A Study on Pitch Recognition for Trumpet

Chuchi S. Montenegro[1]
chuchi.montenegro@gmail.com

Al Crimson D. Alcaide[2]
balakiki23@yahoo.com

Airus Lee T. Cabrera[(2)]
airuslee_cabrera@yahoo.com

Ian Joseph Timothy A. Selda[(2)]
ianzelda@gmail.com

Abstract

Pitch recognition is the ability to identify a pitch from a sound. This research focuses on recognizing the pitch of a trumpet much like the ability of a skilled human being capable of recognizing an exact pitch or note. A software application using Java language is developed in order to explore on how to extract a pitch of a sounding trumpet and able to recognize the pitch played. Two experiments have been carried out in this study, one was to find out trumpet recognition rate as performed by novice and experienced players; and second, it was tested if it can and only can recognized pitch aside from other musical instruments. The first experiment generates a pitch recognition rate of 78%, while the second generates pitch recognition rate of only 13%.

Keywords : digital signal processing, pitch recognition

1. INTRODUCTION

The three properties of sound are *pitch*, *loudness*, and *timbre*. It is on the first property of sound that this study focuses. Specifically, the group deals mainly on pitch recognition. Since real time pitch recognition programs are very rare, most people do not even think that pitch recognition using computer applications can be done at all.

Pitch recognition is the ability to identify notes contained in an audio signal through the highness and lowness of the sound. The ability to identify the highness or lowness of the sound, however, is not inherent in every individual. Those who are gifted can recognize pitch in an instance. They can easily respond to the pitch by naming the corresponding note. For most people, the common pitch recognizer is the human ear. This requires a keen sense of hearing, skill and practice. Since most people do not possess the ability to recognize pitch as easily as the more gifted individuals do, a pitch recognition program will solve the problem. The researchers focus on the trumpet because pitch recognition programs for the instrument are very rare, if not at all absent.

[1] *Computer Science Department, Silliman University, Dumaguete City, Philippines*
[2] *Information Technology Department, Silliman University, Dumaguete City, Philippines*

A trumpet belongs to the family of brass instruments. The most common type of trumpet is B*b*. It is composed of a brass tubing material that is bent twice in an oblong shape. It is played by blowing air through the tube along with the three piston valves that are used for increasing and lowering the pitch (Baines , 1976). Playing the trumpet requires a lot of effort and dedication to learn and master it .

The main objective of the study is to recognize and distinguish the pitch produced by a trumpet in a real time. An application is developed to automatically segment continuous audio signal, extract features of the sound and decide the likelihood of an appropriate pitch in which results are essential to those who are musically inclined to help determine correct hit of a pitch. Nevertheless, it is also especially important to people who have just started studying music and who, oftentimes, encounter difficulties in recognizing pitches. This study is beneficial in several ways. Foremost, pitch recognition for trumpet is made easy and once the pitch recognizer is used, the player will know what note he or she is playing in real time. Concurrently, the application can be used as a tuner for the trumpet and a guide to those who want to learn to play the instrument which can be used to gauge whether the note they are playing is the same as the standard note (in frequency) stored in real time. Since audio capturing and recording can be done in either monophonic or stereo environment, the application in this study is only capable of processing monophonic sound of an instrument playing one at a time. An example of this case is a soloist trumpet player practicing his piece which is, sounds are easier to capture because they belong to a single thread. Otherwise, it will be, according to Martin, K.D., especially difficult to analyze two pitches where another octave is present (for example a middle *C* and the next highest *C* played together). This is because all of the frequency components found in the higher note in an octave will also be present in the lower note.

The range of pitch covered in the study is from lower *C* to higher *G*, which is the normal range for a trumpet. Further, the application is designed not to recognize other musical instruments, specifically, wind instruments.

2. LITERATURE REVIEW

Pitch can be high or low. It can be processed so that its frequencies can be analyzed and can be transformed into signals. A study of Naotoshi Seo, Pitch Detection via Cepstral Method, found that through cepstral analysis assessment of pitch is possible. The system, Cepstral Method, is a way to estimate the frequencies of a pitch by analyzing and estimating its wavelets (Seo, 2008).

