ON FINDING REPEATED STANZAS IN FOLK SONG RECORDINGS

Ciril Bohak, Matija Marolt
University of Ljubljana, Faculty of Computer and Information Science
{ciril.bohak, matija.marolt}@fri.uni-lj.si

ABSTRACT

In this paper we present our current work on an approach for finding repeated stanzas in folk song recordings. We improve our previous work, which relied on detection of vocal pauses to find repetitions, by relying less on prior knowledge of folk song collections. Instead, we calculate similarities between several chunks of the audio signal with the whole signal to obtain repetition patterns. We align the obtained similarity curves and calculate the average similarity curve that represents repetitions in the whole track. Distances between peaks in the obtained curve represent lengths of individual repetitions. Repetitions are aligned with the audio to yield the final segmentation.

1. RELATED WORK

Many approaches for segmenting and finding repeating parts in music recordings were developed in recent years. Most of them are using various audio features such as MFCCs or chroma vectors to calculate self-similarity matrices (Foote (1999)) and were developed for commercial music (Bartsch & Wakefield (2001); Goto (2006); Cooper & Foote (2002); Foote & Cooper (2003); Peeters (2002)). With increased interest in computational folk music analysis, several approaches for segmentation of these recordings were also introduced, based on algorithms such as DTW (Müller et al. (2009); Bohak & Marolt (2012)) and a special fitness measure (Müller et al. (2011)).

2. METHOD

Most of current segmentation methods are developed for commercial music and do not take into the account features of folk music such as inaccurate singing, presence of noise and tempo deviations. To find repeating stanzas in folk song recordings, we modified our previous algorithm (Bohak & Marolt (2012)). In the original approach we used dynamic time warping in combination with shifted chroma vectors to find similarities of short excerpts of audio starting at vocal pauses with the entire song. We assumed that vocal pauses will occur at beginnings of segments where the signal will either have low magnitude or no detectable pitch. Detection of vocal pauses turned out to be unreliable, and was also problematic for choir and instrumental recordings, so with our new approach, we decided to omit this step. Our new approach is presented in Figure 1. Before applying the method we average the audio channels into a single channel and normalize it.

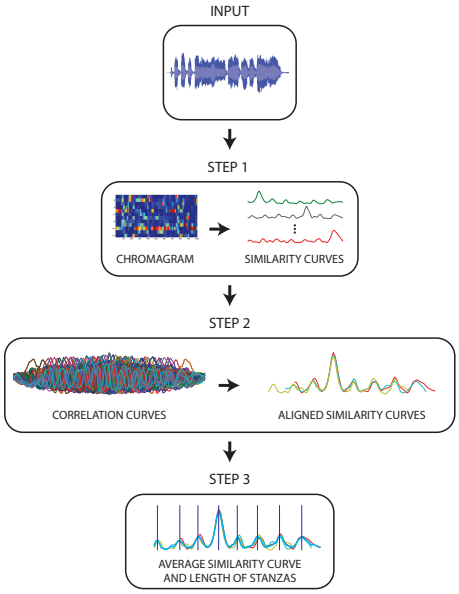

Figure 1: Outline of dual domain method for finding repeating stanzas.

2.1 Step 1 - Extracting chomagram and calculating the similarity curves for randomly selected parts

In the first step we calculate the CENS chromagram (Müller (2007)) for the whole audio track. Next we randomly select n locations in the audio track that are approximately equally distributed throughout the whole track. We use m seconds long chunks of the chromagram at selected locations and calculate similarity between the selected chunks and the entire song with Dynamic time warping and shifting chroma vectors. This results in n similarity curves that represent the similarity between selected chunks and the whole track. An example is given in Figure 2. Peaks in these curves represent repetitions of chunks in the track.

