Brandon Wright

ECE 6950

Fall 2023

**WAV-to-MIDI Audio Processing MATLAB Program**

**Abstract –**

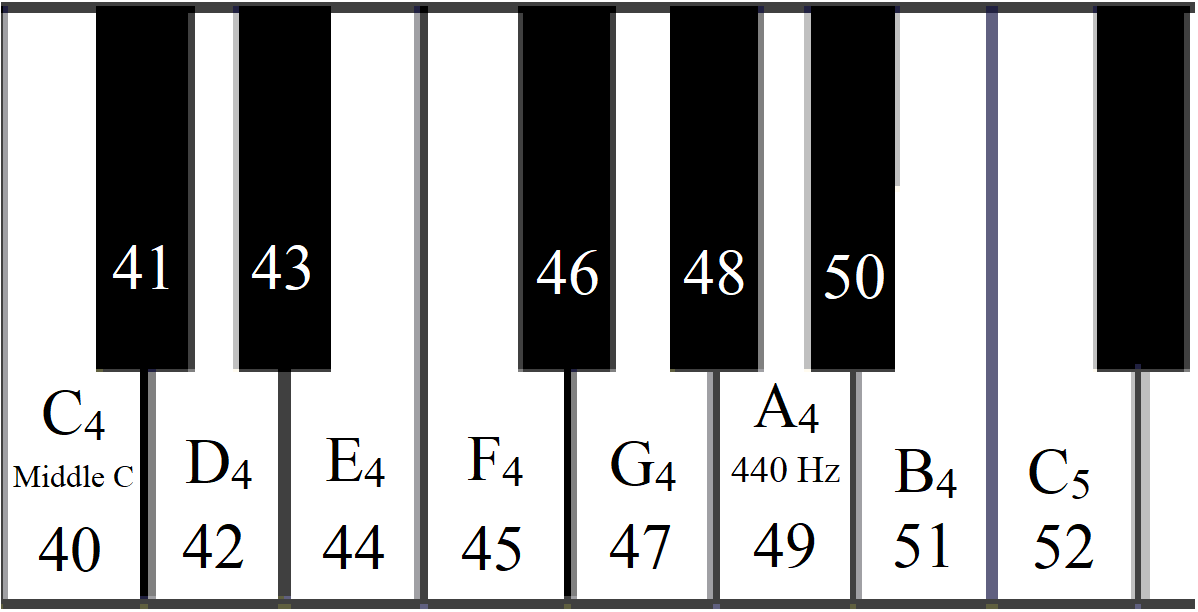
**INTRODUCTION**

*Musical Notes and Equal-temperament Tuning*

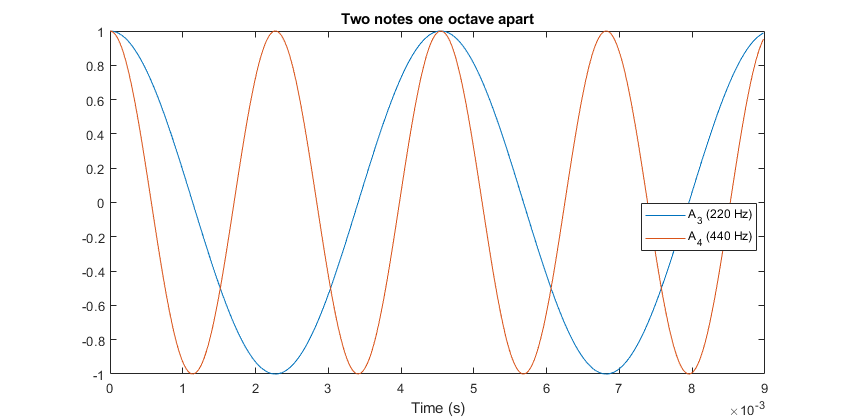
Music theory is based around the pitch or frequency of individual notes. The most fundamental interval in music is the octave. When given a base note, or root, with frequency f0, a second note an octave higher has a frequency of 2f0. A note that is an octave lower than the root has a frequency of ½ f0. An octave can be divided up into as many notes as desired, but an octave comprised of 12 notes was canonized during the inception of Western music. The piano keyboard, as shown in **Figure 1**, is a common way of depicting the 12 notes in an octave. As music theory became more standardized, alphabetic letters (A through G) were assigned to white keys on the keyboard, as seen in **Figure 1**. An octave begins one on letter, such as A, and by the time all the letters have repeated, a new octave has been reached. Notes that are an octave apart can be intuitively heard because the frequency of the higher note is perfectly divisible by the frequency of the lower note, as seen in **Figure 2**. Put another way, one period of a note contains exactly two periods of a note one octave higher than it.

