SIGNAL PROCESSING AND STATISTICAL ANALYSIS OF AUDIO FILES

Roger LaCroix, Michael Nagler and Kevin Wang

Brandeis University

ABSTRACT

This project will introduce the various applications of signal processing. This project describes algorithms that effectively analyze and quantify the tempos, keys, and lengths of songs over the course of the past fifteen years in order to find patterns. We will first give an in depth explanation of the theory behind tracking key and tempo in an audio signal. Afterwards, we will discuss the steps we have taken over the course of this semester to complete our objectives, and then discuss the results of our analysis. Next, we will go on to talk about the issues weve encountered over the course of this semester and how we surmounted them. Finally, we will go on to discuss any correlations or data we found in our analysis.

***Index Terms*—** Signal processing, audio analysis, tempo, key

1. INTRODUCTION

For over 50 years Billboard has tracked the most-listened-to music in the U.S. Since its conception, the music industry has changed drastically. By using signal analysis to analyze the time, tempo, and key of songs, one can distinguish certain listener patterns, which would provide a competitive edge to artists to predict the future of the music industry. Length indicates how much time listeners care to devote to a track, tempo indicates the mood and energy that listeners prefer, and key provides insight to which musical notes are favored by the human ear.

Our final product is able to run an analysis on a set of data, determining its average tempo and length, its mode key and, given an input song, can predict whether or not that song is similar to other songs in the set based on the statistics were measuring. This is done using the averagestats script, which takes in data from our analysis and reports the shortest song, the longest song, and the average song length in the set, as well as the fastest song, slowest song, mean tempo, and finally, the mode key of the set. It then prompts the user for an audio file and, given this input song, reports whether or not it belongs in the set. Our major analysis was performed on sets of top country songs from the years 2000, 2005, 2010, and 2015, and we were able to run our program on this set and graph the average length and tempo of each year in order to view trends in listener patterns over this time period.

2. TEMPO ANALYSIS

2.1. Algorithm

The Tempo algorithm proposed by Eric D. Scheirer[4] and implemented by Sam Drazin[5] with some alterations done by us to change the scope and pick the middle portion of the song was used to obtain the tempo of a song in beats per minute. However, we had previously altered it to choose the second highest peak due to accuracy errors and false positives. The algorithm works by reading an audio file and reading it with Matlabs built in audioread function, which returns the sampling rate and the number of samples. Next, six frequency bands are created and then envelope extraction is implemented to outline the signal. Then, we find the variation in the bands using Matlabs built in diff function. Afterwards the signal is filtered and we use Matlabs max function to find the highest peak on the graph which signifies the highest energy reading. The corresponding bpm value on the x axis under this reading is the most likely tempo.

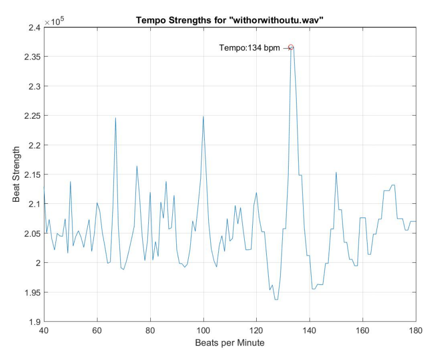

Fig. 1. Beat strengths for With or Without You

Fig. 1 is the result of running the tempo analysis algorithm. The tempo of this song is 134 bpm, and on the y-axis is beat strength, and on the x-axis is beats per minute. To find the tempo given a graph, one must find the value with the highest beat strength.

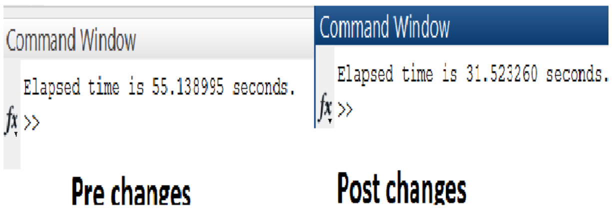

Fig. 2. Runtimes pre- and post-change

Figure 2 above is the result of our changes in the tempo analysis algorithm. We managed to maintain almost all of its accuracy while significantly reducing its runtime by reducing the scope of the analysis by emptying parts of the signal matrix.

2.2. Results

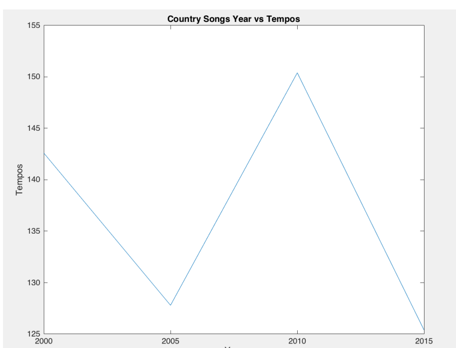

Fig. 3. Tempo of country songs from 2000-2015

Figure 3 above is the result of our tempo analysis on the top 5 country songs from 2000, 2005, 2010, and 2015. Based on this sample size, there appears to be no trend, but we believe that if we were able to expand the scope of our analysis and increase the size of our database to perform a more comprehensive analysis, we would see more of a trend in tempo. We think the results look so cyclical now because of how small the sample is.

