Automatic Tuning Saxophone Mouthpiece

Author: Zhang Yifan

Supervisor: Koen Langendoen

Contents:

Chapter 1: Introduction

The saxophone is a single-reed woodwind instrument. Like all other woodwind instruments, the saxophone tends to be out of tune. Even saxophone players with more than ten years of experience struggle with pitch accuracy when they need to play harmony with other instruments or vocals. Traditionally, musicians tune their instruments before the performance by matching the instrument's frequency with the frequency of a sound fork. They adjust their instrument until they are in perfect harmony with the sound fork. This tuning process costs anywhere from a few seconds to a few hours, depending on specific instruments. From our personal experience, the saxophone tuning process takes roughly 30 seconds. However, finishing such a tuning process only means that the central note of the saxophone is in tune, while other notes can still be out of tune. Because of the saxophone’s structure, it is only possible to tune one note to the accurate frequency before the performance. The saxophonist can only tune other notes during the performance by first determining the note by ear and then tuning the note by mouth. Tuning by ear and mouth is a difficult skill that needs years of practice to learn and a lifetime to perfect. We want to create a device that can provide a more efficient way for musicians to tune their saxophones in real time.

1. 1 Motivations

1.1.1 For bands:

Many pop or rock bands have a saxophone player, and that saxophone player will frequently act as an accompaniment to the vocals. In such a scenario, a loud and out-of-tune saxophone can be disastrous for the performance. Because of the saxophone’s inherent tendency to go out of tune and loudness, a saxophone player normally has to reach intermediate or even advanced levels to be accepted by amateur bands. As a result, many amateur saxophone players cannot play in bands with their friends. If we can help amateur saxophone players to play more in tune, then saxophone players in general will have more opportunities to play in bands and make more friends.

1.1.2 For orchestra:

Although our current device is developed just for saxophone, similar products can easily be made for other woodwind instruments. This can greatly improve the pitch accuracy of many non-professional orchestras, which are dominantly student orchestras. This product can shorten the training time needed for student orchestra woodwind members. The implementation of such a device may significantly improve the quality of high school orchestras’ performances.

1.1.3 For personal recording:

Most music YouTubers, including ourselves, edit their soundtracks before posting a video. A major part of this editing is to make everything in tune so their videos don’t easily get roasted by viewers. Our new device can decrease, or even avoid the workload of editing

1.2 Research questions

1.2.1 Possibility:

The exact frequency of a note played on a saxophone is determined by multiple factors. Whether it is possible to adjust the frequency of that note by manipulating one or more of those factors is yet to be explored. The mouthpiece is the part of a saxophone that has the biggest influence on the sound of the saxophone. Thus, we decided to start from the mouthpiece. Our device should change the frequency of the sound in real time. Such a device entails knowledge not only of embedded systems, but also of mechanical engineering and fluid dynamics. We want to see whether it is possible to develop such a device, despite our relative lack of knowledge in many related disciplines.

1.2.2 Accuracy:

The pitch of musical notes is a subjective perception. The accuracy of our device has to make people perceive the note as in tune. Different people have different levels of sensitivity regarding sound frequencies. Musicians with trained ears are generally more sensitive to frequency changes. We want to explore how accurate our device can be.

1.2.3 Speed:

Notes only last a certain amount of time. Our device must be fast enough to finish the tuning process a certain amount of time before the note ends so the note can be perceived by the audience as being in tune. The exact length of each note varies, depending on the speed of the music piece. We want to know the minimum achievable reaction speed of our device, and consequentially the shortest duration of a tunable note.

1.2.4 Ergometry:

Our device is developed with the eventual goal of helping musicians play their music better. We highly suspect that it will not be as comfortable as a normal mouthpiece. We want to know if the discomfort caused by our device is tolerable, how big an impact it will have on the sound, and whether we can reduce those negative impacts.

In conclusion, we aim to make a device that can help saxophone players tune their saxophones to a certain accuracy within a certain time window, without incurring too much discomfort.

1.3 Development Approach

We developed our device with the following steps:

1. Pick an idea. We have many ideas that all seem promising. Our first step is to pick the currently most reasonable idea and try to verify it.

2. Build a static prototype. To verify an idea, we build a prototype of it using 3D printing technology.
3. Test our prototype. If our static prototype can prove the viability of the idea behind it, go to the next step; otherwise, go back to step one and start again from the beginning.
4. Build a complete prototype. This prototype is supposed to be capable of all designed functionalities.
5. Test and evaluation. We test whether the prototype behaves as expected. If not, is it still acceptable? If not acceptable, we find flaws. Depending on the testing outcome, we can either call it a success, go back to step three, or even go back to step one.

