition we are examining relationships of salient frequencies and contour in examples of the same sura texts, as exemplified in *mujawwad* and *murattal* versions. In comparing these versions we hope to find further correlations between the rules of *tajwīd* and use of *maqamat* within Qur'an recitation.

Acknowledgements. We are grateful to Imam Mesrur Coşkun for allowing us to record his recitations of Qur'an. Peter van Kranenburg is supported by the Computational Humanities Programme of the Royal Netherlands Academy of Arts and Sciences, under the auspices of the Tunes & Tales project.

8. REFERENCES

al Faruqi, Lois Ibsen. (1987). "The Cantillation of the Qur'an" *Asian Music*, 19 (1), 1-25.

D.P. Biró, P. van Kranenburg, S.R. Ness, G. Tzanetakis, and A. Volk (2012) "Stability and Variation in Cadence Formulas in Oral and Semi-Oral Chant Traditions – a Computational Approach." *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*. Thessaloniki. pp. 98-105.

De Cheveigne, A. & H. Kawahara. (2002). YIN, a Fundamental Frequency Estimator for Speech and Music. *Journal of the Acoustic*
Society of America, 111 (4), 1917-1930.

Kodály, Zoltán. (1960). *Folk Music of Hungary*. Budapest: Corvina Press.

Krumhansl, Carol L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Levy, Kenneth. (1998). *Gregorian Chant and the Carolingians,* Princeton: Princeton University Press.

Nelson, Kristina. (1985). *The Art of Reciting the Qur'an,* Austin, University of Texas Press.

Ness, Steven R., Dániel P. Biró, and George Tzanetakis (2010). "Computer-Assisted Cantillation and Chant Research Using Content-Aware Web Visualization Tools," in *Multimedia Tools and Applications* 48 (1) 207-224.

Neubaurer, Eckhard and Veronica Doubleday. 'Qur'anic Recitation,' *Grove Music Online* ed. L. Macy (Accessed 6 May 2014), <http://www.grovemusic.com>

Sells, Michael (2007). *Approaching the Qur'an. The Early Revelations*, Ashland: White Cloud Press.

Somfai, László. (1981). Hungarian Folk Music: Gramophone Records With Bartók's Transcriptions, edited by László Somfai. Magyar népzenei hanglemezek Bartók Béla lejegyzéseivel, szerkesztette, Somfai László, Budapest: Hungaroton.

Treitler, Leo. (1982). "The Early History of Music Writing in the West, *Journal of the American Musicological Society,* 35 (2), 237-280.

Van Kranenburg, P, D.P. Biró, S.R. Ness, and G. Tzanetakis (2011), "A Computational Investigation of Melodic Contour Stability in Jewish Torah Trope Performance Traditions". In: *Proceedings of the International Society on Music Information Retrieval (ISMIR2011) Conference*, pp. 163-168.

A COMPREHENSIVE COMPUTATIONAL MODEL FOR MUSIC ANALYSIS, APPLIED TO MAQAM ANALYSIS

Olivier Lartillot
University of Jyväskylä
olartillot@gmail.com

Mondher Ayari
University of Strasbourg
mondher.ayari@ircam.fr

ABSTRACT

We introduce a new computational framework for music analysis decomposed into a set of modules. Each module addresses a core aspect of music analysis and offers some innovative breakthrough compared to the state of the art. In order to overcome the limitations of local segmentation, we propose an alternative paradigm based on hierarchical local grouping. New mechanisms for ornamentation reduction based on local grouping enable to build a syntagmatic network for the search for ornamented patterns. We propose an approach for modal analysis based on comparison of the local context (defined by the current and recent notes, and taking into account ornaments reduction) with all possible modes and key scales. We show how this could be applied in particular for the analysis of Maqam music. Pattern mining is applied for the search for motives, for mode-related patterns as well as metrical analysis. The integration of the modules into a single framework enables to model interdependencies, which play a major role in music.

1. TRANSCRIPTION FROM AUDIO

The analysis can be carried out on MIDI, score representations or audio files. For audio recordings, a first step of transcription attempts to locate temporal position of notes through a combined detection of significant increase of energy and of stabilised pitch. This results in a "proto-symbolic" representation where notes are characterised by temporal location and duration, pitch and dynamics. Pitch quantisation and spelling is carried out with interaction with the modal analysis described in section 4. Rhythm quantisation is planned to be carried out using the metrical analysis module evoked in section 5.

2. LOCAL GROUPING

There has been significant research around the concept of *local segmentation*, studying the emergence of structure related to the mere variability in the succession of musical parameters. These research, notably by Tenney & Polansky (1980) or Cambouropoulos (2006), focus on the analysis of monodies, and model this structural phenomenon as a segmentation of the monody, which cuts the temporal span at particular instants, resulting into a linear succession of segments [1]. We previously showed that in these approaches the heuristics for segmentation is based on a mixture of several constraints related to what happens *both before and after* each candidate segmentation point (Lartillot et al., 2013). We presented instead a simpler approach focused only on what happens before each candidate segmentation: this enables to reveal a more complete set of segmentation points, indicate more precisely the temporal locations of the segmentation points, and could also reveal a segmentation hierarchy at multiple structural levels.

We introduce a new formulation of our proposed approach that reveals a much clearer structural description and that can be explained with simple principles. The approaches focuses on *grouping* instead on *segmentation*. In other words, what needs to be characterised are not the segments between notes, but instead the groups of notes that are progressively constructed in a hierarchical framework. The approach is applied uniquely to the temporal domain (to the characterisation of the monody as a succession of inter-onset intervals, or IOIs), and does not apply therefore to the pitch domain (the succession of inter-pitch intervals). This is because pitch-based grouping is more related to streaming, i.e., to the construction, from a given monody that features pitch gaps, of internal monodic lines.

This new model for segmentation along the IOI description can be explained as follows:

- A *local group* G is characterised by a maximal IOI parameter I: the IOIs between the successive notes are all lower or equal to that parameter.

- Local groups form *strict hierarchies*: if one local group G_1 has a maximal IOI I_1 that is higher than the parameter I_2 of another local group G_2, and if both groups coincide in the monody, then necessarily G_2 is strictly included into G_1.

- Local grouping is computed through a single chronological pass of the monody. For a given note n_k, we consider the groups $\{G_i\}$ containing the previous note n_{k-i} from larger groups (higher maximal IOI) to smaller groups (lower maximal IOI). Each group G_i is compared with the new IOI $I_{k-1,k}$ between n_{k-1} and n_k.

- If the new IOI is equal to or smaller than the maximal IOI of the given group G_{i_0}, the group is *extended* with this new note n_k. In order to tolerate slight slow down, even IOIs that are a little higher than the maximal IOI are accepted. More precisely,

[1] Pearce et al. (2010) research also enters into that linear segmentation paradigm, but the underlying principles are not based on the local texture of music, but instead on stylistic rules.

the condition for group extension is the following:

$$\log \frac{I_{k-1,k}}{I_{i_0}} < \delta \quad (1)$$

where δ is fixed to .3 in our current tests.

- If the new IOI is larger than the maximal IOI of the given group G_{i_0}, i.e. if

$$\log \frac{I_{k-1,k}}{I_i} > \delta \quad (2)$$

the group is *closed*: it is followed by a longer silence so will not be extended any more. All the other groups $\{G_i\}_{i>i_0}$, with even lover maximal IOIs, are closed as well.

- Hence only the larger groups with higher maximal IOIs, $\{G_i\}_{i<i_0}$, have been extended. G_{i_0-1} is the smallest group that has been extended. If the next IOI $I_{k-1,k}$ is even smaller than the maximal IOI I_{i_0-1} of that smallest group, i.e. if:

$$\log \frac{I_{k-1,k}}{I_{i_0-1}} < \epsilon \quad (3)$$

where ϵ is fixed to -.4 in our current tests, then a *new* group G_i is created, whose maximal IOI is set to the current IOI, i.e. $I_i = I_{k-1,k}$.

This approaches generates a hierarchical structuration of the monody that is very intuitive to understand, as shown in the example in Figures 2 to 5.

Each local group starts immediately at the onset of its first note. Concerning the temporal location of the closure of the group – that can be called *group offset* –, we can apply the very same heuristics we introduced previously in the context of local segmentation (Lartillot et al., 2013): In order to detect that a given group G_i is closed, we need to wait at least the temporal interval I_i after the last note's onset in order to check whether a new note appear during that temporal span or not. If no new note appear, the group is closed. Thus the group offset can be assigned to that moment at I_i after the last note's onset.

3. ORNAMENTS REDUCTION

Previous computational attempts to model processes related to melodic reduction in music primarily (Gilbert & Conklin, 2007; Marsden, 2010) formalize general aspects without detailing concrete conditions for reduction. We present a set of rules founding the detection of ornaments, based on local grouping.

3.1 Local group's head

By definition, a local group terminates with a note that is followed by a duration (before the next note) that is significantly longer than the IOIs within the group. As such, the local group can be perceived as a phrase that terminates with a concluding note that has a more structural importance. This hypothesis might not be always valid, in particular in the presence of particular accentuations at particular

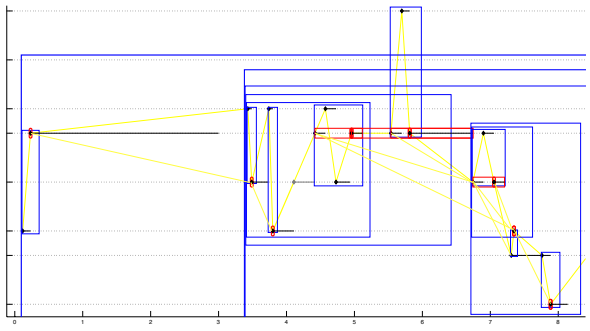


Figure 3: Analysis of the first stave of the improvisation displayed in figure 1 using the same convention as in figure 2.

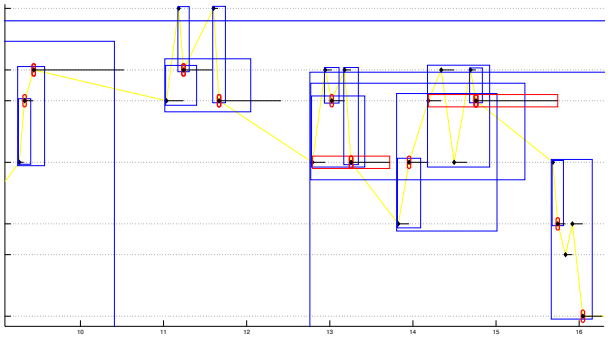


Figure 4: Analysis of the second stave of the improvisation displayed in figure 1 using the same convention as in figure 2.

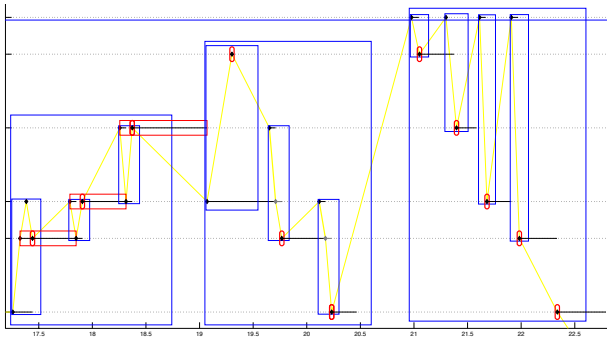


Figure 5: Analysis of the third stave of the improvisation displayed in figure 1 using the same convention as in figure 2.

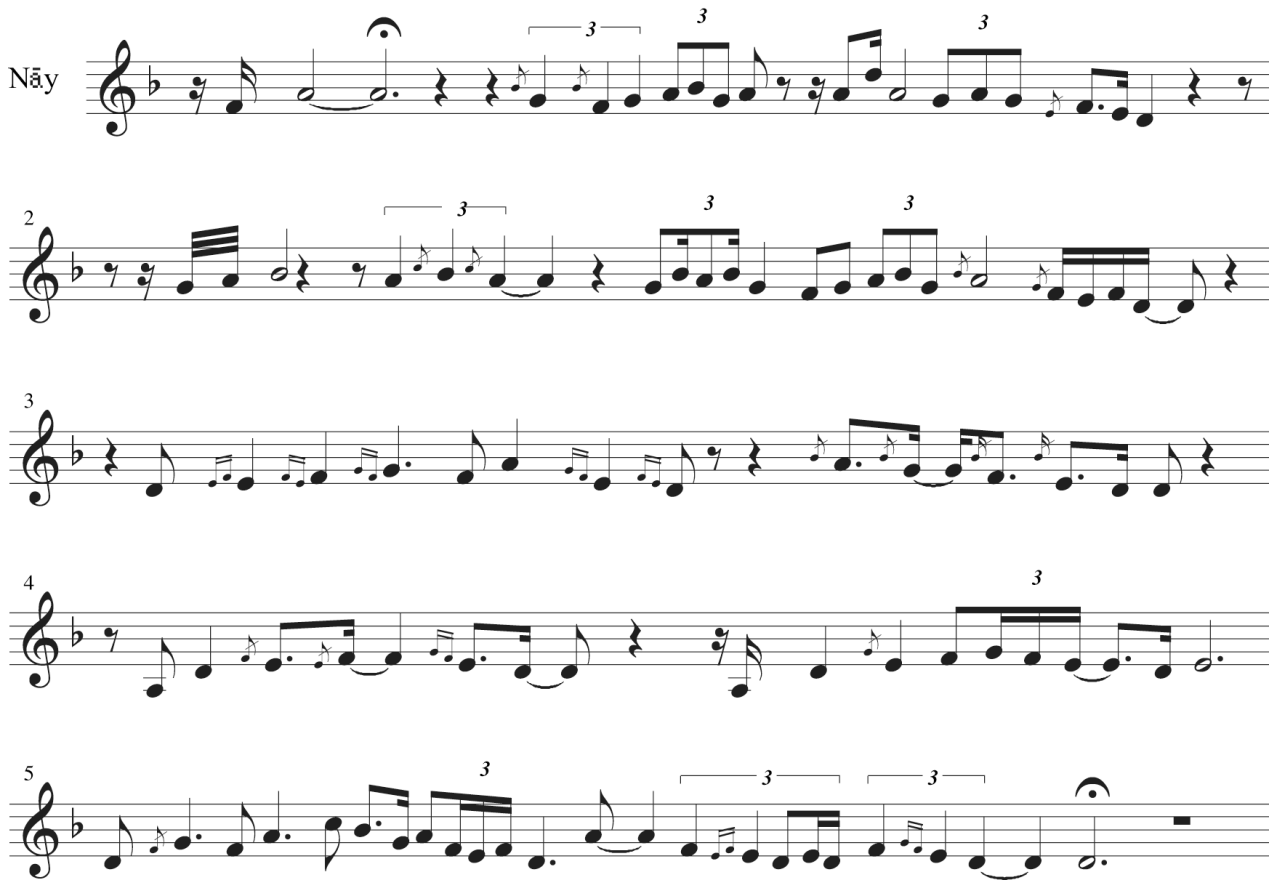


Figure 1: Beginning of a traditional Tunisian modal improvisation *Istikhbâr* played by flute master Mohamed Saâda on the *Mhayyer Sîkâ* maqam.

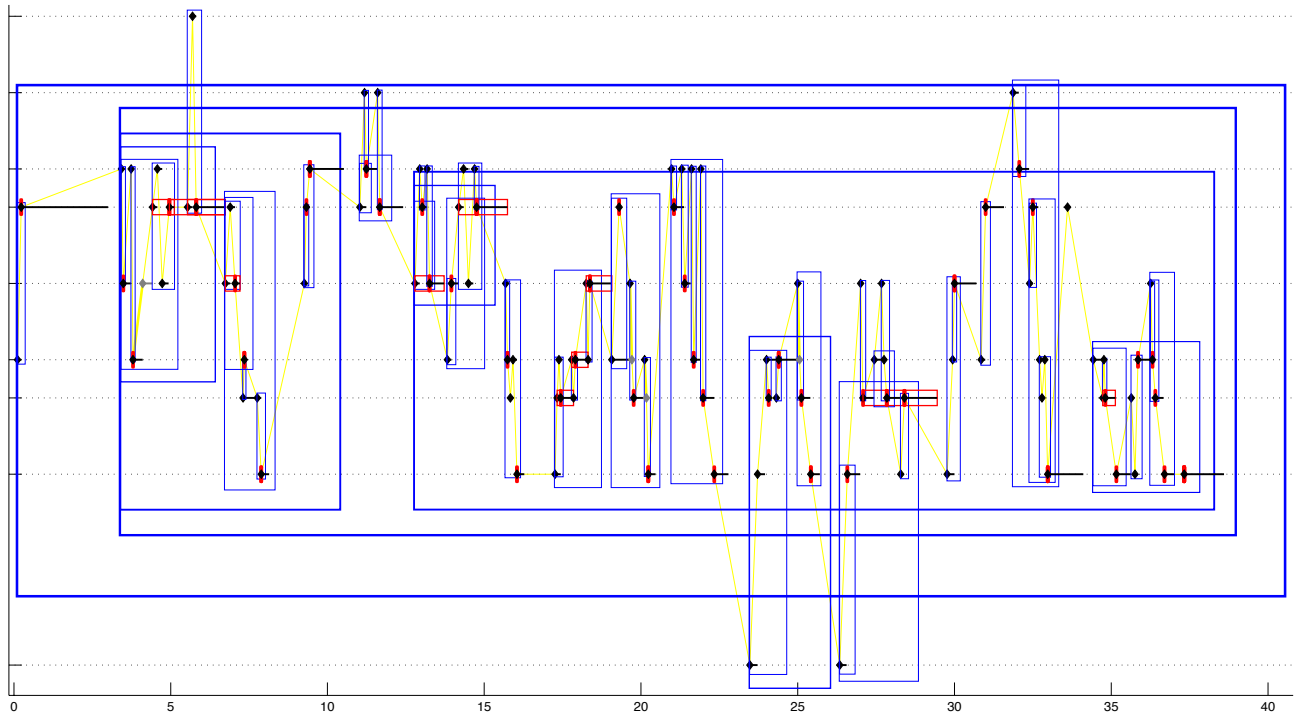


Figure 2: In black: piano roll representation of the beginning of a traditional Tunisian modal improvisation (*Istikhbâr*). In blue: local groupings. In red: Local groups' heads. In grey: passing note.

notes within the group. But in more general case, it seems to offer some general interest. Following this observation, we propose to formalise this hierarchy of notes in local groups by associating with each local group a main note, or "*head*", to follow Lerdahl & Jackendoff (1983)'s Time-Span Reduction terminology, which would in the simple case be the last note of the group, as circled in red in Figures 2 to 5. The other notes would be considered as "*subordinate events*", or – why not – as the "tail" of the group.

When subordinate events in a local group have same pitch than the final note of the group, they all form a single note – a "*meta-note*", as we called in Lartillot & Ayari (2012) –, which will become the actual head of the group. The subordinate events of the groups can be considered as forming an ornementation – such as a *cambiata* or a *trill* – of the group's head. They are highlighted by red rectangles in Figures 2 to 5.

Meta-note can form is any hierarchical level: For instance, in Figure 3, the meta-note is formed on the smallest local group, but in Figure 4 the meta-note around time 13 second in the improvisation is formed on a intermediary hierarchical level.

So far, these subordinate events consist of the notes forming the local group. But a richer understanding of the structural configuration is that the subordinate events of a local group G consist actually of the smaller local groups that belong to that group. And this hierarchical structuration continues recursively for the smaller local groups. Thus when searching for the subordinate events that have same pitch than the last note of group G, we don't need to remember all the notes in the group, but only to check the pitch of the heads of the local groups one level down in the hierarchy.

3.2 Passing note

Within a local group, all notes do not have the same importance. A monotonous and uniform conjunct melodic motion is a series of notes such that:

- inter-pitch intervals between success notes are all of 1 or 2 semi-tones and in the same direction (up or down),
- inter-onset intervals between successive notes are very similar,
- no note is particularly accentuated.

In such configuration, the intermediary notes form *passing notes*: these subordinate elements play mainly a role of filling the interval gap between the starting and ending points of the line. For that reason, these intermediary notes are generally not perceived as note that play a more global role outside that particular melodic line. Intermediary notes are shown in grey in Figures 2 to 5.

This can have an impact in different modules of the integrated analysis framework. For instance, in the previous paragraph dealing with local group's head (section 3.1), a note within a local group that has same pitch than the last note of the group cannot be included in the group's head if it is a passing note.

Passing notes can exist in melodic lines of any length, so for a given note n_k, its "passingness" can be defined based simply on the previous note n_{k-1} and next note n_{k+1}, according to the following conditions:

- the inter-pitch intervals between n_{k-1} and n_k and between n_k and n_{k+1} are of 1 or 2 semi-tones and in the same direction (up or down),
- their inter-onset intervals are very similar:

$$\left|\log \frac{I_{k-1,k}}{I_{k,k+1}}\right| < \beta \tag{4}$$

  where β is fixed to .1 in our current tests.

