Making Searchable Melodies: Human vs. Machine

NORTHWESTERN UNIVERSITY

Mark Cartwright, Zafar Rafii, Jinyu Han, Bryan Pardo
Department of EECS, Northwestern University, Evanston, IL
music.cs.northwestern.edu mcartwright@u.northwestern.edu

interactive audio lab

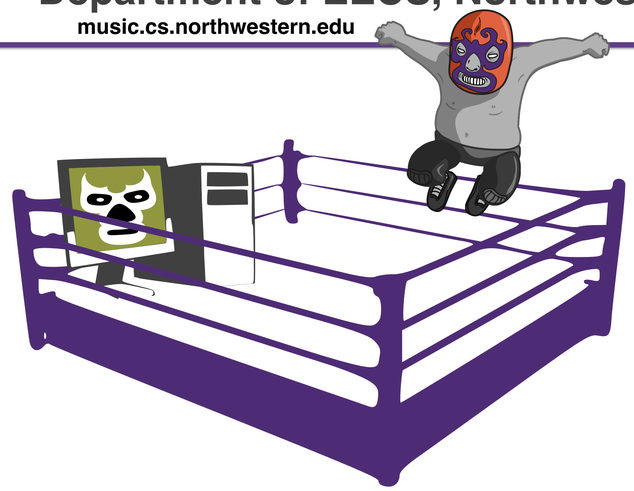
Introduction

Systems that find music recordings based on hummed or sung, melodic input are called Query-By-Humming systems. Such systems employ search keys that are more similar to a cappella singing than the original recordings. Successful deployed systems use human computation to create these search keys: hand-entered MIDI melodies or recordings of a cappella singing. Tunebot is one such system. In this work, we compared search results using keys built from two automated melody extraction system to those gathered using two populations of humans: local paid singers and Amazon Turk workers.

Four Methods for Generating Search Keys

We generated search keys for a target set of 100 songs with four different search key generation methods. The 100 target songs were from popular music genres.
The four methods are described below:

1. **REpeating Pattern Extraction Technique (REPET) (Machine Generated)**
   - REPET is a novel and simple approach for extracting the repeating musical background from the non-repeating musical foreground in an audio signal. For more information see [3].
   - **Search key duration**: length of entire song
   - **Cost**: Less than $1 per song (cost of MP3 download + computer + electricity)
   - **Time**: 15272 songs per week (implemented in Matlab on an Intel Core2 Quad 2.66 GHz CPU)
2. **Probabilistic Latent Component Analysis (Machine Generated)**
   - This method models the spectrogram of a piece of polyphonic music as a two-dimensional distribution in time and frequency. A statistical model of the non-vocal segments of the signal is learned adaptively and employed to remove the accompaniment from the mixture, leaving mainly the vocal components. For more information see [2].
   - **Search key duration**: length of entire song
   - **Cost**: Less than $1 per song (cost of MP3 download + computer + electricity)
   - **Time**: 831 songs per week (implemented in Matlab on an Intel Core Quad 2.4 GHz CPU)



3. **Local Paid Singers (Human Generated)**
   - Solicited through flyers and student job postings
   - Search key almost always contained verse, chorus, or both
   - **Mean search key duration**: 24.46s
   - **Cost**: $3.60 per song
   - **Time**: 50 songs per week
4. **Amazon Turk Singers (Human Generated)**
   - Solicited by posting a $0.10 Human Intelligence Task on Amazon Mechanical Turk
   - Search key almost always contained the verse or chorus
   - **Mean search key duration**: 26.02s
   - **Cost**: $0.60 per song (6 contributions per song)
   - **Time**: 35 songs per week

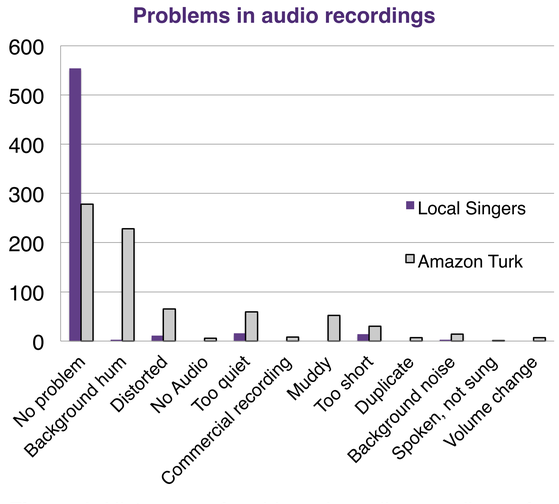


Figure 1. Histogram of problems in audio recordings of paid local singers vs. audio problems in recordings of Amazon Turk workers.