The development of new devices is never without setbacks. We went back and forth multiple times. We made the following diagram to further illustrate our development approach.

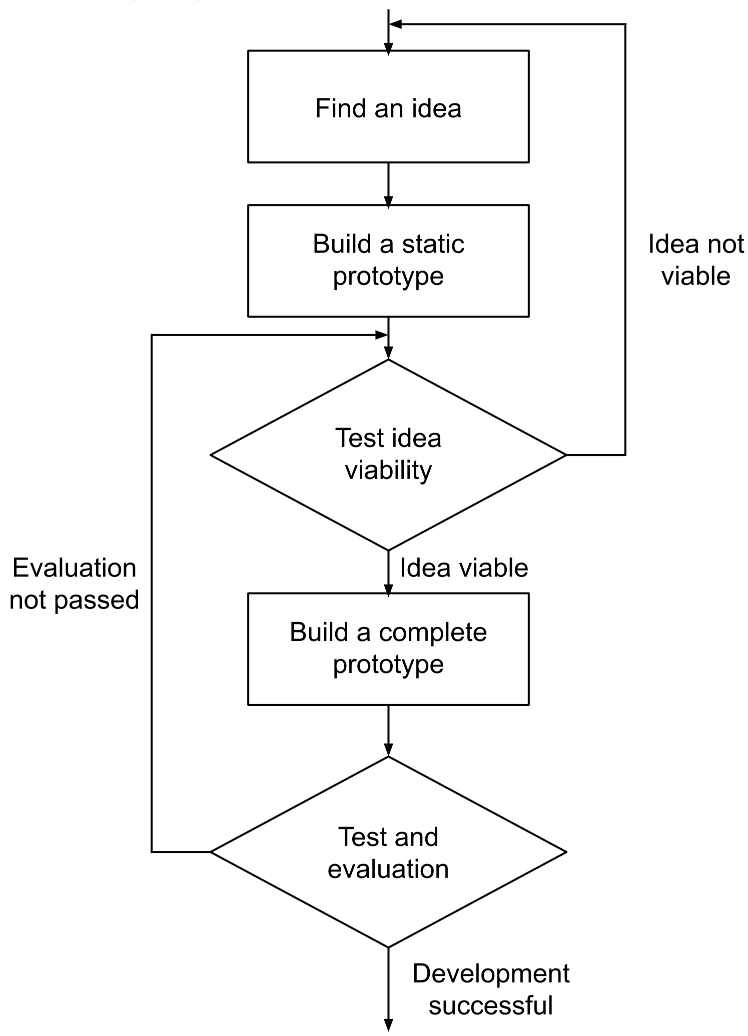


Fig. 1. Development approach diagram

Chapter 2: Background and Related Work

Scientists try to view music as mechanical waves with different frequencies, while musicians view music as perceived feelings. This results in a different set of words that describe similar concepts, but with some unignorable nuance. We made the following table to clarify the differences and similarities among jargon of both scientists' objective world and musicians' subjective world.

Musicians' language	Scientific definition
Pitch: The perceived frequency of a sound, determining its highness or lowness and corresponding to the musical attribute of tone. Pitch is a subjective human perception.	Frequency: The number of oscillations or cycles per second of a sound wave, measured in hertz (Hz), determining the pitch of the sound. Frequency is an objective fact.
Note: Symbols representing a specific pitch in musical notation. In Western music, there are 12 different notes, which correspond to 12 keys (7 white and 5 black) of each interval on a piano.	Note: A preset collection of frequencies. Each of the 12 notes has a specific frequency. Normally, it is not allowed to change these frequencies. Being in tune means right on one of these frequencies.
Sharp: A sharp note means that it has a higher pitch than the specified pitch of its note.	Sharp: A sharp note means that it is in a higher frequency than the specified frequency of its note.
Flat: A flat note means that it has a lower pitch than the specified pitch of its note.	Flat: A flat note means that it is in a lower frequency than the specified frequency of its note.
Intonation: Accuracy of the pitch when playing. Good intonation means that notes are perceived as being in tune with the intended musical scale or key.	Intonation: Intonation is purely a subjective feeling based on perception and has no scientific definition.
Timbre: The unique quality or color of a sound that distinguishes it from other sounds, even when they have the same pitch and intensity.	Frequency spectrum: The distribution of different frequencies within a given sound, showcasing the various components that contribute to its overall timbre.