3.3 Syntagmatic network

Melodic "reduction" is commonly understood as a process of eliminating the ornamentation (here, the subordinate elements) of the monodic surface in order to keep the deeper structure (on various hierarchical levels) made of the more important notes. Our conception of melodic reduction, however, does not impose such reduction of information, but on the contrary, integrates the deeper structure information with the monodic surface. More precisely, the monodic surface is formalised as a chain of connection between successive notes, i.e., a chain of *syntagmatic connections*, or a *syntagmatic chain*, following Saussure's terminology. The deeper structure can be represented by adding new syntagmatic connections between successive elements in the deeper hierarchical levels. We obtain hence a *syntagmatic network* presenting a set of possible alternative syntagmatic chains. We are currently formalising heuristics ruling this construction of syntagmatic chains. They could include for instance:

- The head of any local group is syntagmatically connected to the note n_i preceding the group as well as to the head of any local group closed by n_i.
- The head of any local group is syntagmatically connected to the note n_j succeeding the group as well as to the head of any local group started by n_j.
- In a series of passing note, there is a direct syntagmatic connection between the notes just before and after that series.
- A syntagmatic chain of notes $\{n_{k-1}, n_{k-2}, \ldots\}$ of same pitch form a single meta-note (Lartillot & Ayari, 2012) whose onset is set at the first note n_{k-1}. This meta-note can be syntagmatically connected to any note succeeding any note $n_{k-2}, n_{k-3}, \ldots$ constituting the meta-note.

This concept of syntagmatic network follows Lartillot & Ayari (2012), but we propose here a much simpler network with connections justified by stronger perceptual heuristics, based on local grouping and ornamentation reduction.

The interest of this network is that motivic pattern can be searched for on any path, which enables to detect pattern repetition with or without ornamentation.

Example of syntagmatic networks is shown in yellow in Figures 3.

4. MODAL ANALYSIS

4.1 Scale identification

Previous methods in tonal and modal analysis traditionally compute global pitch statistics (of either whole pieces or on successive arbitrary time frames) that are compared to mode or key templates (Krumhansl, 1990; Gomez, 2006; Gedik & Bozkurt, 2010). The main limitations are that these arbitrary time frames often encompass complex modulations or spurious note events that are foreign to the main mode, and that the templates force a single stereotypical representation of each mode.

In contrast, we are conceiving a new paradigm for modal analysis carried out for each successive note in the proto-symbolic representation, and based on a comparison of the local context (defined by the current and recent notes, and taking into account ornaments reduction) with all possible modes and key scales. A scale can also be identified while recognizing only subset of it, by taking also into account pivotal notes in the scale and longer and main notes in the local context.

Below is a more precise description of the approach, currently under development, specialised here on the analysis of Arabic Maqam music.

- Each Maqam mode is defined by:
  - A scale: a series of pitches, indicated relatively to the origin of the scale.
  - A juxtaposition of *ajnas* (plural of *jins*), as shown in Figure 6. A *jins* is defined as a group of 3 to 5 successive notes such that one (or two) of those notes is considered as pivotal, i.e., melodic lines tend to rest on such notes. The pitch scale of that *jins* is also expressed relatively to the pivot, to which is associated the value 0. One of the *ajnas* is considered as the main *jins* of the Maqam mode: it is typically the one that starts at the *tonic* of the scale.

- For each successive note n_k being analysed, we keep track of any combination of set of pitches $\{S_i\}$ that have been recently played. For each set of pitches S_i, one pitch is selected as *pivot*, it is the pitch of the note with longest duration, among the notes associated with that set of pitch. The pivot is fixed as the pitch origin, with value 0, and the other pitches of the scale are expressed with respect to the pivot.

- The combinatory of possible subset of pitches can be reduced by taking into account the local grouping structure discussed in section 2. For a given note n_k of pitch p_k, a pitch that is not expressed by head of previous local groups, but that appears only as subordinate event(s) of one or several groups G_i will not form a pitch subset with p_k unless the subset also includes the pitch expressed by the head of the group G_i. This represents the hierarchical structure of pitch scale, where the less important pitches, appearing as ornaments, can be considered in the scale only if the whole ornament including the group head is included.

- For each set of active pitches S_i, we can measure the degree of fit $SJ_{i,j}$ with each different possible *jins* J_j, which is defined as the number of pitches in S_i that can be associated with a pitch in the *jins* scale, divided by the total number of pitches in J_j. Hence this score $SJ_{i,j}$ indicates the proportion of the *jins* scale covered by S_i. Note that scales related to S_i and J_j are both expressed relatively to their pivot, which is in both case equal to 0. The alignment of pivots means that the *jins* J_j compared to S_i is at a specific transposition.

- For each possible transposition of each Maqam mode, and for each constituting *jins* J_j, we can find the set of active pitches S_i that fits best, i.e. the one with highest score $SJ_{i,j}$. For each Maqam mode, and for each possible transposition, we obtain hence a table of scores that shows how much its constituting *ajnas* are covered throughout the piece.

- The development of a Maqam mode can be considered as a succession of *ajnas*. But it appears that an important part of ornamentation transgresses somewhat the border fixed by the *ajnas*. These ornamentation could be understood as transitory *ajnas* that appear locally and are superposed to the longer-term logic of the *ajnas* developed in a larger scale.

- Again, the local grouping structure can be used to infer the temporal scope of the different ajnas discovered. Particular notes that are subordinate events of local groups can infer particular *ajnas* that do not exceed the scope of the local group.

- We are conceiving methods for choosing the actual Maqam mode, among the different possible candidates. For a new Maqam to be detected at a given note n_k, its main *jins* needs to be activated for that note.

4.1.1 Example

Here there ideas are illustrated through the analysis of the beginning of an improvisation based on *Mhayyer Sîkâ maqam* (Lartillot & Ayari, 2011). To make the analysis more challenging, we ignore the pedal note D played from the beginning of the improvisation to the end, emphasising the tonic of the mode.

- The improvisation starts with a short note with pitch F.

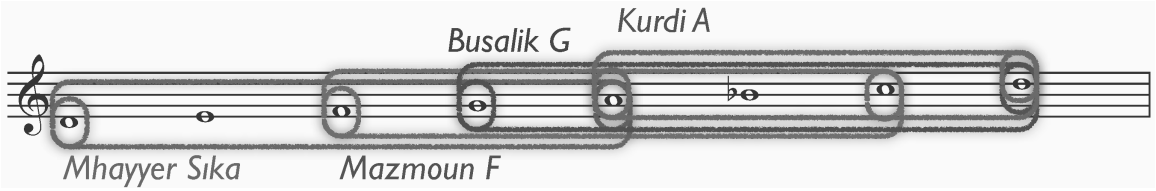


Figure 6: Modal scheme of the *Mhayyer Sîkâ maqam* developed in the improvisation shown in Figure **??**.

- The next note (pitch A) is very long. We obtain the subset (F,*A*) with pivot on A. Evidently, a huge number of possible *ajnas* from various scale at various transposition can be associated to this minimalistic description, so no particular modal context is inferred for the moment.
- The third note (pitch Bb) is very short. Subset (*A*,Bb) with pivot on A is inferred. The combination of the subset (F,*A*) and (*A*,Bb) leads to an identification of the *Mhayyer Sîkâ* scale with (F,*A*) belonging to the main *jins* while (*A*,Bb) belongs to the *Kurdi A jins*.
- The next note (pitch G) is longer than the previous one, closing a 2-note local group whose pitch subset (*G*,Bb) belongs to the *Busalik G jins*. But in a larger scale, the subset (F,G,) belongs to the main *jins*.
- The next two notes (pitch Bb and F) form a local group related to the *Mazmoun F jins*. In larger scale, F develops further the main *jins*.
- The analysis continues similarly.

4.2 Modal pattern identification

But beyond scales, *ajnas* and pivotal notes, Maqam modes are also characterised by particular melodic lines. We conceived an innovative method (Lartillot, 2005) for listing sequential patterns and tracking their possible cyclical repetition. We are currently integrating the detection of these modal patterns through an interaction between local grouping, ornamentation reduction and the motivic analysis module.

For instance, the main characteristic melodic line associated with *Mhayyer Sîkâ maqam* is the pitch series A F E D. The modal pattern can be explicitly found along particular paths of the syntagmatic nework constructed based on the method presented in section 3.3.

5. FURTHER WORKS

Adequate rhythmical description of notes requires the inference of the underlying pulsation, and more generally of multiple pulsation levels forming a metrical structure. Current beat tracking methods are based on global description of periodicity (such as autocorrelation function) (Dixon, 2007), which, here also, fails to grasp particular idiosyncrasies of music, such as ornaments or sudden changes of tempo. Besides, periodicities can also be expressed as successive repetitions of sequential patterns. This method is applied among others to detect pulsation and construct the metrical structure. The metrical structure cannot always be inferred from a mere search for periodicity in the signal, but may sometimes require the recognition of learned rhythmical patterns.

Our aforementioned method for sequential pattern mining is used to detect not only metrical periodicities, but also any type of sequential repetitions such as melodic themes and motifs (Lartillot, 2005). Ornaments reduction enables to detect varied repetitions; modal and metrical analyses add further musical representations (such as diatonic scale and rhythmic values) along which the sequential pattern mining is carried out as well.

Following Lerdahl & Jackendoff (1983), grouping structures are founded not only on local discontinuity (cf. local grouping module), but also on higher-level aspects, such as mode, metre and motifs. We are extending our previous works (Lartillot & Ayari, 2009) focused initially on linear segmentation in order to integrate multi-level grouping. Contrary to the purely hierarchical representation in Lerdahl & Jackendoff (1983), we propose a model that allows overlapping and multiple segmentation alternatives. One main application of this grouping structure construction is the detection of musical forms. This would be modeled as a mapping between the global grouping structure constructed on a given piece with form patterns that constitute the predefined cultural knowledge.

6. CONCLUSION

One main objective of the *CréMusCult* project [2] was to study the impact of cultural knowledge in listeners' understanding of music and in our proposed cognitive modeling of music analysis. Cultural knowledge is implemented as specification of list of modes, metrical structures, motifs, forms, but could also take the form of particular parametrization such as in the local grouping module.

The complete framework presented in this paper forms a package, called *MusMinr*, of *MiningSuite*, a new Matlab environment for audio and music analysis. A standalone graphical user interface, also called *CréMusCult*, offers to musicologists the possibility of easily visualize the complete analysis of musical pieces of their choices.

[2] Funded by the French research agency (ANR) during the years 2011-2013.

7. ACKNOWLEDGMENTS

This study has benefited from stimulating discussions with Mathieu Giraud, Petri Toiviainen and Funda Yazici.

8. REFERENCES

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

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

Gedik, A. C. & Bozkurt, B. (2010). Pitch frequency histogram based music information retrieval for turkish music. *Signal Processing*, *10*(1049-1063).

Gilbert, E. & Conklin, D. (2007). A probabilistic context-free grammar for melodic reduction. In *Proceedings of the International Workshop on Artificial Intelligence and Music, IJCAI-07*.

Gomez, E. (2006). *Tonal description of music audio signal*. PhD thesis, Universitat Pompeu Fabra, Barcelona.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.

Lartillot, O. (2005). Multi-dimensional motivic pattern extraction founded on adaptive redundancy filtering. *Journal of New Music Research*, *34*(4), 375–393.

Lartillot, O. & Ayari, M. (2009). Segmentation of tunisian modal improvisation: Comparing listeners' responses with computational predictions. *Journal of New Music Research*, *38*(2), 117–127.

Lartillot, O. & Ayari, M. (2012). Retentional syntagmatic network, and its use in motivic analysis of maqam improvisation. In *Proceedings of the 2nd International Workshop of Folk Music Analysis*.

Lartillot, O., Yazici, F., & Mungan, E. (2013). A pattern-expectation, non-flattening accentuation model, empirically compared with segmentation models on traditional turkish music. In *Proceedings of the 3rd International Workshop on Folk Music Analysis*.

Lerdahl, F. & Jackendoff, R. (1983). *The Generative Theory of Tonal Music*. Utrecht: MIT Press.

Marsden, A. (2010). Schenkerian analysis by computer: A proof of concept. *Journal of New Music Research*, *39*(3).

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, *39*(1367-1391).

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

THE TRANSFER/ADAPTATION STAGES OF TURKISH FOLK MUSIC PHONETIC NOTATION SYSTEM/TFMPNS TO VOICE EDUCATIONAL/DOCTRINAL APPLICATIONS: CANTOVATION SING & SEE™

Gonca Demir
gnc.dmr@windowslive.com

ABSTRACT

Turkish Folk Music Phonetic Notation System/TFMPNS is a notation system example which aims to initiate a parallel application to the international linguistic/musicological application foundations of which were laid under the scope of Istanbul Technical University Institute of Social Sciences Turkish Music Post Graduation Program thesis, which will be developed under the scope of Istanbul Technical University Institute of Social Sciences Musicology and Music Theory Doctorate Program thesis, which is configured in phonetics, morphology, syntactic, vocabulary and lexical axis of together with traditional/international attachments based on Turkish Linguistic Institution Transcription Signs/TLITS and International Phonetic Alphabet/IPA sounds. Through the announcement that will be made under the scope of the 4th International Workshop on Folk Music Analysis/FMA 2014; CantOvation Sing & See™ usage of which is foreseen for the transfer and adaptation stages of Turkish Folk Music Phonetic Notation System/TFMPNS (developed in the axis of local/universal structural/generative/transformation linguistic theories, linguistic/written science/rhetorical/phonological approaches in ethnomusicology and it is consist of Turkish Folk Music Phonetic Notation System Alphabet Database/TFMPNS AD & Turkish Folk Music Phonetic Notation System Sound Database/TFMPNS SD & Turkish Folk Music Phonetic Notation System Dictionary Database/TFMPNS DD & Turkish Folk Music Phonetic Notation System Works Database/TFMPNS WD & Turkish Folk Music Phonetic Notation System Phonotactical Probability Calculator Database/THMFNS PPCD ect) to vocal training applications.

INTRODUCTION

It has been accepted that the basic factor in the traditional musics that reached our modern day through bush telegraph completely as a result of oral culture for centuries; whose theoretical aspect was formed much later; and which has a wholly-performative qualifications is not theory but application and vocalization techniques. It is known that theories related to folk music as notation consist of some rules and classifications that were detected by the musicologists long after their emergence. In the local musics, the theory is tried to be taken from the performance, the notation that is used to detect the music aren't known sufficiently and it isn't understood in all aspects. (Gedikli, 1997: 58-63). Note writing to stabilize the fundamental properties of a music first emerged in the form of theoretical instructions. Throughout history it has been emphasized that the unique tone image should be investigated and vocalized so as to reveal and understand the tone existence of the note writings mediating between theory and practice completely. Nowadays, the high qualifications that traditional notation technique, one of the note writing techniques, has are being ignored. It has been emphasized that the improvement of the writing quality of the traditional notation (Pekoglu, 2007: 1, 21, 53) should be transferred from the debate field based on the protection and vocalization of the traditional music to the planning and implementation field (Gunay, 1988: 65). While putting the musics in a written form, different writing techniques have been used by the musicologists who started a parallel application to the linguistic developments and it has been identified that transcription technique, one of the notation technique of the folk music, has provided many opportunities in the process of putting the local musics in written forms (Duygulu, 2006: 210). Turkish folk music has a priviliged place in music types due to regional dialect varieties. The future of Turkish folk music depends on protection of its attitude originating from dialect differences and its resistance against change. Turkish folk music regional dialect properties are transcripted by Turkish Linguistic Instituion Transcription Signs/TLITS depending on linguistic laws in axis of phonetics/morphology/parole existence. On the other hand, depending on musicological laws, regional dialect properties of Turkish folk music which is a verbal/artistic performance type structured in axis of linguistic approaches in ethnomusicology/performance display theory are also transcripted by Turkish Linguistic Institution Transcription Signs/TLITS. It is determined and approved by linguistic/musicology source and authotrities that this reality which is also present in other world languages can be transferred to notation and vocalized again and again in accordance with its original through International Phonetic Alphabet/IPA existence and usability of which have been registered by local and universal standards through the notification that will be submitted (Demir, 2011) (see Figure 1).

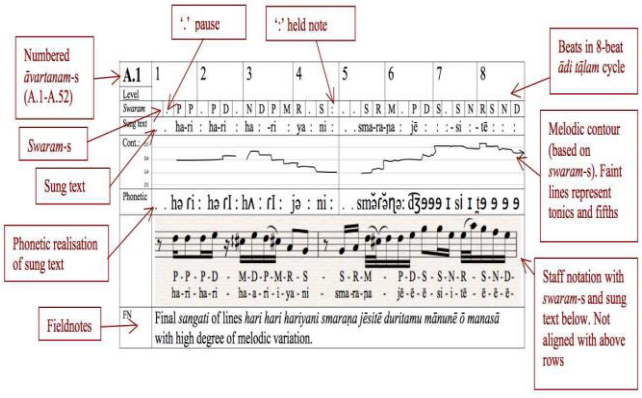


Figure 1. Musicolinguistic graphic sample: Radhakrishnan, 2011: 423-463

1. TURKISH FOLK MUSIC PHONETIC NOTATION SYSTEM/TFMPNS

Turkish Folk Music Phonetic Notation System Alphabet Database/TFMPNS AD: transcription system of Turkish Language Institution/TLI dialect researches (Ercilasun, 1999: 43-48), transcript in dialect studies (Sağır, 1999: 126-138), vowel and consonant changes of Anatolia dialects (Caferoglu, 1964-1965: 1-33), Urfa/Kerkuk/Tallafer Dialects Turkish Language Institution Transcript Signs/UKTD TLITS (Ozbek, 2010: xviii, 11-19), IPA provisions of the words in Turkish alphabets and TDK-IPA provisions of voice descriptions-transcription signs (Pekacar & Guner Dilek, 2009: 584-588), phonology ABCs of Turkey Turkish Pronunciation Dictionary/TTPD: IPA provisions of vowels and consonants (Ergenç, 2002: 46-47), IPA tables (URL <http://www.langsci.ucl.ac.uk/ipa/ipachart.html>),International Phonetic Alphabet/IPA Turkish vowel/consonant letter tables (IPA, 1999: 154-156), extra-IPA symbols for irregular speaking (URL <http://www.langsci.ucl.ac.uk/ipa/extIPAChart2008.pdf), IPA number table (URL <http://www.langsci.ucl.ac.uk/ipa/IPA_Number_chart_(C)2005.pdf>), IPA unicode character codes (URL <http://www.langsci.ucl.ac.uk/ipa/phonsymbol.pdf>),IPA X-SAMPA equivalency table (URL <http://www.kreativekorp.com/miscpages/ipa/ipa-x.html>).

Turkish Folk Music Phonetic Notation System Sound Database/TFMPNS SD: International Phonetic Alphabet/IPA sound records (URL <http://www.langsci.ucl.ac.uk/ipa/sounds.html>),International Phonetic Alphabet/IPA Turkish vowel/consonants tables sound records (IPA, 1999: 154-156), Turkish Language Institution Turkish Audio Dictionary/TLI TAD (URL <http://www.tdk.gov.tr/>), 128 pieces of Turkish folk music texts sound records transcripted with the Urfa/Kerkuk/Tallafer Dialects Turkish Language Institution Transcript Signs/UKTD TLITS (URL <https://tez.yok.gov.tr/UlusalTezMerkezi/>).

Turkish Folk Music Phonetic Notation System Dictionary Database/TFMPNS DD: Turkish Language Institution Current Turkish Dictionary/TLI CTD (URL <http://www.tdk.gov.tr/index.php?option=com_gts&view=gts>), Turkish Language Institution Turkish Audio Dictionary/TLI TAD (URL <http://www.tdk.gov.tr/index.php?option=com_seslissozluk&view=seslissozluk>), Turkish Language Institution Big Turkish Dictionary/TLI BTD (URL <http://www.tdk.gov.tr/index.php?option=com_bts&view=bts>), Turkish Language Institution Search Dictionary/TLI SD (URL <http://www.tdk.gov.tr/index.php?option=com_tarama&view=tarama>), Turkish Language Institution Turkey Turkish Dialects Dictionary/TLI TTDD (URL <http://www.tdk.gov.tr/index.php?option=com_ttas&view=ttas>), Turkish Language Institution Folk Dialects Compilation Dictionary in Turkey/TLI CDFDT (TDK, C. I-VI), Turkey Turkish Pronunciation Dictionary/TTPD (Ergenç, 2002: 91-486), Urfa/Kerkuk/Tallafer Dialects Index and Dictionary/UKTD ID (Ozbek, 2010: 113-253).

Turkish Folk Music Phonetic Notation System Works Database/TFMPNS WD; 128 pieces of Turkish folk music texts transcripted with the Urfa/Kerkuk/Tallafer Dialects Turkish Language Institution Transcription Signs/UKTD TLITS (Ozbek, 2010: 254-329), IPA Turca: Rule-Based Turkish Phonetic Converter Program/RBTPCP (Bicil & Demir, 2012).

