A GRAPHICAL USER INTERFACE FOR UNDERSTANDING AUDIO RETRIEVAL RESULTS

Stefan Balke, Meinard Müller

International Audio Laboratories Erlangen, Friedrich-Alexander-Universität (FAU), Germany
{stefan.balke, meinard.mueller}@audiolabs-erlangen.de

ABSTRACT

In 1948, Barlow and Morgenstern released a collection of about 10,000 themes of well-known instrumental pieces from the corpus of Western Classical music [1]. These monophonic themes (usually four bars long) are often the most memorable parts of a piece of music. Using a musical theme as a query, the objective is to identify all related music recordings from a given audio collection. In this demonstration, we describe a graphical user interface which we developed to systematically evaluate the matching results. The goal is to identify the challenges of this particular retrieval scenario and gain more insights into the used data.

1. MATCHING PROCEDURE

An overview of the retrieval procedure is shown in Figure 1. In this example, we use the famous "Fate Motif" from Beethoven's Symphony No. 5 as query. The objective is to retrieve the corresponding documents in a database consisting of audio recordings. Both, the query given in some musical notation and the audio recordings are transformed to chroma features using the *Chroma Toolbox* [3]. Subsequence Dynamic Time Warping (SDTW) is used to compare the query with local section in the audio recordings. This is a well-known approach, and we refer to the literature for further details see, e. g. [2]. As a result, the matching procedure delivers a list of candidates which we use as an indicator for the performance of our matching procedure.

In the following, we assume to have $N \in \mathbb{N}$ musical themes as queries $\mathcal{Q}_1, \ldots, \mathcal{Q}_N$ and $M \in \mathbb{N}$ audio recordings as database documents $\mathcal{D}_1, \ldots, \mathcal{D}_M$, whereas each query matches to a single database document. The objective of the retrieval task is to identify for a given query $\mathcal{Q}_n$, $n \in [1 : N]$, the related audio recording in the database. The retrieval result is reflected by a rank value $r_n(m) \in [1 : M]$ for each database document $\mathcal{D}_m$, $m \in [1 : M]$, for example in Figure 2, we receive for query $\mathcal{Q}_9$ and database document $\mathcal{D}_6$ a rank of $r_9(6) = 1$.
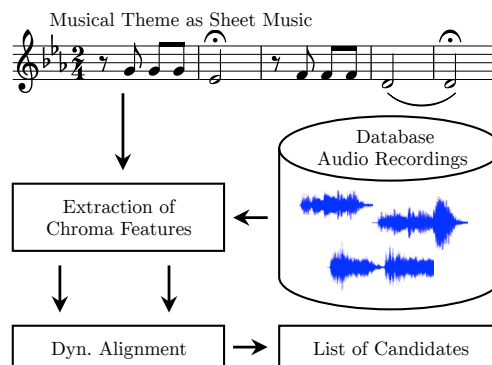
Figure 1. Overview of the retrieval procedure. The sheet music representation of a musical theme and the audio recordings from the database are transformed to chroma features. A DTW-based technique is used to locate the musical themes in the database which results in a ranked list of candidates.

2. EVALUATION GUI

Figure 2 shows the main window of the graphical user interface (GUI), which we implemented using MATLAB®. The top row shows the audio recordings $\mathcal{D}_m$ contained in the database and the first left column lists the used queries $\mathcal{Q}_n$. By pushing one of the oval rectangles, one can inspect the calculated feature representation and listen to the audio or to a sonified version of the musical theme, respectively. For example, Figure 3 shows the chroma feature representation of query $\mathcal{Q}_9$ and the blue bar indicates the current position of the playback.

In the middle of Figure 2, we show all retrieval results $r_n(m)$ as a grid of boxes. Additionally, green background indicates the most relevant match as obtained from manual annotations. By pushing one of the boxes, e. g., $r_9(6)$ (red rectangle), the cost matrix of the corresponding best matching segment is visualized. Additionally, the warping path between the query and this segment is shown as a red line. The green bar at the bottom incorporates the exact postition of the query from the manual annotations. In the case of the shown example, the retrieval result is correct as the relevant database document is identified as the first element of the ranked list

Furthermore, by sonifying the retrieval results, we get a feeling for the problems and challenges the algorithm faces when dealing with this kind of music. We do this by playing back the audio recording at the position of the

Figure 2. Main GUI window. The retrieval results in form of ranking values $r_n(m)$ are mapped to a grid of boxes. The columns represent the audio recordings from the database and the rows the musical themes which were used as query. A green background is used to indicate ground truth annotations (the most relevant document).

estimated match and additionally acoustically overlay this recording with a sonified version of the time-aligned query.

In this way, the graphical user interface can make results from a retrieval system more accessible and also audible. Poorly performing matches can be analyzed and the gained knowledge can possibly be integrated into the algorithm.

3. ACKNOWLEDGMENT

4. REFERENCES

[1] Harold Barlow and Sam Morgenstern. *A Dictionary of Musical Themes*. Crown Publishers, Inc., revised edition, 3. edition, 1975.

[2] Meinard Müller. *Fundamentals of Music Processing*. Springer Verlag, 2015.

[3] Meinard Müller and Sebastian Ewert. Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 215–220, Miami, Florida, USA, 2011.
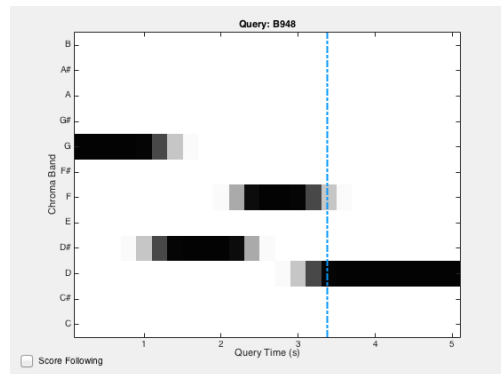
Figure 3. Chroma feature representation of the monophonic theme. The blue bar indicates the playing position.
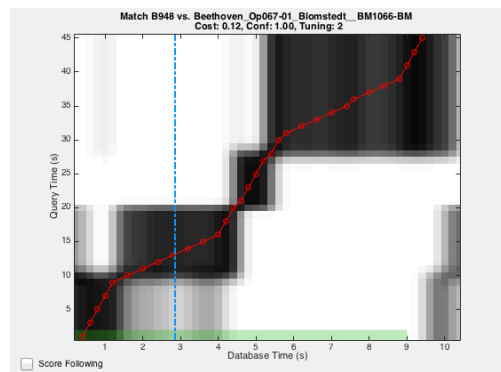

Figure 4. Visualization of the match in the audio recording. The plot shows the cost matrix with the actual warping path obtained from the SDTW. The green bar indicates annotations from the ground truth.