2.1 Introduction to Saxophone

A saxophone is composed of 6 major parts as shown in the picture below.

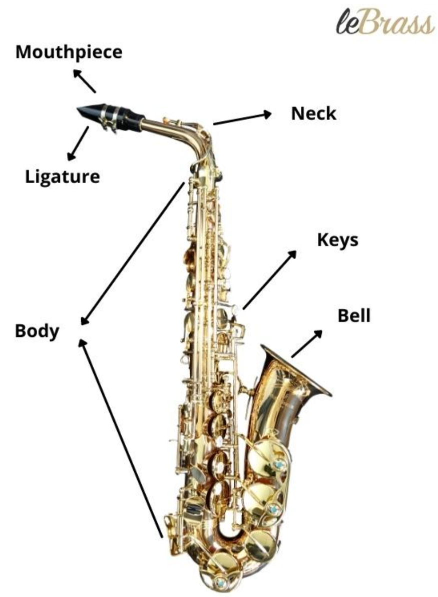


Fig. 2. Saxophone composition.
Taken from leBrass

The saxophone relies on a single-reed mouthpiece and a system of keys and tone holes to produce a wide range of notes. The player's breath causes the reed to vibrate against the mouthpiece, generating sound waves that travel through the instrument's body and exit through the bell. The pitch of the saxophone is controlled by the length of the closed part of the body of the instrument. The bigger this closed portion, the lower the note, and vice versa. Pressing different keys will open or close different holes in the body to change the effective length of the closed body. As shown in Fig.3, the number of holes closed is directly related to the frequency of the standing wave and, thus the pitch of the sound. The right saxophone has fewer holes closed, thus producing a higher frequency, and thus a higher pitch. There are many different kinds of saxophones (soprano, alto, tenor, etc.). In this essay, for simplicity, the word saxophone refers to the alto saxophone.

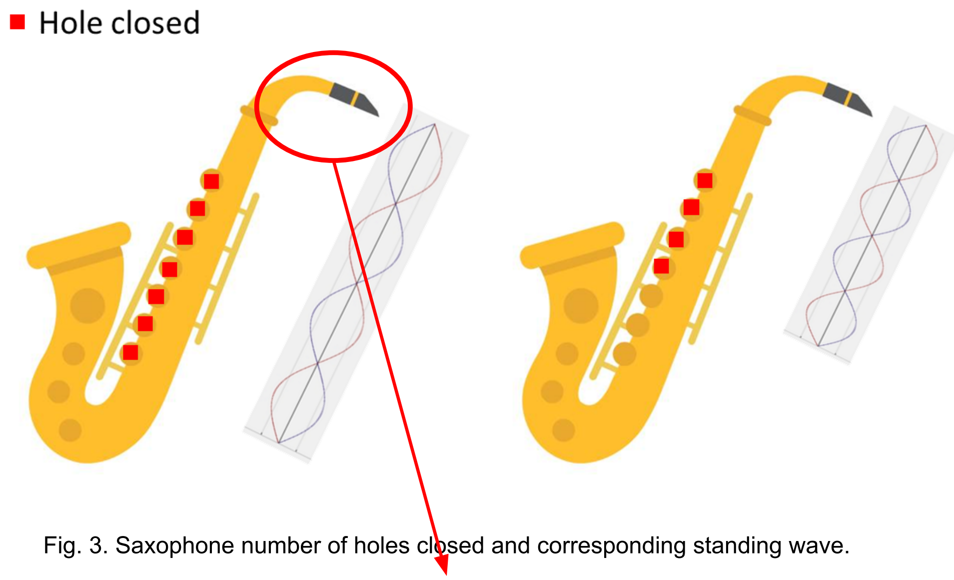


Fig. 3. Saxophone number of holes closed and corresponding standing wave.

Cork and mouthpiece | Backward mouthpiece position | Forward mouthpiece position

Fig. 4. Saxophone mouthpiece positions

However, the process above can only determine the approximate pitch to be played. The exact pitch also depends on the following 5 factors:

