Predicting Key Recognition Difficulty in Music Using Statistical Learning Techniques

Ching-Hua Chuan, School of Computing, University of North Florida, Jacksonville, FL, USA
Aleksey Charapko, School of Computing, University of North Florida, Jacksonville, FL, USA

ABSTRACT

In this paper, the authors use statistical models to predict the difficulty of recognizing musical keys from polyphonic audio signals. The key recognition difficulty provides important background information when comparing the performance of audio key finding algorithms that often evaluated using different private data sets. Given an audio recording, represented as extracted acoustic features, the authors applied multiple linear regression and proportional odds model to predict the difficulty level of the recording, annotated by three musicians as an integer on a 5-point Likert scale. The authors evaluated the predictions by using root mean square error, Pearson correlation coefficient, exact accuracy, and adjacent accuracy. The authors also discussed issues such as differences found between the musicians’ annotations and the consistency of those annotations. To identify potential causes to the perceived difficulty for the individual musicians, the authors applied decision tree-based filtering with bagging. By using weighted naïve Bayes, the authors examined the effectiveness of each identified feature via a classification task.

Keywords: Audio Key Finding, Key Recognition Difficulty, Multiple Linear Regression, Proportional Odds Model, Weighted Naïve Bayes

INTRODUCTION

Automatically identifying keys in music recordings is an integral step in many content-based music indexing and retrieval tasks. In Western tonal music, the key creates a system that defines the roles of fundamental music elements such as pitch and chord. Pioneering research work (Pauws, 2004; İzmirli, 2005; Gómez, 2006b) on audio key finding heavily relies on key templates developed via the theoretical and perception-based approaches by Krumhansl (1990) and Temperley (1999). Systems based on geometrical models (Chew, 2000) can also perform real-time key extraction (Chuan & Chew, 2005). Currently, rule-based approaches based on music theory (Weiß, 2013) remain popular. Data-driven approaches have also been applied to audio key finding. For example, hidden Markov models have been used to determine key (Peeters, 2006) as well as key and chord simultaneously (Burgoyne & Saul, 2005; Chai & Vercoe, 2005; Noland & Sandler, 2007; Papadopoulous & Peeters, 2012). Support vector machines have also been used to determine

DOI: 10.4018/ijmdem.2014040104
Although various audio key finding algorithms have been proposed, evaluating their effectiveness is difficult. Because music compositions are complex, it is not straightforward to determine a ground truth (the key considered correct in this case) to be compared with the system-generated answer. For example, it is not unusual for a composition to use more than one key or for musicians to disagree about the key or keys it uses. To ensure that the ground truth is determined objectively, many researchers focus on classical pieces from the common practice period and use the title key as the ground truth (Pauws, 2004; Chuan & Chew, 2005; Peeters, 2006). However, a particular composition may not use the title key throughout the entire piece (Chuan & Chew, 2012). Another way to obtain a ground truth is to ask musicians to manually annotate the key throughout a composition (Chai & Vercoe, 2005; Schuller & Gollan, 2012; Papadopoulous & Peeters, 2012; Weiβ, 2013). However, this process is extremely time-consuming, which limits the size of the research using it, and can be problematic when musicians’ annotations disagree (Chuan & Chew, 2012). To enable large-scale evaluations without sacrificing the accuracy of the ground truth, it would be desirable to have an automated process that identifies pieces with less uncertainty in their ground truths, leaving musicians with fewer pieces in question that would require manual examination.

In this paper, we propose methods to predict key recognition difficulty from audio signals by using statistical learning. Modeling the key recognition difficulty rather than directly detecting the key provides another way to select pieces that need manual annotation. Not only will correctly predicting the difficulty levels quicken the key finding process by filtering out the easy-to-judge tracks, it will also provide important background information about the private data set used to evaluate the algorithm. With this information, the reported accuracy can be compared and interpreted more meaningfully when evaluating systems that use different private data sets.

To predict the key recognition difficulty, we first extracted low-level features from audio recordings using Mel-frequency cepstral coefficients and features related to harmony and timbre. We then applied multiple linear regression and proportional odds model to predict the perceived key recognition difficulty. We conducted experiments using 1083 classical pieces, each manually annotated by three musicians. In addition to examining the prediction accuracy, we also examined the differences between the musicians’ annotations and the consistency of their annotated labels.

Just as knowing the key recognition difficulty can provide useful information for music education, so can knowing the potential cause for such perceived difficulty. To be able to meaningfully identify the cause, we extracted high-level features related to various musical elements, including tuning, pitch, and chord. To select attributes, we then applied decision tree–based filtering with bagging. We used the depth of the attribute in the decision tree to compute its weight in naïve Bayes, which was trained to classify audio tracks into difficult and easy categories. The computed attribute weight indicates how much the participating musician relied on a certain musical element to determine the key, while the classification output demonstrates how effectively the selected elements can be used to describe the difficulty.

The following sections first describe our work in predicting the perceived key recognition difficulty level of a recording, followed by methods to find the potential contributors to that difficulty for individual musicians.

**PREDICTING KEY RECOGNITION DIFFICULTY**

In the first part of this paper, we focus on predicting the level of difficulty perceived by musicians when they attempt to determine the key while listening to a piece of music. We aim to answer the following questions related to key recognition difficulty:
JIRL: A C++ Toolkit for JPEG Compressed Domain Image Retrieval
www.igi-global.com/article/jirl/84022?camid=4v1a