

Saxophone Sonority: An Acoustic Explanation of What Differentiates the Tone Quality of Beginner and Professional Alto-Saxophonists

Vidur Sanghi

Received August 28, 2023

Accepted November 04, 2023

Electronic access December 15, 2023

This paper explores the difference in the tone quality of Alto Saxophone samples by professionals and beginners. Beginners are often guided to render “warmer” or “fuller” tones but such abstract guidance is confusing as it does not offer feedback or a tractable path to emulate a professional’s tone quality. Students eventually achieve advanced tone quality after months or even years of practice, often enduring significant frustration and demotivation along the way. Even after beginners develop a superior tone quality, it remains a feeling rather than a tangible attribute they can describe or work to achieve consistently. By comparing Fast Fourier Transforms, Mel Spectrograms, and Root Mean Square $RMS = \sqrt{\frac{1}{n} \sum_i x_i^2}$ values, this paper analyzes physical differences between beginner and professional players. The RMS analysis yields a Pearson’s Correlation of 0.704 and Spearman’s Correlation Coefficient of 0.631, indicating that tone quality produced by professionals correlates strongly to higher RMS values of their samples, compared to beginners. Further, the correlations indicate that a ‘Linear Regression’ model can be used to identify recordings by professional and beginner saxophonists.

Introduction

The Saxophone was first created in 1846 by Adolphe Sax. It is the only woodwind instrument made of brass, and it employs a Reed (thin strip of material secured to the mouthpiece with a ligature) that vibrates to produce sound¹. In order of ascending size, weight, and descending pitch, the four primary types of Saxophones are Soprano, Alto, Tenor, and Baritone.

Currently, students are guided to perform exercises to improve tone quality without an explanation for how these exercises help them achieve that goal. Since human perception of quality is subjective, defining the differences between good and bad tone quality is beyond the scope of this paper. Instead, the focus of this paper is identifying and explaining acoustic differences between samples by professional and beginner players. The following parts of the introduction define physics concepts used in the analysis, explain how the saxophone works, and clarify concepts around tone quality.

Sound, Tone (Timbre), and Frequency: The human ear detects changes in pressure as sound, in three categories:

1. Pitch refers to the frequency of a sound and allows the ear to discern how high or low a sound is,
2. Loudness is the intensity at which the ear perceives a sound, and
3. Tone (also known as Timbre) is the attribute that lets you differentiate sounds with the same pitch and frequency.

Change in pressure is represented by frequency and measured in Hertz (1 Hertz = 1 Cycle/Second). The lowest frequency of a note is called the ‘Fundamental’, and its whole number multiples are called ‘Harmonics’. Any frequency above the Fundamental, including Harmonics, is an ‘Overtone’.

Factors that influence Tone (Timbre): There are two kinds of tone: Pure (also known as simple) and Complex. Pure tones consist of a single, repeating frequency in the shape of a sine wave, while complex tones comprise periodic waves with patterns that repeat². Combining two (or more) pure sine waves (i.e., pure tones) results in a complex tone. Three factors affect tone: 1) Harmonic Content, which is determined by the presence and strength of overtones, 2) Attack/Decay, which represents the intensity with which a note is started and the rate at which it diminishes, and 3) Vibrato/Tremolo, which are periodic changes in the pitch of a note³.

Saxophone instrument design: The saxophone is shaped like a conical bore and is considered a reed-pipe system. The sound produced by a saxophone is the result of vibrations from a reed, which does not do much to alter the frequency but creates a positive pulse that travels through the “pipe” until it reaches the bell, where excess pressure drops to zero and a negative pulse travels back toward the reed. This results in a conversion from Steady Power (DC) to Acoustic Power (AC). The amount the reed oscillates is based on the pressure difference created by the air⁴.

Level of Saxophone Players: For this research, Beginners

are those who have played for under five years and do not play music as their primary source of income, and Professionals are those with at least an undergraduate degree in Saxophone performance, who teaches at a school or college, and rely on music as their primary source of income⁵.