1. Mouthpiece position(see Fig.4). A mouthpiece placed in a more forward position results in sharper sounds while a mouthpiece placed in a backward position results in flatter sounds. Changing the mouthpiece position is also the most typical way that saxophone players tune their saxophones.
2. Embouchure: Embouchure refers to the way musicians shape their lips and control oral muscles while playing the saxophone. It significantly impacts pitch. Increased lip pressure raises the pitch by causing the reed to vibrate faster, producing higher frequencies. Conversely, relaxing the embouchure lowers the pitch. Altering tongue position and oral cavity shape can further modify pitch. Varying air pressure through embouchure also affects pitch, with stronger air pressure yielding higher notes and softer air pressure producing lower ones. Skilled saxophone players use precise embouchure adjustments to achieve accurate intonation and expressive musical phrasing, while most saxophone players' unrefined embouchure control makes their notes out of tune easily.
3. Temperature and Humidity: Environmental conditions, such as temperature and humidity, influence the pitch of the saxophone. As metal expands and contracts with temperature

changes, it can affect the instrument's overall pitch. Saxophone players often need to adjust their playing or reed selection in response to temperature and humidity variations.
4. Reed Strength: The thickness and strength of the reed influence the pitch. A softer reed produces slightly lower pitches, while a harder reed results in higher pitches. Reed's strength is also influenced by humidity and the degree of use.
5. Instrument imperfectness: Instrument imperfectness can be slightly improved by buying a more expensive saxophone, but even the best saxophone is not perfect in terms of pitch accuracy.

2.2 ISO International Standard Pitch

To tune a musical instrument, we need to first define what is being in tune. All civilizations on earth shared interests and pursuits in music and consequentially developed many different scale systems. Alternatives to the Western 12 equal temperament include the Byzantine/Arabic scale, Chinese scale, and many more. For this essay, we only focus on the Western 12 equal temperament.

In 1975, the International Organization for Standardization (ISO) adopted A4 tuned to 440 Hz as the international standard pitch. (*ISO 9001:2015(en), Quality Management Systems — Requirements*, n.d.) This decision was influenced by several factors, including the fact that 440 Hz had already been widely accepted in various European countries. The choice of 440 Hz was also influenced by the convenience of having a pitch that is easily divisible by 2 (to produce octaves) and by 5 (to produce semitones).

It's worth noting that while 440 Hz is the international standard, some orchestras and musicians still use different pitch standards. For example, some historical performance groups may use lower pitches that were more common in earlier periods of classical music. Additionally, there has been some debate and experimentation with different tuning systems and pitches in contemporary music. Nevertheless, the adoption of 440 Hz as the international standard pitch aimed to provide a common reference point for musicians around the world, as well as our device.

With A4=440Hz as the anchor point, we get Table I after applying the 12 equal temperaments:

Note Frequency Chart

	Octave 0	Octave 1	Octave 2	Octave 3	Octave 4	Octave 5	Octave 6	Octave 7	Octave 8
C	16.35	32.70	65.41	130.81	261.63	523.25	1046.50	2093.00	4186.01
C#	17.32	34.65	69.30	138.59	277.18	554.37	1108.73	2217.46	4434.92
D	18.35	36.71	73.42	146.83	293.66	587.33	1174.66	2349.32	4698.64
D#	19.45	38.89	77.78	155.56	311.13	622.25	1244.51	2489.02	4978.03
E	20.60	41.20	82.41	164.81	329.63	659.26	1318.51	2637.02	5274.04
F	21.83	43.65	87.31	174.61	349.23	698.46	1396.91	2793.83	5587.65
F#	23.12	46.25	92.50	185.00	369.99	739.99	1479.98	2959.96	5919.91
G	24.50	49.00	98.00	196.00	392.00	783.99	1567.98	3135.96	6271.93
G#	25.96	51.91	103.83	207.65	415.30	830.61	1661.22	3322.44	6644.88
A	27.50	55.00	110.00	220.00	440.00	880.00	1760.00	3520.00	7040.00
A#	29.14	58.27	116.54	233.08	466.16	932.33	1864.66	3729.31	7458.62
B	30.87	61.74	123.47	246.94	493.88	987.77	1975.53	3951.07	7902.13

TABLE I. Note frequency chart. Saxophone range boxed in red.
Taken from Artificial Tunes on Tumblr

In short, being "perfectly in tune" means that the frequency of sound produced is one of the numbers in Table I. We aim to either increase or decrease the frequency of the note played by the musician so that it is the same as one of the numbers in Table I. Different instruments have different ranges. The lowest note playable on a saxophone is C#3, while the highest note normally playable on a saxophone is A5. This range has been boxed in red.

2.3 Frequency spectrum

Each instrument or note in a musical composition generates a unique combination of frequencies, resulting in a spectrum that showcases the music's tonal qualities. The same notes on different instruments' different tonal qualities lead to different time domain diagrams. To analyze the frequency of those instruments, we need to first convert the time domain diagram into frequency domain diagrams.