Turkish Folk Music Phonetic Notation System Phonotactical Probability Calculator Database/THMFNS PPCD structured with the local correlations in the axis of local structural/generative/transformational linguistic theories-linguistic/writing scientific/phonological/ rhetorical in ethnomusicology sound knowledge/phonetic-morphological/syntactic-word assets/lexical extent: IPA provisions and sound definitions of the letters in Turkish alphabets IPA provisions of transcript signs (Pekacar & Guner Dilek, 2009: 584-588), Urfa/Kerkuk/Tallafer Dialects Turkish Language Institution Transcript Sings/UKTD TLITS vowel/consonant distinctive signs (Ozbek, 2010: xviii, 11-19), IPA Turca: Rule Based Turkish Phonetic Translator Program/RBTPTP character codes (Bicil & Demir, 2012), Turkey Turkish Pronunciation Dictionary/TTPD phonology ABCs: Standard Tukey Turkish/STT IPA provisions of vowels and consonants (Ergenç, 2002: 46-47), UCLA phonetics lab archive/Turkish language section (URL <http://archive.phonetics.ucla.edu>).

Turkish Folk Music Phonetic Notation System Phonotactical Probability Calculator Database/THMFNS PPCD structured with the universal correlations in the axis of universal structural/generative/transformational linguistic theories, linguistic/written scientific, rhetorical/phonological approaches in ethnomusicology-sound information/phonetic-figure information/syntactic-word assets/lexical extent: International Phonetic Alphabet/IPA(URL<http://www.langsci.ucl.ac.uk/ipa/IPA_chart_(C)2005.pdf>), extra-IPA symbols for irregular speech(URL<http://www.langsci.ucl.ac.uk/ipa/extIPAChart2008.pdf>), IPA number table (URL <http://www.langsci.ucl.ac.uk/ipa/IPA_Number_chart_(C)2005.pdf>), IPA X-SAMPA equivalency table (URL <http://www.kreativekorp.com/miscpages/ipa/ipa-x.html>), International Phonetic Alphabet/IPA Turkish vowel/consonant letter tables (IPA, 1999: 154-156), IPA unicode character code charts (URL <http://www.langsci.ucl.ac.uk/ipa/phonsymbol.pdf>), IPA fonts (SIL Encore IPA and SIL IPA93 fonts (doulos/sophia/manuscript fonts: base characters/diacritics/tone and punctuation)-phonetic fonts for macintosh/windows-adobe fonts for macintosh/windows-the four stone phonetic fonts in GIF form (stone sans/stone sans alternate/stone serif/stone serif alternate)-rogers fonts (IPAPhon) for macintosh/windows-phonetic fonts for TeX/LaTeX etc. (URL <http://www.langsci.ucl.ac.uk/ipa/ipafonts.html>). (see Table 1-2).

Standard Turkey Turkish/STT	International Phonetic Alphabet/IPA
Gele gele geldik bir kara taşa	ɟelɛ ɟelɛ ɟeldic biɾ kaɾa taʃa
Yazılanlar gelir sağ olan başa aman efendim	jaʑɯłanłaɾ ɟeliɾ saː ołan baʃa aman efɛndim
Bizi hasret koyar kavim kardaşa	biʑi hasɾɛt kojaɾ kavim kaɾdaʃa
Bir ayrılık bir yoksulluk bir ölüm aman efendim	biɾ ajɾɯłɯk biɾ joksułłuk biɾ ølym aman efɛndim
Nice sultanları tahttan indirir	nidʒɛ sułtanłaɾɯ tahttan indiɾiɾ
Nicesinin gül benzini soldurur aman efendim	nidʒɛsinin ɟyl benzini sołduɾuɾ aman efɛndim
Niceleri dönmez yola gönderir	nidʒɛleɾi dønmɛʑ joła ɟøndɛɾiɾ
Bir ayrılık bir yoksulluk bir ölüm aman efendim	biɾ ajɾɯłɯk biɾ joksułłuk biɾ ølym aman efɛndim
Note 1. Transcription systems in Anatolia dialect researches: transcripted with Standard Turkey Turkish/STT in the axis of standard writing/transcription/variation method (Demir, 2010: 93-106).	**Note 2.** IPA Turca: IPA provisions and sound description (Pekacar & Guner-Dilek, 2009: 575-589) of the letters in Turkish alphabet in Rule-Based Turkish Phonetic Converter Program/RBTPCP (Bicil &Demir, 2012). Turkey Turkish Pronunciation Dictionary/TTPD phonology ABC's: transcripted with International Phonetic Alphabet/IPA (IPA, 1999) by the IPA correspondences of vowel and consonants (Ergenç, 2002, 1-496).

Table 1. Turkish Folk Music Phonetic Notation Syastem/TFMPNS provisions of transcription signs are transcripted with Standard Turkey Turkish/STT > International Phonetic Alphabet/IPA.

Turkish Language Institution Transcription Signs/TLITS	International Phonetic Alphabet/IPA
Gele gele geldĭm bir ḳara daşa	Gɛlɛ gɛlɛ gɛldŭm biɾ kaɾa daʃa
Yazılanlar gelĭr saġ olan başa aman efendĭm	jazɯlanlaɾ gɛlŭɾ sag ɔlan baʃa aman ɛfɛndŭm
Bĭzĭ ḥesret ḳoydı ḳavım ḳardaşa	Bŭzŭ ḫɛsɾɛt kɔjdɯ kavɯm kaɾdaʃa
Bir ayrılıḫ bir yoḫsıllıḫ bir ölŭm aman efendĭm	Biɾ ajɾɯlɯχ biɾ jɔχsɯllɯχ biɾ œl̡im aman ɛfɛndŭm
Nice Sülėymanları taḫtan ėndĭrĭr	Nidʒ͡ɛ ʂœlejmanlaɾɯ taχtan endŭɾŭɾ
Nicesĭnĭn gül benzĭnĭ soldırır aman efendĭm	Nidʒ͡ɛsŭnŭn gyl bɛnzini sɔldɯɾɯɾ aman ɛfɛndŭm
Nicesĭnĭ dönmez ėle gönderĭr	Nidʒ͡ɛsŭnŭ dœnmɛz elɛ gœndɛɾŭɾ
Bir ayrılıḫ bir yoḫsıllıḫ bir ölŭm aman efendĭm	Biɾ ajɾɯlɯχ biɾ jɔχsɯllɯχ biɾ œl̡im aman ɛfɛndŭm
Note 3. Linguistic approaches in ethnomusicology (Stone, 2008: 51-53): phonetic writing usage in data recording in musicology: necessity of dialect documentation in linguistic and musicological axis: Urfa/Kerkuk/Tallafer Dialects Turkish Language Institution Transcription Signs/UKTD TLITS (Ozbek, 2010: iii-338) in the axis of phonetic notation method of local dialect features of Turkish folk music: transcripted with vowels-consonants-distinctive signs (Demir, 2011: v-294).	**Note 4.** International Phonetic Alphabet/IPA usage in dialect researches of Turkish language: written dialect texts in Turkey by using IPA (TDK-IPA) provisions of transcription signs are transcripted with Standard Turkey Turkish/STT- Turkish Language Institution Transcription Signs/TLITS-International Phonetic Alphabet/IPA (Pekacar & Guner-Dilek, 2009: 575-589).

Table 2. Turkish Folk Music Phonetic Notation Syastem/TFMPNS provisions of transcription signs are transcripted with Turkish Language Institution Transcription Signs/TLITS > International Phonetic Alphabet/IPA.

2. CANTOVATION SING AND SEE™ (SINGING SOFTWARE FOR REAL-TIME VISUAL FEEDBACK OF THE VOICE IN VOCAL TRAINING) FEATURES/PRODUCT RANGE

CantOvation Sing & See™ features; pitch trace: is indicated in the comprehensive presentation of ivories reminder as lines constantly moving. In the processes of passing from a musical note to another, it indicates changes in intonation and control/domination effectiveness/deficiency on the notes. Piano keyboard: pitch is indicated by active ivories reminder being determined as red. Stave view: pitch can be directly seen at stave display. Spectrogram display: brightness and power of each voice harmonics indicated in colour spectogram is directly proportional. Tone balance, tone changes and tone interval can be seen in the process of singing. The singer editor can be observed between horizontal lines at 2kHz-4kHz and be varied between narrow band- broad band. Real-time spectrum: volume of lines/graphic curves indicates the power of each harmonic in voice. Song editor spectrum enables to find tone balance of voice in the process of singing and present sudden visual reaction/feedback related to changes in singing performance. Level meter: helps adjusting balance tone throughout tessitura as directly proportional with the volume of green bar. (URL <http://www.singandsee.com/features.php>).

CantOvation Sing & See™ product range; student version: real time intonation was designed in three modes as ivories, musical stave display and intonation as a practical tool in order to provide feedback. Professional version: united display which can present integrated spectographic and intonation images was designed as a software having advanced level of features as recording voice records and replaying. Teacher's manual: Singing Pedagogy at Digital Period prepared by Jean Callaghan and Pat Wilson was exclusively designed for the instructors. Visual display explanations on manual screen includes question discussions related to voice and exercises with the number more than 200 to be taken into consideration in individual applications. Teacher's package: the product includes user manual, professional singing software and instructor manual. Multi-Packs: includes 10 licences for student version in order to be able to distribute students. (URL <http://www.singandsee.com/products.php>) (see Figure 2-8).

Figure 2 CantOvation Sing & See™ : URL <http://www.singandsee.com/>

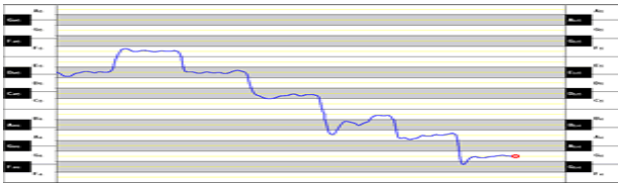


Figure 3 CantOvation Sing & See™ - pitch trace: URL <http://www.singandsee.com/features.php>

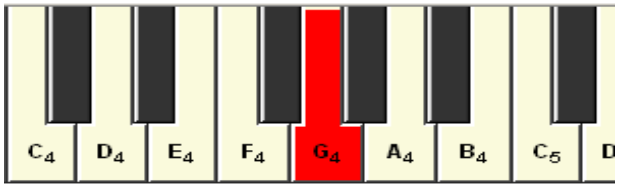


Figure 4 CantOvation Sing & See™ - piano keyboard: URL <http://www.singandsee.com/features.php>

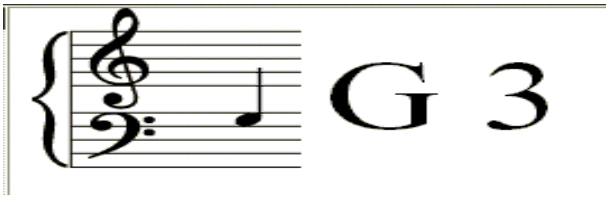


Figure 5 CantOvation Sing & See™ - stave view: URL <http://www.singandsee.com/features.php>

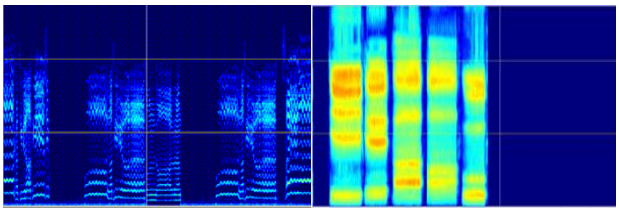


Figure 6 CantOvation Sing & See™ - spectrogram display: URL <http://www.singandsee.com/features.php>

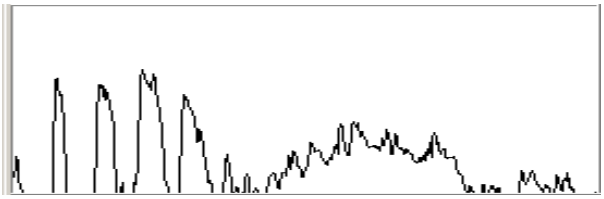


Figure 7 CantOvation Sing & See™ - real-time spectrum: URL <http://www.singandsee.com/features.php>

Figure 8 CantOvation Sing & See™ - level meter: URL <http://www.singandsee.com/features.php>

4. CANTOVATION SING AND SEE™ (THE LATEST ADVANCES IN TECHNOLOGY FOR VOICE TRAINING VOCAL SOFTWARE FOR SINGERS) IPA SECTION

International Phonetic Alphabet/IPA is a type of standard alphabet consisting of signs and symbols for representing sounds on paper, coding sounds in all languages in an exemplary form, avoiding confusions caused by several transcription system with inconsistent and arbitrary software by enabling languages to accurately pronounced. Latin characters are basically used in formation processes of International Phonetic Alphabet/IPA; characters borrowed from another alphabets have been changed in a way that will conform to Latin characters. (IPA, 2009: vii-viii). IPA consists of six main parts as pulmonic consonants, non-pulmonic consonants, vowels, supresegmentals, diacritics, other symbols. In addition to this, it is included in Turkish vowel and consonant character tables (IPA, 1999).

Use of IPA in musical applications: it has been emphasized that language contains several dialects and this content is existing in each language and thus in music of each language, accurate pronunciation of this different articulation features can only be accessed through IPA use and pronunciation contributing melodic memory development with invoking effect, form of singing based on pronunciation features of vowels provides immense facilities in the issues such as enabling performer to find its personal dialectic, yielding beautiful-accurate sound etc. To what extent use of IPA within music directs explicitly practical behaviors and performances of musicians during adaptation and the reasons underlying level-details of perceptibility-intelligibility by musicians for IPA forms (Kurt, 1967: 4) (see Figure 9).

Voice therapy programmes/applications defined as changing voice through behavioural methods and making a new behaviour pattern by determining targeted voice within physio anatomic boundaries (being able to be steady during verifying/speaking process in determining standard/targeted voice-producing targeted voice/change vocal behaviour patterns) are structured by using physiopathology mechanism. Moreover, most of pedagogic approaches providing basis for voice therapy programmes and used in performing arts are used as voice therapy technique, and traditional therapy methods related to the area of vocal pedagogy which is a type of verbal-artistic performance direct phoniatry by structuring at the accent of multidisciplinary approaches. Voice therapists/vocal experts emphasized that practical scientific researches should be improved and extended within the scope of experimental voice studies in voice laboratories along with voice/speaking pathologists through concentrating on more contemporary methods instead having off applied in the area of singing pedagogy. (Denizoglu, 2008: 1-16 URL <http://fonomed.net/pages/sesterapileri.pdf>) & (Denizoglu, 2005: 1-11 URL <http://www.fonomed.net/ders_notlari.asp>).

It was emphasized by the phonoligsts that many speaking/voice appraisal tools developed upon therapy applications on language/speaking disorders being foreseen to use at the axis of International Phonetics Alphabet/IPA; Turkish Folk Music Phonetic Notation System Phonetic Awareness Competence Development Processes/TFMPNS PACDP (ensuring transitivity among getting sensitive to speaking voices/being able to manipulate-healthy vocal cord requirement-technique ergonomics-behaviour modifications/symptoms-standard kinaesthetic voice therapy techniques-carrying intonation exercises-melodic memory development-personal performance dialectics-accurate voice production-oral motor control competences-articulation-notation and repertory transfer- phonetic structure particular to language and voice educational/doctrinal applications-reaching the level of standardization by merging therapy applications for language/speaking disorders with traditional therapy methods particular to the area of vocal pedagogy etc.) are among the materials survived in the infrastructure of theoretical/operational database (Koçak, URL <http://drismailkocak.com/tr/ses-terapisi.html>).

In order to initiate an application in parallel with methods and techniques of singing accepted as cult in the world, without changing the articulation centre in the process of speaking and singing, establishing direct relation between Turkish language voice formation provisions and singing voice formation provisions, defined Turkish language as a language having phonologic features which may help to develop and enrich Turkish voice cult, causing the emergence of voice cults different from each other by the influence of vowel and consonant phonemes used by communities on the method and techniques of singing; the requirement for developing method and techniques for singing as in compliance with the structure of Turkish language phonetic and articulation mechanic were emphasized by singing pedagogues (Toreyin, 2000: 83-91) & (Sazak, 2001: 82-88) & (Çevik, 1988: 79).

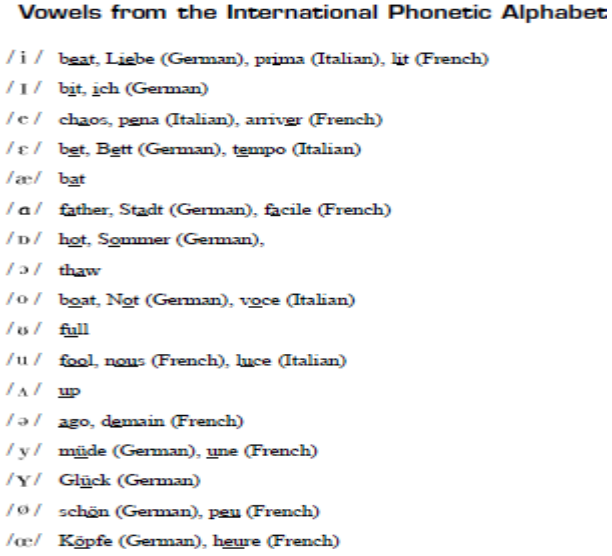


Figure 9 CantOvation Sing & See™ IPA section: vowels from International Phonetic Alphabet/IPA: URL <Callaghan & Wilson, 2003: 275>

3. SUMMARY

Which was developed by company of innovative voice technology CantOvation as a research project under the leadership of vocal trainers/software engineers who make detailed studies in Sydney University School of Communication Sciences and National Vocal Center on singing pedagogy/analysis of voice projections/software development in the axis of high level/recent technology opportunities, existence/usability in musicological literature of which is registered in various fields by national/international standards which is used by many vocal trainers/vocal performers actively, database of which continues to exist in the fictional/executional infrastructure by including vocal/musical properties such as International Phonetic Alphabet pitch labeling, tone quality (voice timber/harmonics/analysis ect), nuance properties (vibrato/legato/glisendo ect), syllabic/pitch frequency regions (ting/metallic voice/power conveying ect), complex real-time linguistic/musicological operation algorithms, piano key reminders, notational/differential confirmation providing perspective schematic graphs/feedbacks (spectrogram/spectrum ect), which can provide teacher/student/professional versions in its product range, sequential/modal/pitch differential application/activation forms, vocal/musical performance assessment sheets ect within its hand book, identified as "Singing Software for Real-time Visual Feedback of The Voice in Vocal Training" and "The Latest Advances in Technology for Voice Training Vocal Software For Singers" CantOvation Sing and See™; in the theoretical/operational infrastructure of Turkish Folk Music Phonetic Notation System/TFMPNS (Turkish Folk Music Phonetic Notation System Alphabet Database/TFMPNS AD-Turkish Folk Music Phonetic Notation System Sound Database/TFMPNS SD-Turkish Folk Music Phonetic Notation System Dictionary Database/TFMPNS DD-Turkish Folk Music Phonetic Notation System Work Database/TFMPNS WD-Turkish Folk Music Phonetic Notation System of Phonotactic Awareness Skill Development Processes/TFMPNS PASDP-Turkish Folk Music Phonetic Notation System of Phonotactic Therapy Applications/TFMPNS PTA-Turkish Folk Music Phonetic Notation System of Phonotactic Probability Calculator Database/TFMPNS PPCD) along with local/universal relations phonics/phonetic-morphology/syntactic-vocab/phonetic awareness competences at the level of vocable scientific dimensions, it is a model foreseen to use in transfer and adaptation processes to voice training/doctrinal applications of International Phonetics Alphabet symbols/voices subsisting at the axis of development processes (requirement of healthy vocal chord/technical ergonomics/correct sound production/behavior modifications/articulation/articulatory features of vowels/consonants/development of melodic memory/personal performance dialectic/notation and implementing on repertory etc. and supra-segmental features/piece sound units like sound quality/level/height/emphasis/tonal/timbral variation ratios etc.)

4. REFERENCES

ANU, (2013). World phonotactics database. *Australian National University Press*, URL <http://phonotactics.anu.edu.au/index.php> (access date: 14.11.2013).

Beckman, Ö. A., Munsonb, B. & Edwardsc, J. (2011). Methodological issues in the analysis of phonotactic probability effects in nonwords, *ICPhS XVII*, Hong Kong, 300-303.

Bicil, Y. & Demir, G. (2012). IPA Turca: Rule-based phonetic converter program/KTTFDP. *TUBITAK National Electronics and Cryptology Research Institute of Multi-Media Technology Research and Development Laboratory*, Gebze/Istanbul.

Caferoglu, A. (1964-1965). Vowels and consonant changes of Anatolian dialects. *Turkish Language Studies Yearbook-Belleten Offprint From 1963-1964, Turkish Historical Society Printing House*, Ankara, 1-33,

Callaghan, J., Thorpe, W., Van Doorn, J. & Wilson, P. (2003). Sing and see. *Proceedings of the 4th Asia-Pacific Symposium on Music Education Research*, 9-12 July 2003, Hong Kong.

