TARSOS - A PLATFORM TO EXPLORE PITCH SCALES IN NON-WESTERN AND WESTERN MUSIC

Joren Six and Olmo Cornelis
Royal Academy of Fine Arts & Royal Conservatory,
University College Ghent
joren.six@hogent.be - olmo.cornelis@hogent.be

ABSTRACT

This paper presents Tarsos [1], a modular software platform to extract and analyze pitch and scale organization in music, especially geared towards the analysis of non-Western music. Tarsos aims to be a user-friendly, graphical tool to explore tone scales and pitch organization in music of the world. With Tarsos pitch annotations are extracted from an audio signal that are then processed to form musicologically meaningful representations. These representations cover more than the typical Western 12 pitch classes, since a fine-grained resolution of 1200 cents is used. Both scales with and without octave equivalence can be displayed graphically. The Tarsos API [2] creates opportunities to analyse large sets of - ethnic - music automatically. The graphical user interface can be used for detailed, manually adjusted analysis of specific songs. Several output modalities make Tarsos an interesting tool for musicological analysis, educational purposes and even for artistic productions.

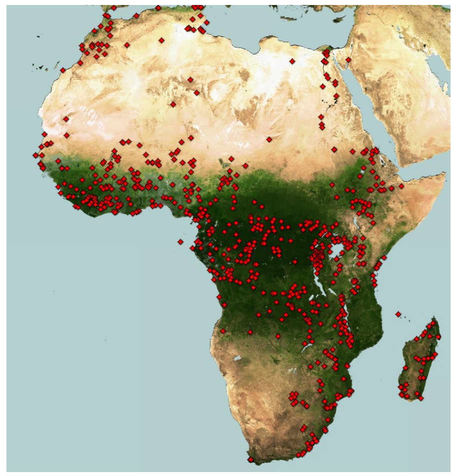

Figure 1. A visualization of the music archive of the Royal Museum for Central Africa. The dots mark places where recordings have been made.

1. INTRODUCTION

A 2007 f(MIR) article by Cornelis et al. [3] argued that access to ethnic music could become one of the next big challenges for the MIR community. It gives an overview of the difficulties of working with ethnic music: i) There is an enormous variety of styles, timbres, moods, instruments falling under 'ethnic music' umbrella. ii) The absence of a theoretical framework and a different attitude towards music imply that western music-theory concepts do not always apply. iii) A third factor that complicates access to ethnic music is its distribution. Archives of ethnic music are often not or not yet digitized, badly documented, and when metadata is available terminology and spelling may vary. These three elements cause a lack of ground truth, which makes (large scale) research challenging.

There are difficulties working with ethnic music but there is also a lot of potential. While some archives with ethnic music are being digitized, the need for specialized MIR applications becomes more apparent. Ethnic music offers a unique environment of characteristic timbres, rhythms and textures which could use adapted or completely new, innovative tools. The potential of computational research within the context of ethnic music has been stressed by the introduction of the term *Computational Ethnomusicology* [13]. Hopefully this new interdisciplinary (sub)field can give an impulse to the study and dissemination of a rich heritage of music that is now hidden in archives and aid or even stimulate new musicological field work.

This research focuses on one of those unique collections of ethnic music: the audio archive of the Royal Museum for

[1] Tarsos is open source and available on http://tarsos.0110.be. It runs on any recent Java Runtime and can be started using Java Web Start.
[2] With the Application Programmers Interface tasks can be automated by programming scripts. For an application see chapter 5.

Central Africa (RMCA) in Belgium. It is one of the largest collections worldwide of music from mainly Central Africa. Figure 1 displays the geographical distribution of the audio collection [3] that consists of about 50,000 sound recordings (with a total of 3,000 hours of music), dating from the early 20th century up to now. A selection of this data set with African music has already been used for a study on pitch organization and tone scales [9].

This paper is structured as follows: After this introduction sketching the background for this research the following chapter identifies the need for precise pitch analysis and the rationale behind the development of Tarsos. Chapter three will provide a view on the method we use, and give a brief overview of related work. Chapter four documents the structure and method of the Tarsos platform. Example applications of Tarsos can be found in chapter five. The final chapter gives a conclusion and ponders on future work.