3. KEY SIGNATURE ANALYSIS

3.1. Algorithm

The algorithm implemented for this project is a loose interpretation of the Maximum Key Correlation (MKC) algorithm described in [3]. The main difference is that this implementation does not compute a chromagram, which maps time, pitch and intensity on the x-, y-, and z-axis respectively. This step is omitted in favor of analysis of a two-dimensional representation of the song. Namely, frequency vs. amplitude. This is represented as a 2-column matrix. The matrix is then trimmed to span 65 Hz to 1100 Hz, which is effectively C2 to B5; a four-octave span. From here several approaches were made. All approaches were based on the 12-note chromatic scale around which Western music is based. The first attempt to find key was based on the idea that the tonic would be the frequencies that appeared the most in the spectrum. The frequency values were hard coded in, and for each important frequency, a window was generated around that value and the amplitudes of all frequencies in the window were summed, and that value was the strength of that note. This was done for all 12 tones over all 4 octaves. Because each octave has twice as many frequencies as the preceding one, the window function that I came up with was:

$$W(f) = 2\frac{f}{B}$$

Where f is the input frequency and B is the frequency of that note in the lowest octave. This allowed me to create a window that grew as I looked higher into the spectrum. With the vector of strengths of all notes, the one with the highest strength would be the tonic, and the tonality of the key would be determined by whether the major or minor 3rd was more prevalent. This method proved to be inaccurate, so a more refined approach was created. This new approach followed more closely the MKC algorithm by looking at important pitches for each key, and finding the probabilities of all keys and picking the most likely. The new approach ranked all pitches in each octave from 1-12, 1 being the weakest and 12 being the strongest in that octave. This was done for each octave. This new way of ranking pitches was used because the lowest pitches had too high of an influence on the results of the first algorithm, possibly because of a bass drum or other similar unimportant sounds (in regards to key). From here a matrix was created that represented all notes in each of the 12 key signatures. The relative strength of notes in each key is added together to create a vector of length 12 representing the relative strengths of all 12 possible key signatures. The strongest one is then picked.

3.2. Results

Both versions of the key-finding algorithm proved to be inaccurate. Results are posted in the table below.

Both results proved to be about 14% accurate. This highlights the difficulty of spectral analysis of live recorded music. Extraneous noise and slightly out-of-tune musicians prove problematic. Although method B and A both had the same types of error, that is, being off by a 4th or a 5th, method B is the better approach because it uses more information from the song. In order to improve this method, one can use four to five important notes in a key, rather than all 7.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Key	E	C	D	Bb	E	E	E	C	D	G	E	E	E	D
A.	D	D	E	Eb	D	D	D	C	C	C	E	D	D	F#
B.	C#	B	G	E	G	C	C	G	G	G	A	Eb	E	G

Fig. 4. Results of methods A and B on the album *Please Please Me*

Since it is difficult, if not impossible, to quantify key signature, talking about an average would make little sense, and not show any significant kind of trend. Therefore the program reports the number of unique keys in groups of songs. In theory, with the rise of electronics in music, musicians will be less bound to real instruments, and thus be able to compose in any key with equal ease. For example, on the piano, it is significantly easier to play in the key of C compared to the key of E, whereas programmable instruments don't have these limitations. It can be expected that there would be a positive trend between key diversity and time. However, these results are gained from country music, which is still composed on guitars and pianos, so no such trend was evident. Results are found below.

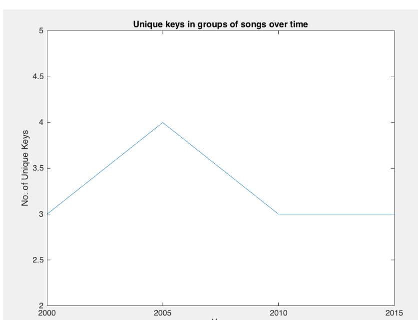

Fig. 5. Key Diversity Over Time

4. LENGTH ANALYSIS

Length of songs were obtained by creating a *getKey* function that had passed parameters of the sampling rate and the number of samples per song. The *getKey* function then returned the length of a song in seconds by dividing the samples by the sampling rate. The length of the songs were then plotted on a graph on a time vs length axis. The graph can be illustrated below:

From this data, it is clear that the length of popular country songs is steadily declining. The peak of the data showed that there was a trend of longer songs becoming popular, but since 2005 the average length has fallen from around 270 seconds to 190 seconds.