Callaghan, J. & McDonald, E. (2007). A comparative study of spoken and sung voice in performance. *In CIM07: 3rd Conference on Interdisciplinary Musicology*, 15-19 August, Tallinn, Estonia.

Callaghan, J., Thorpe, W. & Van Doorn, J. (2004). The science of singing and seeing. *Proceedings of the Conference on Interdisciplinary Musicology (CIM04)*, 15-18 April 2004, Graz, Austria.

Callaghan, J., Thorpe, W. & van Doorn, J. (1999), Computer-assisted visual feedback in the teaching of singing. in Barrett, M. S. & McPherson, G. E. & Smith, R. (Eds.) Children and Music: Developmental Perspectives, *Proc. IMERS* 1999, 105-111.

Callaghan, J., Thorpe, W. & van Doorn, J. (2001). Applications of visual feedback technology in the singing studio, *Australian Association of Research in Music Education Annual Conference*, Newcastle, September 21-24

Callaghan, J. & Wilson, P. (2003). Singing pedagogy in the digital era. *A Cantare Systems Publication. 275*. (ISBN 0-646-42925-6).

Capodieci, F., Hayes, B. & Wilson, C. (2009). UCLA: Manual: phonotactic learning program. URL <http://www.linguistics.ucla.edu/people/hayes/Phonotactics/Manual.pdf> 1-17, (access date: 14.11.2013).

Çevik, S. (1988). Importance of word element in music, effects of music on language and effects of language on music. *First Music Congress, Ministry of Culture and Tourism, Fine Arts General Directorate*, Ankara.

Davidson, L. (2006). Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. *Cognitive Science, vol. 30*, 837-862.

Demir, G. (2011). Phonetic notation of local dialect features of Turkish folk music formed in language-music relationship axis. *Istanbul Technical University Social Sciences Institute of Turkish Music Program*, (published master's thesis, adviser: Assoc. Prof. Erol Parlak), Istanbul/Turkey.

Demir, N. (2010). On variations in Turkish. Ankara University Language and History-Geography Department, *Turcology Journal, vol. 17*(2), 93-106.

Denizoğlu, I. (2005). Vocology lesson notes. URL <http://www.fonomed.net/ders_notlari.asp> (acces date: 29.03.2014).

Denizoğlu, I. (2008). Vocal therapies. URL <http://fonomed.net/pages/sesterapileri.pdf> (acces date: 29.03.2014).

Ercilasun, A. B. (1999). Transcription signals that will be used for dialect researches. *Dialect Research Information Fest, Ataturk Culture, Language and History Institute TLI Publications: vol. 697*, Ankara, 43-48.

Ford, J. K. (2003). Preference for strong or weak singer's formant resonance in choral tone quality. *International Journal of Research in Choral Singing, 1*(1), 29-47.

Gedikli, N. (1997). Theory-practice relation in our traditional music and method problem. *4. İstanbul Turkish Music Days, Education in Turkish Music Symposium May 15-16th, Turkish Republic Ministry of Culture Publications/2058 Publication Unit Directorate Art-Music Masterpieces Series/161-4*, 58-63.

Gorman, K. (2013). Generative phonotactics. *The University of Pennsylvania Linguistic Department*, (published doctor of philosophy thesis), Pennsylvania, 39-64.

Günay, E. (1988). Modern education in Tukish music. *Modern Execution Symposium, XVI. International Istanbul Festival July 4-5-6th, İstanbul Culture and Art Foundation Publications*, İstanbul, 65.

Hayes, B. & White, J. (2012). Phonological naturalness and phonotactic learning. URL <http://www.linguistics.ucla.edu/people/hayes/PhonologicalNaturalness/> 1-32, (access date: 14.11.2013).

Hayes, B. & Wilson, C. (2007). A maximum entropy model of phonotactics and phonotactic learning. URL<http://cogsci.jhu.edu/people/files/_pubs-Wilson/HayesWilsonMaximumEntropyPhonotactics.pdf> 1-67, (access date: 14.11.2013).

Hayes, B. (2012). BLICK: A phonotactic probability calculator, URL<http://www.linguistics.ucla.edu/people/hayes/BLICK/> 1-11, (access date: 14.11.2013).

HEC, Higher education committee local thesis center. URL<https://tez.yok.gov.tr/UlusalTezMerkezi/> Tez No: 263098 (access date: 14.11.2013).

Iclal, E. (2002). Speaking language and Turkish usage dictionary, *Multilingual Foreign Language Publications Baskı Printing,* Istanbul, 1-496.

IPA, International Phonetic Alphabet/IPA chart 2005. URL<http://www.langsci.ucl.ac.uk/ipa/IPA_chart_(C)2005.pdf> (access date: 14.11.2013).

IPA, International phonetic alphabet/IPA chart 2008. URL<http://www.langsci.ucl.ac.uk/ipa/extIPAChart2008.pdf> (access date: 14.11.2013).

IPA, International Phonetic Alphabet/IPA fonts. URL <http://www.langsci.ucl.ac.uk/ipa/ipafonts.html> (access date: 14.11.2013).

IPA, International phonetic alphabet/IPA number chart 2005. URL<http://www.langsci.ucl.ac.uk/ipa/IPA_Number_chart_(C)2005.pdf> (access date: 14.11.2013).

IPA, International phonetic alphabet/IPA sounds records. URL<http://www.langsci.ucl.ac.uk/ipa/sounds.html> (access date: 14.11.2013).

IPA, International phonetic alphabet/IPA unicode characters codes. URL <http://www.langsci.ucl.ac.uk/ipa/phonsymbol.pdf> (access date: 14.11.2013).

IPA, International phonetic alphabet/IPA X-SAMPA. URL<http://www.kreativekorp.com/miscpages/ipa/ipa-x.html> (access date: 14.11.2013).

IPA, International phonetic association/IPA. URL<http://www.langsci.ucl.ac.uk/ipa/ipachart.html> (access date: 14.11.2013).

IPA. (2009). Handbook of the international phonetic association: a guide to the use of the international phonetic alphabet. *Cambridge University Press*, Cambridge.

Dziubalska-Kolaczyk, K. (2009). NP extension: B&B phonotactics, *Poznań Studies in Contemporary Linguistics*, *vol. 45*(1), 55-71.

Koçak, I. (2014). Usage of vocal exercises in vocal therapy: pitch transfer and loading exercises. URL <http://drismailkocak.com/tr/ses-terapisi.html> (acces date: 29.03.2014).

Kurt, A. (1967). Phonetic and diction in singing: Italian, French, Spanish, German. Universty of Minnesota Press, Minneaplolis/USA, 4.

Duygulu, M. (2006). Gypsy music in Turkey (*music culture in western Romani.) Pan Yayıncılık*, Istanbul.

Peykoğlu, M. (2007). Graphic notation usage in 20. Century Music. Post Graduate Thesis, *Dokuz Eylül University Faculty of Fine Arts Music Department*, Izmir, 1-53.

Tansu, M. (1963). Static general vocal knowledge and Turkish. *Turkish linguistic society publications-edition: 222, Turkish historical society press*, Ankara.

Vitevitch, M. (2004). PPC: Phonotactic probability calculator. URL<http://www.people.ku.edu/~mvitevit/PhonoProbHome.html> 481-487, (access date: 14.11.2013).

Miller, R. & Franco, J. C. (1991). Spectrographic analysis of the singing voice. *The NATS Journal, 48*(1): 4-5, 36.

Nisbet, A. (1995). Spectrographic analysis of the singing voice applied to the teaching of singing. *Australian Voice, 1*: 65-68.

Ozbek, M. A. (2010). Language and expression properties of Urfa songs. *Istanbul University Social Sciences Institute of Turkish Language and Literature Department New Turkish Science Branch*, (published PhD thesis), Istanbul/Turkey, iii-338

Pekacar, Ç. & Guner-Dilek, F. (2009). International phonetic alphabet and dialect researches in Turkey. *Dialect Research of Turkey Turkish Workshop*, (25-30 march 2008 Sanlıurfa), *Ataturk Culture, Language and History Institute TLI Publications, vol. 989*, Ankara, 575-589.

Radhakrishnan, M. (2011). Musicolinguistic artistry of niraval in carnatic vocal music. *Australian National University/ANU Research Repository Proceedings of the 42nd Australian Linguistic Society Conference*, 422-463.

Rossiter, D., Howard, D. M. & DeCosta, (1996). Voice development under training with and without the influence of real-time visually presented feedback. *Journal of the Acoustical Society of America, 99*(5), 3253-3256.

Sagır, M. (1999). Transcription in dialect studies. *Dialect Research Information Fest, Ataturk Culture, Language and History Institute TLI Publications: vol. 697*, Ankara, 126-138.

Sazak, N. (2001). Examining the adequacy of vocal training techniques to articulation mechanics and Turkish phonetics. Doctoral Thesis, *Gazi University Science Institute Music Education*, Ankara, Turkey.

Shelley, L. V. (2002). Phonotactic therapy. *Seminars in Speech and Language, vol. 23*(1), 43-55.

Sherer, T. D. (1994). Prosodic phonotactics. *The Graduate School of the University of Massachusetts Amherst Department of Linguistics*, (published doctor of philosophy thesis), Amherst, 1-65.

Stone, R. M. (2006). Theory for ethnomusicology. *Pearson Press*, New Jersey, 51-53.

Thorpe, W., Callaghan, J. & van Doorn, J. (1999). Visual feedback of acoustic voice features: New tools for the teaching of singing, *Australian Voice, 5*, 32-39.

Thorpe, W., Callaghan, J. & van Doorn, J. (2002), Visual feedback of acoustic voice features in singing training. *The 1st International Conference on Physiology and Acoustics of Singing, Groningen*, The Netherlands October 3-5, URL <www.med.rug.nl/pas/>

Thorpe, W., Callaghan, J., Wilson, P. & Van Doorn, J. & Jonathon, C. (2014). CantOvation sing & see™. *Cantovation technology product.* URL <http://www.singandsee.com/> (access date: 28.03.2014).

Thorpe, W. (2002), Visual feedback of acoustic voice features in singing training. *in Proceedings of the 9th Australian Speech Science & Technology Conference*, 3-5 December 2002, Melbourne, 349-354.

Thorpe, W. (2004). CantOvation-innovative voice technology: CantOvation singing software: Sing and see. URL <http://www.cantovation.co.nz/index.html> (access date: 28.03.2014).

TLI, Turkish language institution big Turkish dictionary/TLI BTD. URL <http://www.tdk.gov.tr/index.php?option=com_bts&view=bts> (access date: 14.11.2013).

TLI, Turkish language institution current Turkish dictionary/TLI CTD. URL <http://www.tdk.gov.tr/index.php?option=com_gts&view=gts> (access date: 14.11.2013).

TLI, Turkish Language Institution Dictionary Database/TLI DD. URL <http://www.tdk.gov.tr/> (access date: 14.11.2013).

TLI, Turkish language institution dictionary database/TLI DD. URL <http://www.tdk.gov.tr/> (access date: 14.11.2013).

TLI, Turkish language institution scanning dictionary/TLI SD. URL <http://www.tdk.gov.tr/index.php?option=com_tarama&view=tarama> (access date: 14.11.2013).

TLI, Turkish language institution Turkey Turkish dialect dictionary/TLI TTDD. URL <http://www.tdk.gov.tr/index.php?option=com_ttas&view=ttas> (access date: 14.11.2013).

TLI, Turkish language institution Turkish audio dictionary/TLI TAD. URL <http://www.tdk.gov.tr/index.php?option=com_seslissozluk&view=seslissozluk> (access date: 14.11.2013).

Töreyin, A. M. (2000). Vocal training adequate to Turkey Turkish. *Turkish Language and Literature Magazine*, Edition, 583, 83-91.

UCLA, Phonetic lab archive: Language database. URL <http://archive.phonetics.ucla.edu> 2009. (access date: 14.11.2013).

URL <http://www.singandsee.com/manualpages/p.v.pdf> (access date: 28.03.2014).

URL<http://www.singandsee.com/manualpages/index.php?fs> (access date: 28.03.2014).

Vaden, K. L., Halpin, H. R. & Hickok, G. S. (2013). IphOD: Irvine phonotactic online dictionary. version 2.0. [data file], URL <http://www.iphod.com> (access date: 14.11.2013).

Vitevitch, M. S. & Luce, P. A. A. (2004). Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments & Computers, vol. 36*(3), 481-487.

Welch, G. F., Howard, D. M., Himonides, E. & Brereton, J. (2005). Real-time feedback in the singing studio: an innovatory action-research project using new voice technology. *Music Education Research 7*(2): 225-249.

Welch, G. F., Howard, D. M. & Rush, C. (1989) Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music 17*: 146-157.

Welch G. F., Himonides, E., Howard, D. M. & Brereton, J. VOXed: Technology as a meaningful teaching aid in the singing studio, in Parncutt, R., Kesler, A., & Zimmer, F. (Eds.) *Proceedings of the Conference on Interdisciplinary Musicology (CIM04)*, 15-18 April 2004, Graz, Austria.

Wilson, P, Lee, K., Callaghan, J. & Thorpe, W. (2007) Learning to sing in tune: Does real-time visual feedback help? *In CIM07: 3rd Conference on Interdisciplinary Musicology*, 15-19 August 2007, Tallinn, Estonia.

Wilson, P., Lee, K., Callaghan, J. & Thorpe, W. (2008) Learning to sing in tune: Does real-time visual feedback help? *Journal of interdisciplinary music studies spring/fall 2008, volume 2*, issue 1&2, art. #0821210, 157-172.

Wilson, P., Thorpe, W. & Callaghan, J. (2005) Looking at singing: does real-time visual feedback improve the way we learn to sing? *In 2nd APSCOM Conference: Asia-Pacific Society for the Cognitive Sciences of Music*. 4-6 August 2005, Seoul, South Korea.

Wilson, P. (2006) Does real-time visual feedback improve pitch accuracy in singing? *Master of Applied Science Thesis*, University of Sydney.

Chapter 2

Abstracts

DOES ALWAYS THE PHRYGIAN MODE ELICIT RESPONSES OF NEGATIVE VALENCE?

Manuel Tizon
Composer and guitar player
info@manueltizon.com

Francisco Gomez
Technical University of Madrid
fmartin@eui.upm.es

Sergio Oramas
Pompeu Fabra University
sergio.oramas@upf.edu

ABSTRACT

In this paper the question of whether the Phrygian mode is always associated with perceived emotional responses of negative valence is looked into. To this end, we carried out a series of experiments. Music from two musical traditions where the Phrygian mode is very common, flamenco and Galician music, were chosen for listening tests. Our subjects were 124 children of age 4-7. Some subjects were complete unfamiliar with both traditions and some were familiar with the Galician tradition; none was familiar with flamenco music. Results showed that the perceived emotion was in all cases of positive valence, flamenco music having less valence than Galician music.

1. INTRODUCTION

After a few decades of research, some answers can be offered to the central question of which variables affect emotional response. Among those variables, we encounter intervals (Trainor et al., 2002), melodic contour (Juslin & Sloboda, 2010) (chapter 21, pages 575–604), tonal function and harmony (Sloboda, 1991), Costa et al. (2004), texture (Webster & Weir, 2005), and tempo and mode (Dalla Bella et al., 2001), Caetano et al. (2013). Much of the research on mode has focused and still focuses on the effect of mode, in particular, major and minor modes. Ramos et al. (2011) studied the effect of mode on the emotional response. These authors studied the combined influence of tempo and mode by using the seven Greek musical modes.

Several conclusions in the work by Ramos et al. called our attention. For example, they found that the first and third degree of the mode do not determine the emotional expression of the mode. Also, they reached the conclusion that change in mode is enough to modulate emotional judgments. More interesting was the suggestion that manipulation of mode and tempo do not result in sudden changes of emotion. However, the conclusion that seemed worth deserving further research was that the Phrygian mode is associated with negative valence, in particular, there seems to be an association of Phrygian modes with sadness, which, according to those authors, does not change very much when tempo is increased.

We decided to further investigate the apparent association of Phrygian mode with sadness (or feelings of negative valence). We carried out a study by using Galician and flamenco music, two musical traditions where the Phrygian mode is ubiquitous. Furthermore, the subjects of our experiments were children from age 3 to 7. Previous research by the authors can be found in Tizon et al. (2013). The research contained in this paper is a continuation of Tizon et al. (2013); here experiments were performed on a larger number and more general subjects (with no musical training and from a region whose musical tradition includes the Phrygian mode).

2. THE WORK OF RAMOS ET AL.

The experiments conducted by Ramos and collaborates consisted of the following steps. A piece was composed in major mode and then played in all other musical modes at three different tempi (midi piano timbre was chosen for the stimuli). Adults were used for the listening experiment and they had musicians and non-musicians among their subjects.They only used one piece for the whole experiment. See their paper for further details (Ramos et al., 2011). The Figure below summarizes the main findings (on page 169). This graph shows the seven Greek modes and how the emotional response changes as a function of tempo (dots are connected in increasing tempi).

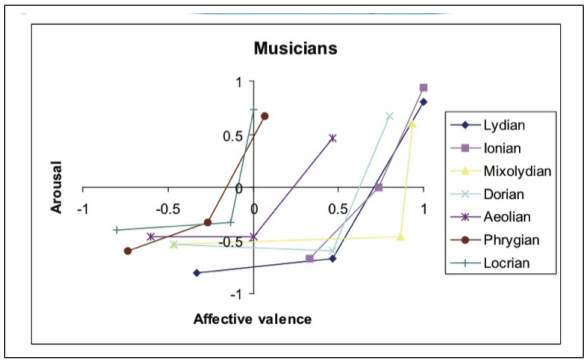


Figure 1: The effect of mode and tempo on the emotional response in Ramos et al. (2011).

It can be noticed how the Phrygian mode is mainly on the quadrant corresponding to sadness and when tempo increases it goes to the first quadrant (that of happiness) but still with low values for the valence.

3. A CLOSER LOOK TO THE PHRYGIAN MODE

As said at the outset, we wanted to further investigate the Phrygian mode. Thus we conjectured that the effect of mode would depend on the particular style, too. Therefore, we set out to carry out further experiments to study the perceptual behavior elicited by the Phrygian mode. We posed ourselves several research questions: (1) Does the Phrygian mode induce the same perceived emotional responses irrespective of the musical tradition?; (2) What is

the perceived emotional response of children to the Phrygian mode?; (3) What is the perceived emotional response of children who are unfamiliar with the given music tradition?

4. EXPERIMENTS

4.1 Participants

124 children of 4-7 years of age participated in the experiment. They listened to pieces written in Phrygian and major mode. The major mode was introduced for the sake of variety of stimuli. Pieces in Phrygian mode were taken from flamenco and Galician musical traditions and pieces in major mode from the Castile tradition. Children were from Galicia and Castile and they were unfamiliar with flamenco music (in the sense that flamenco was not the musical tradition they grew up in).

4.2 Musical stimuli

Pieces were chosen after a careful selection; they should have similar melodic contour (to be sure that was not causing the emotional response). A piano timbre was used for the actual stimuli. The subjects were presented with pieces at three tempi, 72, 104, and 144 bpm. Depending on the age, the subjects listened to 12 pieces or 6 pieces on each session.

4.3 Procedure

As children sometimes have difficulties to verbalize their perceived emotions, we used drawings of faces so that they had only to point to them to give their responses. The Figure below shows an actual response sheet from the experiment. We tested four basic emotions: happiness, fear, sadness, and serenity (read the Figure counterclockwise starting on the first quadrant).

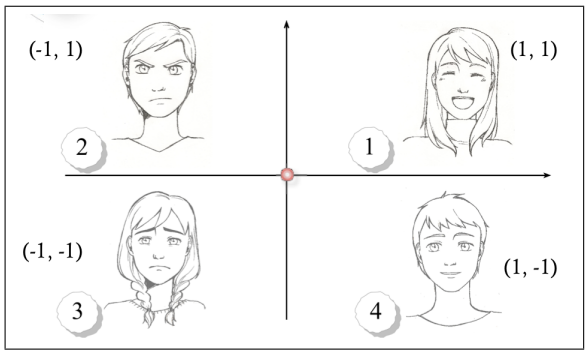


Figure 2: Drawing of faces used in the experiments.

Faces were pre-tested by children and teachers to see if they were adequate. The experiment was carried in groups of 1 to 5 children, depending on age. Also, outliers were identified (some children chose systematically faces by gender). Presentation was randomized by tempo, corpora and faces.

4.4 Results and discussion

In the Table below the emotions most frequently chosen by subjects according to the musical tradition and tempi are shown. For flamenco and Galician music we observe that sadness and fear, the emotions of negative valence, vary very little with tempi, whereas happiness and serenity do increase.

Corpus	Tempo	Happiness	Fear	Sadness	Serenity
Galicia	72	0,27	0,18	0,27	0,27
	104	0,31	0,13	0,25	0,31
	144	0,36	0,14	0,24	0,25
Flamenco	72	0,22	0,16	0,33	0,29
	104	0,24	0,09	0,33	0,34
	144	0,3	0,14	0,23	0,34
Castile	72	0,33	0,19	0,21	0,28
	104	0,42	0,1	0,16	0,33
	144	0,47	0,12	0,16	0,24

Table 1: Emotions most frequently chosen by subjects in different musical tradition and tempi.