2. SCALE ORGANISATION

For *Western music* pitch organization in music relies on a well-defined, historically grown music theory. Nowadays almost all western music relies on a division of the octave in 12 equal parts. Only few composers have explored different divisions of the octave (e.g. Darreg, Partch).

In *non-Western classical music*, tone scale organization leans on an, often very different, theoretical system than the Western equal temperament. The most outspoken difference is that not all pitch intervals have an equal size. This can result in an explicitly sought musical tension. An example is the unequal octave division of the Indonesian gamelan Pelog scale.

Oral music traditions however, rely almost exclusively on musical practice. Without a written musical theory the master-student relationship becomes very important, together with the societal context in which people hear music. An oral culture does not support the development towards a polished music theory but grows more organically. These factors define the specific characteristics of the music itself: less harmonic impact, instruments with varying tuning, no harmonic modulation and a large number of different tone scales. Until now, far too little attention has been paid to this tone scale diversity. There is a need for a system that can extract pitch organisation - scales - from music in a culture-independent manner.

Currently there is software available for pitch analysis but it mainly focuses on Western music and is used for e.g. key-detection in pop music. To fill the need for automated pitch analysis of ethnic music *Tarsos* has been developed. Tarsos creates opportunities to analyse pitch organization in large music archives, document tone scales and find patterns in pitch usage.

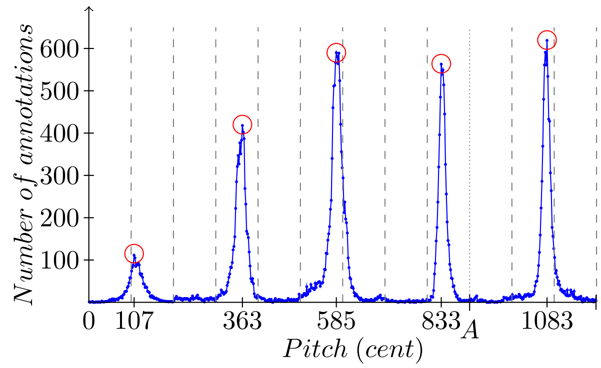

Figure 2. A pitch class histogram that shows how much pitch classes are present in a piece of music. The graph shows absolute pitch annotations collapsed to one octave. The circles mark the most used pitch classes. For reference, the dashed lines represent the Western equal temperament. The pitch class A is marked with a dotted line.

3. PITCH ANALYSIS

The basic idea behind the method we use is simple: count how many times each fundamental frequency is repeated throughout an audio signal and represent this data in a useful way. This method has a long tradition and historically this was done by hand, or, more anatomically correct, by ear. Each tone in a musical piece was compared with a large set of tuned bells and every match was tallied. This method is very labour-intensive and does not scale to large music archives.

Already in the late sixties researchers automated this process to study the tone scale of a Swedish folk instrument [11]. Since then various terms have been introduced to describe this, or closely related ideas: Frequency Histogram [11], Chromavector [9], Constant-Q Profile [10], Harmonic Pitch Class Profile [5] and Pitch-frequency Histogram [6].

Working with ethnic music, and especially African music, it is important that the pitch organization diversity can be captured. In [9] this is done as follows. At first the audio is analysed in blocks of 10ms and for each block a fundamental frequency estimation is made. Secondly, the frequencies are converted to the cents scale with C_0 set to zero cents while maintaining a list with the number of times each frequency occurs. And finally the listed values are reduced to one octave. This results in a quasi-continuous *pitch class histogram* of 1200 values as seen in Figure 2. With Tarsos this method is automated in a flexible way.

Pitch class histograms can be used for various applications. The most straightforward application is tone scale

[3] There is a website featuring complete descriptions and audio fragments, it can be found at `http://music.africamuseum.be`

detection. To extract a scale from a pitch class histogram peak extraction is used: i.e. finding the circles in Figure 2. With the pitch classes identified a pitch interval matrix can be constructed and subsequently used for comparison and analysis.

4. TARSOS PLATFORM

The main contribution of this paper is Tarsos: a platform for pitch analysis. It makes the methods described in [6, 9] easier to use and therefore accessible to a larger audience. Essentially Tarsos tries to make one-off studies of pitch usage easily repeatable, verifiable and scalable to large data sets. The functions of Tarsos will be explained using the block diagram in Figure 3. This should make the information flow clear and provide a feel on how Tarsos can be used.