Methods

Data Collection Phase 1:

This paper focused on Alto Saxophone data for this research since it is the most commonly played saxophone in classical music. A total of 307 samples were collected for analysis, in two phases. During Phase 1, samples from 47 players were recorded, and 114 recordings were downloaded from YouTube. Phase 1 sample data collection criterion included:

1. Audio recorded on an iPhone or Android phone using an application like Voice Memos
2. No external sound was recorded (including other instruments, people talking, metronome ticks, etc.)
3. Recordings exclude any extended technique (like *al-tissimo*, multiphonics, slap tongues, etc.)

Several recordings were captured from students at the local elementary, middle, and high schools. Further, Saxophone professors across the country also volunteered to share recordings of samples by them and their students. To increase the sample size, similar YouTube recordings were also downloaded. These recordings captured a range of Tempo and playing styles across beginners and professionals, to help focus the analysis on Tone (Timbre). To ensure broader representativeness of sample data beyond California, from where the student recordings were collected, the YouTube recordings used were focused on performers from other states.

Data from Phase 1 was used to generate Mel Spectrograms, as detailed in the Methodology section. This initial analysis led us to embark on Phase 2 of data collection in order to support a more clinical research approach, as discussed below.

Data Collection Phase 2:

Data collection in Phase 2 used a more clinical approach, by requesting a subgroup of players in Phase 1 to provide separate recordings for each of the 33 notes in the normal range of the Alto Saxophone. Phase 2 sample data collection criterion included:

1. Record notes with the same dynamic (loudness)
2. Maintain a distance of 3 feet from the bell of the saxophone to the recording device

3. Hold each of the 33 notes recorded, for 4 to 5 seconds
4. Play the note for the entire duration of the recording i.e., only the note is audible throughout the recording)
5. No external sounds are recorded

This round included recordings from one professional and 3 beginners. Samples from 19 other beginners were excluded as they either contained extraneous audio, or they were unable to hold their notes for 4-5 seconds. All players (male and female) who provided sample data were between 14 and 60 years of age and had played the saxophone for at least 3 years. None of the players indicated any medical conditions or activities they participated in that would affect their tone quality.

Although the data collected for Phase 2 was mainly from California, the results of our analysis coincided with the findings in Phase 1 which featured a more diverse data set. The following section discusses the results of the analysis of the data collected.

Results

Mel Spectrogram Analysis

As discussed, changes in pressure manifest as sound to the human ear, and these changes in pressure may be sampled digitally as an Audio Signal and represented as a Waveform in the time domain (Figures 1 & 2). The Fourier Transform (FT) is a mathematical formula that transforms an audio signal from the time domain to the frequency domain, thus revealing its spectral components and providing frequency information about the signal - known as the Spectrum.

The Fast Fourier Transform (FFT), is a widely used signal processing algorithm to compute the Fourier Transform (i.e. Spectrum) of a signal, and it enables visualization of the amplitude (power) at each overtone of the signal (Figures 3 & 4). Most audio signals, including Alto Saxophone samples, are Non-Periodic, in that their frequency content varies over time. Hence, the Spectrum of Non-Periodic signals is represented by calculating Short-Time FFTs of several overlapping windowed segments. This representation called the signal's Spectrogram, includes Time along the X-axis and frequency along the Y-axis, adds a log scale on the Y-axis for signal loudness or amplitude over time at different frequencies, and also uses color to plot the signal's decibel levels. Finally, the Mel Spectrogram transforms the frequency (Y-axis) on a Spectrogram discussed above, to account for the fact that a human ear does not perceive frequency on a linear scale in that it can't discern the difference between 10.0 kHz and 10.05 kHz with the same accuracy as between 0.40 kHz and 0.45 kHz⁶. Hence, the Mel Spectrogram employs a Mel Scale on the Y-axis, where equal

