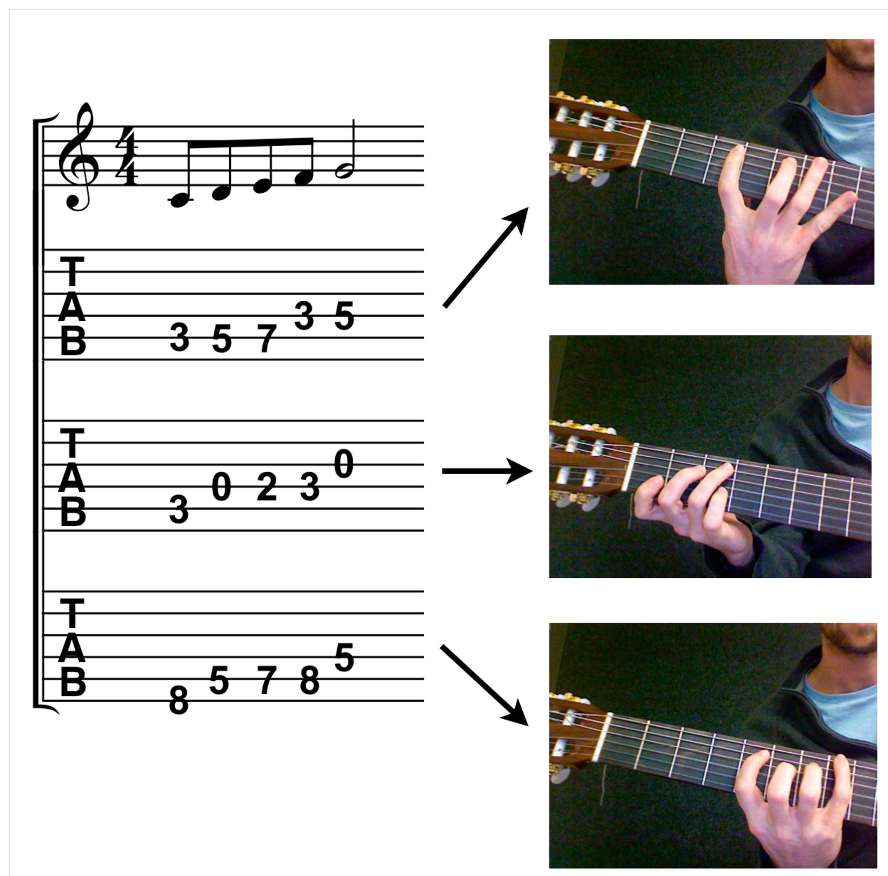
Combined Audio and Video Analysis for Guitar Chord Identification

Alex Hrybyk • ahrybyk@drexel.edu, Advisor: Youngmoo Kim
Electrical & Computer Engineering, Drexel University