A project developed by Stephen Geiger of MIT (Michigan Institute of Technology) called Pitch Recognition with Wavelets investigates the use of wavelets for pitch recognition. This method uses Continuous Wavelet Transform (CWT) at different scales, to identify individual notes. It then transforms the signals from the individual notes into wavelets on small oscillations that are highly localized in time. What comes out is a time frequency representation of a signal offering very good time and frequency localization (Geiger, 2003).

The Polyphonic Audio Finding using the Spiral Array CEG (Center of Effect Generator) Algorithm, developed by Ching-Hua and Elaine Chew of the University of Southern California, Viterbi School of Engineering, is very effective in determining pitch from polyphonic sound in real time algorithm. To extract pitches and pitch strengths from polyphonic sound, it uses Standard Fast Fourier Transform with local maximum detection design (Chuan & Chew, n.d.).

To track musical pitch, Judith C. Brown suggests a pattern recognition algorithm. Plotting the Fourier transform against the log frequency, which would then elicit a pattern that is the same for those sounds that have harmonic components, identifies this pattern. The *Q* Fourier transform that has been determined from this pattern corresponds to a quartertone in music. A method called cross correlation is used to track the musical pitch of the resulting pattern.

3. TECHNICAL ANALYSIS

For the application to be able to recognize pitch, a player is asked to play notes on a trumpet. The sound is then captured by the PC microphone and feed the information to the application which returns an appropriate pitch equivalent to the notes played. The following scenario is described in Figure 1.

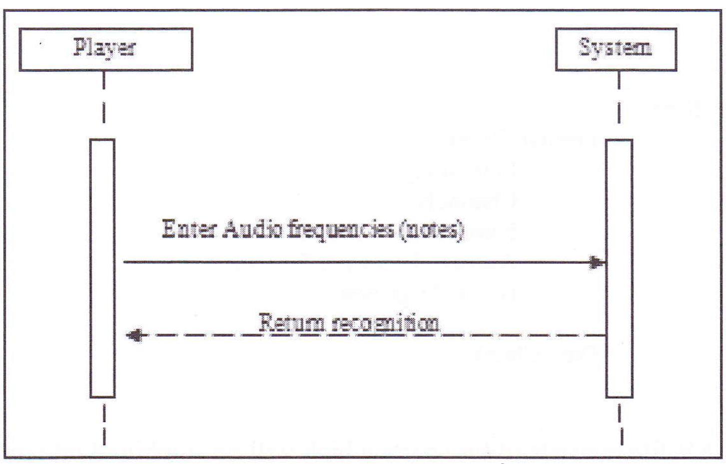

Figure 1. Sequence Diagram for pitch recognition

3.1. Method Analysis

The success of pitch recognition software's is correlated to the processes engaged in the development of the application. This project makes use of the following algorithms in the development modules to generate recognition accuracy. Figure 2 shows the flow of different algorithms used to determine how the pitch from the trumpet is to be recognized and how it (pitch) will be extracted through the sound signals produced. Each of these details is discussed in next sections.

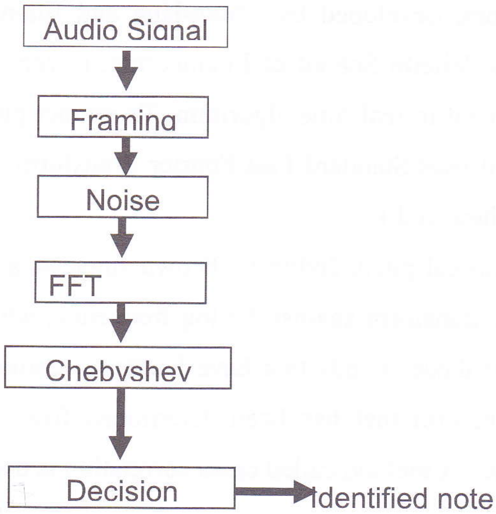

Figure 2. Method Analysis

3.2. Waveform Audio File Format (WAVE)

Conversion of sound to digitized format enables the researchers to analyze its wave form. The study uses waveform audio file format (WAVE) or the .WAV file because it is the simplest way of storing digital audio. WAVE contains parameters describing the waveform such as sample rate and data chunk in which the actual waveform data is contained.

```
WAVE
        RIFF{
                <Format Chunk>
                        Format tag
                        Channels
                        Samples per Second
                        Average Bytes per Second
                        Block Alignment

                <Data Chunk>
                }
```

The .WAV file consists of two parts which will be combined into one called RIFF (Resource Interchange File Format). The most important chunks of the RIFF data are the Format Chunk and the Data Chunk. The former (the Format Chunk), which must occur before the Data Chunk, specifies the format of the latter. It contains fields: format tag, channels, samples per second and average bytes per second, block alignment. The Data Chunk contains the waveform data or the actual sample data which also indicates the actual size of the sound.