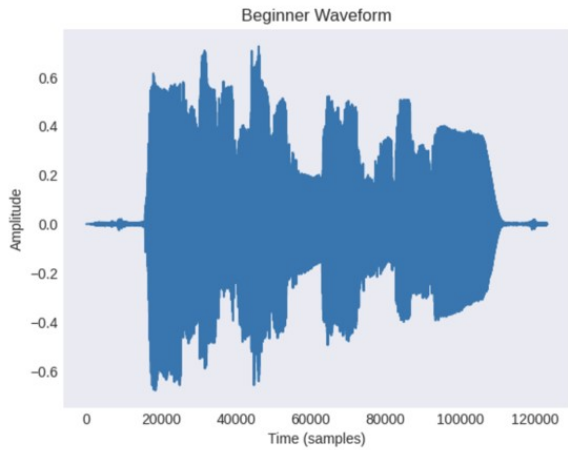


Fig. 1 Waveform of audio sample by Alto Saxophone Beginner (7th grader) playing a scale; from Phase-1 data.

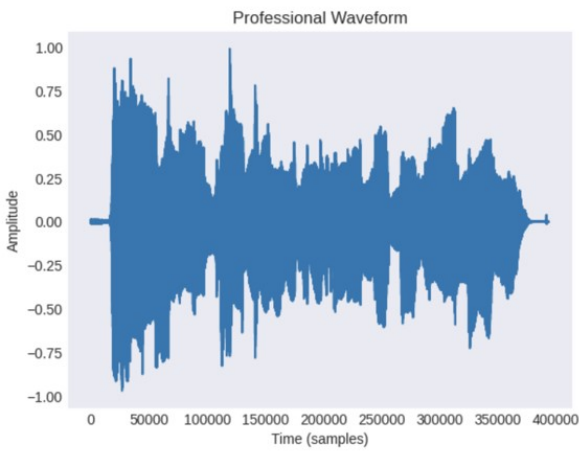


Fig. 2 Waveform of audio sample by Alto Saxophone Professional playing a scale; from Phase-1 data.

Note: The scale of this waveform is different in Figure 1, to account for its greater Amplitude and Time duration.

distances in pitch sound are equally distant to the human ear (Figures 5 & 6)..

Based on the Fast Fourier Transform (Figures 3 & 4), the professional player has stronger harmonics, since the amplitude of their signal at higher frequencies is greater compared to the beginner player. However, looking at the Mel Spectrograms (Figures 5 & 6), it is evident that the professionals have a greater amplitude at harmonics, but not the rest of their overtones, creating a definition between the pockets where there are harmonics and other overtones. On the other hand, there is less definition between the beginner's harmonics.

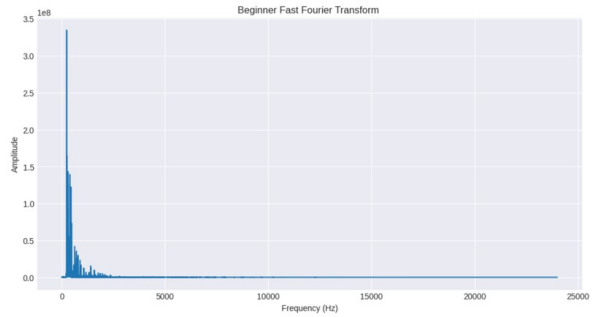


Fig. 3 FFT plot of audio sample by Alto Saxophone beginner (7th grader) playing a scale, from Phase-1 data.

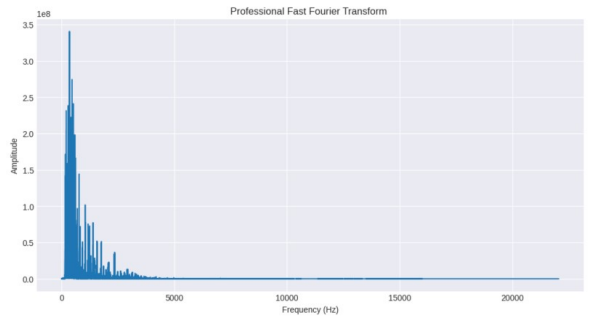


Fig. 4 FFT plot of audio sample by Alto Saxophone Professional playing a scale, from Phase-1 data.

