

# Automatic Transcription and Arrangement Analysis for A Cappella Music



Joseph (Tre) DiPassio

University of Rochester Department of Electrical and Computer Engineering

## Abstract

This work explores the implementation of an automated transcription suite specifically designed for A Cappella singing groups. In this work, a system for recording vocalists and vocal percussionists (VPs) during an improvised musical work will be proposed, as well as a system by which their improvisation is rated against the conventions of classical music theory. The use of cascaded statistical filtering over the passage's beats combined with a priori knowledge about the tendencies of the singing group is explored to improve the accuracy of the automated transcription system, as well as the use of bi-gram methods borrowed from language processing to analyze the passages chord progression.

## Objectives

- Use knowledge of a singing group's tendencies to improve automated transcription of their improvisational passage
- Alleviate the effects of noise, vibrato, and "swooping" onsets on transcription methods
- Use template matching and bi-gram models to identify issues in the observed chord progression and suggest improvements based on music theory

## Data Source

The data used to determine the tendencies of each singer in the experiment came from the repertoire of the RIT Brick City Singers, the A Cappella group the author of this work was a member of at the Rochester Institute of Technology

## References

- [1] Eerola, T. & Toivainen, P. (2004). *MIDI Toolbox: MATLAB Tools for Music Research*. University of Jyväskylä: Kopijyvää, Jyväskylä, Finland.
- [2] Wang, DeLiang, and Guy J. Brown. *Computational auditory scene analysis: Principles, algorithms, and applications*. Wiley-IEEE press, 2006.
- [3] Lu, H. L. (1999). A Hybrid Fundamental Frequency Estimator for Singing Voice.
- [4] Oudre, L., Grenier, Y., & Févotte, C. (2009, October). Template-based Chord Recognition: Influence of the Chord Types. In *ISMIR* (pp. 153-158).
- [5] Suhm, B., & Waibel, A. (1994). Towards better language models for spontaneous speech. In *Third International Conference on Spoken Language Processing*.

For additional resources who's results are not discussed in this poster but are utilized in the paper, please consult the paper's list of references

## Determining A Priori Pitch Information

An important element of any A Cappella group is their repertoire. A great amount of Collegiate A Cappella groups have made the switch to completely digitizing their repertoire. This has the added advantage of allowing MIDI tracks to be stored for each piece.

Using the MIDI Toolbox in [1], the entire repertoire of a group can be imported into the workspace and analyzed, with the result being a probability distribution of all notes within each voice parts range. A MIDI-valued histogram can be created for each voice part over the entire repertoire of the group and normalized to create a probability distribution.