A frequency spectrum provides a visual or analytical representation of a sound or music signal's constituent frequencies and their amplitudes. Converting a signal from the time domain to the frequency domain requires a Fourier Transform. In our case, for the sake of efficiency and digital compatibility, the Fast Fourier Transform (FFT) is applied whenever we need to obtain frequency domain data. The frequency diagram for the saxophone A3=220Hz will look like the picture below (Fig 5). Unless specified otherwise, all saxophone sound spectrums in this essay are produced with the following setup: YAS-82ZII saxophone, Yanagisawa-7 mouthpiece, Harry Hartmann Fiberreed M. All frequency spectrums in this chapter are recorded using the laptop's integrated microphone and processed using ggplot for python. The laptop model is Alienware X15(2021), and for the rest of the essay, we shall refer to this specific laptop simply by saying "the laptop."

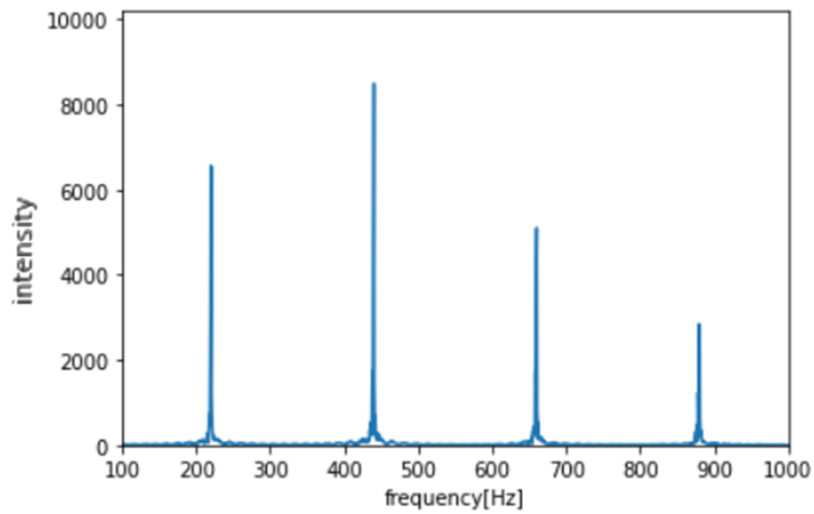


Fig.5. Frequency spectrum of A3 one saxophone

Playing a note of higher pitch will move the entire spectrum to the right while playing a note of lower pitch moves the spectrum to the left. Here's a comparison between A3=220Hz(Blue) and C4=262Hz(Red), both played on the same saxophone(Fig. 6).

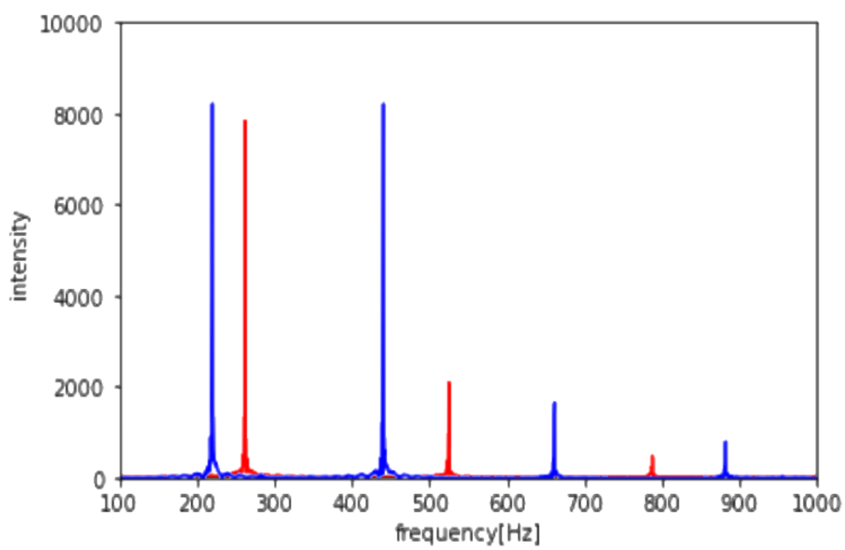


Fig. 6. Comparison between A3=220 Hz(Blue) and C4=262 Hz(Red)

2.4 Fundamental frequency of music

Fundamental frequency, often referred to as the fundamental pitch or the perceived pitch, is the primary and most prominent pitch in a complex sound. In the spectrum below, the fundamental frequency is A=220Hz (highlighted in the red box). It represents the lowest frequency component that our ears perceive when we listen to a sound, not the one with the highest amplitude (440 Hz). Because of the human ears' structure and nervous system, which is completely beyond the scope of this essay, humans perceive the fundamental frequency as the loudest and identify it as the specific pitch of the musical tone. (Benward & Saker, 2003, p. xiii) In the example of Fig. 7, although the actual loudest frequency is 440 Hz, human ears will still consider 220 Hz frequency as the loudest.