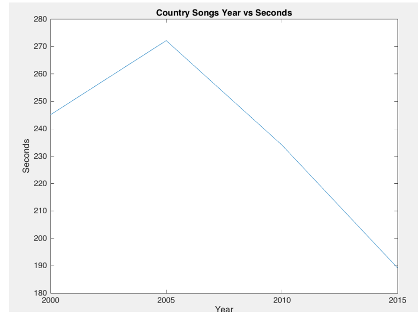

Fig. 6. Figure 2. Length of popular country songs over time

5. ISSUES ENCOUNTERED

When using our algorithms to analyze groups of songs, we ran into a few issues. Because the application of this type of analysis varies based on who is using it, it was difficult to find a balance between speed and accuracy. For a musician who wants to know if their song fits a trend, speed is most likely more important. To music historians, accuracy would be more important. Below are a few issues that we encountered.

- File Format: In our initial goals, we were planning on using .wav files as opposed to mp3 files for the audio analysis. However, we had to use a website that extracted the audio signal from youtube links, which are in the format of mp3. This caused some speculation of the accuracy of certain songs and how well they were able to get the tempo, key, and length of songs from the audio files, since some data is lost during the conversion.

- Speed: The algorithm is relatively slow, and we decided that it took too long to run. We found a way to speed it up by reducing the scope of the analysis.

- Peak Picking: Faulty conversions and a scope that was too small led us to some false positives, which ended in us attempting to pick the wrong peak for a while.

- Variance in Genres: Our initial analysis failed and gave us completely nonsensical results because the songs we chose for the first time were of different genres. Different tempo, length, and key values work better for different types of music, which is why successful songs of different genres dont necessarily have to follow a pattern. For example, successful country songs will have slower tempos than successful pop songs.

- Creating a Database: A major problem weve had throughout the course of this project has been creating a sizeable database in a short time frame. The way weve been collecting songs to analyze has been going to the Billboard site and writing down their hits, and then looking up those songs on youtube and using a converter to turn it into a .wav file. This process is very slow and manual.

6. CONCLUSION

Our recent and most successful analysis was performed on the top 5 country songs from the years 2000, 2005, 2010, and 2015. We took the top songs from the first week of July each year according to Billboards database. As predicted, when time went on, song length plummeted. Starting at 2005, the average song length of top country songs saw a pretty extreme drop, telling us that as time went on, listeners werent willing to devote as much time to a single song. From the data we looked at, we saw no trend in tempo. The results were pretty cyclical, but since our sample size was relatively small, we believe that if we performed this analysis on a larger set of songs, we would see more of an upward trend in average beats per minute.

However, a larger set of data might not be the answer to the mystery of the tempo analysis. It could very well be that, even with a larger set of data, our results would still be cyclical and hover around the same value. Its hard to tell without being able to test it, but it could be that the industry has arrived at a sort of equilibrium for the tempo of hit country songs. It could be that theyve found a tempo value that people like, so they just continue to make songs at that tempo.

7. SONGS USED

- 2000 Songs
  - Lee Ann Woman - I Hope You Dance
  - Clay Walker - The Chain of Love
  - Clay Davidson - Unconditional
  - Chad Brock - Yes!
  - Collin Raye - Couldn't Last a Moment
- 2005 Songs
  - Keith Urban - Making Memories of Us
  - Rascal Flatts - Fast Cars and Freedom
  - Dierks Bentley - Lot of Leavin' Left to Do
  - Geroge Strait - You'll Be There
  - Toby Keith - As Good as I Once Was
- 2010 Songs
  - Luke Bryan - Rain Is a Good Thing
  - Jason Aldean - Crazy Town
  - Brad Paisley - Water
  - Miranda Lambert - The House that B
  - Clay Walker - She Won't Be Lonely Long
- 2015 Songs
  - Little Big Town - Girl Crush
  - Luke Bryan - Kick the Dust Up
  - Sam Hunt - Take Your Time
  - Blake Shelton - Sangria
  - Kelsea Ballerini - Love Me Like You Mean It

8. REFERENCES

[1] P. Desain and H. Honing. Computational models of beat induction: The rule-based approach. Journal of New Music Research, 28(1):2942, 1999

[2] S. Dixon. Automatic extraction of tempo and beat from expressive performances. Journal of New Music Research, 30(1):3958, 2001.

[3] Steffen Pauws. "Musical Key Extraction from Audio." (n.d.): n. pag. Web.

[4] J. New Music Research, Special Issue on Beat and Tempo Extraction, vol. 36 no. 1, March 2007, pp. 51-60. (10pp)

[5]http://www.samdrazin.com/classes/mmi505/tempo_tracker.txt