Panako - A Scalable Acoustic Fingerprinting System
Handling Time-Scale and Pitch Modification

Joren Sij and Marc Leman - joren.sij@ugent.be - IPEM, University Ghent - Belgium

Abstract
This poster presents a scalable granular acoustic fingerprinting system. An acoustic fingerprinting system uses condensed representations of audio signals, acoustic fingerprints, to identify short audio fragments in large audio databases. The system presented here is shown to answer queries quickly and reliably even when queries are subjected to time-scale and pitch modifications. The design of this system is the main contribution of this research.

Introduction
Acoustic fingerprinting systems have many practical uses. They follow the scheme depicted in Figure 1. Ideally, a fingerprinting system only needs a short audio fragment to find a match in a large set of reference audio. One of the challenges is to design a system in a way that the reference database can grow to contain millions of entries. Another challenge is that a robust fingerprinting should handle noise and other modifications well, while limiting the amount of false positives and processing time [1]. These modifications typically include dynamic range compression, equalization, added background noise and artifacts introduced by audio coders and A/D-D/A conversions.

Over the years several efficient acoustic fingerprinting methods have been introduced [2,3]. These methods perform well even with degraded audio quality and with industrial-sized reference databases. However, these systems are not designed to handle queries with modified time-scale or pitch although these distortions can be present in replayed material. During radio broadcasts songs are occasionally played faster to make them fit into a time slot. During a DJ-set pitch-shifting and time-stretching are present almost continuously. To correctly identify audio in these cases as well, a fingerprinting system robust against pitch-shifting and time-stretching is desired.

Some fingerprinting systems have been developed that take pitch-shifting into account [6]. Others are designed to handle both pitch and time-scale modifications [9,3]. To find a match, these modifications are expensive systems iterate the whole database. To the best of our knowledge, a description of a practical fingerprinting system that allows substantial pitch-shift and time-scale modification can only be found in [7], and in this work.

Method
The proposed method is inspired by three works [2,4,6]. Combining key components of those works results in a design of a granular acoustic fingerprinter that is robust to noise and substantial compression, has a scalable method for fingerprint storage and matching, and allows time-scale modification and pitch-shifting.

The method presented here uses local maxima in a spectral representation [2]. It combines three event points, and takes time ratios to form time-scale invariant fingerprints [4]. It leverages the Constant Q transform, and only stores frequency differences for pitch-shift invariance [6]. The fingerprints are designed with an exact hashing matching algorithm in mind [2].

The whole process is depicted in Figure 2.

Results & Conclusions
The system has been evaluated using a freely available data set of 30,000 songs and compared with a baseline system (see Figure 4, 5, 6, 7).

This work presented a practical acoustic fingerprinting system. The system allows fast and reliable identification of small audio fragments in a large set of audio, even when the fragment has been pitch-shifted and time-stretched with respect to the reference audio. If a match is found the system reports where in the reference audio a query matches, and how much time and frequency has been modified. To achieve this, the system uses local maxima in a Constant Q spectrogram. It combines event points into groups of three, and uses time ratios to form a time-scale invariant fingerprint component. To form pitch shift invariant fingerprint components only frequency differences are stored. For retrieval, an exact hashing matching algorithm is used.

Availability & Reproducibility
The Panako software is available on http://panako.info under the AGPL. Panako is limited to Debian and Mac OS X, but works with every platform with a recent Java implementation.

The datasets used in the evaluation is freely available from Javascriptson, a website where artists share their work freely, under various creative commons licenses.

To reproduce the results, scripts are available to download the audio dataset, generate query files, run the reference audio and query the system. Supporting links to download the query results are available as well.

References

15th ISMIR Conference, Taipei, Taiwan, 27-31 October, 2014