4.1 Input

As input Tarsos accepts audio in almost any format. All audio is transcoded to a standardized format. The conversion is done using FFMPEG [4], the default format is PCM WAV with all channels are downmixed to mono.

Another input modality are Scala files. Scala files are standardized text files which contain tone scale information. The file format is defined by the Scala program. Quoting the Scala website: `http://www.huygens-fokker.org/scala/`

> "Scala is a powerful software tool for experimentation with musical tunings, such as just intonation scales and non-Western scales. It supports scale creation, editing, comparison, analysis, storage, ... Scala is ideal for the exploration of tunings and becoming familiar with the concepts involved."

The Scala program comes with a dataset of over 3900 scales ranging from historical harpsichord temperaments over ethnic scales to scales used in contemporary music. Tarsos can parse and export scala files. Their use should become apparent in section 4.5.

4.2 Analysis

During analysis each block of audio is examined and zero, one or more fundamental frequencies are assigned. The block size and the number of extracted frequencies depend on the underlying fundamental frequency detection algorithm. Several detection algorithm implementations are distributed together with Tarsos and thanks to its modular design new ones can be added. For practical purposes platform-independent - pure Java - implementations of YIN [4] and MPM [8] are available without any configuration. Currently there is also support for the MAMI-detector [2] and for any VAMP-plugin [1] that generates frequency annotations. These external detectors are platform dependant and need some configuration but once correctly configured their use is completely transparent: the generated annotations are transformed to a unified format, cached and then used for representation.

[4] FFmpeg is a complete, cross-platform solution to record, convert audio and video. It has decoding support for a plethora of audio formats.

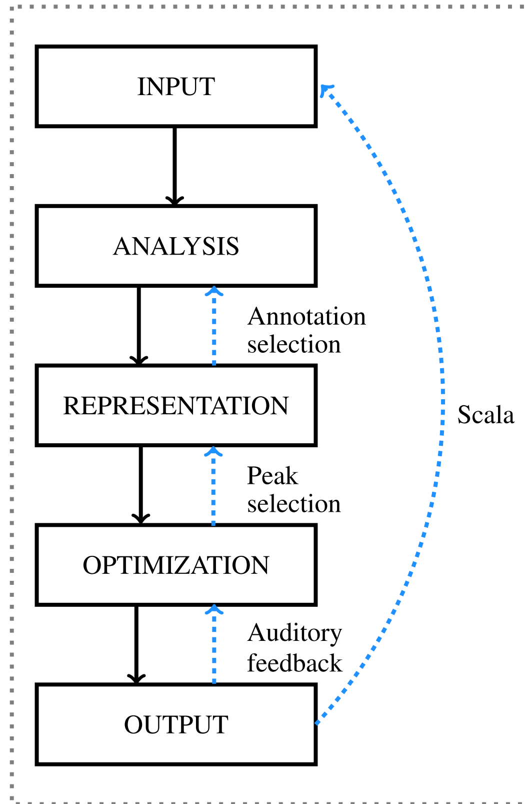

Figure 3. The main flow of information within Tarsos.

4.3 Representation

The most straightforward representation of pitch annotations is plotting them over time. This results in a *piano-roll* like view. In monophonic music this visualizes the melody. In polyphonic music it shows information about harmonic structures and the melodic contour. The piano-roll aids transcription and makes repeating melodic patterns visible. Figure 4 shows a screenshot of Tarsos, the piano roll representation is marked with 3. With the interactive user interface the piano roll representation can be used to select an area of annotations you are interested in. This can be used to ignore annotations below a certain pitch threshold (e.g. pitched percussion) or to compare the first part of a song with the second part. The selection - represented by the upwards arrow between analysis and representation in Figure 3 - influences the next representation.

