Search by lyrical phrases in acapella Turkish Makam recordings

April 27, 2015

Abstract

In this work we propose an approach for locating the exact occurrences of a lyrical query in musical audio, a problem known in speech processing research as keyphrase-spotting. A query is constructed by expanding text to MFCC-based phoneme models, which are trained only on speech. To address the differences of syllable durations, specific for singing, durations inferred from musical score are incorporated in the phonetic query.

In a first step we apply a dynamic time warping to estimate candidate segments. In a second step these audio segments are ranked by means of a novel hierarchical hidden Markov model (HHMM), which models a query as a separate structural section.

The proposed approach is evaluated on a small acapella dataset of recordings from Turkish Makam recordings. Results show that the combination of the good recall of the DTW and the good precision of HHMM is promising, even with the standard phoneme speech models.

A comparison to a s.o.a. keyword-spotting system is done.

Being on of the first methods for searching by lyrics, and the first on non-eurogenetic music in particular, we expect that it can serve as a baseline for further MIR research on the topic.

1 Introduction

TODO: rewrite first paragraph and reorder

In this work we investigate the problem of locating the exact occurrences of a lyrical query from performance recording for a particular composition. We address the case when a query represents an entire structural section or phrase from textual lyrics. The composition is known in advance, but no information about the structure of the particular performance is given. This problem is comparable to phrase-spotting when considering speech recordings ([ref]). We assume that the musical score with lyrics is present for the composition of interest.

It has ben shown the durations of singing voice are quite different than in speech [Anna]. Therefore adopting an approach from speech recognition might lack some singing-specific rules (or semantics) including among others note durations. Hitherto approaches do not rely on temporal information. A lot of this information can be inferred from musical scores.

Why is it important: Search by lyrics has an inherent connection to the problem of structure discovery. For most types of music a section-long lyrical phrase is a feature that represents the corresponding structural section in a unique way.

2 Architecture

Figure 1 presents an overview of the proposed approach.

We propose a two-pass retrieval approach: On the first pass a subsequence DTW retrieves a set of candidate audio segments that roughly correspond to a query. On the second pass each candidate segment is separately fed into the HHMM, for which we run a Viterbi decoding to assure that only one (the most optimal) path is detected for an audio segment. Any query-to-audio fullpath match is considered as a hit and all results are ranked according to their respective Viterbi likelihoods.