\*\*\*\*Source DSP First chapter 4

David Sulzer. (2021). Music, Math, and Mind : The Physics and Neuroscience of Music. Columbia University Press.

****

**Figure 1.** Music notes for one octave on keyboard



**Figure 2.** Frequencies of two notes an octave apart

Because most instruments are capable of playing notes from multiple octaves, a subscript is often placed below the letter to denote which octave that note belongs to. For example, the lowest note on the standard 88-note piano is A0, and the highest note is C8­. The starting point for an octave is arbitrary, but most common tunings set the note A4 to 440 Hz. The frequencies of the other notes are determined relative to A4 using 12-tone equal-temperament (12-TET) tuning, as the octave is broken up into 12 notes that are equally spaced around the log-2 scale and using 440 Hz and A4 as the starting point.

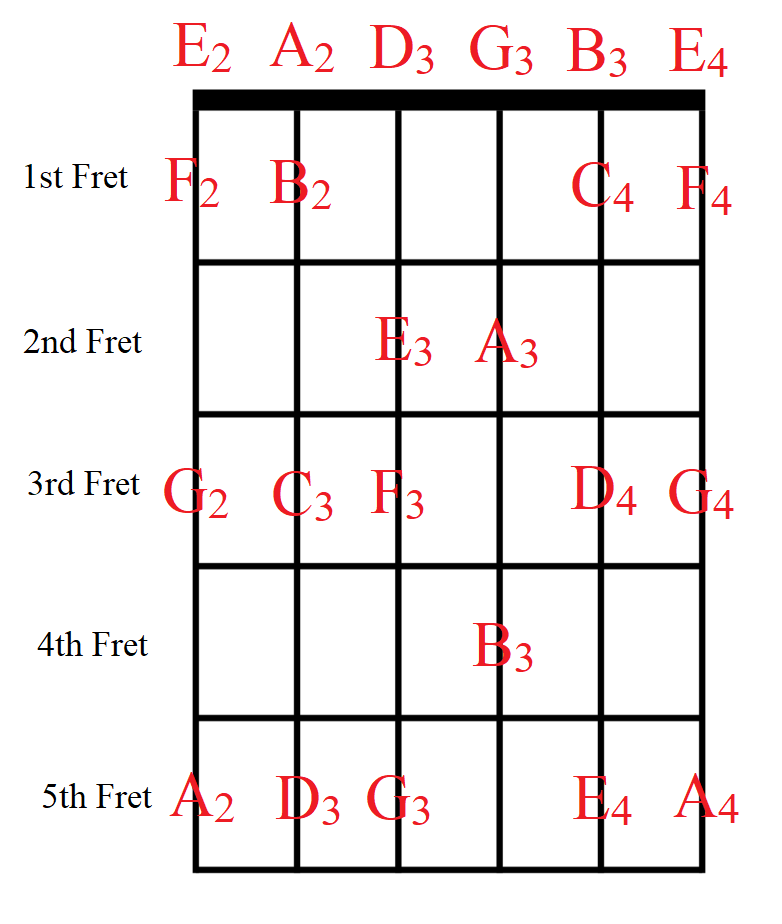
The equation for 12-TET tuning can be used to derive the number of the key (N) on the standard keyboard, from 1 to 88, as seen in **Equation 1**. A4 is the 49th key on the keyboard, and the pitch and thus the key number go up when moving right across the keyboard. The difference in pitch between two adjacent keys is called a semitone. Notes an octave apart are separated by 12 semitones. The black notes on the keyboard are referred to as sharps (♯) and flats (♭) and complete the 12 notes that comprise a 12-TET octave. **Table 1** depicts the frequencies and key number of the octave C4 to C5.