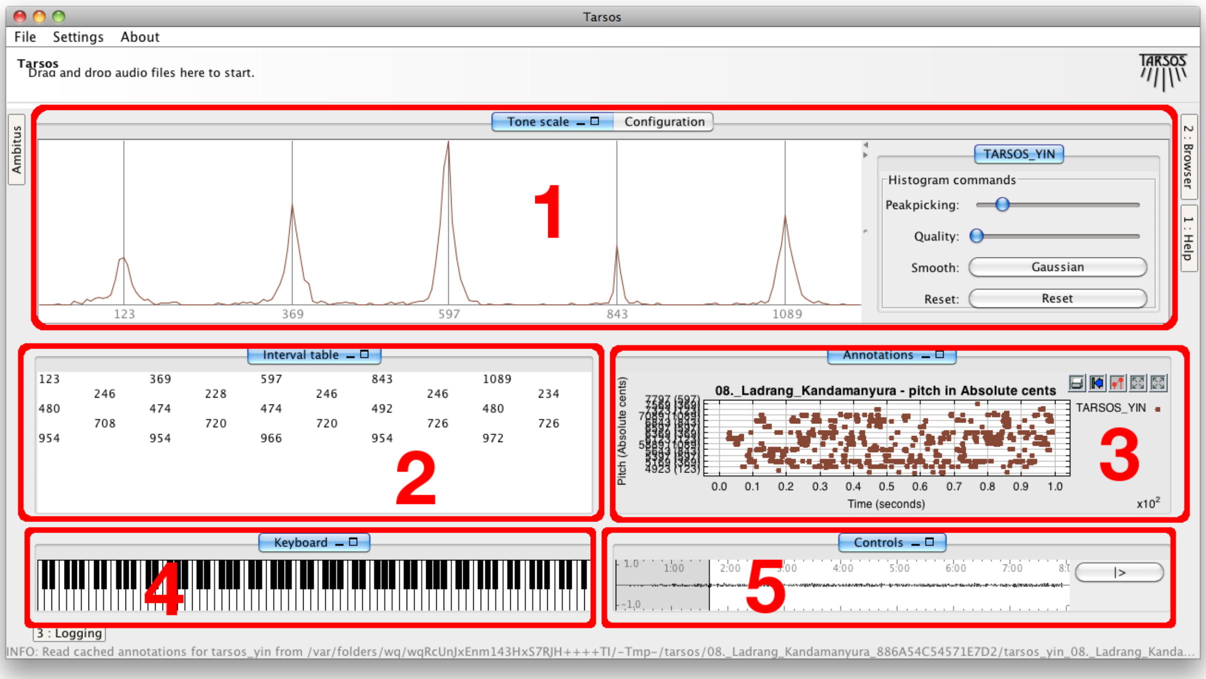

Figure 4. A screenshot of Tarsos: 1) a pitch class histogram, 2) a pitch class interval table, 3) a piano roll like view on annotations, 4) a MIDI keyboard and 5) a waveform. Tarsos is available on `http://tarsos.0110.be`.

Within Tarsos the *pitch histogram* is constructed by assigning each annotation to a bin between 0 and $12 \times 1200 = 14400$ cents, spanning 12 octaves. The height of each peak represents the total duration of a particular detected absolute pitch within a piece. As mentioned in section 3 to transform the pitch histogram to a *pitch class histogram* all values are folded to one octave. In the pitch class histogram a peak represents the total duration of a detected pitch class within a piece. An example of a pitch class histogram can be seen in Figure 2 or the area marked with 1 in Figure 4.

A more high level, musicologically more meaningful representation is the *pitch interval matrix*. It is constructed by applying automatic or manually adjusted peak detection on the pitch class histogram and extracting the positions of the pitch classes. It contains the tone scale of a song and the intervals between the pitch classes.An example of a pitch interval matrix extracted from the pitch class histogram in Figure 2 can be seen in Table 1. In the screenshot, Figure 4 it is marked as 2.

4.4 Optimisation

Automatic peak extraction may yield unwanted results. Therefore there is a possibility to adjust this process manually. Adding, removing or shifting peak locations is possible with the pitch class histogram user interface. Changing the position of a peak has an immediate effect on all other representations: the pitch interval matrix is reconstructed, the reference lines in the pitch histogram and piano roll are adjusted accordingly.

4.5 Output

Tarsos contains export capabilities for each representation, from the pitch annotations to the pitch class interval matrix there are built-in functions to export the data, either as comma separated text files or as image files. Since Tarsos has a scriptable, documented API which can be accessed by any Java Virtual Machine (JVM) compatible programming language - Groovy, Scala [5] , Clojure, Java - there is also a possibility to add new output formats based on the internal object model. Scripting is also the way to go when processing a large number of files.