As Ramos et al. did for their experiments, we also mapped the results onto the Russel's circumplex model of affect, which is formed by four quadrants; see Figure below. Each emotion was assigned a pair of values valence-arousal. Thus, happiness is $(1, 1)$, fear $(-1, 1)$, sadness $(-1, -1)$, and serenity $(1, -1)$; see Figure 2. Notice that, irrespective of the musical tradition, all the emotional responses have positive valence (serenity or happiness) in constrast to the results of Ramos et al.

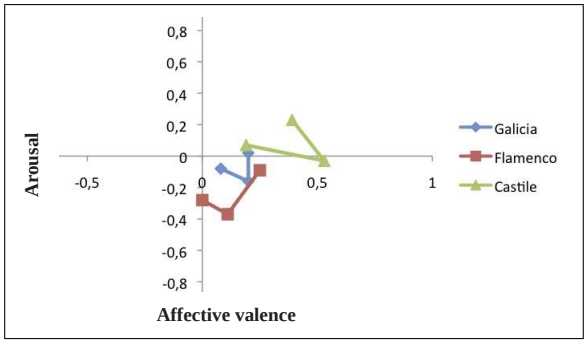


Figure 3: Arousal and valence of emotional responses as function of tempo.

5. CONCLUSIONS

Our results suggest that the Phrygian mode is not necessarily associated to emotions of negative valence. According to our findings, the Phrygian mode seems to be related to serenity. Although it is often the case, tempo is not always associated with large increase of arousal; we run statistical tests (not shown in this abstract because lack of space) to support that fact and they were positive. Arousal increased as tempi increased. It seems that the particular style is important to the emotional response. The musical characteristics of Galician and flamenco music are quite different. Mode modulates emotional response, but it does not determine emotional response on its own.

Our future work will consist of conducting the same experiments with children from Andalucia who are familiar with flamenco music and study how musical enculturation affects the emotional response. We also observed variations on the perceived emotional response across age, which certainly deserves to be further investigated. Also, we would like to repeat all the experiments with adults and compare the results to those obtained with children.

6. REFERENCES

Caetano, M., Mouchtaris, A., & Wiering, F. (2013). The role of time in music emotion recognition: Modeling musical emotions from time-varying music features. In M. Aramaki, M. Barthet, R. Kronland-Martinet, & S. Ystad (Eds.), *From Sounds to Music and Emotions*, volume 7900 of *Lecture Notes in Computer Science* (pp. 171–196). Springer Berlin Heidelberg.

Costa, M., Fine, P., Bitti, P., & Ricci, E. (2004). Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception*, *22*(1), 1–14.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *6*(80), B1–B10.

Juslin, P. & Sloboda, J. (2010). *Music and emotion: Theory and research*. Oxford: Oxford University Press.

Ramos, D., Bueno, J., & Bigand, E. (2011). Manipulating greek musical modes and tempo affects perceived musical emotion in musicians and non-musicians. *Brazilian Journal of Medical and Biological Research*, *44*(44), 165–172.

Sloboda, J. (1991). Music structure and emotional response: some empirical findings. *Psychol Mus.*, *19*, 110–120.

Tizon, M., Gomez, F., & Oramas, S. (2013). Perceived emotion in phrygian mode in musically trained children. In *Proceedings of the Third International Conference on Music and Emotion*, (pp. 14–15).

Trainor, L., McDonald, K., & C., A. (2002). Automatic and controlled processing of melodic contour and interval information measured by electric brain activity. *J. Cogn. Neuroscience.*, *14*(14), 430–442.

Webster, G. & Weir, C. (2005). Emotional response to music: interactive effects of mode, texture and tempo. *Motiv Emot.*, (29), 19–39.

COMPARATIVE STUDY ON THE TIMBRE OF WESTERN AND AFRICAN PLUCKED STRING INSTRUMENTS

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

1. EXTENDED ABSTRACT

Two main approaches of instrument classification methods have been developed so far. On one side, the oldest one is the classical Organology (Tranchefort, 1980), coming from physics and musicology, which aims to define the main classes of instruments based on their sound production modes. Among these classes, chordophones gather any musical instrument that makes sound by way of vibrating strings stretched between two points. It is one of the four main divisions of instruments in the original Hornbostel-Sachs scheme of musical instrument classification. On the other side, computing sciences have developed the field of Machine learning dedicated to music, called Music Information Retrieval (MIR) (Klapuri, 2004; Herrera-Boyer et al., 2006). MIR is the interdisciplinary science (bridging musicology, signal processing, psychology) which aims to retrieve any meaningful information from music, which can be whether learned first (i.e. supervised classification), or discovered without any *a priori* knowledge (i.e. unsupervised classification). Within these two approaches of instrument classification, the timbral complexity and diversity of ethnic music raises major problems, making the definition of robust note templates rather difficult (Cornelis et al., 2010; Lidy et al., 2010), and consequently the definition of instrument classes and their automatic recognition.

In this study, we performed an acoustic characterization of the timbre of six plucked string instruments of Africa (e.g. Mvet from Cameroun, Marovany and Valiha from Madagascar), with the goal of quantifying the timbral complexity within this instrument class. As a first step, a complete set of acoustic descriptors is used to project the timbral signature of each instrument in a multidimensional space integrating many physical components of the timbre (e.g. temporal profile, spectral content). This signature is studied on the whole pitch range of each instrument. Then, methods to reduce the dimensionality of this representation space have been used to conserve only dimensions optimizing certain criteria, such as the inter-class discrimination or the descriptor redundancy. As a second step of analysis, unsupervised analysis of data visualisation/structuration (e.g. *clustering*) have been used to quantify the dispersion of the acoustic timbre. A complexity measure of the timbre has been derived from this dispersion. This study presents inter-note / inter-instrument / inter-cultural (ethnic / Western, where the same methods are applied to six different western plucked string instruments) comparative results. Discussions are eventually proposed on the thematics of automatic instrument classification (in particular on the appropriateness of current methods developed for Western instruments to classify ethnic instruments) and music transcription (in particular on the design of specific note templates and dictionaries able to integrate the acoustic variability between instruments) when considering ethnic music.

2. REFERENCES

Cornelis, O., Lesaffre, M., Moelants, D., & Leman, M. (2010). Access to ethnic music: Advances and perspectives in content-based music information retrieval. *Signal Processing, 90*, 1008–1031.

Herrera-Boyer, P., Klapuri, A., & Davy, M. (2006). *Signal Processing Methods for Music Transcription*, chapter Automatic classification of pitched musical instrument sounds, (pp. 163–200). Springer US, editors Klapuri, A. and Davy, M.

Klapuri, A. (2004). Automatic music transcription as we know it today. *Journal of New Music Research, 33*, 269–282.

Lidy, T., Silla, C. N., Cornelis, O., Gouyon, F., Rauber, A., Kaestner, C. A. A., & Koerich, A. L. (2010). On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-western and ethnic music collections. *Signal Processing, 90*, 1032–1048.

Tranchefort, F.-R. (1980). *Les Instruments de musique dans le monde*. Seuil.

ON THE USE OF SCATTERING COEFFICIENTS IN MUSIC INFORMATION RETRIEVAL. APPLICATIONS TO INSTRUMENT RECOGNITION AND ONSET DETECTION ON THE *MAROVANY* REPERTOIRE

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

Olivier Adam
Université UPMC Paris 6
Equipe Lutheries Acoustique Musicale (LAM)
olivier.adam@upmc.fr

1. EXTENDED ABSTRACT

Within our MIR project on the *Marovany* zither repertoire (Chemillier, 2013), we illustrate here the usefulness of multiscale scattering in Music Information Retrieval (MIR) tasks. As most notes from plucked string instruments, a note of zither is composed of two main phases: note attack (i.e. transients) and note decay phases (which may include intermodulation, corresponding to sympathetic resonances between adjacent strings). This repertoire offers an interesting challenge to the task AMT, with difficulties including note overlap, intermodulation, great dynamics, great speed of playing, polyphonic characteristics (due more to a slow note decaying than to a specific vertical writing), the presence of external musical (e.g. a rattle called *kantsa*, hand-clapping and vocal interactions) and noisy (e.g. ambient noise from the audience) sources.

This paper compares scattering representation with other representations of the signal, namely Fourier transform, MFSC, Wavelets. To prove that a certain representation does provide benefits as a replacement over usual representations, then a natural and scientifically sound approach is to provide a comparative study with a reference dataset and state-of-the-art algorithms, where we explicitly show the differences in performances by replacing successively one transform with another. Two crucial tasks in MIR will be evaluated in this work : onset estimation and instrument recognition. For the onset estimation task, the spectral flux (Bello et al., 2004) will be computed, as characterizing distance measures between consecutive frames of a multi-dimensional representation of a signal (classically, a spectrogram). We will also focus our interests on instrument recognition, using simple classifiers to concentrate on the properties of feature vectors as opposed to a specific classifier.

The scattering coefficients have already been applied to various information retrieval tasks, including musical instrument classification (Anden & Mallat, 2011), note characterization (Anden & Mallat, 2012) and texture classification (Bruna & Mallat, 2013). They have proven useful in many audio classifiers, which can be partly attributed to their stability to deformation (Anden & Mallat, 2012). Based on their results, second-order Cosine Log Scattering (CLS) vectors were shown to achieve significantly higher accuracy than other classical representations (e.g. MFCCs or Delta-MFCCs) since they recover lost non-stationary structure of the signal and provide richer representations. Anden & Mallat (2012) also stated that : "The ability to characterize non-stationary signal structures opens the possibility to capture more sophisticated auditory phenomena such as transients, amplitude and frequency modulations, time-varying filters, chord structure and rhythms with co-occurrence scattering coefficients".

This scattering representation was developed by Mallat (2012) around the notion of invariability, crucial to define robust instrument-specific templates used in supervised classification. To explain the concept of stability, let's consider a feature representation ϕ, mapping an audio signal x to ϕx. ϕ is stable to deformation if the Euclidean distance $||\phi x - \phi \tilde{x}||$, where $\tilde{x}$(t) is a deformed signal defined as x(t -τ(t)), is small for a time-warping function τ small. It is interesting to underline the motivations which originate this method, conceived as an extension of Mel-Frequency Spectral Coefficients (MFSCs). High-frequencies are more sensitive to deformation than low-frequencies, which makes the Fourier-based spectrogram particularly non-adapted to take into account small deformations of a signal. The logarithmic averaging used in a mel-scale removes this instability, providing the MFSCs with a non-variable representation of a signal from one observation to another. However, this averaging naturally loses high-frequency information. Scattering coefficients have then been conceived through a cascade of wavelet decompositions and modulus operators with the goal of recovering the information lost in MFSCs while remaining stable. The acoustic richness of scattering representation emerges from this construction.

To test this method, we chose to use a reduce database with a precise onset annotation, performed robustly using a multichannel sensory system to retrieve musical notes (Cazau et al., 2013). At the opposite to past studies (Bello et al., 2005), which used extensive databases with hand labeling, which is prone to errors. For the task of instrument classification, we present some results on several African zithers of Madagascar and Cameroun, which also offers a very challenging task. From these two studies, scattering coefficients are indeed shown to bring better results, especially in the detection/characterization of musical onsets,

which are explained through the richer acoustic information contained in scattering representation.

2. REFERENCES

Anden, J. & Mallat, S. (2011). Multiscale scattering for audio classification. In *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*.

Anden, J. & Mallat, S. (2012). Scattering representation of modulated sounds. In *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12), York, UK , September 17-21*.

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, *13*, 1035–1047.

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.

Bruna, J. & Mallat, S. (2013). Invariant scattering convolution networks. *IEEE Transactions on Pattern analysis and Machine intelligence*, *35*, 1872–1886.

Cazau, D., Chemillier, M., & Adam, O. (2013). Information retrieval of marovany zither music with an original optical-based system. In *DAFx 2013, Maynooth, Ireland, 2-6 september*.

Chemillier, M. (2013). Web page of our computational ethnomusicology project on marovany zither of madagascar. In `http://ehess.modelisationsavoirs.fr/marovany/index.html`.

Mallat, S. (2012). Communications on pure and applied mathematics. *Wiley Periodicals, Inc.*, *LXV*, 13311398.

A COMPUTATIONAL ETHNOMUSICOLOGY STUDY OF CONTRAMETRICITY IN THE TRADITIONAL MUSICAL REPERTOIRE 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

1. EXTENDED ABSTRACT

The meeting between Music Information Retrieval and Ethnomusicology has given way to a new scientific community called Computational Ethnomusicology (Tzanetakis et al., 2007), which aims to adapt MIR tools and develop specific ones to the corpus of ethnic music. In particular, rhythmical characteristics in music have received a great number of studies (especially on the tasks of pulsation detection, identification of metrics). In this project, we will consider a corpus of African music, and focus on a specific common rhythmical characteristic designated as contrametricity, using a computational ethnomusicology approach. (Arom, 1983, p. 339) defines contrametricity as a rhythm in which onset locations and accents are in conflict with pulsation. For example, in jazz music, hand clapping located on weak beats are contrametric in regards to strong beats of the same measure. Chemillier et al. (2013) studied the contrametricity in different musical repertoires of Africa, including the music of Madagascar.

For this project, our musical dataset consists of traditional tunes played by the *Marovany*, a tall zither in the form of a rectangular box built from recycled wood products, which also takes part in a possession cult called *tromba* in southern Madagascar. A first originality in this work is the use of a multichannel sensory system to constitute our sound database. This technology allows a multi-channel acquisition of the *Marovany* music, with independent signals corresponding to the played strings. Such analytical recordings make the task of note segmentation and labelling much easier (Cazau et al., 2013). A complete set of traditional tunes of *Marovany* has been collected with this technology in Madagascar in last June [1] .

As a preliminary step in this study, we have first performed a qualitative musical analysis aiming to characterize and illustrate the contrametricity in the *Marovany* repertoire. Next, in the framework of Computational Ethnomusicology, we designed a complete set of rhythmical descriptors, including some from the tasks of describing rhythm complexity and syncopation, in order to retrieve automatically the contrametricity in musical excerpts. We will distinguish in particular two classes of contrametricity measures. A first one based on the hierarchical metrics of Lerdahl & Jackendoff (1983) (e.g. LHL by Longuet-Higgins & Lee (1984)). A second one based on the sole pulsation (e.g. WNBD by Gomez et al. (2005), and two original ones based on histograms of time deviations from the pulsation). These two classes of descriptors can be opposed fundamentally considering the perceptual and musicological concepts they follow. Indeed, metrical structure (with its subdivisions of the pulsation and groups of notes) has been subject to strong divergences between researchers for more than thirty years. On one side, Lerdahl & Jackendoff (1983) stated that, cognitively, some universal rules exist, which organize the perception of metrics as a superposition of hierarchical levels. More precisely, pulsations are perceived in groups of two or three, among them one is affected with a particular weight (depending on musical factors such as amplitude or duration) which allows its identification as a strong beat. However, on the other side, ethnomusicologists criticized this principle of alternance between strong and weak beats instaured in Lerdahl & Jackendoff (1983)'s theory, especially when considering African music (Arom, 1983), arguing particularly that African music do not possess such alternating beats.

Within this context, this computational ethnomusicological study aimed at :

1. selecting optimal rhythmical descriptors in the characterization of the contrametricity in the *Marovany* repertoire, with labelled musical sequences took as references. Following this study, a general discussion will be held on the appropriateness of Lerdahl & Jackendoff (1983)'s theory to study contrametricity in African music ;

2. identifying acoustic parameters of influence (e.g. duration, amplitude, onset locations of notes) and quantifying their respective contributions to the contrametric characteristic. Such sensitivity study has helped in the design of more specific descriptors which integer the proper representations and parameters to retrieve the contrametricity ;

3. performing a computational study to make this analysis automatic and objective on a large database, and observe statistic tendencies of this characteristic. Also, we will address the topic of classification of tradi-

[1] Recordings of other plucked-string traditional instruments are ongoing. This dataset will be made available to the community of Computational Ethnomusicology soon.

tional African tunes, and comment on the ability of the contrametricity to discriminate between these different tunes, and also between country-specific tunes.

2. REFERENCES

Arom, S. (1985). *Polyphonies et Polyrythmies Instrumentales d'Afrique Centrale: Structure et Méthodologie,* Paris, SELAF

Cazau, D., Chemillier, M. & Adam, O. (2013). "Information retrieval of marovany zither music with an original optical-based system" In *Proceedings of DAFx 2013, Maynooth, Ireland, 2-6 september*.

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

Gomez, E., Melvin, A., Rappaport, D. & Toussaint, G. (2005). "Mathematical Measures of Syncopation" In *Proceedings BRIDGES: Mathematical Connections in Art, Music and Science, p. 73-84*.

Lerdahl, F. & Jackendoff, G.(1983). *A generative theory of tonal music,* MIT Press

Longuet-Higgins, H. C. & Lee, C. S. (1984) "The Rhythmic Interpretation of Monophonic Music" *Music Perception: An Interdisciplinary Journal, 1,* 424-441.

Tzanetakis, G., Kapur, A., Schloss, W. A. & Wright, M. (2007) "Computatioal Ethnomusicology" *Journal of interdisciplinary music studies, 1,* 1-24.

COMPUTATIONAL ETHNOMUSICOLOGY: A STUDY ON FLAMENCO AND ARAB-ANDALUSIAN VOCAL MUSIC

Nadine Kroher, Emilia Gómez, Mohamed Sordo
Music Technology Group, Universitat Pompeu Fabra
{name}.{surname}@upf.edu

Francisco Gómez-Martín
Universidad Politécnica de Madrid
fmartin@eui.upm.es

Jose-Miguel Díaz-Báñez, Joaquín Mora
Universidad de Sevilla
jbanez@us.es jmora@us.es

Amin Chaachoo
Asmir Center Tetouan
chaachooamin@gmail.com

INTRODUCTION

This interdisciplinary case study focuses on two well-established music traditions, flamenco and Arab-Andalusian music with special focus on the melodic aspects of the singing voice. We apply a hybrid methodology combining a traditional musicological analysis with music information retrieval techniques. Focusing on two representative pieces, a flamenco *martinete* and an Arab-Andalusian *inshad,* we illustrate commonalities and divergences among these traditions, and evidence on how existing technologies allows us to formalize expert's knowledge and complement traditional analysis methodologies by discovering relationships that might otherwise have been unnoticed.

1. BACKGROUND

1.1 Flamenco

Flamenco is an oral music tradition with roots as diverse as the cultural influences of its area of origin, Andalusia, a region in southern Spain. Over the centuries, the area and, of course, its music have been influenced by the ancient Tartessian culture as well as Phoenician and Roman colonizations, notwithstanding later settlements by Visigoths, Arabs, Jews, Christians, and to a large extent gipsies, who decisively contributed to shape its form as we know it today (Blas-Vega & Ríos-Ruiz, 1988; Navarro & Ropero, 1995).

Flamenco music germinated and nourished mainly from the singing tradition (Gamboa, 2005). Accordingly, the role of the singer soon became dominant and fundamental. The flamenco singing voice can be characterized as unstable in pitch, timbre and dynamics (Merchán Higuera, 2008). Melodic movements are composed of conjunct degrees and performances include spontaneous complex, microtonal ornamentations and melisma (Mora et al., 2010).

1.2 Arab-andalusian music

The Arab-Andalusian music tradition can be traced back to the 12th century in Al-Andalus, to the muslims and-christians living in the Moorish Spain (Chaachoo, 2011) and is the result of many influences, including Middle-East Arabic classical music, Hispanic music traditions of the Iberian peninsula, and other classical traditions such as the Gregorian and Byzantine ones. The Andalusian tradition is maintained in quite a few north african regions (Poché, 1995; Guettat, 2000) mainly in Morocco, Algeria, and Tunisia.

Andalusian music is organized around the concept of *nawba* (Poché, 1995; Guettat, 2000), a collection of melodies belonging to the same melodic mode, which defines not only the pitch content but is also linked to specific emotions or states of mind and is consequently associated with certain social occasions. Furthermore a mode is defined by its pitch range (tessitura), a tonal center around which the melody gravitates, important scale degrees which provide the mode's characteristic flavor and a set of melodic cells, known as *centones*. The singing voice being the key element, Andalusian music is characterized by the use of sung poems (*san'as*) (Cortés García, 2003), taken from Arabic classical poetry. As in the case of flamenco, Andalusian music tradition has been preserved and kept alive as an oral tradition.

2. MUSICOLOGICAL ANALYSIS

1.3 2.1. The flamenco *martinete*

The *martinete* is a traditional monophonic flamenco singing style characterized by a common melodic skeleton, slow tempo, solemn performance, free rhythmic interpretation and a large amount of melismatic ornamentation. In the present study we provide a detailed analysis of a representative *martinete* performance by the renowned singer *Tomás Pavón*, treating musical form, characteristic melodic movements, underlying tonality and the use of ornamentation. We furthermore introduce different concepts of symbolic representation and analysis schemes.