3.3. Framing

Framing is applied to get the frequencies of a tune by segments. The concept behind framing is that it places all signals produced by a sound into a container. It takes as input the raw audio signal and outputs a framed equivalent of the input signal. Every signal represents a single data and is stored into a container, which can hold several numbers of signals. After going through this method, framed audio signal outputs are produced.

Figure 3 shows three notes played in succession within one second. The frames, which contain the raw frequencies, are moved to the next process where the unwanted frequencies or noises are filtered.

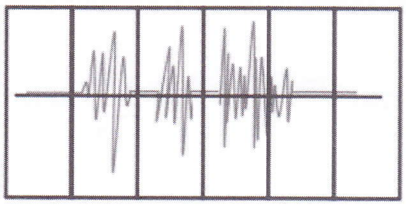

Figure 3. Framed Audio Signal

3.4. Noise Filtering

Noise filtering is a very important process in audio processing. The absence of this stage takes high effect on the success of the recognition rate. Room noise (i.e. air conditioning, etc) for example is assimilated into the audio signal during recording.

In this stage, the framed audio signals are filtered so that only the appropriate audio signals of the trumpet are remained. The audio signals that are considered noise are discarded; and only the relevant (trumpet) audio signal are remained to undergo feature extraction process. The filtering process plays a crucial and important job because as the player plays continuously, the recorded audio signal not only contains the frequencies of the notes but also other irrelevant noises.

Because it is important to be able to have data of relevant noises, a room sound recording is performed that will be used as a reference for the filtering stage. The allocated recording time of the room is 30 seconds long which is enough to provide complete details on the characteristic of the noise. The purpose of this recording is to determine the maximum peaks of the noise audio signal. Audio signals below the maximum peaks are considered unimportant and will be classified as noise and treated as a signal of value zero.

These signals then will be omitted and will be used as input for feature extraction. Likewise, audio signals exceeding the maximum peak is assumed to be music playing. Figure 4 and Figure 5 illustrates the differences on the audio signals after filtering stage.

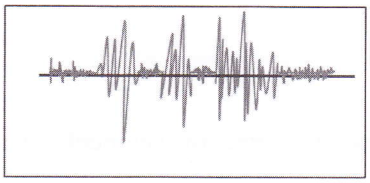

Figure 4: Audio Signals with Noise

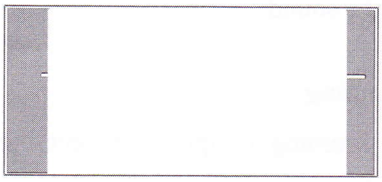

Figure 5: Filtered Audio Signals

3.5. Fast Fourier Transform

The main key factor in audio processing is the extraction of sound features from the original audio files which will later on be used for classification of pitch. In this work, the filtered framed audio signals from the previous stages are converted to frequencies so that every single frame corresponds to a frequency denoting an identified set of note. The formula used in getting the frame's frequency uses the Fast Fourier Transform algorithm with the general representation as shown in the FFT Formula here.

$$X_k = \sum_{m=0}^{M-1} x_{2m} e^{-2\pi i \frac{mk}{M}} + e^{-2\pi i \frac{k}{N}} \sum_{m=0}^{M-1} x_{2m+1} e^{-2\pi i \frac{mk}{M}} \qquad [1]$$

The FFT in this manner is used in identifying the pitch. It generates a frequency spectrum that gives the pitch class and pitch strength. The Fourier transform decomposes or separates a waveform or function into sinusoids of different frequency which sum to the original waveform. It identifies or distinguishes the different frequency sinusoids and their respective amplitudes [5,13]. Since a frame can contain a number of audio signals equivalent to a period of time, FFT computes the audio signals over a period of time. Xk is the result of computed framed audio signal. M is equivalent to a period of time, which is equal to the length of a frame. The small x represents the audio signal inside a frame and k represents the actual frame being processed. The output of this formula is the computed frequency from an audio signal for a given period of time.

