A brief project summary
SoundTracer

Benedikte Wallace

July 9, 2018

1. Introduction

The SoundTracer project is a collaborative effort between the Norwegian National Library and the Department of Musicology at the University of Oslo. The goal of the project is to use the audio recordings collected by the folk music department of the National Library to create a query-by-gesture application that is able to search and retrieve specific pieces of music using motion detected by a mobile device. This brief summary will outline the objectives, the methods that are used, as well as the results of the SoundTracer project.

1.1 The Folk Music Dataset

For use in this project the Norwegian National Library has allowed access to their FIOL database, an audio visual filing system for cataloguing and retrieving folk music and dance. FIOL is currently used by 13 folk music archives in Norway, its development and support is run by Folkmusikens hus in Rättvik (Sweden). The folk music dataset used for this project consists of 343 concatenated field recordings gathered from across Norway and parts of Sweden over several decades. The tapes used for recording these songs have been digitalised by the Norwegian National Library and are available through the FIOL database.

The majority of these field recordings are of vocal or fiddle performances. Each song is separated by a short introduction for the upcoming piece of music as well as a brief period of silence. In addition to the audio itself, the database contains metadata for each piece of music. The metadata used in this project is listed in figure 1. In 2018 work has been done in order to move the recordings over to a new, semantic EBU Class Conceptual Model (CCDM) database in collaboration with the Norwegian Public Broadcasting Corporation, NRK.

Table 1: Metadata from FIOL database used in the SoundTracer project

Metadata	Example
Instrument	Gitar, munnspill, fløyte, vokal etc.
Region	Sogn og Fjordane, Lappland, Skåne etc.
District	Follo og Aker, Bærum, Fors etc.
Type	Joik, Kontradans, Nystev etc
Identity	Name and other details of performer
Occurrence	Details such as song title and start and end times of each song.

1.2 Objectives

The three main goals outlined for the SoundTracer project were as follows:

1. How is it possible to extract features from the movements of a mobile phone?
2. How is it possible to map movement features to sound features?

3. How is it possible to search in the audio content of a large music Library?

Each of these objectives are addressed in the development of the SoundTracer application, which is available for iPhone and iPad.

2. The SoundTracer App

The SoundTracer application allows the user to generate up and down movements by moving their device during a set time frame to imitate the change of pitch in the melody of a song. The app, developed by Olivier Lartillot, uses a small subset of the folk music data set (50 distinct music pieces) to demonstrate how we can use gestures to query these folk music recordings. In order to extract melodies from these distinct songs, the recordings first need to be segmented into the separate pieces of music they contain.

2.1 Segmentation

The metadata found under *Occurrence* in the FIOL database contains information regarding start and end times of each piece of music within a recording, but these are only approximate and therefore cannot be used to segment the songs directly. Instead segmentation of the recordings has been done by a combination of looking at the dynamic evolution of the audio signal and examining the pitch content. Higher, sustained energy indicates music, while low energy indicates silence. By looking at the modulation of pitch in the audio signal we can extract further information on its contents. The pitch of a singing voice or musical instrument will modulate less than the pitch of speech, thereby this can also be used as a parameter for detecting the beginning of a new piece of music.

2.2 Transcription

The melodies of each song are transcribed in order to create a melody representation which can be mapped to the up/down movements generated by the user of the SoundTracer app. Transcription of vocal melodies was performed using autocorrelation functions available in the MIRtoolbox. This strategy, together with pitch transition detection, allows for transcription of monophonic melodies in a vocal recording. A more complex problem arises in transcription of polyphonic melodies generated by instruments such as the Hardanger fiddle. The algorithm developed to extract the melody of such songs for this project involves the use of spectrogram representations of the audio. Recordings of fiddle music usually contain a large amount of overtones or harmonics which complicate the melody extraction process. Although these overtones can have relatively low energy they play an important role in our perception of the note.

By filtering the audio in such a way as to be sensitive to frequencies around 2-4 kHz, similar to how the human ear would focus on certain frequency ranges, spectrograms can be created which give a representation closer to our perception of the sound. In order to create a monophonic melody representation of a piece of music played on an instrument such as a Hardanger fiddle several heuristics are added to the algorithm. These include disregarding drone notes which are typical for music performed with Hardanger fiddle, as well as including a preference for notes in higher pitch registers and alternating to lower frequencies when no activity is registered in the higher ranges. The algorithm developed to handle transcription of Norwegian folk music played on a Hardanger fiddle [1] detects pitch based on the dynamic evolution of partials. The proposed method tracks the pitch contours, or dynamic evolution of the magnitude of each partial. These contours are used to track the perceived melody across time steps. Currently, only the first 15 seconds of each tune is transcribed.

2.3 Gesture Mapping & Queries

The gestures are extracted using the Apple ARKit which is available for the most recent iOS devices. The ARKit registers inertia as well spatial position using the camera. The first version of the SoundTracer app which was demonstrated at the initial SoundTracer workshop in December of 2017 used simply the gross contour of the transcribed melodic line and compared it to the ascending and descending parts of the recorded gesture using a block-based dynamic time warping approach to select the transcribed melodic contour which results in the shortest path. In the second version, exhibited at the 2018 SoundTracer workshop, the app allows the gesture to encompass the onset of a new note by moving the device downwards in a sharp motion. This vertical movement will be interpreted as the onset of a new note in the overall melody contour. Thereby the gestures are matched against both the gross contour of the transcribed melody as well as to the transcribed note onsets, securing that the duration of the notes in the tune match those communicated through the gesture.

3. Review of Objectives

1. How is it possible to extract features from the movements of a mobile phone? Using the ARKit for iOS devices the SoundTracer app can access information regarding spatial position and inertia over a sequence of time thereby modelling ascending and descending motion and velocity.

2. How is it possible to map movement features to sound features? The SoundTracer matches ascending and descending movement to the ascending and descending melodic lines transcribed from the music.

3. How is it possible to search in the audio content of a large music Library? Using the frequency at each of the 15 seconds of audio the melodic contour, note onsets and duration can be extracted and used to match against the gestures recorded by the SoundTracer app. These frequency sequences can be stored in a collection and searched using dynamic programming strategies such as block-based dynamic time warping in order to select a possible match.

4. Future Work

In future versions of the SoundTracer application it would be possible to consider mapping new gestures to other audio features such as rhythmic pulse, musical ornamentation and so on. For simplicity the app uses only the first 15 seconds of a tune to match the contour of the transcribed melody to the ascending and descending movements of the gesture. In future work it may be interesting to experiment with using motivic analysis to generate the contours used for gesture mapping instead of using only the beginning of each tune.

For these first versions of the application the 15 second audio clips are held locally in the app. A possible way forward in order to scale up the application could be to create a database to host the recordings from the folk music database together with the current gesture representation. The database would not need to host audio segments longer than 15 seconds which would decrease the spatial constraints. If suitable, each entry could instead contain a connection to the complete audio file and record in the new, semantic web EBU database or to any web archives that may become available through this resource.

References

[1] Olivier Lartillot, Hans-Hinrich Thedens and Alexander Refsum Jensenius. *Computational model of pitch detection, perceptive foundations, and application to Norwegian fiddle music*, 2018.