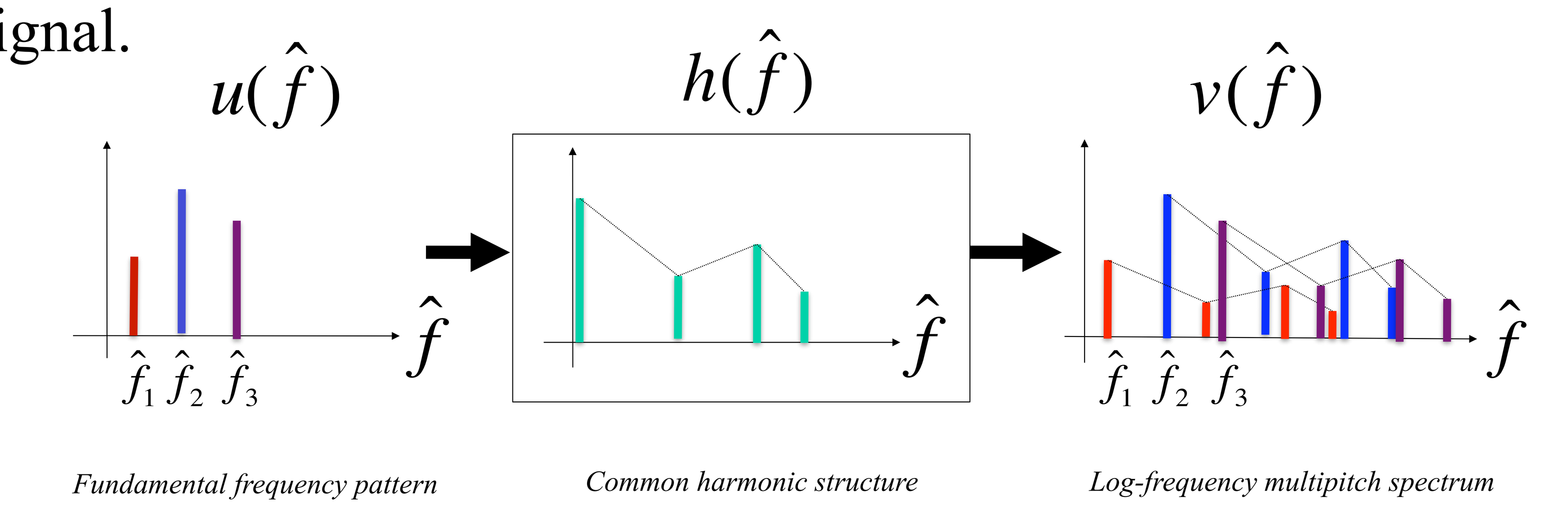
Introduction

This research presents a multi-modal approach to automatically identifying guitar chords using audio and video of the performer. Chord identification is for stringed instruments adds extra ambiguity as a single chord or melody may be played in different positions on the fretboard. Preserving this information is important, because it signifies the original fingering, and implied "easiest" way to perform the selection. This chord identification system combines analysis of audio to determine the general *chord scale* (i.e. A major, G minor), and video of the guitarist to determine *chord voicing* (i.e. open, barred, inversion), to accurately identify the guitar chord.