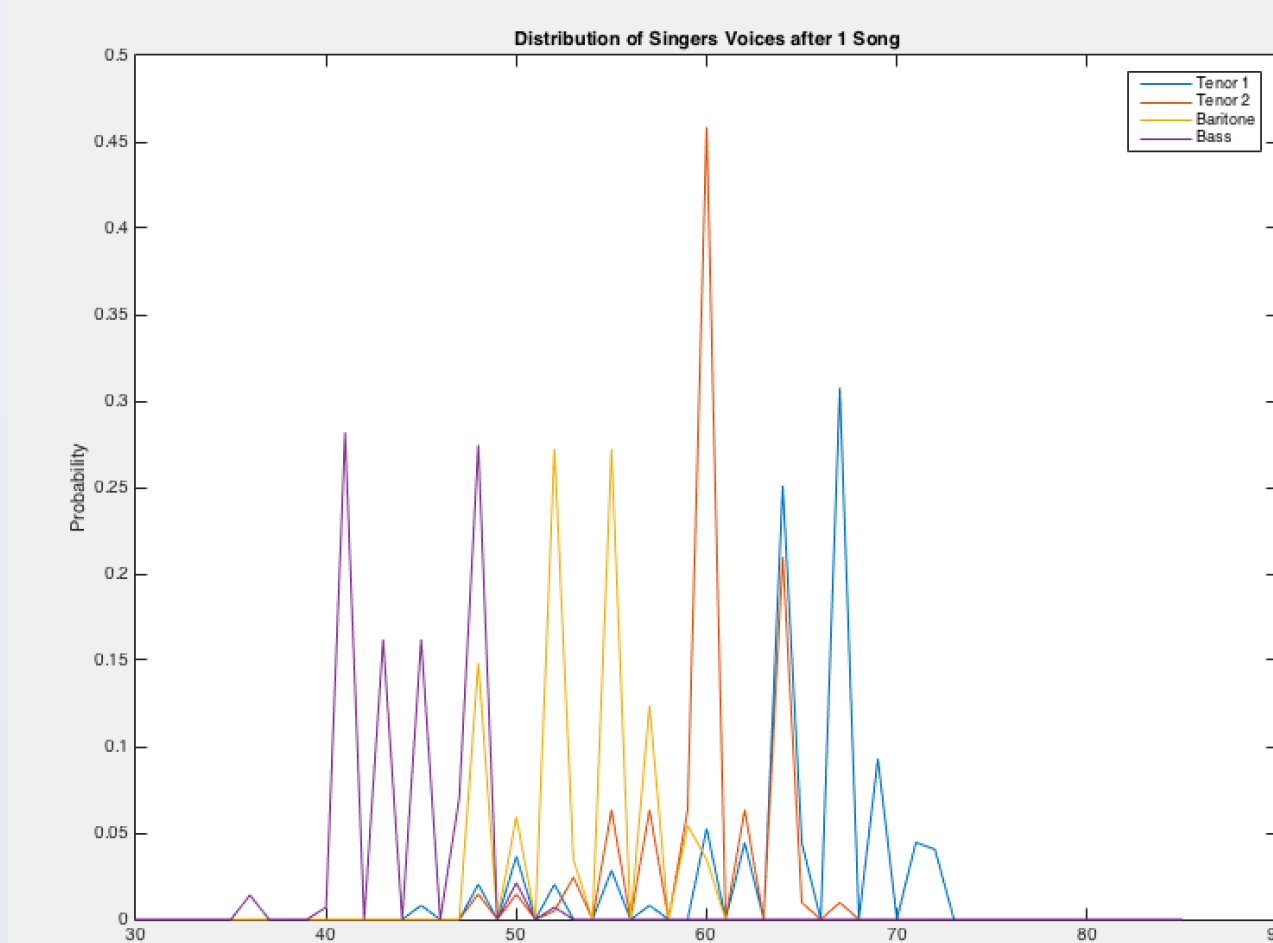


Figure 1: Distribution after One Song of Training



Figure 2: Distribution over Entire Repertoire

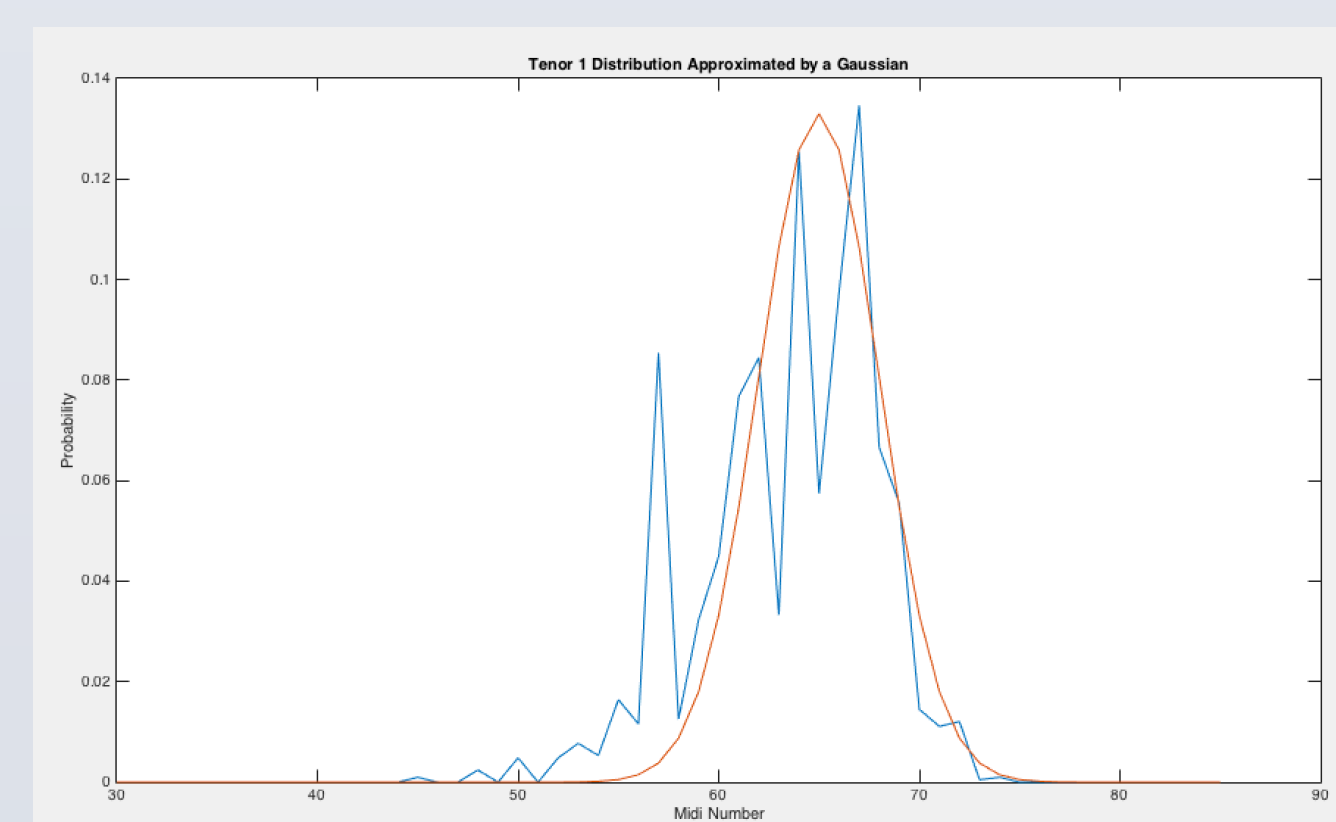


Figure 3: Approximate Gaussian for Tenor 1 Voice

By the principle of the central limit theorem, and assuming each piece in the repertoire is relatively independent and independently distributed, the resulting probability function should resemble a Gaussian. As seen in Figure 3, a Gaussian can be used a relatively good fit for the singer's distribution, hence providing good preliminary results

## Pitch Detection

The pitch detection algorithm used in this work is a spectral approach involving pattern matching, as discussed in [2]. The following shows the governing equation:

$$Bin_f = \sum_j \sum_i I[(Peak_j / i) = f]$$

The resulting histogram will be weighted with the A Priori information. Median filtering will be performed across the sequence of estimated fundamental frequencies. This type of filtering should alleviate some noise concerns. Quantizing each beat of the passage via mode filtering will not only allow a standard subdivision of the audio to be defined, but it will be tested for it's ability to filter out swooping effects and partially eliminate the artifacts caused by singer vibrato.