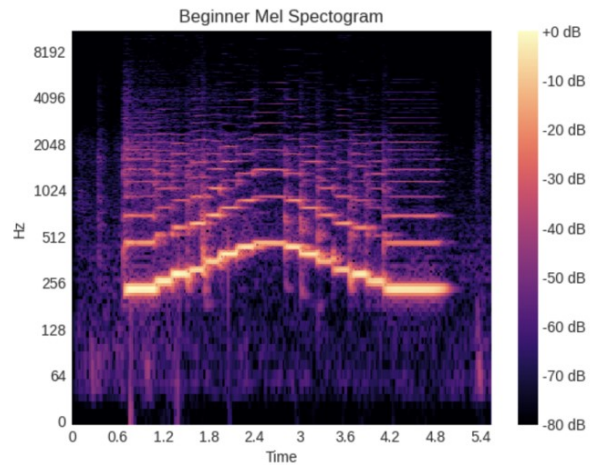


Fig. 5 Mel Spectrogram of the audio sample by Alto Saxophone Beginner playing a scale, from Phase-1 data.

Root Mean Square:

The energy of samples from professional and beginner players helps quantify their differences as seen in the preceding figures. A Root Mean Square (RMS) operation yields the av-

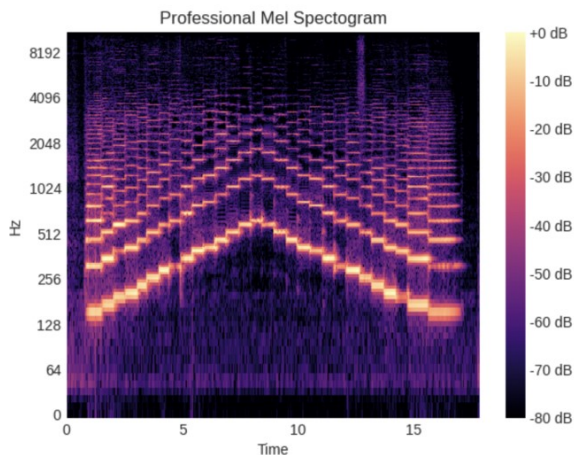


Fig. 6 Mel Spectrogram of the audio sample by Alto Saxophone Professional playing a scale, from Phase-1 data.

erage of a set of samples by adding the square of each sample, taking the average of the sum of squares, and taking the square root of the averages $RMS = \sqrt{\frac{1}{n} \sum_i x_i^2}$. When the resultant RMS value is squared, it is proportional to the power of a signal. This means the RMS will indicate how much power is contained in the waveform, and all of its harmonics⁷⁻⁹. Using data from Phase 2, the average RMS value was calculated, by dividing the total RMS Value by the length of the audio sample (Figure 7). Further, the RMS across all Beginner samples is 0.071 amperes, and the average RMS across all Professional samples is 0.132 amperes (Figure 8).

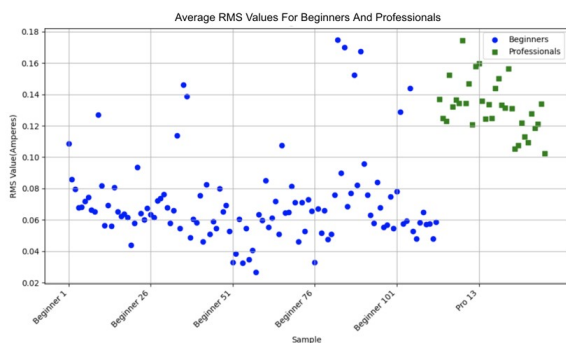


Fig. 7 Scatter Plot with 146 RMS values from Beginner and Professional samples in Phase 2.