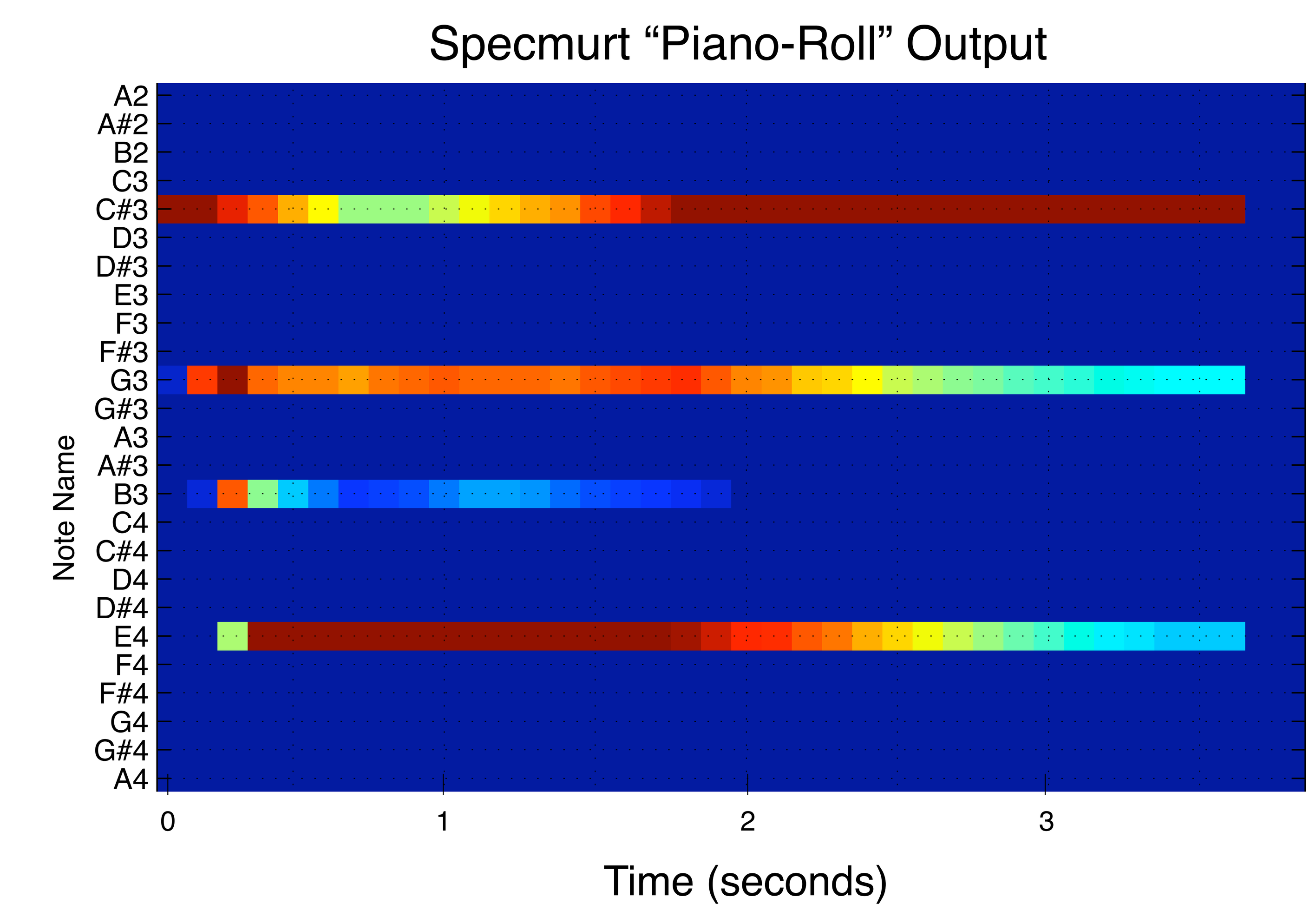
Audio Analysis

When playing a single note, the guitar, and many other instruments produce natural harmonics (overtones) in addition to the note's fundamental frequency. When playing multiple notes, the frequency spectrum of the audio appears cluttered, making detecting the fundamental frequencies (the actual notes) hard to locate. Using a technique known as *Specmurt* analysis [1], the notes of the guitar chord can be extracted from the audio signal.