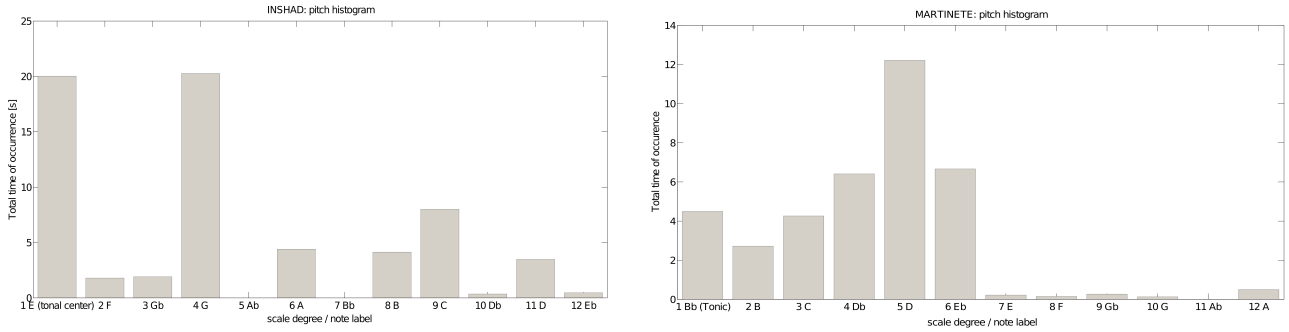


Figure 1. Pitch histograms of the *inshád* (left) and the *martinete* (right) performance.

2.2 The Arab-andalusian *inshád*

The *inshád* is a sung poem of two verses. Forming part of the Arab-Andalusian musical heritage, the melodic progression is based on set motives which are spontaneously ornamented (Chaachoo, 2011). We analyze a performance by singer *Zohra Abbetiw* in the *Al-Sika* mode. After giving an overview of the general characterstics of the mode, we discuss rhythmic interpretation, the tonal importance of the scale degrees, formal aspects and the use of ornamentation in the performance under study.

3. COMPUTATIONAL ANALYSIS

Manual analysis is complemented by the extraction and interpretation of melody-related features on various abstraction levels. We extract automatically-computed fundamental frequency and energy envelopes and analyze their fine-structure regarding the use of ornamentation, frequency and volume vibrato as well as melisma. In a next step, an automatic transcription algorithm quantizes the fundamental frequency into single note values.

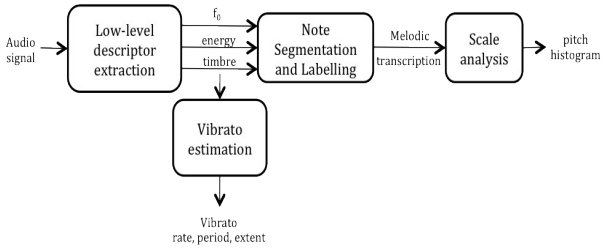


Figure 2. Computational audio analysis scheme

The so obtained symbolic representation is the basis for the computation of pitch histograms, which display the temporal of occurrence of each pitch class wrapped into a single octave. We analyze the pitch histograms regarding the relation between temporal occurrence and tonal importance of the different scale degrees. Based on both, automatic transcriptions and fundamental frequency envelopes, we compute statistical performance descriptors related to pitch content, vibrato and note onsets and durations.

4. RESULTS AND CONCLUSION

Analyzing the pitch histograms, we illustrated that the tonal importance of the tonic in the *martinete* does not correspond to an elevated temporal occurrence, since long melismatic ornamentation and insistence on recitative note are usually limited to scale degrees other than the tonic. The histogram furthermore shows a concentration of the pitch content in the first six scale degrees and a comparatively high temporal occurrence of pitch classes not contained in the underlying scale. In contrast, the temporal pitch distribution of the notes of the *inshád* does reflect their corresponding tonal importance, is spread more equally over all pitch classes and is mainly limited to the notes contained in the *Al-Sika* mode. Analyzing fast pitch fluctuations we furthermore observe a simultaneous overlay of melisma and vibrato in the *martinete* performance.

In this case study, we have presented a methodology for the analysis of music recordings that combines manual and automatic descriptions of music recordings by using

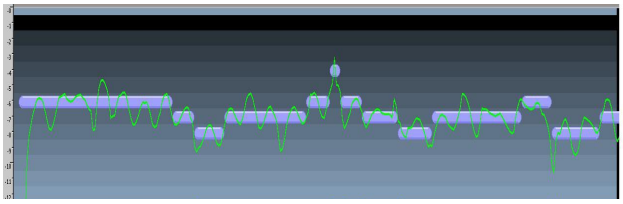


Figure 3. Automatic transcription and fundamental frequency envelope of a *martinete* excerpt.

state-of-the-art techniques. We detected similarities and differences on the two analyzed pieces, *martinete* and *inshád*, in terms of tonality, ornamentation, melodic line, scale, and vibrato. Although these properties refer to specific pieces, some of the traits are representative of flamenco and Arab-Andalusian music traditions. In the future, we intend to carry out a more extensive analysis that will include many pieces of each music tradition. Our methodology can be applied to other pieces and traditions, but proper adaptation should be carried out.

REFERENCES

Blas Vega, J. & Ríos Ruiz, M. (1988). Diccionario enciclopédico ilustrado del flamenco. Madrid: Cinterco.

Chaachoo, A. (2011): La Música Andalusí. Historia, conceptos y teoría musical. Córdoba: Almuzara.

Cortés García, M. (2003): Kinnas al-Haik. Centro de documentación musical de Andalucía.

Gamboa, J. M. (2005): Una historia del flamenco. Madrid: Espasa-Calpe.

Guettat, M. (2000): La musique arabo-andalouse. L'empreinte du Maghreb. El Ouns / Fleurs sociales.

Merchán Higuera, F. (2008): Expressive characterization of flamenco singing. *Master thesis*. Universitat Pompeu Fabra, Barcelona.

Mora, J., Gòmez, F., Gómez, E., Escobar-Borrego, J. & Díaz-Bañez, J. M. (2010): Characterization of Melodic Similarity of A Cappella Flamenco Cantes. In *Proceedings of the 11th International Society for Music Information Retrieval (ISMIR)*.

Navarro, J. L. & Ropero, M (1995): Historia del flamenco. Sevilla: Tartessos.

Poché, C. (1995): La musique arabo-andalouse. Paris: Cité de la Musique / Actes Sud.

CROSS-CULTURAL COMPARISONS OF EXPRESSIVITY IN RECORDED ERHU AND VIOLIN MUSIC: PERFORMER VIBRATO STYLES

Luwei Yang, Elaine Chew
Centre for Digital Music
Queen Mary University of London
{l.yang,elaine.chew}@qmul.ac.uk

Khalid Z. Rajab
Antennas & Electromagnetics Group
Queen Mary University of London
k.rajab@qmul.ac.uk

1. INTRODUCTION

Many traditional Chinese pieces have been transcribed for, and are frequently performed on, Western Classical instruments. Our study focuses on a cross-cultural comparison of erhu music performed on the violin and on the original instrument. Performances of erhu music on violin differ qualitatively, both in the pacing and sonic shaping of the music, from that on the original instrument. It is our goal to quantify this difference so as to create better models and representations for folk music analysis.

The use of vibrato in modern erhu playing can be traced back to the violin. Western Classical music played an important role in the development of erhu playing from the beginning of the twentieth century. Liu Tianhua (刘天华) (1895-1932), a Chinese musician, erhu teacher and composer, is especially noted for his innovative work on erhu. He studied violin, piano and Western music composition theory in his younger years, and adopted the violin vibrato and trills for erhu playing. He also introduced tremolo, spiccato, and pizzicato into erhu playing (Wang, 2002). These diverse playing styles gave a new life to erhu, and made it stand out in Chinese music. These techniques are now widespread and used extensively in erhu playing.

Our study of an erhu piece played on both erhu and violin seeks to measure playing styles when the violin may be emulating the erhu. In fact, the influence goes full circle and can be traced back to the violin because the vibrato techniques on erhu originated in erhu players borrowing from violinists. Nevertheless, the quantitative study of vibrato in these two instruments reveals interesting differences.

1.1 Motivation

Scientific study of expressivity in performances of Western Classical music has been a subject of study since the beginning of the twentieth century (Seashore, 1932, 1938). While music technology research focused on Chinese music has increased in recent years—for example, Yang & Hu (2012)'s work on automatic classification of Chinese and Western music instruments and Tian et al. (2013)'s work on emotion categorisation of singing in Chinese songs. However, studies on expressive features of non-Western music has received comparatively less attention, and the field is wide open for exploration. Ozaslan et al. (2012) has showed a pioneering coss-cultural comparative analysis between Turkey Mamkan music and Western Classical music.

As is true for transcription of music of diverse cultures (Ellingson, 1992), expressivity in the performance of Chinese music is poorly captured by Western common music notation (CMN). Chinese music was traditionally notated using *Gongche* (工尺谱) notation, which comprises of Chinese characters representing notes, and sparse rhythm signs to the right of these characters. Modern Chinese music notation borrows from CMN, and uses primarily numbers representing scale degrees, augmented by dots above or below the symbol denoting register, and lines and dots to represent note durations similar to that in CMN.

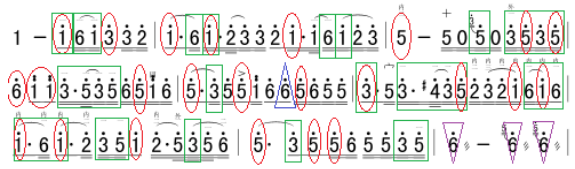


Figure 1: Modern Chinese music notation for first nine bars of *The Moon Reflected on the Second Spring*.

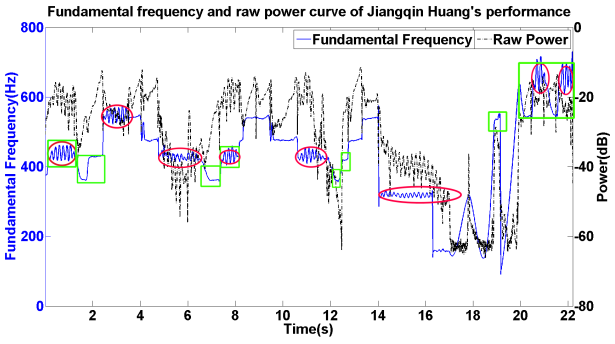


Figure 2: Fundamental frequency and raw power curve of Jiangqin Huang's performance of *The Moon Reflected on the Second Spring*.

Figure 1 shows the first nine bars of *The Moon Reflected on the Second Spring*(《二泉映月》), considered to be a traditional Chinese piece for erhu, composed by Abing (aka Hua Yanjun 华彦钧) (1893-1950). Figure 2 shows the fundamental frequency and the raw power curve of Jiangqin Huang's recorded performance of the first 3 bars of Figure 1. In both Figures 1 and 2, red ellipses mark vi-

bratos, green boxes indicate portamentos, blue upright triangles mark notes elaborated with trills, and purple upside down triangles indicate tremolo notes. As can be seen by the dense markings, very little of these common expressive devices are indicated in the Modern Chinese music notation. But these expressive devices clearly stand out in the fundamental frequency and raw power curve.

1.2 Aim

The piece has been transcribed for violin and piano and is frequently performed throughout East Asia. The violin transcription closely mirrors the notes and structure of the original erhu composition. When performed on the violin, the music sounds different even when some violinists attempt to emulate idiomatic erhu expressive gestures. Our ultimate goals are: (1) to quantify the differences between erhu and violin playing of the same music; and, (2) to determine the expressive devices employed by erhu players.

We began our investigation into the differences between erhu and violin playing by first considering vibrato. In (Yang et al., 2013), we presented summary statistics for vibrato in erhu and violin playing focussing on the instrument. We found that the physical form of the instrument and how it is played may be the most dominant factors affecting the differences in vibrato style between erhu and violin playing.

In the present study, we delve deeper into the vibrato differences between individual erhu and violin players. The methodology also presents a way to perform cross-cultural vibrato analysis especially for world folk music.

2. METHODOLOGY

To compare the vibrato styles, we used the vibrato rate, extent, and sinusoid similarity as parameters. Details of how we extracted the vibratos and obtained these quantities are outlined below and can be found in (Yang et al., 2013).

2.1 Data

We collected twelve performances of *The Mood Reflected on the Second Spring*—six on erhu and six on violin—as shown in Table 1.

Erhu			Violin		
#	Performer	Nationality	#	Performer	Nationality
1	Guotong Wang	China	7	Laurel Pardue	U.S.A
2	Jiangqin Huang	China	8	Lina Yu	China
3	Wei Zhou	China	9	Baodi Tang	China
4	Jiemin Yan	China	10	Nishizaki Takako	Japan
5	Huifen Min	China	11	Yanling Zhang	China
6	Changyao Zhu	China	12	Yangkai Ou	China

Table 1: Selected performances

2.2 Vibrato rate and extent

The vibrato rate and extent are both calculated from he peaks and troughs of the vibrato fundamental frequency. We assume the interval between one peak and one trough is the half cycle of the vibrato period. The averaged inverse of the intervals for all half cycles results in the vibrato rate. Similarly, the extent for one half cycle is half the difference between the peak and the trough. The averaged extents for all half cycles is the vibrato extent.

2.3 Vibrato Sinusoid similarity

We use the normalised cross-correlation of the shape of f0 and the relevant sinusoid having the same frequency to determine the vibrato's similarity to a sinusoid. The procedures are:

1. Convert the f0 to a MIDI scale.
2. Apply smoothing to obtain the average f0.
3. Subtract the average f0 from the MIDI scale f0 to block the DC component.
4. Compute the FFT of the 0-centred f0.
5. Pick the peak from the spectrum to get the vibrato frequency.
6. Use this vibrato frequency to create a sine wave with amplitude 1.
7. Calculate the normalized cross-correlation between the 0-centred f0 and the sine wave.
8. Set the vibrato sinusoid similarity to the maximum of the normalized cross-correlation results.

3. RESULTS

Figure 3 shows the vibrato rate for each player. In general, violinists tended to apply faster vibratos. However, the Player 11 (a violinist) had vibrato rates similar to erhu players, and Player 12 (another violinist) had vibrato rates lower than all erhu performers. Thus, the violin vibrato rates varied more widely.

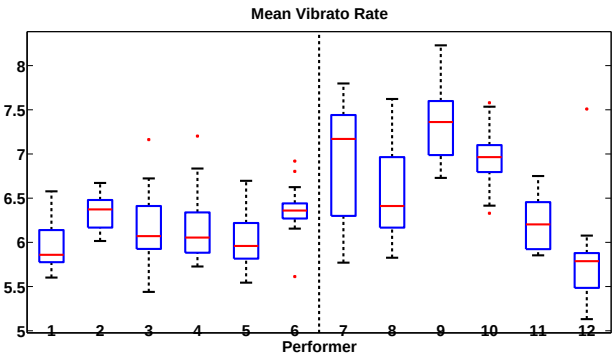


Figure 3: Box plots of vibrato rates for all performers.

Figure 4 presents the vibrato extent for each player. The results show that erhu performers' vibratos had much greater extents than that of violinists. The standard deviation of the erhu vibratos were also markedly larger than that for violin. The 2nd erhu player showed the largest vibrato extent, almost 1 semitone, which is twice that of the 3rd erhu player. In contrast, all violinists maintained relatively small vibrato extents, with low standard deviations.

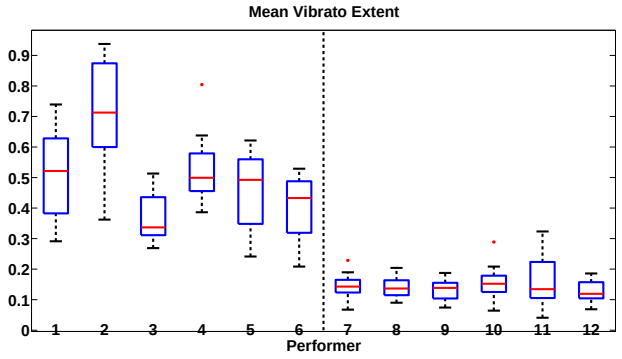


Figure 4: Box plots of vibrato extents for all performers.

The vibrato sinusoid similarity for all performers is represented by Figure 5. All erhu vibratos have greater sinusoid similarity than that of the violins. Player 7 (the US violinist) showed the widest sinusoid similarity range, and lowest sinusoid similarity values.

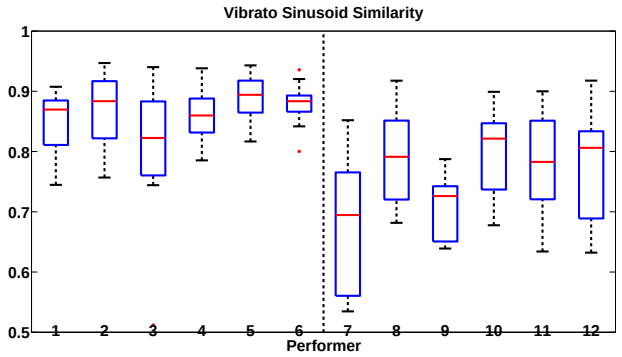


Figure 5: Box plots of vibrato sinusoid similarity values for all performers.

4. CONCLUSIONS

In general, violin performers had marginally higher vibrato rates, and varied the vibrato rate more widely than erhu performers, but the differences are not significant. All erhu performers had significant larger vibrato extents than violin performers. Furthermore, the erhu players also varied the vibrato extents more widely than violin players. The shape of the erhu vibrato samples was closer to that of a sinusoid than the violin samples.

Thus, even though the erhu borrowed vibrato techniques from violin, and the violin may be emulating the erhu in playing an erhu piece, the differences between their vibrato styles can still be identified quantitatively; this is especially true for the vibrato extents.

To determine if the only factor leading to this difference in vibrato styles is the instrument, the ideal experiment would analyze the vibratos of the same performer playing both the erhu and violin. This represents future work as and when such a player can be found. A carefully designed experiment will be required to obtain a systematic analysis of vibrato performance style that eliminates the effect of habit or practice.

5. ACKNOWLEDGEMENTS

This research was funded in part by the China Scholarship Council.

6. REFERENCES

Ellingson, T. (1992). *Helen Meyers (ed.): Ethnomusicology: an Introduction*, chapter Transcription, (pp. 110–152). New York: W.W. Norton.

Ozaslan, T. H., Serra, X., & Arcos, J. L. (2012). Characterization of embellishments in ney performances of makam music in turkey. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*.

Seashore, C. E. (1932). *University of Iowa Studies in the Psychology of Music (Vol. 1: The Vibrato)*. Iowa City, Iowa: The University Press.

Seashore, C. E. (1938). *Psychology of music*. New York: Dover Publications.

Tian, M., Black, D. A., Fazekas, G., & Sandler, M. (2013). An content-based emotion categorisation analysis of chinese cultural revolution songs. In *Proceedings of the Third International Workshop on Folk Music Analysis (FMA)*.

Wang, Y. (2002). *Zhongguo Jinxinadai Yinyueshi(The Chinese contemporary Music History)*. People's Music Publishing House. in Chinese.

Yang, L., Chew, E., & Rajab, K. Z. (2013). Vibrato performance style: A case study comparing erhu and violin. In *Proc. of the 10th International Conference on Computer Music Multidisciplinary Research(CMMR)*.

Yang, Y.-H. & Hu, X. (2012). Cross-cultural music mood classification: A comparison of english and chinese songs. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*.

EXPLORING PHRASE FORM STRUCTURES. PART I: EUROPEAN FOLK SONGS.

Klaus Frieler
Liszt School of Music Weimar
{klaus.frieler}@hfm-weimar.de

ABSTRACT

In this explorative study, we investigate the phrase form structures of 7821 folk songs taken from the EsAC database. The main purpose was to test whether phrase form structure is a useful feature for computational folk song analysis, e. .g, classification tasks. To this end, we examined the self-similarity of phrase sequences in folk songs with regard to semitone intervals and duration classes as well as in combination. Phrase form structure can be characterised by coherence values measuring the amount of repetition. Duration-based form structures show a coherence about twice as high than interval-based form structures, i. e., rhythmic models are much more likely to be repeated than interval sequences. The most common interval-based form is ABCD, whereas the the most frequent duration-based form model is AAAA. Folk songs show a strong preference for even-numbered form lengths, with 4, 6, and 8 phrases being the most common. In comparing sub-collections, significant differences between folk songs of different origin were found, particularly, between Irish and Polish tunes on one hand, and Central European songs on the other hand. Interestingly, the sub-collection of German children songs showed the most diverse set of different form structures, whereas the *Kolberg* collection of Polish song was the most homogeneous set. All in all, this shows that phrase form structure might indeed be a useful starting point to derive features for other computational applications.