**Equation 1.** Converting frequency to musical key number

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Note** | **C4** | **C♯4/D♭4** | **D4** | **D♯4/E♭4** | **E4** | **F4** | **F♯4/G♭4** | **G4** | **G♯4/A♭4** | **A4** | **A♯4/B♭4** | **B4** | **C5** |
| **Number** | **40** | **41** | **42** | **43** | **44** | **45** | **46** | **47** | **48** | **49** | **50** | **51** | **52** |
| **f0 (Hz)** | ~262 | ~277 | ~294 | ~311 | ~330 | ~349 | ~370 | ~392 | ~415 | 440 | ~466 | ~494 | ~523 |

**Table 1.** Frequencies of octave beginning with Middle C (C4)

Other instruments have different ways of manipulating the pitch produced by the instrument to match the notes in 12-TET tuning. The standard guitar has six strings which are tuned to different pitches, and the fingers are placed upon certain points on the neck of the guitar, which are divisible by sections called frets. **Figure 3** depicts the frets on the guitar which correspond to white keys on a piano keyboard. Going up the frets on a single string on the guitar is equivalent to moving up the keyboard one semitone.



**Figure 3**. Notes on guitar fret board

The guitar, like other stringed instruments, exhibits a series of harmonics beyond the pitch of the open string, known as the fundamental frequency f0. \*\*\*physics\*\*\* The combination of the fundamental frequency with the series of harmonics contributes to the timbre, i.e. character of the note to distinguish it from a pure sinusoidal sound wave. The first harmonic is 2f0. Other harmonics exist at , , and all integer multiples of f0. The higher the harmonic, the less it contributes to the overall sound of the plucked string. The human brain is adept at picking up f0 out of the harmonic series. \*\*physics\*\* The fundamental frequency of a note can be detected by a computer and computed in a multitude of ways, each with its own tradeoffs. \*\*\*CREPE paper\*\*\*

*Electric Guitar*

The electric guitar uses electromagnetic induction to produce a pitch, which is sent to an amplifier and speaker to produce a tone. <https://www.purdue.edu/science/science-express/labs/labs/Physics%20of%20the%20Guitar.pdf> Fundamentally, the tone is created by the interaction between the metallic strings and a series of pickup coils, which are comprised of a cylindrical magnetic core surrounded by copper wire. When a voltage is supplied to the pickup, and electric field is created. The pickups are placed beneath the strings on the body of the guitar. When a string vibrates, it disturbs the pickup’s electric field at the frequency of the vibration of the string, which is detected by magnetic sensors also on the guitar body. The frequency of this disturbance is then fed into an amplifier and speaker to create the sound of the note.

*Musical Instrument Digital Interface (MIDI)*

Musical Instrument Digital Interface (MIDI) is a technology used to transmit and store musical notes in a digital format that can be used by a variety of synthesizers and other digital instruments. <https://www.midi.org/midi-articles/about-midi-part-1-overview> MIDI is not an audio file, but rather a sequence of data bytes that encode information about the desired musical notes. MIDI is a universal technology standard that can be used by a variety of hardware and software from virtually every brand or manufacturer. MIDI is flexible and is not beholden to the sound of a single instrument. Using digital synthesizers, the timbre of any instrument can be mimicked and modified and even changed on the fly.

The MIDI standard is comprised of three parts. \*\*\*\*id\*\*\* Firstly, MIDI messages are the software protocol that is used to encode music in a digital format. MIDI messages are comprised of a sequence of bytes that contain information about which note is played, when and how long the note is played, and the velocity at which the note is struck. The velocity of the note can be interpreted by the synthesizers as the volume, brightness, or length of the note, depending on the instrument the synthesizer is programmed to replicate. The units and values of the parameters of a MIDI message are found in **Table \*\*\***.

|  |  |  |
| --- | --- | --- |
| Parameter | Value range | Description |
| Note | 1-127 (C-1-G9) [[1]](#footnote-1) | Musical note |
| Velocity | 1-127 | How hard note was played |
| Duration | Any value in seconds | Length of note |
| Timestamp | Any value in seconds | Time when note begins |

**Table \*\*\*. Midi message parameters**

Secondly, the MIDI standard provides a universal physical transport interface that allows a variety of MIDI-equipped devices to communicate with one another. \*\*\*id\*\*\*. MIDI messages can be streamed between two devices in real time without any perceptible delay. MIDI devices include digital keyboards, electric drum pads, or digital audio workstations (DAWs) that run on most standard operating systems, include PC, Mac, and Linux. DAWs can be used to record MIDI messages and added to the mix of a multi-track digital audio recording.