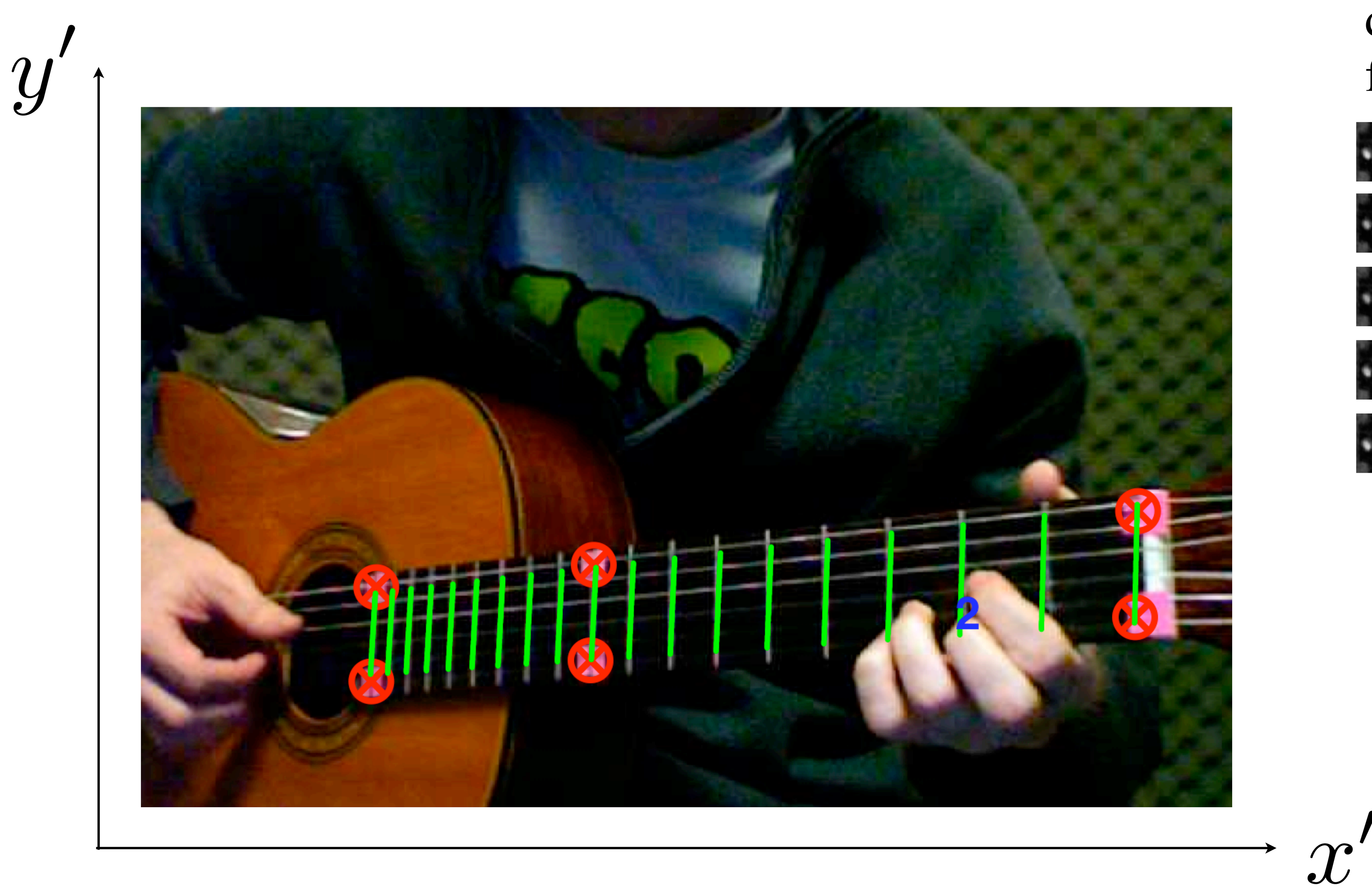
Through a process of de-convolving the log-scaled frequency spectrum of a signal with a known harmonic structure, the resulting spectrum will contain only peaks at the fundamental frequencies, making it easy to locate the notes being played.

Since de-convolution is difficult, we use the time/frequency duality of convolution/multiplication to make finding $u(f)$ much easier, taking the frequency data to the "Specmurt domain". The resulting spectrum $u(f)$ contains only peaks at the fundamental frequencies.