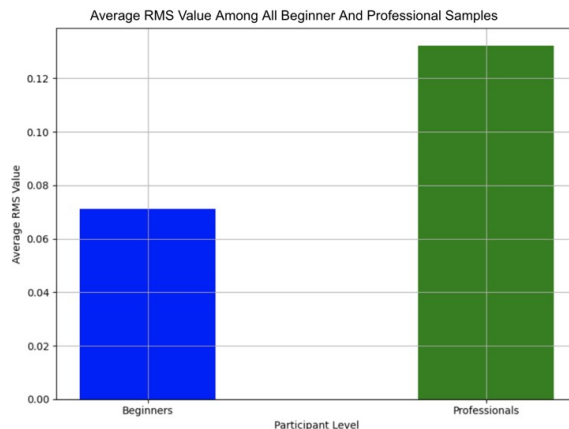


Fig. 8 Average RMS Value Among Beginner and Professional Samples. Average of All of the scatter plot data from Figure 7.

Discussion

Data Collection Part 1 Analysis:

Based on the findings from the Fast Fourier Transform analysis, it is evident that differences between Professional and Beginner Alto Saxophone players are not solely driven by the total energy with which they play. Instead, these differences are based on the spread of the energies in the spectrum of their respective samples. Professionals have the ability to control their airflow and create sounds at the base frequency and harmonics of the notes they are playing, resulting in a warm tone. Also, professionals are able to control their airflow to produce minimal energy outside the base frequency and overtones of the note they are playing. Therefore a student or professional shall not be able to produce a better tone quality by playing the Alto Saxophone louder with a lot of energy.

Based on the initial waveforms of the beginners and professionals, the amplitude of the beginners' sound is consistently less than the amplitude of the professional player. This is also consistent with the Fast Fourier transform, where the professional's upper harmonics are significantly louder than the beginner's upper harmonics. This is also evident in the Mel Spectrogram, which is visibly brighter for the professional's signal since there is more energy.

To validate the correlation above, 50 randomly chosen Mel Spectrograms were manually sampled without their labels. Based on differences in the strength of overtones, the quality of the attack and sustain, and the amount of vibration/tremolo, 47 of the 50 samples were classified accurately.

Data Collection Phase 2 Analysis:

In addition to confirming that the average RMS values suggested that beginners' signals had less power to professionals, Pearson's and Spearman's correlation between the ground truth (average RMS values) and labeled values confirmed a correlation between the power of a signal and skill level. 1) Pearson's correlation measures the strength of the linear relationship between two variables. 2) Spearman's correlation measures the monotonic relationship (when one variable directly increases or decreases with another) between data points^{7,10}. In this case, each of these correlations indicates whether a sample is played by a beginner or professional based on the amplitude. The Pearson's Correlation Coefficient is 0.704 and the Spearman's Correlation Coefficient is 0.631. Finally, the calculated p-value for these correlations is 3.579e-23, which indicates a statistically significant correlation¹¹. These results are a confirmation of what is visible through the Fast Fourier Transforms and Mel Spectrograms: Professional players have clearer and stronger harmonics. The following section explains what the values above mean.

According to industry standards, a Pearson's correlation value between 0.5 to 1, a Spearman's correlation of 0.6 to 0.79, and a p-value less than 0.05 indicates accuracy and consistent correlations^{7,11,12}.

Combined Analysis:

Audio recordings only capture a waveform of the output of the sound a saxophonist plays. This means it captures data regarding the energy used to create the sound in an audio sample. However, this is not necessarily an indication of the loudness of a sound (in decibels).