Thirdly, the MIDI standard provides a file format for the storage of MIDI messages that can be saved. \*\*\*id\*\*\* These files can include multiple tracks for different instruments and contain information for the tempo or speed of the song. These files can be opened in any DAW and some hardware instruments and the notes and tempo can be further manipulated as desired by the musician. Some music software can be used to create sheet music that can be converted to or from a standard MIDI file.

\*\*\*Bartlett, Bruce., and Bartlett, Jenny. Practical Recording Techniques : The Step-by-Step Approach to Professional Audio Recording. 7th ed. 2016. Web.

*Digital Audio Sampling*

Analog audio is recorded into a digital format through digital sampling of the analog signal at a sampling rate fs. Nyquist theorem states that fs of a periodic signal must be twice as high as the highest frequency present in the analog source without introducing aliasing into the digital sample. Because the human ear can detect frequency up to around 20 kHz, digital audio is usually sampled at around 44.1-48 kHz. An ADC for digital audio encodes the amplitude of a sound wave as a binary number, usually with a 16- or 24-bit depth. Because a higher bit depth is capable of encoding more quantized levels in one sample, recording at a higher bit depth produces a more digital recording that is capable of higher signal-to-noise ratio.

Digital audio can be stored in different file formats with different values for fs and bit depth. The WAV file format is the most commonly used uncompressed audio file format. During digital audio recording and playback, the binary values of the audio source are written and read into the buffer memory of the soundcard of the computer at a constant rate. Direct boxes (DI) are external set top boxes that can be connected to an analog sound source (such as a microphone or electric guitar) and contain a ADC that streams the sampled digital audio to the soundcard’s memory buffer through USB or other I/O interfaces.

\*\*\*Bartlett, Bruce., and Bartlett, Jenny. Practical Recording Techniques : The Step-by-Step Approach to Professional Audio Recording. 7th ed. 2016. Web.

*Pitch Detection with CREPE Neural Network*

There are a multitude of techniques to determine the fundamental frequency of a note, each with its own tradeoffs. \*\*\*CREPE paper\*\*\* Many algorithms and models exist to process a digital audio file and determine its f0. Kim, et. al propose a deep convolution neural network operating in the time domain to more accurately determine f0. Kim, et. al refer to their neural network as a Convolutional Representation for Pitch Estimation (CREPE). CREPE takes a 1024-sample slice of a time-domain audio signal with a 16 kHz sampling rate. The data is run through six convolutional layers to produce a 360-dimensional output vector comprised of f0 with respect to time.

CREPE was trained with two data sets. The first data set was a simple synth-based set of audio recordings with very little variation in timbre. Next, CREPE was trained with a set of real-world recording stems from a multitude of different sources. The end result of training led to a pitch detection accuracy of over 90%. CREPE was able to operate effectively for samples with a signal-to-noise ration below 10 dB.

**METHODS**

The audio from an electric guitar was recorded and processed using a custom script in MATLAB. **Code Snippet 1** shows the program that runs while the script is live. The script records ten seconds of audio from the Windows audio input, and then processes the recording and plays back the MIDI messages. First, the MIDI devices are initialized using functions from MATLAB’s Audio Toolbox plugin. There are multiple MIDI outputs in Windows, and they can change based on software or hardware configurations. MIDI mapper plays the MIDI message back in the default piano sound over the current Windows audio output. Loop MIDI takes the outputted MIDI message and loops it back into Windows as a MIDI input. Loop MIDI is chosen as the output when you wish to record to a MIDI message with a DAW.

After the MIDI outputs are initialized, the MATLAB audio recorder is configured. A sampling frequency Fs of 44,100 bits/s was chosen to obtain a high-quality recording. The bit depth of the recording was set to 16 bits, and the recording was performed on one channel, i.e. in mono. The recording duration was set to 10 seconds. In the main while loop, audio is recorded for ten seconds and audio recording is processed by the create\_midi script to create the MIDI signal. The script then uses the Audio Toolbox MATLAB function midisend to send the MIDI message to the selected MIDI output device. Finally, the while loop in the main script pauses for 11 seconds while the MIDI message plays, and the while loop runs again until the script is terminated by the user.