3.6. Chebyshev Distance Algorithm

Chebyshev distance is a metric defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension as represented in the following formula:

$$D_{\text{Chebyshev}}(p, q) := \max_i(|p_i - q_i|). \qquad [2]$$

The algorithm is being used to compute the differences of frequencies from the frequencies returned after feature extraction to the frequencies defined in the stored knowledge of frequencies in the codebook. It is used to help decide the likelihood of a note by evaluating the lowest difference of the computation. The stored frequency which has the minimum difference is the appropriate corresponding to a note.

4. TESTING AND EVALUATION

Testing of the systems' recognition rate is done in two ways. First, the system is tested to determine recognition rate of single notes and note combinations as played by trumpet players. Second, the system is assessed to verify trumpet instrument recognition only and no other musical instruments.

To determine notes recognition rate of the system, experiments were done by asking 6 different trumpet players to play live the notes described in Table 1 for a span of 1-3 seconds. If the note played by the player equals the note recognized by the system, a check is marked on the appropriate column on the table. This process is repeated two more times to revalidate the recognition. To further evaluate the recognition capability, the researchers test it on two to three-note combinations in succession to see whether the combination of notes can be recognized.

Table 1.
Notes recognition as performed by a player who's been playing trumpet for 5 years

	1st try		2nd try		3rd try	
List of Notes	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
Low do		✓	✓		✓	
Low re	✓		✓			✓
Low mi	✓		✓		✓	
Low fa	✓		✓		✓	
Low sol	✓		✓		✓	
Low la	✓		✓		✓	
Low ti	✓			✓	✓	
Middle do	✓		✓		✓	
High re	✓		✓		✓	
High mi		✓	✓		✓	
High fa	✓			✓	✓	
High sol	✓		✓		✓	
Two-note combination						
Low do – middle do	✓		✓		✓	
Middle do – high sol		✓	✓		✓	
Low do – low fa	✓		✓		✓	
Low mi – low la		✓	✓		✓	
Low re – low ti	✓			✓		✓
Low la – low fa	✓		✓		✓	
Low sol – high sol	✓		✓		✓	
Middle do – high sol	✓		✓		✓	
Three-note combination						
Low fa – low do – low la	✓			✓		✓

Another experiment has been conducted to make sure that the system can only recognize trumpet and no other instruments and subjecting other sounds into the system; whistling, a human voice, and other musical instruments performed the second experiment. Each instrument plays three sets of the same notes (Table 2). If the system recognizes the notes of the new instrument, a check is placed on the column corresponding to the note being played.

Table 2.
Notes recognition as tested on Piano

List of Notes	1st try		2nd try		3rd try	
Low do to High sol	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
Low do		✓		✓		✓
Low re		✓		✓		✓
Low mi		✓		✓		✓
Low fa		✓		✓		✓
Low sol		✓		✓		✓
Low la		✓		✓		✓
Low ti		✓		✓		✓
Middle do		✓		✓		✓
High re		✓		✓		✓
High mi	✓		✓			✓
High fa		✓		✓		✓
High sol		✓		✓		✓

The first experiment employs respondents whom the group classifies into novice trumpet players and expert or seasoned trumpet players. All five of them are either new members of the Silliman University band or have been playing for more than two years already. Because the objective of this experiment is to determine the accuracy of the notes recognition, highly skilled trumpet players are selected to perform the experiment.

Table 3 summarizes the recognition percentage of notes from lower do to high-sol played by the individual trumpet players. The percentages were arrived by inspecting the number of Yes over the number of trials. Two-note and three-*notes* tests were also summarized on the same manner generating recognition rate of 94% and 67% respectively.