The fundamental frequency is the foundation of a sound and serves as the reference point for people's perception of pitch in a given auditory experience. In the example, all sounds that are not 220Hz are called non-fundamental frequencies, or in musicians' language, harmonies. All non-fundamental frequencies are multiples of the fundamental frequency. In Fig. 7, non-fundamental frequencies are 440 Hz, 660 Hz, 880 Hz, and higher, all the way to the infinity. Those sounds give the saxophone its specific tonal quality. When the frequency of the fundamental frequency changes, the frequency of all non-fundamental frequencies also changes accordingly. When musicians talk about the pitch of a musical note, they mean the fundamental pitch, which is essentially just another way to say fundamental frequency. If the fundamental frequency is tuned accurately, all non-fundamental frequencies will also be tuned automatically. Tuning a saxophone, or any other musical instrument, means tuning the fundamental pitch.

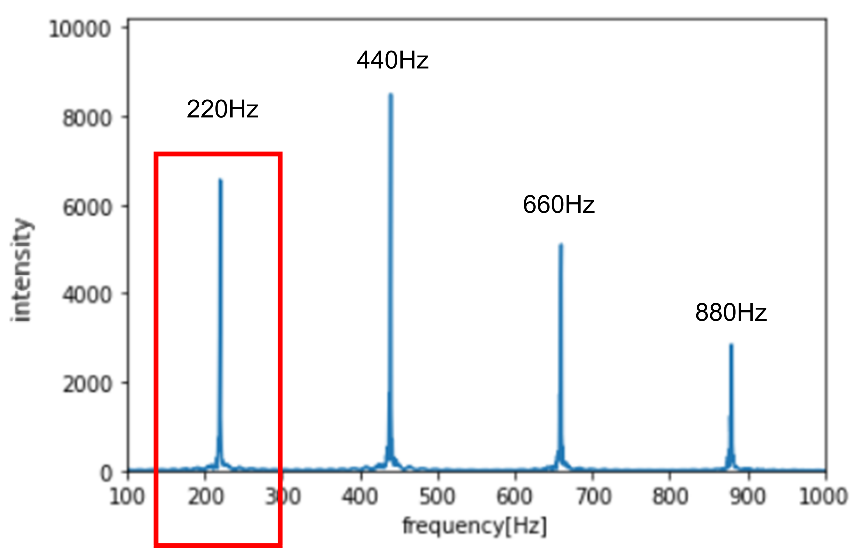


Fig. 7. Redbox highlights the fundamental frequency(220 Hz)

2.5 The pitch of the saxophone

In the first section, we explained that while a key pattern determines which note is being played, the other five orthogonal factors also play an important role in determining the exact frequency of that note. We have generalized these factors in the following list:

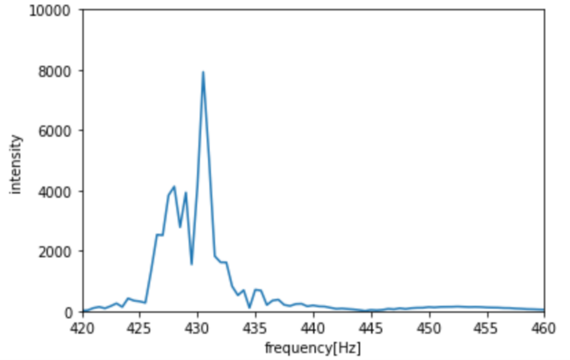
- *Key pattern*
- *Embouchure*
- *Instrument imperfectness*
- *Temperature and humidity*
- *Reed strength*
- *Mouthpiece position*

Key pattern is how keys on the saxophone are pressed and the saxophone players have full control over it. Embouchure is controlled by the saxophone player's mouse muscles, and can only be improved by years of training. Instrument imperfectness was settled when the saxophone was produced, so there's nothing we can do about it. Controlling the temperature and humidity of open space around the saxophone is unrealistic. Reed strength is determined by its material and thickness, which is also impossible to change in real time. Among all those 5 factors, the only option manipulable in a realistic way is the mouthpiece position.
To further explain the impact of different mouthpiece positions, we measured the frequency of A4 on an alto saxophone with four different mouthpiece positions.