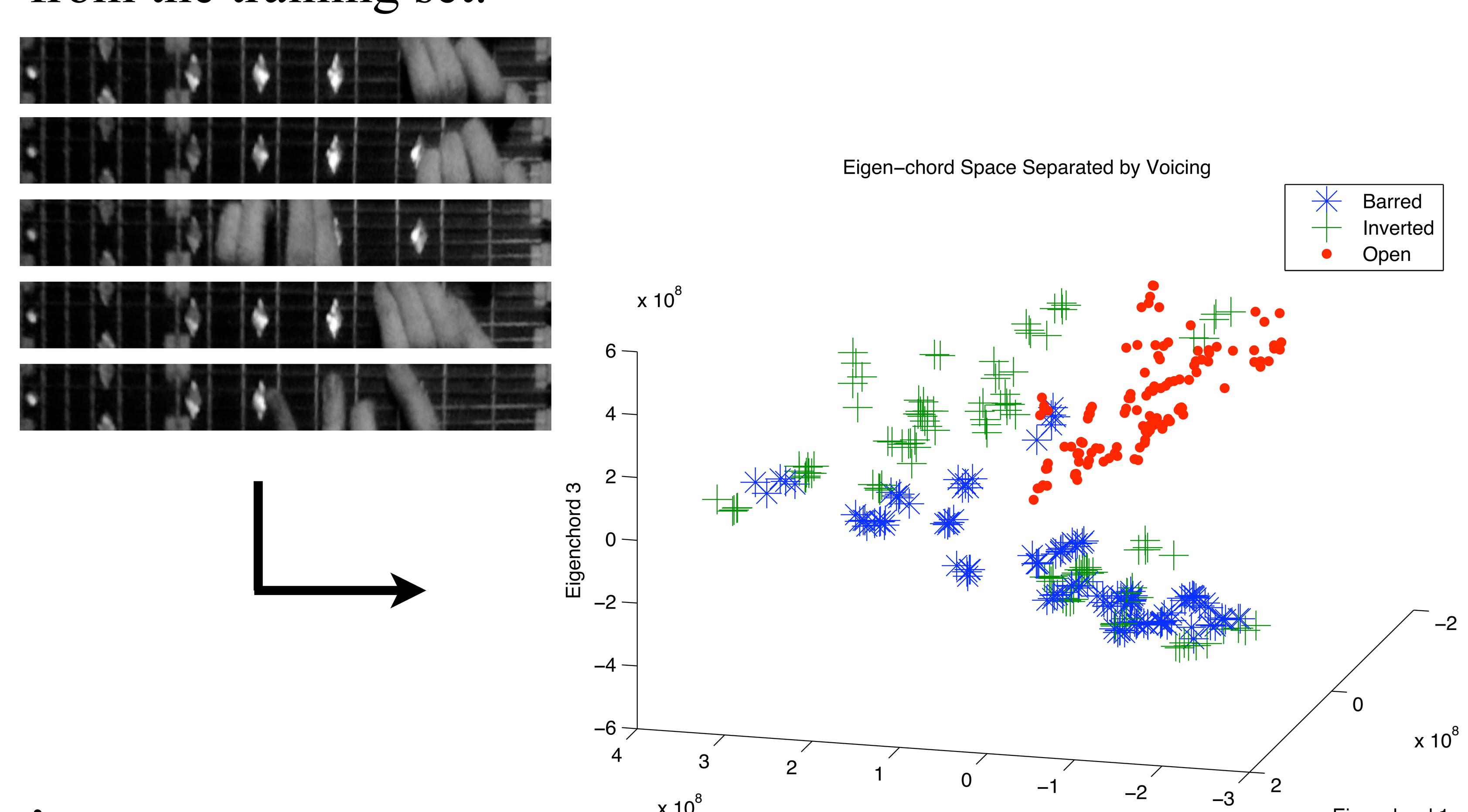


Video Analysis

While performing, the guitar can be held in many different orientations relative to the camera, making it difficult to find the locations of frets of the guitar. Homography is used to rectify or warp our original image to fit the ideal fretboard making it easy to locate the fretboard in the image [2].



Once the fretboard has been rectified and extracted from the image, it can be reduced into its "eigen-chord" components, using many images drawn from a training set. The various voicings of a chord tend to group together in the eigen-chord space. By projecting an unknown image into the space, we can determine which voicing it belongs to using the closest centroid from the training set.



Results

The system that performs the best in terms of correctly identifying the overall chord (scale and voicing) utilizes the strengths of audio and video results. Since Specmurt analysis yielded extremely high accuracy for determining scale, it was used as a preprocessing step to voicing identification via video.

	Audio Only	Video Only	Combined System
Scale	**98.6%**	34.8%	**98.6%**
Voicing	62.0%	**94.4%**	**94.4%**
Both	61.1%	32.8%	**93.1%**

References

[1] S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto, and S. Sagayama, "Specmurt analysis of polyphonicmusic signals,"Audio, Speech, and Language Processing, IEEETransactions on, vol.16, no.3, pp. 639–650, February 20

[2] X. Wang and B. Yang, "Automatic image registration based on natural characteristic points and global homography," in Computer Science and Software Engineering, 2008 International Conference on, vol. 5, dec. 2008, pp. 1365 –1370.