*Create MIDI Algorithm*

The following algorithm was used to convert the WAV file into a MIDI signal. First, the fundamental frequency of the notes were found using the CREPE neural network MATLAB plugin. Next, the notes of the frequencies were found using **Equation 1**. Next, the index of the note array where the notes changed were found and a for loop was run for every one of those change indeces.

In the for loop, the note at the change index of the current iteration of the for loop was found, along with the max sound pressure during which the note was played (note peak). Next, the length of the note was determined by finding the number of samples between the current change index and the next index.

Next, another for loop put together a midi message using the midi notes, used the note peak to find the note volume, and determined the duration and timestamp of the note using the sampling frequency Fs. An array of these midi messages was then returned by the function to be played by the main script.

**Algorithm:** WAV-to-MIDI Conversion

**Input:** N samples of soundwave amplitude for n = 1, 2, 3, … N

Find f0[n]

Find notes corresponding to f0[n]: note = 12\*log2(f0/440) + 49

Find sound pressure levels Lp[n]

Find values of n where notes change (change\_indeces)

Find note\_peaks, note\_lengths, and midi\_notes

for (i = 1; i++; i < length(change\_indeces))

if i == 1

midi\_notes(i) = notes(change\_indeces(i));

note\_peaks(i) = maximum peak of Lp(i:change\_indeces(i));

note\_lengths(i) = change\_indeces(i);

else if i == length(change\_indeces)

note\_peaks(i) = = maximum peak of Lp(change\_indeces(i):change\_indeces(i-1));

midi\_notes(i) = notes(change\_indeces(i));

note\_lengths(i) = change\_indeces(i) - change\_indeces(i-1);

else

midi\_notes(i) = notes(change\_indeces(i));

note\_lengths(i) = change\_indeces(i) - change\_indeces(i-1);

peak = maximum peak of Lp(change\_indeces(i):change\_indeces(i-1));

if(peak is empty); peak = 0; end

note\_peaks(i) = peak;

end

end

Time t = 0:1/Fs:length(notes)/Fs

Generate MIDI signal

for(i = 1; i++; i < length(change\_indeces)

if(midi\_notes(i) == 0)

volume(i) = 0

else

volume(i) = floor of 127\*note\_peak(i)/100;

end

midi\_message =

{

command = “Note”;

channel = 1;

note = midi\_notes(i);

velocity = volume(i);

duration = note\_lengths(i)/Fs

timestamp = t(change\_indeces(i - 1) + 1);

}

end

**Algorithm 1.** WAV-to-MIDI conversion

*Determining Pitch and Note*

The fundamental frequencies of the recording were estimated using the pretrained CREPE convolution neural network. The CREPE network is available as part of the Audio Toolbox plugin for MATLAB. CREPE makes estimating pitch very easy, as can be seen in **Code Snippet 2**. The CREPE training data was downloaded from the MathWorks website and unzipped in the same path as the main script. However, the length of the array of frequencies created by CREPE is not the same length as the input array. Accordingly, the array of frequencies needs to be stretched out to reach the right number of samples, as can be seen in **Code Snippet 2**.



**Code Snippet 2**. Determining the pitches of the recording

Next, the musical notes that correspond to the pitches are calculated using **Equation 1**, as can be seen in **Code Snippet 3**. Because a guitar must be manually tuned, every string might not be exactly have the right pitch as the desired note. Therefore, the floor of the note calculated by **Equation 1** must be found because a MIDI message must have a note that is an integer. Additionally, because the CREPE network assigns a value of NaN to a pitch it cannot detect, all instances of NaN in are set to 0, which is a valid MIDI note value.



**Code Snippet 3.** Determining musical note from frequency

*Determining note length and timing*

The note length and timing is determined by determining what indeces in the recording array correspond to a change in the note (change\_indeces), and using the difference between concurrent values of change\_indeces along with Fs to determine the length of the note in seconds. **Code Snippet 4** shows the code that implements this procedure. Next, an array of time values the same length as the recording was created using Fs to come up with a timeline in seconds. In the last for loop, the value of the change index for a particular note was used as the index for the value in the time array (t) to get the moment in time the note was first played. Both the timestamp and the note duration were saved in the MIDI message for each particular note.