As previously mentioned, for pitch class data there is a special standardized text file format defined by the Scala program: the scala file with the `.scl` extension. Scala files can be used to compare different tone scales within Tarsos or with the Scala program. When used as input for Tarsos, these files provide a reference for the pitch class histogram extracted from audio. A scala file e.g. extracted from Figure 2 with pitch classes $(107, 363, 585, 833, 1083)$

[5] Do not confuse the Scala programming language with the Scala software tool for scale analysis. Information about the programming language can be found at `http://scala-lang.org`

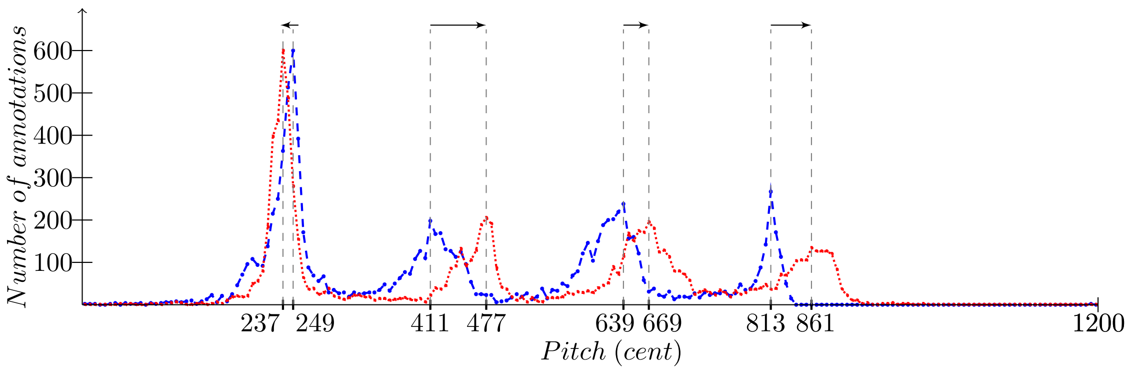

Figure 5. MR.1973.9.41-4, the second minute of the song is represented by the blue, dashed line, the seventh by the red, dotted line. Comparing the second with the seventh minute shows that during the performance the fiddle player changed hand position. The lowest, most stable pitch class is the result of the open string which lost tension during the piece and started to sound lower, in stark contrast with the other pitch classes.

Pitch Class (cent)	Interval (cent)			
107				
	255			
363		478		
	222		725	
585		470		976
	248		720	
833		498		
	251			
1083				

Table 1. A pitch interval matrix with pitch classes and pitch class intervals, both in cents. The peaks detected in Figure 2 are used.

can be used to compare pitch use of a song recorded in the same geographical area: do they both use the same absolute pitch or the same pitch intervals?

The pitch annotations can also be synthesized. This results in an audio file which can be used to check if the annotations make sense. Overlapping the original sound with the synthesized annotations makes clear when no, or incorrect annotations were made and, conversely when annotations are correct. This auditory feedback can be used to decide if the annotations are trustworthy (the upwards arrow starting from output in Figure 3).

A completely different output modality is MIDI. The MIDI Tuning Standard defines MIDI messages to specify the tuning of MIDI synthesizers. Tarsos can construct `Bulk Tuning Dump`-messages based on extracted pitch class data to tune a synthesizer enabling the user to play along with a song in tune. Tarsos contains the Gervill synthesizer, one of the few (software) synthesizers that offer support for tuning messages.

5. APPLICATIONS

This section illustrates how Tarsos enables or facilitates research on pitch use. The examples given could inspire third party users - musicologists - to try Tarsos and use it to solve their own research questions.

A first example is an analysis on a single African fiddle piece. In African music pentatonic scales are common but this piece uses a tetratonic scale as seen in Figure 5. The scale is a result of a playing style with three - more or less equally spaced - fingers and an open string. The graphical interface of Tarsos was used to compare the second minute of the song with the seventh, this can be accomplished by selecting the annotations in the piano roll window. This shows that the open string lost tension during the performance - it started to sound lower - in stark contrast with the other pitch classes. The results were exported using the LaTeX-export function and are shown in Figure 5.