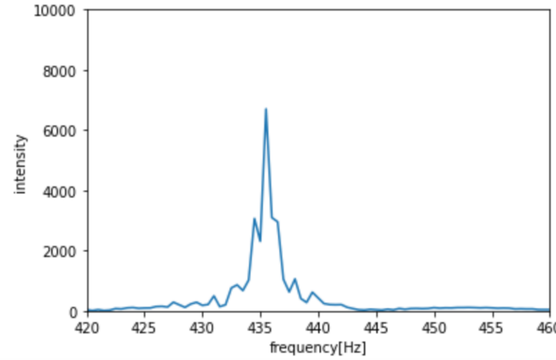
The position of the mouthpiece is measured by the distance D from the bottom of the mouthpiece to the end of the corkwood, as shown in the picture below (Fig. 8). In this case D = 11.7mm.

Fig. 8. Mouthpiece position D=11.7mm

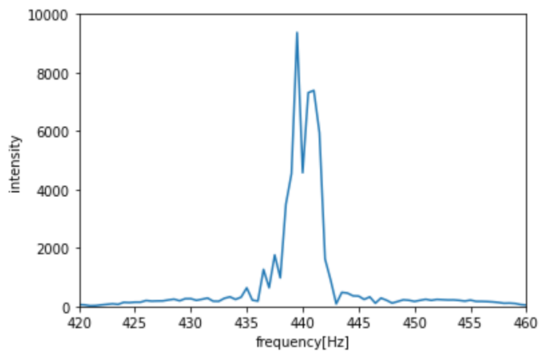
We measured the note A4 frequency F at four different positions with different D, obtaining the following results:



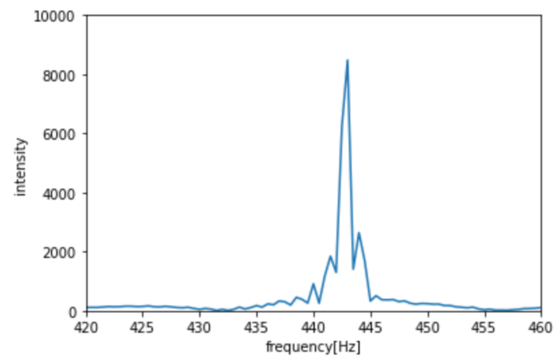
D=14mm, A4=430.5 Hz



D=11mm, A4=435.5 Hz



D = 8mm, A4=439.5 Hz



D=5mm, A4=443 Hz

As shown in the four pictures above, the note A4 frequency decreases as D increases. The relationship between D and F can be further illustrated in the following diagram(Fig. 8):

Fig. 9. Mouthpiece position versus A4 frequency

The diagram (Fig. 9) shows that by changing D from 5mm to 14mm, we can effectively decrease A4 from 443 Hz to 430.5 Hz. To be in tune, as shown in the frequency note chart, A4 needs to be 440H z, and D will roughly equal 8mm.

2.6 Tuning the saxophone

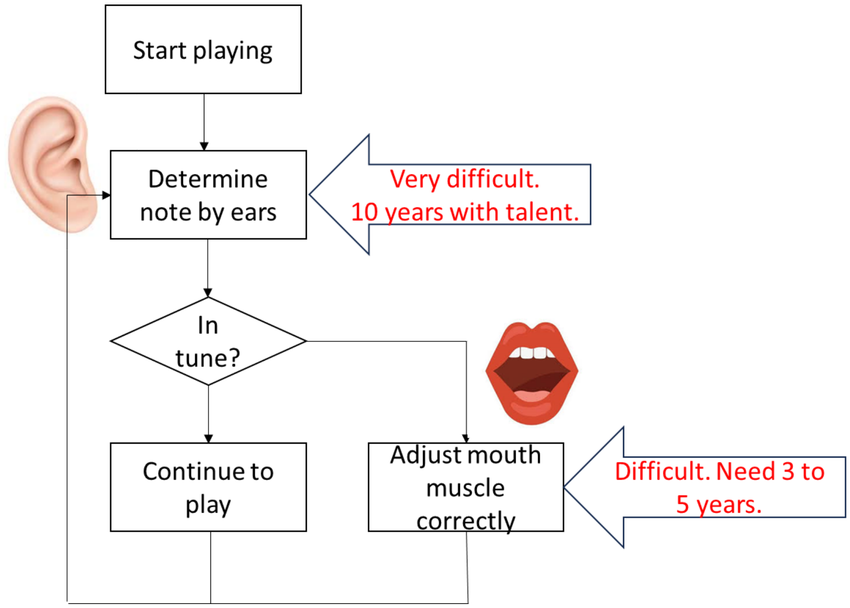
Now we know that we need to tune the saxophone’s A4 fundamental frequency to the ISO international standard pitch of 440Hz. We know how the mouthpiece position affects the fundamental frequency of the saxophone. We can finally introduce how to tune a saxophone.