There are two ways to alter the amount of air traveling through the saxophone: 1) blowing more air and 2) increasing the cross-section. However, even if you keep the amount of air that travels through the saxophone constant (i.e. blowing faster air through a smaller opening and blowing slower air through a larger opening), the sound and loudness are not necessarily the same. Generally, slow air through a larger cross-section will mean a deeper, louder sound, and faster air will be a higher-pitched, softer sound¹³.

Based on the description above, it can be concluded that the professional can play at a softer volume, while still cycling an equal (or greater amount of air) through the instrument as a beginner. This is what is evident in the recording of professional players. A couple of factors that might account for this include: 1) Professionals have a more focused airstream, projecting all their air through the instrument, ensuring none of it leaks outside. 2) Professionals have a faster airstream, cycling more air through the instrument in the same amount of time, producing more energy while maintaining the same loudness. Either of these differences, or both, would result

in the difference observed based on the RMS values and Mel Spectrograms.

Conclusion

This paper explains why professionals and beginners sound different and have different timbres. By understanding the physical difference, students can isolate areas where they are not as strong as professionals and practice strengthening those factors, rather than others which may not be as helpful. Further exploration into different practice techniques meant to improve tone quality, with a physical explanation for why they work, would be beneficial for students to understand how they are improving. Further, an expansion of this research could help students have a tool they can use to visualize the difference between their sound and the sound of a professional player. The following are plans and ideas to expand this research:

1. Measure the decibels of an audio signal, and the audio signal to ensure that all recordings have the same loudness
2. Explore other woodwind and brass instruments to explore if a similar correlation lies between them
3. Understand the influence of different equipment, such as mouthpieces, reeds, and instrument neck/body
4. Explore tone quality in the extended range of the saxophone (altissimo)
5. Exploration of how recommended practice techniques help students improve
6. Building and training a Machine Learning model based on Mel Spectrograms using image recognition
7. Use acoustic metrics like Spectral Centroid, Spectral Bandwidth, and Harmonic-to-Noise Ratio in order to analyze the energy of the signal at specific bands

References

- 1 C. R. Nave, *The Saxophone*, Georgia State University – Hyperphysics concepts, <http://hyperphysics.phy-astr.gsu.edu/hbase/Music/sax.html>.
- 2 D. W. Roberston, *Sounds and Tones (Pure and Complex)*, Middle Tennessee State University.
- 3 *Sound is a pressure wave*, The Physics Classroom. <https://www.physicsclassroom.com/Class/sound/u1111c.cfm>.
- 4 D. Hinze, *Spectral Analysis of an Alto Saxophone*, University of Illinois, Physics.
- 5 T. Fountain, *Professional musician vs. hobby musician*, <https://www.openmicuk.co.uk/advice/professional-musician-hobby-musician/>, Which one are you? Open mic.

-
- 6 L. Roberts, *Medium*.
 - 7 D. I. Weir, *Spearman's correlation*, Stats Tutor.
 - 8 R.M.S., *Sweetwater*, <https://www.sweetwater.com/insync/rms-root-mean-square/>.
 - 9 D. Vaidya, *Root Mean Square*, Wall Street Mojo.
 - 10 *The Pearson's r*, *Virginia Commonwealth Univeristy*, <https://www.google.com/url?q=http://www.people.vcu.edu/~pdattalo/706SuppRead/Pearson%27s%2520r.html&sa=D&source=docs&ust=1692558725747781&usg=AOvVaw3UiOU-XPaUOkU0aGBjUZiP>.
 - 11 N.-F. S, *The Korean Journal of Pain*, **30**, 241–242.
 - 12 *Pearson's Correlation Coefficient*, *Statistics Solutions*, <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/pearsons-correlation-coefficient/#:~:text=High%20degree%3A%20If%20the%20coefficient,to%20be%20a%20small%20correlation>.
 - 13 J. Wolfe, *Air Speed and blowing pressure in woodwind and brass instruments: how important are they?* *Australian Research Council*, <https://newt.phys.unsw.edu.au/jw/air-speed.html>.