2.2 Step 2 - Aligning the similarity curves

In the second step we calculate cross-covariances between each pair of similarity curves to find those that are most similar. We select the similarity curves with above mean maximum cross-covariance value and smooth them with a low pass filter. Amongst these, we select the curve with the highest similarity to all other curves as the most representative one. As the curves are shifted in time, due to the randomly selected audio chunks, we then calculate time shifts between the selected most representative curve and

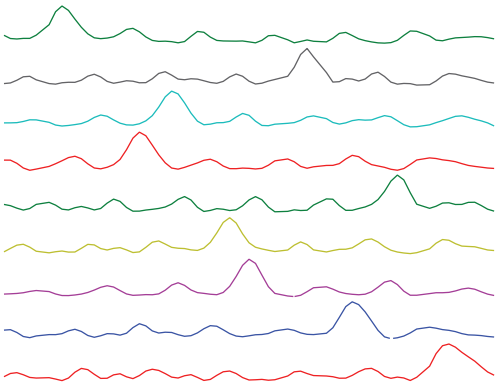

Figure 2: Similarity curves for random locations in a track.

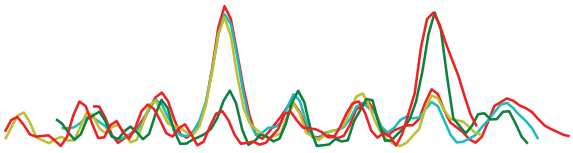

Figure 3: Curves aligned according to the calculated time shifts. Not all the curves were selected for alignment due to thresholding by mean covariance.

all the others. We use the calculated time shifts to align the curves as shown in Figure 3.

2.3 Step 3 - Calculating the average similarity curve and length of repeating stanzas

In the last step we calculate the average similarity curve that represents repetitions in the audio track. The average curve is calculated as the average of all aligned similarity curves as shown in Figure 4. By calculating the distances between peaks in the obtained average similarity curve we can calculate lengths of individual stanzas. We determine the beginning of the first stanza by cutting the silence from beginning of the track and use the calculated distances to find beginnings of all other stanzas.

3. CONCLUSIONS AND FUTURE WORK

In this paper we presented our current work in progress on a method for finding repeated stanzas in folk song recordings. In the future we are planning on extending our method to a double domain approach, in which we will augment results from the presented approach with an algorithm that works on the symbolic domain. Symbolic data will be obtained with polyphonic transcription. We plan to use approximate string matching approaches on the obtained data to find the repeating parts and merge both approaches into a robust segmentation algorithm.

Figure 4: The average curve (left) and calculated lengths of stanzas (right)

4. REFERENCES

Bartsch, M. A. & Wakefield, G. H. (2001). To catch a chorus: Using chroma-based representations for audio thumbnailing. In *Applications of Signal Processing to Audio and Acoustics*, (pp. 15–19)., New Platz, NY , USA.

Bohak, C. & Marolt, M. (2012). Finding repeating stanzas in folk songs. In *Proceedings of the 13th International Conference on Music Information Retrieval (ISMIR)*, (pp. 451–456).

Cooper, M. L. & Foote, J. (2002). Automatic music summarization via similarity analysis. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, (pp. 81–85)., Paris, France.

Foote, J. T. (1999). Visualizing music and audio using self-similarity. In *Proceedings of the 7th ACM international conference on Multimedia*, (pp. 77–80).

Foote, J. T. & Cooper, M. L. (2003). Media segmentation using self-similarity decomposition. In *Proceedings of SPIE Storage and Retrieval for Multimedia Databases*, (pp. 167–175).

Goto, M. (2006). A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech & Language Processing*, *14*(5), 1783–94.

Müller, M. (2007). *Information Retrieval for Music and Motion*. Springer Verlag.

Müller, M., Grosche, P., & Jiang, N. (2011). A segment-based fitness measure for capturing repetitive structures of music recordings. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*.

Müller, M., Grosche, P., & Wiering, F. (2009). Robust segmentation and annotation of folk song recordings. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, (pp. 735–740)., Kobe, Japan.

Peeters, G. (2002). Toward automatic music audio summary generation from signal analysis. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, (pp. 94–100)., Paris, France.