**Code Snippet 4**. Determining note length and timestamp

*Determining note volume*

The volume of a note was determined by calculating the sound pressure generated by a sound wave. Because a sound wave has peaks a troughs, volume is correlated to the total pressure created in the air. As seen in **Code Snippet 5**, MATLAB’s Audio Toolbox has a built-in function to calculate sound pressure from a recording. Because the sound pressure might not reach a maximum until after the note is initially strummed, the peak value of the pressure was found during the duration of the note, and then normalized to the scale required by the velocity parameter of a MIDI message, i.e. an integer between 0 and 127.



**Code Snippet 5**. Determining note volume.

CREPE is very sensitive, and it is possible to change the pitch of the guitar without strumming a note. For example, when lifting a finger off a guitar fret, it is possible that a harmonic could ring out before another note is strummed on that string. CREPE picks up these harmonics very well, so the volume must be set to zero for notes that do not create a peak in the sound pressure. If only notes that trigger sound pressure have volumes, the harmonics will not make it into the MIDI message and the guitar riff will sound more like what the human ear hears.

*Creating MIDI Messages*

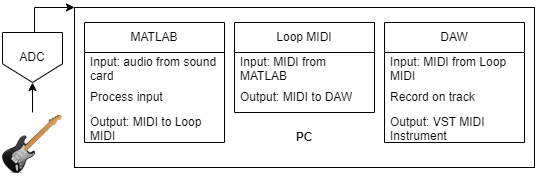
MIDI messages can be created using the MATLAB Audio Toolbox plugin. Different MIDI instructions can be encoded, including note, note on, note off, pitch bend, channel pressure aftertouch, or all sound off. Each MIDI command has a value for channel, note (pitch), velocity (volume), and timestamp. “Note” MIDI messages also have a value for duration. **Code Snippet 6** demonstrates how a MIDI message was created in MATLAB. The comment in the code shows the format of a MIDI message

**Code Snippet 6**. Creating the MIDI messages

*Hardware and Software Setup*

Audio was recorded by plugging a Fender Telecaster electric guitar into the ¼ inch line input for an M-Audio M-Track Solo USB recording interface. The M-Track solo was connected through a USB to a personal computer. Audio processing was performed in MATLAB using plugins from MathWork’s Audio Toolbox library. A new MIDI track was created and set to record in the DAW, and a new audio recording was started.

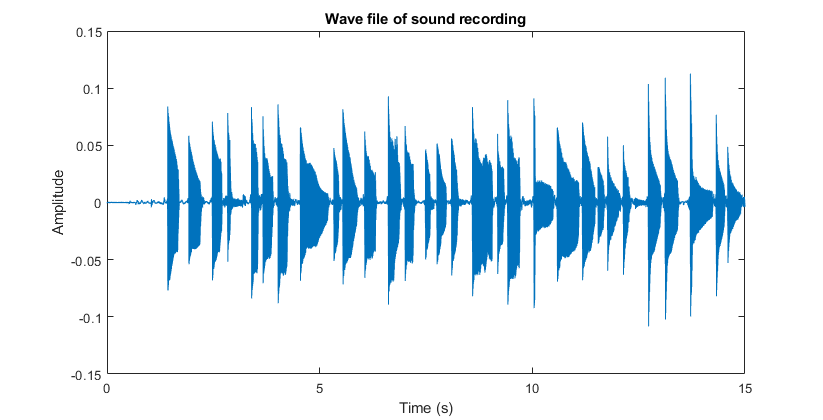
A Guitar short guitar riff was recorded in MATLAB and run through a custom script to determine the melody’s change in pitch, their corresponding musical note values along with the timing and volume information of all the notes. The script then generated a matrix of bytes from a MIDI signal, which encoded the note, its volume, its duration, and its timestamp. The MIDI signal was then streamed into Loop MIDI, which took the signal from the selected MIDI output port and looped the MIDI back into a Windows MIDI input port. Next, the DAW took the input from Loop MIDI and recorded the signal into the track as it was played in MATLAB. The recording session on the DAW was ended, and a track was successfully recorded. Next, by changing the VST MIDI instrument plugin, and desired digital instrument could be used to play back the notes of the guitar riff. Additionally, when opening up the MIDI track in the DAW, individual MIDI notes could be dragged and dropped to change pitch or timing.