1. INTRODUCTION

Exact and varied repetitions of musical units constitute musical form, which can be viewed as self-similarity within a piece of music. Since musical units are often hierarchically structured, musical form can be found on different hierarchical levels as well. One reasonable level to consider is that of the musical phrase, which can roughly defined as coherent Gestalt-like units. Hence, the investigation of phrase form structure (PFS) (also called phrase relationships, c.f. Sagrillo, 1999) of melodies is a good starting point for examing musical structure. This is the topic of the current paper, more specifically, the phrasal structures in a certain set of Central, Eastern, and Western Europe folk songs as contained in the well-known EsAC database (Schaffrath, 1995; Dahlig, 2000). The EsAC database is unique compared to most other folk song collections, since it provides phrase information. The annotation was mostly done by the transcribers, either of the original sources or the coders of the source, and are hence not fully objective. Furthermore, experiments in segment perception (e.g. Pearce et al., 2008; Spevak et al., 2002) showed that perceived phrase boundaries are subject to individual variation—a fact already well-known to folk song researchers. However, in the case of the EsAC database one can safely assume that the transcribers were highly trained experts who chose the most likely and reasonable segmentation, particularly in the presence of lyrics (see Sagrillo, 1999 for a discussion, also Spevak et al., 2002).

The next higher level of form of parts such as verse, refrain, bridges etc, is beyond scope of the present paper, since the EsAC creators only coded one verse and one refrain for a songs if more than one were present. Unfortunately, they did not indicate it anywhere how many verses there were and which part might be considered a verse or a refrain. Hence, it is hardly possible to discern larger form structures (e. g. such as proposed in Lomax & Grauer, 1967) from the current data.

The next lower level, that of motifs would also be a good candidate, but the distinction between phrases and motifs is often hard to draw. Sub-phrases are most likely to be perceived if some smaller unit, i. e., a motif, is repeated within the same phrase or in another phrase of the song. Hence, the existence of motifs is already tightly connected to similarity, and thus a topic for pattern mining. In contrast, phrases boundaries are often determined by longer (breathing) rests and cognitive load (Temperley, 2001), as well as – particularly in the case of folk songs – by semantical and grammatical language units in the lyrics.

Sagrillo (1999) carried out a very similar investigation of phrase form structure for the *Luxemburg* collection, using the old EsAC analysis software ANA, which does not easily run anymore on modern computers. The software provided pitch-based and rhythmic form strings in the same manner as used here, but it is not known on which algorithm(s) the automated detection was based. Moreover, Sagrillo was considering variations and sequentiation of phrases, which is not done in this study, but nevertheless some of his results could be roughly reproduced as a subset of our results.

2. METHOD

The analysis was carried out using the version of the EsAC database included in the MeloSpySuite [1] (Frieler et al., 2013), which contains currently 7821 folk songs in 13 collections. To extract the form information, similarity values between each phrase of a song were calculated using edit distance (Levenshtein, 1965) on a interval-based or duration-class-based representation of the melody. Intervals are signed semitone differences, and durations classes range from "very short" to "very long" with the beat level

[1] Available at http://jazzomat.hfm-weimar.de

as the mid point. Form strings (such as AAAA, ABCD etc.) were extracted from the resulting self-similarity matrices using fixed numerical thresholds. Two phrases p_i, p_j were deemed identical if

$$\sigma_I(p_i, p_j) \geq 0.6$$

for intervals and

$$\sigma_D(p_i, p_j \geq 0.7$$

for duration classes, where $\sigma(p_i, p_j)$ is the normed edit similarity taking values in $[0, 1]$ (Müllensiefen & Frieler, 2004a). Edit distances are generally the most successful algorithms in modeling perceived similarities and were applied in folk song research earlier (e. g., Müllensiefen & Frieler, 2004b; van Kranenburg et al., 2013). This particular thresholds were chosen after some informal testing while trying to replicate manual phrase form analysis as carried out by the author and as found in Sagrillo (1999). As any fixed thresholds, these are of course not perfect, and could surely be optimised further with more strict procedures. However, it turned out that these values are high enough to capture "true" similarities, since any value higher than ≈ 0.3 is already very unlikely to occur by chance (Müllensiefen & Frieler, 2004b), and low enough to still find a reasonable amount of PFS variation. Moreover, due to the smaller event space of duration classes spurious similarities are more likely to occur, the rhythm form threshold is slightly higher.

The analysis was carried out using the `melfeature` commandline tool from the MeloSpySuite (Frieler et al., 2013). Actually, the software allows to set two independent thresholds for labelling (nearly) exact repetitions and variied repetition, which were set equal here to disregard variied repetition to simplify further analysis. The resulting form strings were imported into R (R Core Team, 2013) for further analysis.

3. RESULTS

In total, 1008 different interval form strings (IF) and 1385 duration-based forms (DF) were found, which gave rise to 990 distinct combined form classes (CF). Combining was done by enumerating each unique pair of IF and DF symbols of a song with a new form symbol. The most common forms are ABCD for IF (17.7%) as well as CF (19.4%), and AAAA for DF (5.8%) (c.f. Sagrillo, 1999). The most common DF going along with the IF ABCD is AAAA (14.1%).

Form lengths range from 1 to 20 elements, with a median of 5 elements and a 75%-quartile of 7 elements. Very long forms are relatively rare, only 276 (3.5%) tunes have more than 10 phrases. The majority of songs consists of 4 phrases, followed closely by songs with 6 and 8 phrases. These results are very similar to Sagrillo (1999). Generally, a clear preference for an even number of form parts can be found; there are nearly twice as many forms with an even length than with an odd length. The number of different parts in a form is often smaller than the total form length. IFs and CFs have a median of 4 and DFs a median of 3 different parts per form. One can define the *coherence* (or *redundancy*) of a form as the amount of contained repetition, i. e. the number of unique elements divided by the total length of the form (subtracted from 1 for better interpretation). A coherence of 0 means that no form part is repeated, whereas a coherence of 1 can only be reached in the limit of a single, infinitely repeated part. We found a median coherence for IFs of 0.2, for CFs of 0.1667, and for DFs of 0.5. Thus, DFs contain about more than twice as much repetitions than IFs. This means, that songs often exhibit roughly constant rhythmic models over more varying tonal content, often in consecutive phrases, since the probability for an IF part to be repeated is only 5%, but 35% for a DF part. Finally, we defined the *coherence ratio* as the ratio of IF and DF coherence. We found a median coherence ratio of 0.67, which means that in the average the coherence of IFs is about two-thirds of the DF coherence. Only in 277 songs (3.5%), the IF coherence was larger than the DF coherence.

On a global level, many differences with regard to PFS can be found between the collections. All ANOVAs using form length and the four coherence variables as dependent variables became highly significant ($p < 0.0001$) due to the large dataset, but effect sizes were rather low (all $R^2_{adj} < 0.05$, except for form length with $R^2_{adj} = 0.21$). Generally, the Central European songs, mostly from Germany, Lorraine, and Luxembourg, which make up approx. 90% of the database, were more similar to each other than to the 62 Irish songs and the 676 songs from the two Polish sub-collections. Interestingly, the most diverse spectrum of forms as measured by the entropy of the form distribution can be found in the *Kinderlieder* (German children songs), which has a nearly flat distribution. The most common IF for *Kinderlieder* is ABCD with a relative frequency of only 8%. In contrast, the *Kolberg* collection from Poland has the most homogeneous distribution. Here the most common IF is AB used in over more than half of the songs (54%). Furthermore, the *Irland* (Ireland), *Warmia* (North-East Poland), and *Kolberg* collection are the only collections in which the most frequent IF type is *not* ABCD, but AABA, AABC, and AB respectively. On the other hand, there are only two collections, *Lothringen* (Lorraine) and *Kolberg* (Poland), in which AAAA is *not* among the first two most frequent DF form types.

4. CONCLUSION

In this explorative study, we found interesting statistical details of phrase form types in a certain set of European folksongs. Collections from different cultural origin exhibit clear differences with respect to form types and derived features. Hence, we suggest that descriptors of phrasal form structure might be useful features for manual and automated classification and other types of folk song research.

5. REFERENCES

Dahlig, E. (2000). An integrated system of encoding, analysing and processing of one-part melodies. In und Joachim Stange-Elbe, B. E. (Ed.), *Musik im virtuellen Raum*, volume 3 of *Musik und Neue Technologie*, (pp. 427–433)., Osnabrück.

Frieler, K., Abeßer, J., Zaddach, W.-G., & Pfleiderer, M. (2013). Introducing the Jazzomat Project and the Melo(S)py Library. In van Kranenburg, P., C. Anagnostopoulou, C., & Volk, A. (Eds.), *Proceedings of the Third International Workshop on Folk Music Analysis, Meertens Institute and Utrecht University Department of Information and Computing Sciences*, (pp. 76–78).

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, *163*(4), 845–848. Englische Übersetzung in: Soviet Physics Doklady, 10(8) S. 707-710, 1966.

Lomax, A. & Grauer, V. (1967). *Cantometrics Coding Book*.

Müllensiefen, D. & Frieler, K. (2004a). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology*, *13*, 147–176.

Müllensiefen, D. & Frieler, K. (2004b). Optimizing measures of melodic similarity for the exploration of a large folksong database. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR2004), Universitat Pompeu Fabra*.

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.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Sagrillo, D. (1999). *Melodiegestalten im luxemburgischen Volkslied: Zur Anwendung computergestützter Verfahren bei der Klassifikation von Volksliedabschnitten*. Bonn: Holos.

Schaffrath, H. (1995). The Essen Folksong Collection in Kern Format. In Huron, D. (Ed.), *Computer Database*, Menlo Park, CA.

Spevak, C., Thom, B., & Hoethker, K. (2002). Issues in Melodic Segmentation. In *Proccedings of the Second International Conference on Music and Artificial Intelligence, Edinburgh, Scotland, 2002*, Springer's LNCS/LNAI series.

Temperley, D. (2001). *Cognition of Basic Musical Structures*. Cambridge: MIT Press.

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*, 1–18.

COGNITIVE FEATURES FOR COVER SONG RETRIEVAL AND ANALYSIS

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

This article presents the first results of a study on large-scale automatic identification of derivative works in early popular music. We present a music cognition-inspired approach to (audio) version detection that is (1) tailored for early 20th century recordings, (2) based on simple, indexable feature representations and (3) allows for interpretation of the resulting musical descriptions. For this purpose, two new descriptor types are introduced: pitch bihistograms and chroma correlation coefficients.

2. MOTIVATION

Musical heritage collections such as folk music archives may contain large numbers of closely related documents, such as exact duplicates of a recording, different renditions of a song, or loose variations on a theme. The particularities of such variations are of great interest in the study of music genealogies, oral transmission of music, and other aspects of music studies (Volk et al. [2012]).

The field of Music Information Retrieval has produced a lot of research on automatic systems for cover version identification. The large majority of these systems focuses on popular music from the last 50 years. An overview of version detection techniques is found in Serrà [2011].

Also, a large majority of systems is based on alignment. This means that, for the system to assess the similarity between two document, it needs to compute the optimal way of aligning a pair of descriptive time series for those two documents. In other words, when analyzing a complete collection of recordings, a computationally expensive comparison of time series is required for every pair of songs in that collection. For large datasets, this is far from practical, and on the scale of the type of corpus required to study patterns and trends in musical, infeasible: a full-blown comparison of 10,000 documents at 1 pairwise comparison per second would take over a year (578 days).

Outside of MIR's audio community, interesting work has been done on the analysis of folk songs and tune families based on their scores (Kranenburg [2010], Bohak & Marolt [2009]). Unfortunately, symbolic transcriptions of music are not always available.

In our study, we aim to construct and test a cognition-inspired type of music features that is global and indexable, can be computed from (polyphonic) audio, and allows to reliably identify derivative works in a collection of recordings. Additionally, we want the features to be interpretable and easy to extend for use in analysis (e.g. clustering songs into genres, creating a phylogenetic tree of music pieces or quantifying the schematic musical expectations in a corpus). Interpretability is an important necessary requirement for applications in analysis, and a well-known issue with many currently used features Aucouturier & Bigand [2013]. In this study, we aim to address this by drawing inspiration from descriptors used in music cognition.

3. FEATURES

We propose a pair of simple, audio-based analogues to the well-known bigram representation often used in symbolic music processing. The first representation relates to melody and approximates a histogram of pitch bigrams (pairs of consecutive pitches) weighted with duration. The second, harmony-related representation is an abstraction of the co-occurrence matrix of chord notes in a song.

Many authors have proposed analyses based on pitch bigrams, most of them from the domain of cognitive science or otherwise interested in the information dynamics at play in music (Li & Huron [2006], Müllensiefen & Frieler [2006], Rodriguez Zivic et al. [2013]). This is not surprising: distributions of bigrams effectively encode a form of first-degree expectations: if the relative frequency of bigrams in a piece is conditioned on the first pitch in the bigram, we obtain the conditional frequency of a pitch given the one preceding it. Expectations of this kind have been linked to melodic complexity, familiarity and even preference ratings (Huron [1999]).

Marginalized and non-marginalized variants have been proposed for pitch events corresponding to pitch heights, durations, pitches with height and duration, pitch intervals, and scale degrees. The first new feature we introduce will follow the latter paradigm: it is a distribution over pairs of (chromatic) scale degrees $\{1, 2, \dots 12\}$. It will be referred to as the **pitch bihistogram**, a bigram representation that can be computed from continuous pitch data.

Assume that a pitch time series $P(t)$, quantized to semitones and folded to one octave, can be obtained. If a pitch histogram is defined as:

$$H(p) = \sum_{P(t)=p} \frac{1}{n},$$

with n the length of the time series and $p \in \{1, 2, \dots 12\}$,

the pitch bihistogram is then defined:

$$B(p_1, p_2) = \sum_{\substack{P(t_1)=p_1 \\ P(t_2)=p_2}} w(t_2 - t_1)$$

where

$$w(x) = \begin{cases} \frac{1}{d}, & \text{if } 0 < x < d. \\ 0, & \text{otherwise.} \end{cases}$$

and d is the size of the look-ahead window. In this study we will consistently use pitch contours that have been aligned to an estimate of the piece's overal tonic.

The second feature representation we propose focuses on vertical rather than horizontal pitch relations.

$$C(p_1, p_2) = \text{corr}(c(t, p_1), c(t, p_2)),$$

where $c(t, p)$ is a 12-dimensional chroma time series (also known as pitch class profile) computed from the song audio (Gomez [2006]). From this chroma representation of the song $c(t, p)$ we compute the correlation coefficients between each pair of chroma dimensions to obtain a 12×12 matrix of **chroma correlation coefficients** $C(p_1, p_2)$. Again, the chroma features are all transposed to the same tonic (e.g. A) based on an estimate of the song's overall key.

For key detection, a global chroma feature is computed from a full chroma representation of the song. This global profile is then correlated with all 12 modulations of the standard diatonic profile to obtain the tonic. The binary form (ones in the 'white key' positions and zeros in the others) is used. For this study, pitch contours and chroma features where computed using Melodia (by Salamon & Gomez [2010]), and HPCP (Gomez [2006]) respectively, with default settings. [1] For efficiency in computing the pitch bihistogram, the pitch contour was median-filtered and downsampled to 10 frames / s.

4. DATASET AND METHOD

The above features were tested on a set of 150 recordings digitized especially for this study: 100 45-rpm records from the 50's and 60's, and 50 78-rpm records, most of them from before 1950. The corpus they belong to is a real-world 'popular music heritage' collection that was, until recently, only accessible through manually transcribed metadata (such as titles, artists, original title, composer).

Amongst these records are 50 pairs of songs that correspond to the same composition. All of these tunes are translated covers or interpretations of melodies with a different text. Such songs are especially interesting since they garantee some deviation from the source, which is desirable when models of music similarity are tested.

A retrieval experiment on these songs was carried out: for each song that is part of a pair, the song was taken out of the corpus and used as a query to which all the remaining 149 songs were ranked (e.g. using a cosine distance). The rank r of the other song of the pair then determined the precision $p = \frac{1}{r}$. This was done 100 times to obtain the Mean Average Precision (MAP). A random baseline for this task was established at MAP $= 0.036$, with a standard deviation of 0.010 over 100 randomly generated distance matrices.

[1] mtg.upf.edu/technologies

Weights:	H	B	C	MAP
	1	0	0	0.27
	0	1	0	0.43
	0	0	1	0.42
	0	1	1	0.53

Table 1: Overview of results. H = pitch histogram, C = chroma correlation coefficients, B = pitch bihistogram.

The main experiment parameters for the features described above were d, the look-ahead window of the bihistogram, and the weighting of each of the three features when all are combined. d was set to $0.500s$ after a quick optimisation. The weights were restricted to 0 and 1 (feature used or not used) as can be seen in the results summary in Table 1.

5. RESULTS

The precision obtained using only pitch histograms is already substantial, about 0.27. However using only the pitch bihistogram feature, a MAP of around 0.43 can be obtained, compared a very competitive 0.42 for using just the chroma correlations. Finally, when the last two features are combined, the MAP goes up to 0.53. In the latter configuration, 44 of the 100 queries retrieve their respective cover version in first place: the 'precision at 1' is 0.440.

A thorough comparison with existing version detection systems will be performed later in this study, but a survey of reported performances of cover detection systems shows that precisions of this order of magnitude make a promising start. For example, the intervalgram method by Walters et al. [2013] achieves a precision at 1 performance of 0.538 on the covers80 dataset (160 songs). This method uses an indexable representation for pruning the candidate set but relies on alignment for the steps thereafter. Methods relying only on pairwise comparison have reported both higher and lower precisions on the same dataset.

In the current experiments, no indexing or look-up was performed. For actual indexing, the proposed features would have to be quantized and reduced in dimensionality. Locality Sensitive Hashing does precisely this, as also described in Walters et al. [2013]. We refer to Slaney et al. [2012] for an in-depth account of the relationship between LSH and high dimensional distances. The features we propose are indeed very suitable for such applications.

6. CONCLUSIONS AND FUTURE WORK

In this abstract, we have proposed two new features for the description and retrieval of early popular music: pitch bihistogram features and chroma correlation coefficients. The features are evaluated on a newly compiled dataset

of variations and translations in early popular music. We demonstrate a promising performance of MAP $= 0.53$. A quantitive comparison to state-of-the-art indexable representations and alignment-based retrieval systems will be carried out in the future. Other future work includes the separate evaluation and optimisation of the pitch extraction and key extraction components. Finally, we hope to include a representation of rhythm in the model.

7. ACKNOWLEDGEMENTS

The authors would like to thank Dimitrios Bountouridis and Marcelo Rodriguez.

8. REFERENCES

Aucouturier, J.-J. & Bigand, E. (2013). Seven problems that keep MIR from attracting the interest of cognition and neuroscience. *Journal of Intelligent Information Systems*, *41*(3), 483–497.

Bohak, C. & Marolt, M. (2009). Calculating similarity of folk song variants with melody-based features. In *Proc Int Society Music Information Retrieval Conf (ISMIR)*, number Ismir, (pp. 597–601). Citeseer.

Gomez, E. (2006). *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra.

Huron, D. (1999). Musical Expectation. In *The 1999 Ernest Bloch Lectures*.

Kranenburg, P. v. (2010). *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*. PhD thesis, Utrecht University.

Li, Y. & Huron, D. (2006). Melodic modeling: A comparison of scale degree and interval. In *Proc. of the Int. Computer Music Congerence*.

Müllensiefen, D. & Frieler, K. (2006). Evaluating different approaches to measuring the similarity of melodies. *Data Science and Classification*.

Rodriguez Zivic, P. H., Shifres, F., & Cecchi, G. a. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(24), 10034–8.

Salamon, J. & Gomez, E. (2010). Melody extraction from polyphonic music signals using pitch contour characteristics. In *IEEE Trans. on Audio, Speech and Language Processing*.

Serrà, J. (2011). *Identification of versions of the same musical composition by processing audio descriptions*. PhD thesis, Universitat Pompeu Fabra.

Slaney, M., Lifshits, Y., & He, J. (2012). Optimal parameters for locality-sensitive hashing. *Proceedings of the IEEE*, *100*(9), 2604–2623.

Volk, A., de Haas, B., & van Kranenburg, P. (2012). Towards modelling variation in music as foundation for similarity. In *Proc. 12th Int. Conf. Music Perception and Cognition*.

Walters, T. C., Ross, D. A., & Lyon, R. F. (2013). The intervalgram: An audio feature for large-scale cover-song recognition. In *From Sounds to Music and Emotions: 9th International Symposium, CMMR 2012, London, UK, June 19-22, 2012, Revised Selected Papers* (pp. 197–213).

THE COVER SONG VARIATION DATASET

Dimitrios Bountouridis, Jan Van Balen
Utrecht University, Department of Information and Computing Sciences
{d.bountouridis,j.vanbalen}@uu.nl

1. INTRODUCTION

As digital music collections grow larger, music similarity becomes one of the most prominent concepts in the field of Music Information Retrieval (MIR). Modelling similarity between music pieces allows efficient retrieval and organizing of such collections. Studies have shown that the concept of *variation* is closely related to similarity, since listeners tend to cluster together musical patterns that are repeated, transformed but still recognizable. Subsequently, musical pieces or segments that contain such patterns are considered similar. Such structural variations are notably present in oral-transmission processes. Folk songs are a standing example of such a process, capturing a huge amount of varying patterns moulded through time. Variations in cover songs in western popular music are also very interesting examples, since *a)* they can be considered products of a "modern" oral-transmission procedure and *b)* covers themselves are typically well documented with rich metadata.

