

Introduction to Statistical Modeling in Transcription Using One Example of a Humming Transcription Model

Michael Murray

February 6, 2004

1 Introduction

Humming transcription represents a manifestation of the most basic form of melodic transcription : monophonic transcription. Much work has been done on the subject and I chose to research one relatively new and untested method which was published for ICME 2003. The title of the paper describing the method is: *A Statistical Multidimensional Humming Transcription Using Phone Level Hidden Markov Models for Query by Humming Systems. The paper was written by by Hsuan-Huei Shih, Shrikanth S. Narayanan and C.-C. Jay Kuo* [1] This paper will serve as an example of transcription using statistic modeling.

The purpose of statistical modeling in transcription is to build a database of information of the type of music you wish to transcribe, hopefully as similar as possible to the music you wish to transcribe, in order to use prediction as a tool in the transcription. Prediction based on a statistical database will ideally help the accuracy of your transcription. The complete transcription process can be broken down into the data collection phase and the analysis and transcription phase.

2 Data Collection

The first step in data collection is to choose a statistical model to base the database on. Some examples of statistical models are nearest neighbor approximation, gaussian estimation and the model used in our example, hidden markov modeling. A statistical model will take data collected from repeated

trials, in this case human humming of melodies, and placing the results in a pattern which can be compared to further trials. Hidden Markov Modeling is summarized by Narada Warakagoda in his webpage on the subject: “The Hidden Markov Model (HMM) is a finite set of states, each of which is associated with a (generally multidimensional) probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are “hidden” to the outside; hence the name Hidden Markov Model” [3] In our example basic musical lines were hummed by test subjects repeatedly and the model was built upon each note as a state and the interval between notes as the change of states. This way the model addresses the inaccuracy of human humming. Since you are only concentrating on the intervals between notes to identify melody, subjects which sing incorrect notes can still have their melodies identified. Further, in this model, incorrect intervals can still be characterized using only their direction and relative magnitude. The choice of HMM for this application is a good one and reveals that for different transcription methods, one statistical method will probably be more appropriate than another.

3 Transcription

The actual transcription in a method that includes statistical modeling must somehow refer to its statistical database. In our example model, after a statistical database is collected using HMM, only melodies collected in the database phase can be properly transcribed. When the end user hums into the system, from the second note hummed (the first interval recorded) the melody being hummed can be compared to the models in the collected database and matching melodies can be returned. Because of the statistical models, even inaccurate melodies can be guessed quickly, (the authors claiming realtime). In the transcription phase, however a given method refers to the statistical data it is built upon it should increase accuracy and speed of the transcription. This is because using a statistical model, neither the data collected to build the database, nor the input to be transcribed need be one hundred percent accurate, and in practice this is often the case

4 Conclusion

There are many transcription methods but those that use statistical modeling have a distinct advantage in the accurate transcription of non-accurate sources. For further studies on transcription methods it is suggested that you navigate your web-browser to Ansi Kalpurri's review of transcription studies, <http://www.cs.tut.fi/~7Eklap/iiro/literature.html> [2].

References

- [1] Shrikanth S. Narayanan Hsuan-Huei Shih and C.-C. Jay Kuo. A statistical multidimensional humming transcription using phone level hidden markov models for query by humming systems. *Proceedings of the ICME 2003*, 2003.
- [2] Ansi Kalpurri. Automatic transcription of music: Literature review. *Tampere University of Technology Website*.
- [3] Narada Warakagoda. Hidden markov models. *Budapest University of Technology and Economics Physics Department Student Page*, 1996.