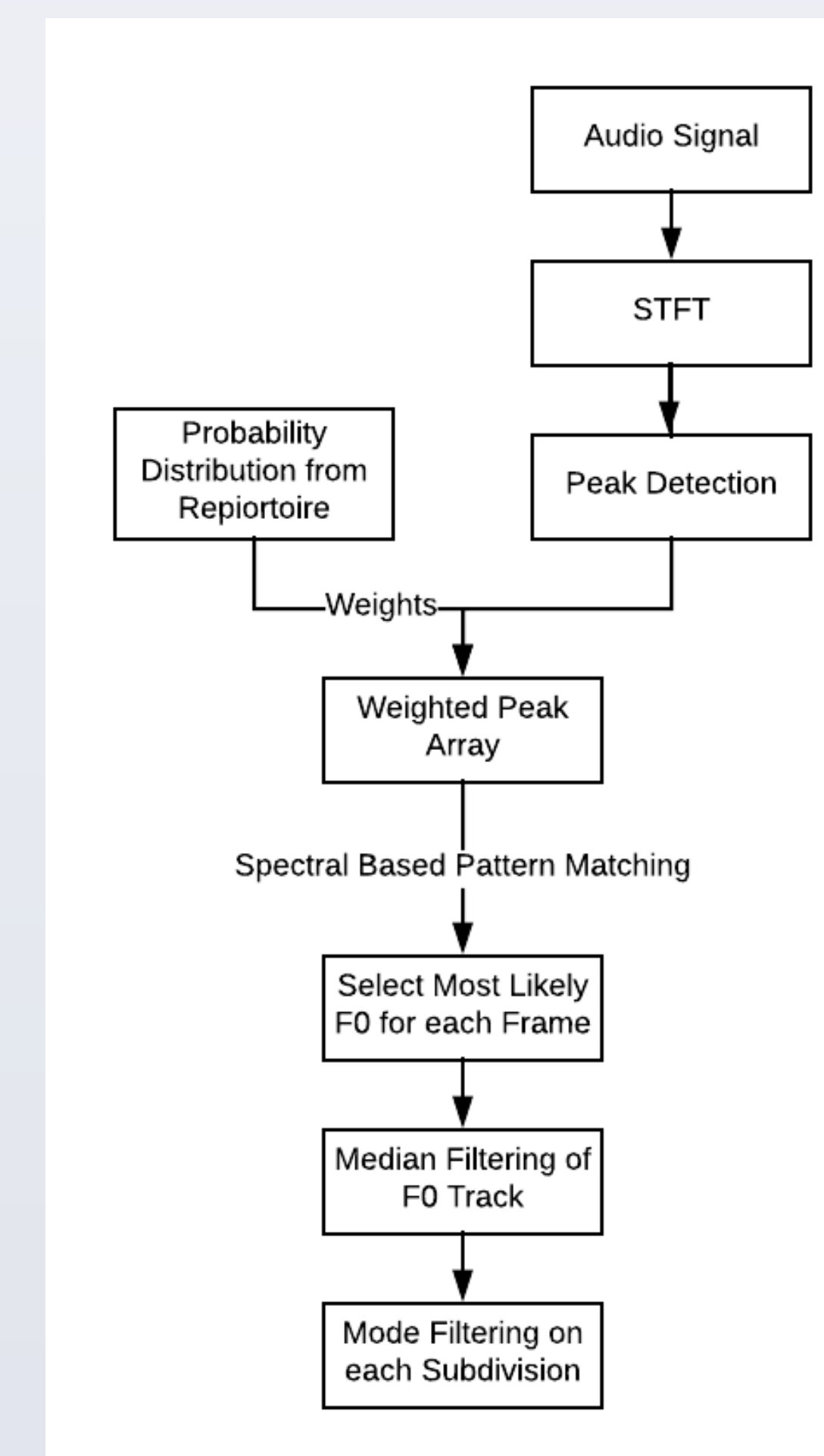


Figure 4: Cascade Filtering Algorithm

The cascading filters show promising results in terms of eliminating the impacts of swoops, vibrato, and noise. The filtered signal in Figure 5 proved to be a perfect transcription, as the short passage sung was [C4 B3 A3 G3], which is exactly what is seen in the frequency domain.