Table 3.
Percentage of recognition from lower do to high-sol

List of Players	Percentage of Recognized Notes from Lower do to High sol											
	Low do	Low re	Low mi	Low fa	Low sol	Low la	Low ti	Middle do	High re	High mi	High fa	High sol	Average
Al Crimson Alcaide	67%	67%	100%	100%	100%	100%	67%	100%	100%	67%	67%	100%	**86.25%**
Gerna Bacayo	0%	33%	33%	100%	100%	67%	100%	33%	67%	0%	100%	33%	**55.5%**
Jahziel Gako	100%	67%	100%	67%	100%	67%	67%	100%	33%	33%	100%	0%	**69.5**
JR Quizo	100%	33%	100%	100%	100%	33%	100%	100%	100%	100%	100%	100%	**88.83%**
Jem Talaroc	67%	100%	100%	100%	67%	100%	33%	67%	100%	67%	67%	67%	**77.91%**
Norman Valencia	100%	100%	100%	100%	100%	67%	100%	33%	100%	100%	100%	100%	**91.6%**
												Recognition Rate 78.27%

The table below (Table 4) summarizes the experiment performed to recognize notes as compared to other instruments. It was found out to generate lower-do to high-sol notes recognition of only 13%.

Table 4.
Summary test of other wind instruments, whistle and voice.

	Percentage of Recognized Notes from Three Different Instruments, Whistle, and Voice												
	Low do	Low re	Low mi	Low fa	Low sol	Low la	Low ti	Middle do	High re	High mi	High fa	High sol	Average
Alto Saxophone	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**0%**
Trombone	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**16.67%**
Whistle	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	**0%**
Voice	33%	33%	67%	33%	0%	0%	0%	33%	67%	100%	67%	67%	**41.67%**
Paino	0%	0%	0%	0%	0%	0%	0%	0%	0%	67%	0%	0%	**5.58%**
	Recognition Rate 12.78%												

5. CONCLUSION

The experiments were performed to test recognition rate of the system not only to trumpet instrument but also to other musical instruments and even to voice. The tests show that the system has better recognition of trumpet sound as opposed to other instruments, enough to conclude that the system can differentiate the musical instruments used. A closer look at the tests result lead to some observation that the data does not affect experienced of a user. However, the small numbers of unrecognized notes were usually high-mi or low-mi notes. It is recommended for further research to investigate the properties and features of the note for better understanding and to try testing on different microphone as the group were only using PC microphone to conduct the test which may not be so effective and has the quality to add additive noise to the signal.

Though the system during tests were not able to recognize some of the notes, the 78% recognition rate it generates is still good enough to be used as a gauge in learning the instrument. It is however recommended that for better results, room noise should be minimized and trumpet properly tuned-up.

References

___, Short-Time Fourier Transform (STFT).Retrieved May 4, 2009 from http://www.originlab.com/index.aspx?s=8&lm=323&pid=1101

Baines (1976), "Brass Instruments: Their History and Development."

Brown, j. (1989), "Musical pitch tracking based on a pattern recognition algorithm"

Chan, K., Erjongmanee, S. and Tay, C. (2000). "A Realtime Automated Transcription of Live Music in to Sheet Music Using CMN".

Chuan, C. & Chew E. (n.d.). "Polyphonic Audio Key Finding Using the Spiral Array CEG Algorithm". University of Southern California Viterbi School of Engineering. LA.

Clement, I. (2003), Modular Audio Recognition Framework and text-Independent Speaker Identification v.0.2.0

Geiger, S. (2003). "Pitch Recognition with Wavelets".

Gerhard, D. (2003). "Pitch Extraction and Fundamental Frequency: History and Current Techniques". Department of Computer Science University of Regina.

Kuhn, W. (1990). "A Real-Time Pitch Recognition Algorithm for Music Application"

Martin, K. D. (1996). "Automatic Transcription of Simple Polyphonic Music: Robust Front End Processing". M.I.T. Media Lab Perceptual Computing Technical Report #399, November 1996, presented at the Third Joint Meeting of the Acoustical Societies of America and Japan, December, 1996.

Seo, N. (2008). "ENEE632 Project4 Part I: Pitch Detection".

Szczerba, M. & Czyzewski, A. "Pitch detection enhancement employing music prediction".

Terhardt, E. (2000). "Definition of Pitch". Feb. 12, 2000.

Weisstein, Eric W.(n.d.). "Fast Fourier Transform." Etrieved may 2, 2004 from http://mathworld.wolfram.com/FastFourierTransform.html.