Although variations in folk songs have been fairly studied [2–4], cover songs have remained merely the subject of interest for the cover song identification task (see the MIREX [1] competition). To our knowledge, there have been no studies focusing on the variations between corresponding segments of cover songs, which can safely attributed to the lack of proper expert annotations and transcriptions. This paper introduces the Cover Song Variation (CSV) dataset, a publicly available set of annotated melodies corresponding to different versions of the same song-section (e.g. chorus, verse). Its creation process, content and two applications are presented.

2. THE DATASET

2.1 Creation and content

The Second Hand Songs (SHS) dataset, an extension of the Million Song Dataset [5] containing cover song metadata, was used as our main input source. Having in mind that the notion of long-term memory salience would be interesting to study, we intersected SHS with a list of the "greatest songs of all time" from Top2000 [2] resulting in a set of 1706 songs. Those were grouped into sets of same song covers denoted as "cliques". From those we filtered out cliques of size 4 or less, resulting in 80 cliques of around 400 songs.

For the annotation we first used the Echo Nest [3] service to retrieve rough structural segmentations for each song. We developed a Spotify [4] application for iPad devices in order to manually cluster and label the different versions of a section inside a clique. For each song, we labelled at most three distinctive sections as 'A', 'B' or 'C' (see Figure 1). During that process a number of songs were disregarded for two main reasons: *a)* the audio was not available in our country and *b)* the Echo Nest segmentation was erroneous. This dropped the clique number to 45.

At the next step, we manually MIDI annotated the main vocal melody of the 'A', 'B', 'C' sections. Avoiding annotating the underlying harmony was a concious decision; folks songs are usually monophonic and vocal oriented, thus any results from our dataset would be easily applied or compared to that context.

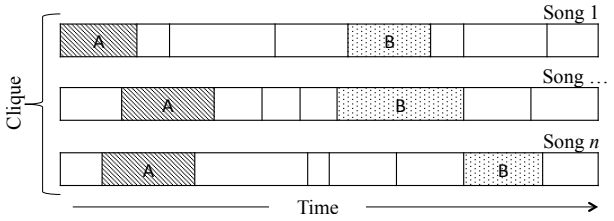


Figure 1. A clique of songs with two sections clustered and labelled across all versions.

Currently, the CSV dataset [5], which is under development and expansion, contains the following for a set of 60 different sections belonging to 45 cliques: *a)* 240 MIDI annotations, transposed to the same key for each clique, *b)* Echonest analysis for the aligned audio sections and *c)* Echonest analysis for each song. Figure 2 presents the distribution of length in terms of number of notes and seconds for the whole CSV corpus.

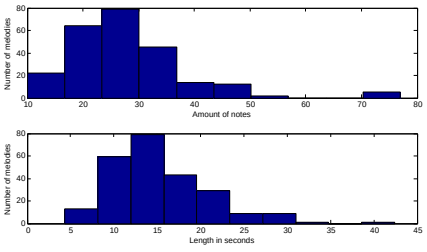


Figure 2. Top: the distribution of amount of notes. Bottom: the distribution of length (in seconds).

[1] www.music-ir.org/mirex/
[2] www.radio2.nl/top2000
[3] echonest.com
[4] www.spotify.com
[5] Avalaible at: www.projects.science.uu.nl/COGITCH/CSV

3. APPLICATIONS

CSV could be used for a range of MIR and musicological tasks. In the following sections we present two applications that relate to topics of folk music analysis.

3.1 Similarity as a stability measure

Stability, alongside variation, is a central concept in musicology with a number of dimensions (e.g. melodic, rhythmic, structural stability). For the sake of this demonstration we will consider an oral-transmission scenario where a similarity value between different variants of a melody can be also considered a stability value; as a consequence, any similarity algorithm that ranks variations of the query on top (in a retrieval context), can act as a stability measure.

For our demonstration we employed Melody Shape [6] [7], a set of symbolic melodic similarity algorithms that were ranked first on the related MIREX competition during the last years. In addition, we used a stripped down version of pairwise alignment [6] (denoted PWA) that uses information only from the pitch-duration domain (and no onsets). The Mean Average Precision (MAP) results of our retrieval experiment are shown in Figure 3.

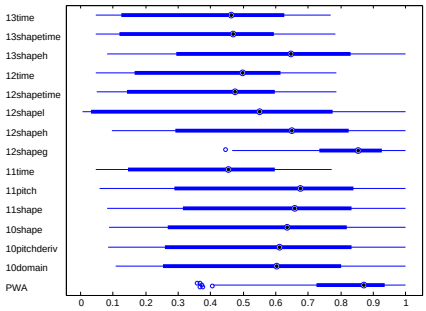


Figure 3. Mean Average Precision for the MelodyShape methods and the pairwise alignment (PWA).

It is worth pointing out the following: *a)* all methods that employ note onset information (those with suffix "time") are ranked last while *b)* PWA and spline transformation with global alignment (12shapeg) outperform the rest. Considering our initial assumptions, it is safe to say that *a)* onset time is not robust against variations and *b)* alignment, as a basic similarity method, captures melodic stability to a great extent.

3.2 Analysing stability features

Similar to [1] we make the intuitive assumption that the investigation of stable features across known music sets is beneficial for melodic variation retrieval. Therefore, symbolic annotations of different versions of a section offer the unique opportunity to examine stability across different features. In this demonstration we aim at identifying stability patterns with regard to the note's duration and onset position inside the encompassing section. This helps to answer the following questions: which note durations are more subjected to variation and which part of a melody is the most stable (e.g. beginning, ending)?

Such an analysis requires first and foremost aligned melodic sequences. To our knowledge, there have been no studies in MIR for aligning more than two sequences. We therefore, employed a bioinformatics method, called Multiple Sequence Alignment [8] that extends the pairwise alignment to a higher number of sequences.

Figure 4 presents stability across the length of a melody, where stability translates to the inter-agreement between the aligned sequences, meaning the number of matching notes at each position. A prominent pattern emerges; the first half of a melody is more stable than the second, which additionally exhibits the lowest stability at 75% of the melody's length. Figure 5 presents stability with regard to note duration. It is illustrated clearly that notes of smaller and larger duration are subjected to more variation.

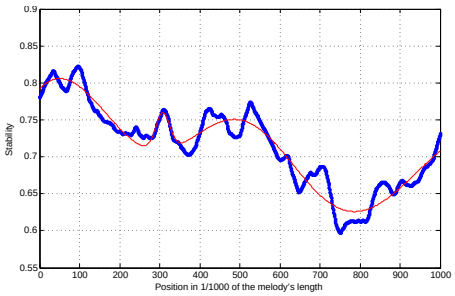


Figure 4. Probability of an event being stable given its position in the melody.

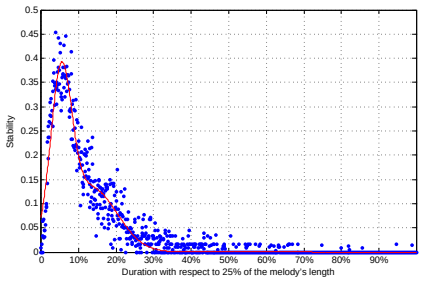


Figure 5. Probability of an event being stable given its duration value with regard to the 25% of the melody's length.

4. CONCLUSIONS

In this paper we introduced the Cover Song Variation dataset: a set of annotated variations of cover song sections. Although the dataset is currently under development and expansion (e.g. addition of key information), we presented a series of applications that would have been impossible without its existence. The emerging patterns from our analysis, although rough, show a promising direction for the study of variation and stability not only in the context of Western pop but also in folk music.

[6] code.google.com/p/melody-shape/

5. REFERENCES

[1] Volk A., de Haas W.B. and Kranenburg P. (2012). Towards modeling variation in music as foundation for similarity. *Proceedings of the International Conference on Music Perception and Cognition*, (pp. 1085-1094).

[2] Garbers J.,Volk A., Kranenburg P., Wiering F., Grijp, L. and Veltkamp, R.C. (2009). On pitch and chord stability in folk song variation retrieval. *Mathematics and Computation in Music, Communications in Computer and Information Science*, 37, 97-106.

[3] Biro D. P., Kranenburg P., Ness S., Tzanetakis G., Volk A. (2012). Stability and variation in cadence formulas in oral and semi-oral chant traditions a Computational Approach, *Proceedings of the International Conference on Music Perception and Cognition*, (pp. 98-105).

[4] Bohak C., Marolt M. (2009). Calculating similarity of folk song variants with melody-based features. *In Proceedings of the International Society for Music Information Retrieval Conference*,(pp. 597-600).

[5] Bertin-Mahieux T., Ellis D.P.W, Whitman B. and Lamere P. (2011) The Million Song Dataset. *In Proceedings of the International Society for Music Information Retrieval Conference*, (pp. 591-596).

[6] Kranenburg P., Volk A., Wiering F. and Veltkamp R.C. (2009). Musical models for folk-song melody alignment. *In Proceedings of the International Society for Music Information Retrieval Conference*, (pp. 507-512).

[7] Urbano J., Llorns J., Morato J. and Snchez-Cuadrado S. (2011). Melodic Similarity through Shape Similarity. *In Exploring Music Contents*, Springer, pp. 338-355.

[8] Carrillo H. and Lipman D.J. (1988). The Multiple Sequence Alignment Problem in Biology. *SIAM Journal of Applied Mathematics*, 48(5), pp. 1073-1082.

WHAT TO DO WITH A DIGITIZED COLLECTION OF WESTERN FOLK SONG MELODIES?

Peter Van Kranenburg and Berit Janssen
Meertens Institute, Amsterdam
{peter.van.kranenburg,berit.janssen}@meertens.knaw.nl

1. INTRODUCTION

This contribution aims to suggest some items for the research agenda of Computational Folk Song Research, based on a selective historic overview of the research tradition of Western Folk Song Research and on the current methods and foci of Computational Musicology and Music Information Retrieval. We specifically focus on folk songs from Western Europe, such as the Dutch or German. For example, the *Meertens Tune Collections*, consisting of thousands of digitized recordings, symbolic representations and metadata of folk songs from Dutch oral tradition, [1] or the *EsAC Folksong Databases*, which currently includes over 20,000 digitized song melodies from various countries in Europe and beyond (Schaffrath, 1995).

2. ELEMENTS OF THE HISTORY OF FOLK SONG RESEARCH

During the major part of the 20th century an enormous amount of work has been carried out on structural research on folk song melodies. As a starting point we take the contest that was organized by Dutch musicologist Daniel F. Scheurleer (1900) for the best way to provide a lexical ordering for a collection of folk song melodies. The solution of one of the contestants, Ilmari Krohn (1903), has had a profound influence on many approaches to the same question during the 20th century. Krohn proposed a classification system in which the number of phrases and the sequence of cadence tones determine the ordering of the melodies. Bártok devised a classification system for Hungarian folk melodies taking Krohn's system as starting point (Bartók, 1981, p. 6). The parallel work of Zoltán Kodály and what followed in Hungarian folk song research is summarized by László Dobszay (1988).

In American folk song research, Bayard's (1950) definition of tune family and Bronson's (1950) analysis of stable melodic elements reflect similar interests in ordering and classifying folk melodies within a broader interest in the study of variants. In the course of the century, a plethora of classification systems came into existence, none of which provides a general solution to the explicit or implicit aims of facilitating retrieval of melodies, and demonstrating the relations between melodies.

A remarkable research project that still captures the imagination of many minds is the Cantometrics project, led by Alan Lomax Lomax (1968). One of the objectives of this ambitious empirical project was to devise a descriptive system that is applicable to all of the world's folk song styles in order to to connect styles of folk song performance with other aspects of culture. Several thousands of songs were analyzed using 37 performance features. One of the results was a 'world song style map'.

All of these studies show an interest in the melodic material as object of research. However, the focus of researchers has shifted away from this topic towards other directions, as has been noted by Bruno Nettl (2005, p. 130) at the end of his overview of the history of research on melodic identity and oral transmission of melodies. The last decades of the 20th century show a decrease of interest in folk song melodies as such. There are several more or less interconnected causes for this. An important factor is the fading away of such underlying ideological motivations as were initially advocated by Johann Gottfried Herder (1744–1803) in the second half of the eighteenth century (Schepping, 1994). According to Herder, who introduced the term *folk song* ("Volkslied"), 'authentic' folk songs show the *soul* ("Geist") of the people. Therefore, collecting and studying folk songs was a meaningful endeavor from the perspective of people's identity, which also got connected with national identity ("one country, one language, one people"), and with history: through studying authentic folk songs, investigators hoped to find traces of ancient, pre-Christian culture that was supposed to still live on in rural areas, far away from emerging urban cultures. In the course of the 20th century, this paradigm was abandoned by ethnologists (Bendix, 1997).

Another cause is the parallel shift in other areas of Musicology away from 'positivistic' research on scores and recordings towards more anthropological and social approaches, which is indicated as 'New Musicology' (Kerman, 1980).

Currently, new approaches in Digital and Computational Humanities are quickly gaining ground. For music, the Music Information Retrieval (MIR) community is one of the most important catalysts for computational research on music. Given the nature of this kind of research, a focus on the musical 'material' rather than its cultural context or social function is inherent to the type of studies in this field. This stimulates a renewed interest in the contents of many of the ethnomusicological archives that were established

[1] The public release of this data set is in preparation for 2014.

during the 20th century, including those with Western folk song melodies. However, the underlying motivation is of a kind that is very different from the 20th-century folk song researchers. An illustration of this is the way in which the EsAC collection is used in MIR research. Virtually all papers in the proceedings of the yearly conference on Music Information Retrieval (ISMIR), in which this set of melodies are used, do not show an interest in folk music as such. Instead, the melodies are taken as just a collection of labeled musical data to test segmentation algorithms, melodic similarity measures, pattern discovery algorithms, and the like. The meta-data that comes with the collections (e.g., region of origin, tune family membership, segment boundaries), are used as ground-truth data for corresponding MIR tasks.

The 'mis-match' between the nature of the EsAC-collection and the purposes for which it is employed raises the question whether there are research questions that could be addressed properly using such a collection, taking a computational approach. More precisely: given the rich history of classification systems, given the state-of-the-art in Music Information Retrieval, given the shifting ideological backgrounds and aims, what are sensible steps to take in computational research on folk song collections?

It seems that Scheurleer's question of classification has lost its relevance because of the many possible ways in which a collection of digitized music can be queried in a modern music information retrieval system. The Herderian aim of revealing a nation's identity by studying its folk songs, is obviously left behind as well. Since it is hard to regard folk songs such as the Dutch or German as a continuously developing tradition — as opposed to vibrant oral traditions such as found in Africa or the Near-East — there is also no strong motivation from the perspective of understanding current musical culture.

3. WHAT TO DO WITH A COLLECTION OF DIGITIZED WESTERN FOLK SONGS?

We present various ways in which a digitized collection of folk songs can be used as research data.

An important property of folk melodies from Western oral tradition is the fact that such melodies were sung by ordinary people with little or no formal musical training. Therefore, these collections offer a rich source of material to study human musicality, including memory for melody, strategies of lyrics-placement, properties of singing, common errors, etc. These are all research interests that are relevant in the field of music cognition. An example of this can be found in the work of David Temperley (2008), who extensively used the EsAC collection to support his theory of melodic perception. Another example is the study by Von Hippel and Huron (2000), who used a corpus of folk song melodies to show that the gap-fill rule, which states that a melodic leap should be continued stepwise in the opposite direction, can be fully explained by the statistical phenomenon of regression to the mean. Yet another example is a forthcoming study from our research group that shows that memory for absolute pitch plays a role in oral transmission by comparing the pitch-contents of recorded variants of a tune family.

Since Narmour (1990) presents his implication-realization theory of melodic structure to be independent of specific style, a collection of folk melodies can be used as empirical data to challenge this theory. Indeed, Schellenberg (1996) employed British folk songs to assess and refine various aspects of Narmour's theory.

Another, though related, set of research questions comes from Folk Song Research itself: investigating oral variation. A collection of recordings such as the Dutch collection *Onder de groene linde* (Grijp, 2008) contains a number of variants of a tune, which allows for investigating stability in oral transmission, which, in turn, sheds light on aspects of human cognition of music. Nettl (2005, 295ff) discusses the constituting musical elements of oral traditions. He proposes that the study of an oral tradition should be directed at the basic 'unit of transmission'. Bohlman (1988) lists some common units of transmission from an analytical perspective. He observes that the whole piece can function as the smallest unit of transmission, but smaller melodic elements such as formulae, conventions, motifs and phrases can also be found as the units of transmission. The study of units of transmission could be addressed with pattern discovery algorithms. The understanding of what units of transmission are can be improved by studying and interpreting the discovered patterns and adapting the discovery algorithm according to sensible hypotheses. Such an approach of confronting algorithmic output with musical and cognitive theories or hypotheses could contribute to the understanding of human musicality. In our own work, we set out to test several hypotheses about stability in oral transmission that relate to harmonic, rhythmic, and melodic aspects.

An emerging field of research has recently been established by increased interest in digitization, curation and unlocking of cultural heritage. One example is the Europeana.eu portal,[2] which aims to unlock Europe's cultural heritage, including folk song audio collections. To unlock and search digitized artifacts, adequate models and tools are needed.

Finally, as has been done in Music Information Retrieval, a collection of monophonic melodies can be used as ground-truth for testing various kinds of MIR-tasks as mentioned in the previous section. We would advocate, though, to interpret algorithmic results on these data sets from a musical or cognitive point of view as much as possible (Van Kranenburg, 2013). This implies putting the 'ground-truth' into question as well, rather than taking it for granted, which goes beyond merely assessing classification accuracy, and which will lead to an increase of knowledge of music.

Acknowledgments The authors are supported by the Computational Humanities Programme of the Royal Netherlands Academy of Arts and Sciences, under the auspices of the Tunes & Tales project.

[2] `http://www.europeana.eu`, accessed 18 May 2014.

4. REFERENCES

Bartók, B. (1981). Introduction. In B. Suchoff (Ed.), *The Hungarian Folk Song by Béla Bartók*, volume 13 of *New York Bartok Archives Studies in Musicology* (pp. 1–11). Albany: State University of New York Press.

Bayard, S. (1950). Prolegomena to a study of the principal melodic families of british-american folk song. *Journal of American Folklore*, *63*(247), 1–44.

Bendix, R. (1997). *In Search of Authenticity: the Formation of Folklore Studies*. University of Wisconsin Press.

Bohlman, P. (1988). *The Study of Folk Music in the Modern World*. Bloomington: Indiana University Press.

Dobszay, L. (1988). Folksong classification in hungary – some methodological conclusions. *Studia Musicologica Academiae Scientiarum Hungaricae*, *30*(1/4), 235–280.

Grijp, L. P. (2008). Introduction. In Grijp, L. P. & van Beersum, I. (Eds.), *Under the Green Linden — 163 Dutch Ballads from the Oral Tradition*, (pp. 18–27)., Amsterdam. Meertens Institute + Music & Words.

Kerman, J. (1980). How we got into analysis, and how to get out. *Critical Inquiry*, *7*(2), 311–331.

Krohn, I. (1903). Welche ist die beste Methode, um Volks- und volksmässige Lieder nach ihrer melodischen (nicht textlichen) Beschaffenheit lexikalisch zu ordnen? *Sammelbände der internationalen Musikgesellschaft*, *4*(4), 643–60.

Lomax, A. (1968). *Folk Song Style and Culture*. New Brunswick (U.S.A) and London (U.K.): Transaction Publishers.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures - The Implication-Realization Model*. Chicago and Londen: The University of Chicago Press.

Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-one Issues and Concepts* (2nd ed.). Urbana and Chicago: University of Illinois Press.

Schaffrath, H. (Ed.). (1995). *The Essen Folksong Collection*. Stanford, CA: Center for Computer Assisted Research in the Humanities.

Schellenberg, E. G. (1996). Expectancy in melody: tests of the implication-realization model. *Cognition*, *58*(1), 75–125.

Schepping, W. (1994). Lied- und Musikforschung. In R. W. Brednich (Ed.), *Grundriss der Volkskunde* (2. ed.). (pp. 467–492). Berlin: Dietrich Reimer Verlag.

Scheurleer, D. (1900). Preisfrage. *Zeitschrift der Internationalen Musikgesellschaft*, *1*(7), 219–220.

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, *32*, 418–444.

Van Kranenburg, P. (2013). On computational modeling in ethnomusicological research: Beyond the tool. In *Proceedings of the Third International Workshop on Folk Music Analysis*, (pp. 109–110)., Amsterdam and Utrecht.

Von Hippel, P. & Huron, D. (2000). Why do skips precede reversals. *Music Perception*, *18*(1), 59–85.