A second example illustrates what can be done with a script that processes a lot of audio files in batch and the Tarsos API. In an article by Bozkurt [6] pitch histograms are used for - amongst other tasks - makam [6] recognition. The task is to identify which of nine makams is used in a specific song. A simplified, generalized implementation of this task was scripted. With this script it is possible to correctly identify 39% of the makams using a dataset of 800 files. Some makams look very much alike: if the first three guesses are evaluated the correct makam is present in 75% of the cases. The example is fully documented in the Tarsos manual available on the website `http://tarsos.0110.be`, also the source code is available there. This method is very general and directly applicable to e.g. harpsicord tuning estimation as done, using another approach, by Tidhar et al [12].

[6] A maqam defines rules for a composition or performance of classical Turkish music. It specifies melodic shapes and pitch intervals.

6. CONCLUSION, DISCUSSION AND FUTURE WORK

In this paper Tarsos was presented, a modular software platform to extract and analyze pitch organization in music. After an introduction explaining the background and the needs for precise pitch analysis, chapter two provided some context about the method used and points to related work. Chapter three gave a high level overview of the different components of Tarsos.

Currently Tarsos offers a decent foundation for research on pitch but it also creates opportunities for future work. One research idea is to reintroduce time domain information. By creating pitch class histograms for a sliding time-window and comparing those with each other it should be possible to detect sudden changes in pitch usage: modulations. Using this technique it should also be possible to detect and document pitch drift in choral or other music on a large scale. Automatic separation of speech and music could be another application.

Another research area is to extract features on a large data set and use the pitch class histogram or interval data as a basis for pattern recognition and cluster analysis. Using Tarsos' scripting abilities with a timestamped and geotagged musical archive it could be possible to detect geographical or chronological clusters of similar tone scale use.

On the longer term we plan to add comparable representations of other musical parameters to Tarsos as well. In order to compare rhythmic and instrumental information, temporal and timbral features will be included. Our ultimate goal is to develop an objective, albeit partial, view on music by combining those three parameters.

During this type of research one should keep this quote in mind:

> "Audio alone might not be sufficient to understand ethnic music. What does it mean to describe music from a culture where the word "music" exists only in connection to body movement, smell, taste, colour. The idea of separating sound from the rest of its physical environment (movement, smell, taste, colour) may well be a weird "invention" of the West. We cannot understand ethnic music correctly without its social function and context [7]."

However, we do can gain interesting insights and alleviate accessibility problems, which is what we are aiming for.

7. REFERENCES

[1] Chirs Cannam. The vamp audio analysis plugin api: A programmer's guide. `http://vamp-plugins.org/guide.pdf`.

[2] L. P. Clarisse, J. P. Martens, M. Lesaffre, B. De Baets, H. De Meyer, and M. Leman. An auditory model based transcriber of singing sequences. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 116–123, 2002.

[3] Olmo Cornelis, Dirk Moelants, and Marc Leman. Global access to ethnic music: the next big challenge? In *Proceedings of 9th ISMIR Conference*, 2009.

[4] Alain de Cheveigné and Kawahara Hideki. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

[5] Takuya Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. In *Proc. Int. Comput. Music Conf*, pages 464–467, 1999.

[6] Ali C. Gedik and Barış Bozkurt. Pitch-frequency histogram-based music information retrieval for turkish music. *Signal Processing*, 90(4):1049–1063, 2010.

[7] Micheline Lesaffre, Olmo Cornelis, Dirk Moelants, and Marc Leman. Integration of music information retrieval techniques into the practice of ethnic music collections. In *Proceedings Unlocking Audio 2*, 2009.

[8] Phillip McLeod and Geoff Wyvill. A smarter way to find pitch. In *Proceedings of International Computer Music Conference, ICMC*, 2005.

[9] Dirk Moelants, Olmo Cornelis, and Marc Leman. Exploring african tone scales. In *Proceedings of 9th ISMIR Conference*, 2009.

[10] Hendrik Purwins, Benjamin Blankertz, and Klaus Obermayer. Constant Q profiles for tracking modulations in audio data. In *International Computer Music Conference*, pages 407–410, 2001.

[11] J. Sundberg and P. Tjernlund. Computer measurements of the tone scale in performed music by means of frequency histograms. *STL-QPS*, 10(2-3):33–35, 1969.

[12] Dan Tidhar, Matthias Mauch, and Simon Dixon. High precision frequency estimation for harpsichord tuning classification. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 61 –64, march 2010.

[13] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2), 2007.