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.

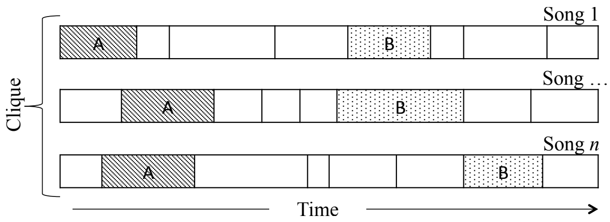

Figure 1. A clique of songs with two sections clustered and labelled across all versions.

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.

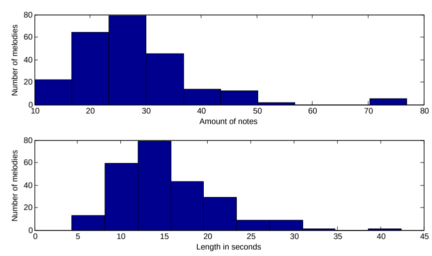

Figure 2. Top: the distribution of amount of notes. Bottom: the distribution of length (in seconds).

[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.

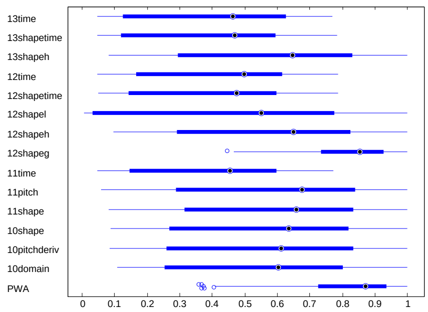

Figure 3. Mean Average Precision for the MelodyShape methods and the pairwise alignment (PWA).

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.

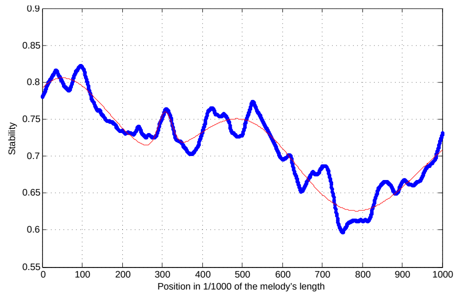

Figure 4. Probability of an event being stable given its position in the melody.

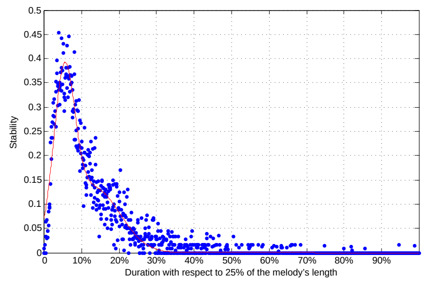

Figure 5. Probability of an event being stable given its duration value with regard to the 25% of the melody's length.

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.