Figure 5: Simple Transcribed Passage

## Chord Recognition

The chord recognition system is based on a simple template matching algorithm based on the work in [4]. The chords in each beat are then treated as a bi-gram model, similar to results from language processing, which is explored in [5]

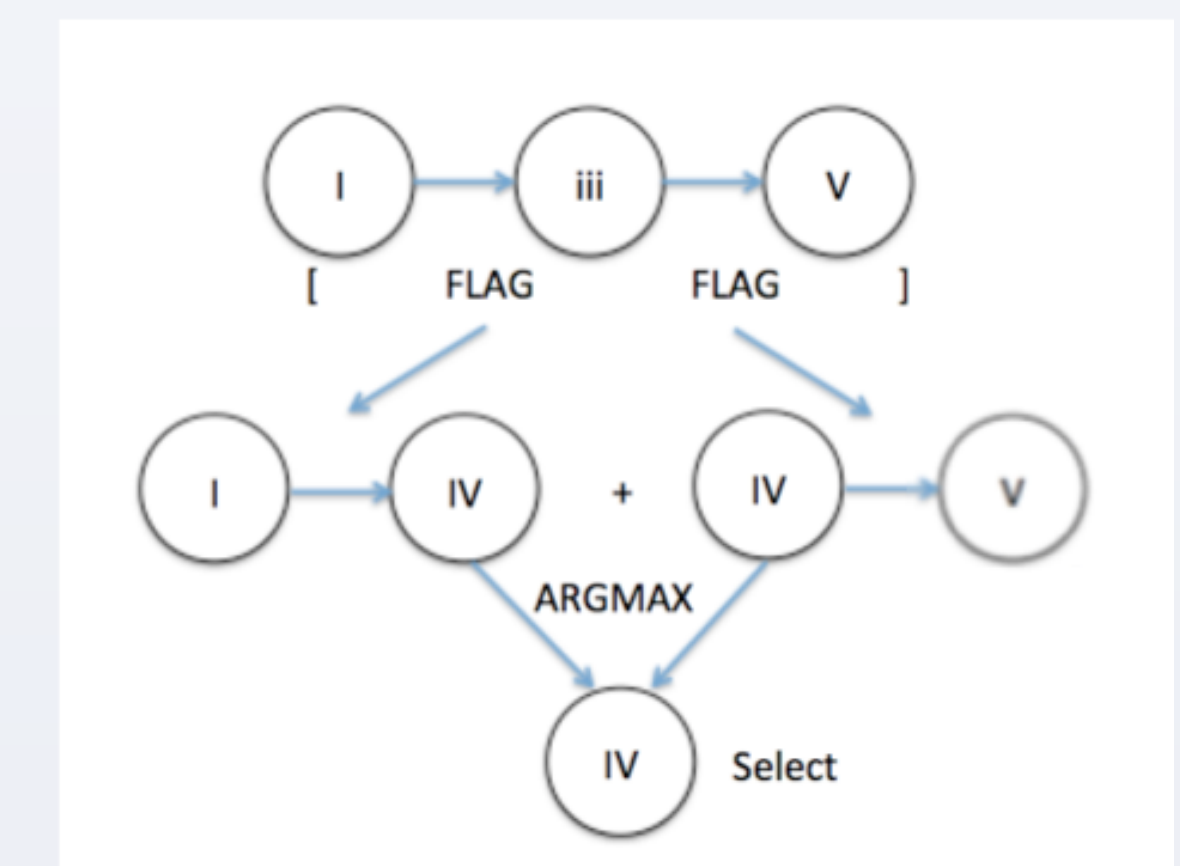


Figure 6: Chord Recommendation Example

## Results

Input Notes	Possible Chords	% Match	Selected?
C, E, G	C Major	100%	Y
	E Minor	66%	
	A Minor	66%	
	G Major	33%	
D,F#,A,B	B Minor 7	100%	Y
	D Major	63%	

Table 1: Template Matching Results



Figure 7: Selected Baritone Passage Transcription

Part	% of Beats Correctly Transcribed	
	Without Weights	With Weights
Tenor 1	72.91%	75.00%
Tenor 2	81.25%	85.41%
Baritone	81.25%	87.50%
Bass	85.41%	79.20%

Table 2: Transcription Results Based on 4 Sets of 16 Beats

## Discussion

- Promising results when using the a priori information, larger dataset needed
- Irregular bass singer distribution might explain regression in their results
- More robust baseline transcription needed for widespread distribution
- Incorporation of a trained dictionary to separate and transcript vocal percussion would be useful