**Figure \*\*\*.** Audio hardware and software setup

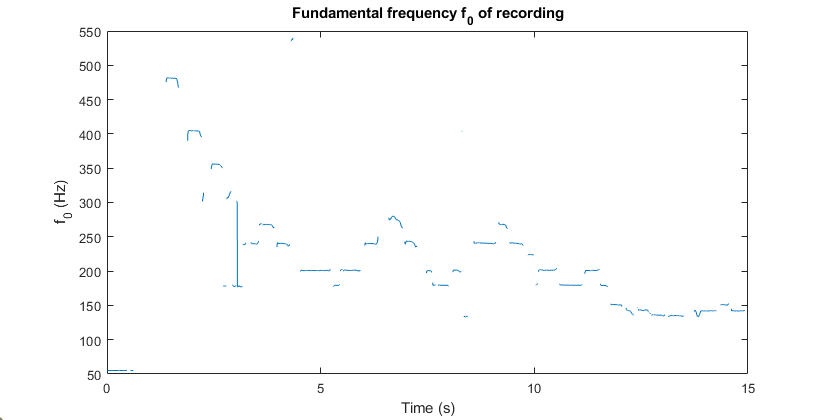
**RESULTS**

*A. Setting up MIDI Ports and Recording Audio*

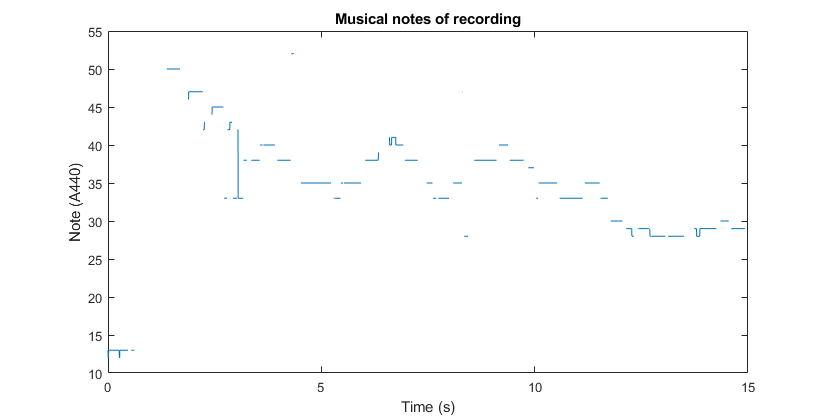


**Figure \*\*\*.** Sound wave of recording

*B. Detecting Pitch and Note*

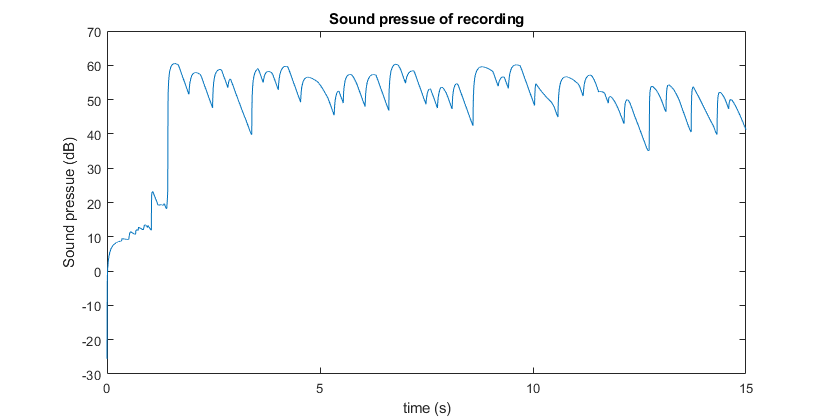


**Figure \*\*\***. Fundamental frequency of open strings on guitar



**Figure \*\*\*.** Keys of open strings on guitar

*C. Detecting sound pressure*



**Figure \*\*\*.** Sound pressure of recording

\*\*\*Insert pictures of track from DAW.

*D. Description of how well it worked*

**IV. CONCLUSION**

* Summary
* Possible improvements
* Lessons learned

**Appendix**

**References**

1. MIDI represents A4 as note 69 instead of 49 [↑](#footnote-ref-1)