Evaluating the Four Methods

To evaluate each method, we inserted keys generated by that method into an existing database with more than 10,000 search keys used by Tunebot. We then took a set of sung queries drawn either from the Amazon Turkers or from the local singers and queried the database. The search key generation method that yielded better search rankings was deemed better.

Results

Table 1 shows the mean reciprocal rank (MRR). This measurement ranges from 1 to 1/N, where N is the number of items in the database. Higher MRRs are better. The performance of all of the automatic key generation methods was extremely weak. When using the Amazon Turk contributions as queries, there was not a significant difference (p=0.29) between the performance of the runs with local singer generated targets and runs with Amazon Turk generated targets. However, when using local singer contributions as queries, the performance of the runs with local singer generated targets is better than with the Amazon Turk generated targets (p=1.4e-5).

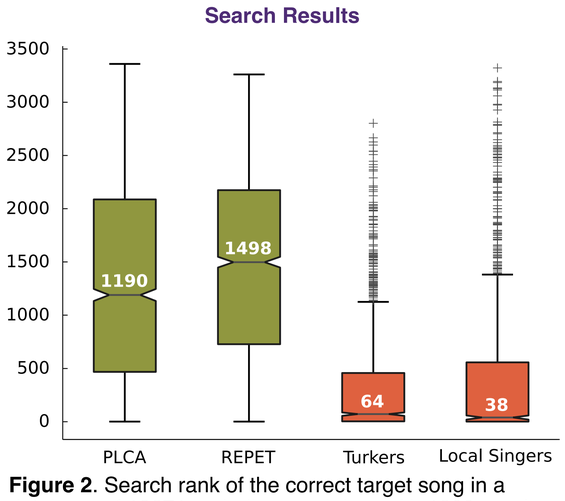


Figure 2. Search rank of the correct target song in a database of 13271 melodic search keys. N=1200 queries per boxplot. Lower numbers are better. Values in boxes are medians. Each boxplot shows search results using search keys generated with the specified method.

Target Source	Query Source	Mean Reciprocal Rank (95% CI)
PLCA	A. Turk	0.0100 (CI [0.0066,0.0165])
Repet	A. Turk	0.0119 (CI [0.0072,0.0191])
A. Turk	A. Turk	0.2140 (CI [0.1909,0.2367])
L. Singers	A. Turk	0.2453 (CI [0.2229,0.2705])
PLCA	L. Singers	0.0128 (CI [0.0086,0.0193])
Repet	L. Singers	0.0087 (CI [0.0060,0.0132])
A. Turk	L. Singers	0.2942 (CI [0.2710,0.3218])
L. Singers	**L. Singers**	**0.3781 (CI [0.3467,0.4047])**

Table 1. Mean reciprocal rank of the 8 conditions

Conclusions

It appears that automated vocal melody extraction methods are still not ready for real world scenarios. The system works best when local singers are used to create both the searchable keys and the queries. It's likely however that the Amazon Mechanical Turk generated queries are more like real world queries. In that case there is no difference in performance between generating search keys by outsourcing to Amazon Mechanical Turk workers and generating search keys by paying local singers. Since the outsourced keys cost 16.67% of the cost of the local paid singers, this seems promising. However, it currently on average takes 40% more time to generate the keys with the outsourced method, so more work will need to be done to increase the throughput of this method while keeping the price down.

Acknowledgements

This work was funded by National Science Foundation Grant number IIS-0812314.

References

[1] J. Han and C.-W. Chen: "Improving Melody Extraction Using Probabilistic Latent Component Analysis," Proceedings of the IEEE ICASSP 2011.

[2] Z. Rafii and B. Pardo: "A Simple Music/Voice Separation Method Based on the Extraction of the Repeating Musical Structure," Proceedings of the IEEE ICASSP 2011.