Traditional tuning is composed of two parts: before the performance and during the performance. Before performance tuning is done with the help of a reference source of accurate A=440Hz. Traditionally, the reference source is a tuning fork (Fig. 10). Nowadays, the reference is mostly electronic tuners (Fig. 11). Saxophone players play A4 on their saxophone and compare the frequency of their instrument with the frequency of the reference source. Either move the mouthpiece position upward or downward to adjust the frequency of the instrument until it is close enough to the reference A4 value of 440Hz. Once tuned, the mouthpiece position is fixed can cannot be further changed during the performance.

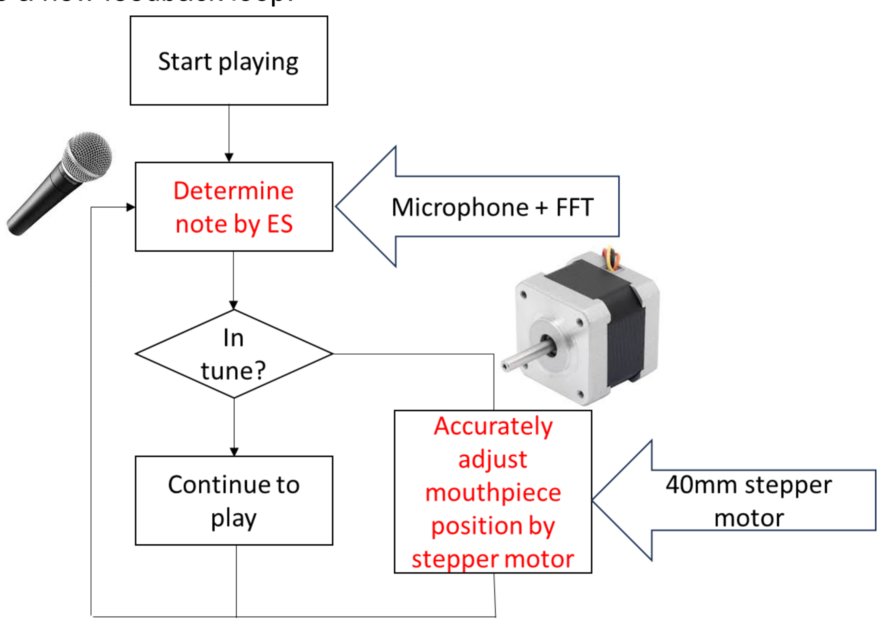
Fig. 10. A tuning fork. Taken from amazon.com

Fig. 11. A digital tuner. Taken from korg.com

Even without considering saxophone imperfectness, before-performance tuning can still make the saxophone roughly in tune since embouchure, temperature, and humidity can change during the performance. Thus, expert saxophonists will also need to tune their instruments during the performance, and this can only achieved by changing embouchure. However, such skills need years of practice to gain and musical talent to perfect. I have personally been playing saxophone for 12 years, and still can only partially tune by embouchure. The flow chart of traditional during-performance tuning below further illustrates this process:



Training a pair of ears that can distinguish note frequency needs more than ten years, but measuring the frequency with a microphone and microcontroller needs only a few milliseconds. Ear training alone, however, is not sufficient because we still need to control mouth muscles and apply the adjustment. Training a set of mouth muscles that can be controlled very accurately takes a few years, however, in this case, the function of a set of painstakingly trained mouth muscles can be achieved by a motor. We have devised an idea to make a normally fixed part move: make the mouthpiece retractable, powered by a stepper motor. A stepper motor is a motor that converts electrical pulses into precise and incremental rotational movements, making it suitable for applications requiring accurate control and positioning. Replace the ears with a microphone controlled by an embedded system(ES), and replace the mouth with a stepper motor, we have a new feedback loop:



With this new feedback loop, we should be able to tune the saxophone automatically. We call it the Automatic Tuning Mouthpiece (ATM).

2.7 Evaluation tool: Melodyne

We evaluate our sound waves using Melodyne. Melodyne is a powerful pitch and time analysis software that allows users to analyze and manipulate individual notes within audio recordings with exceptional precision. Here is an example(Fig. 12).

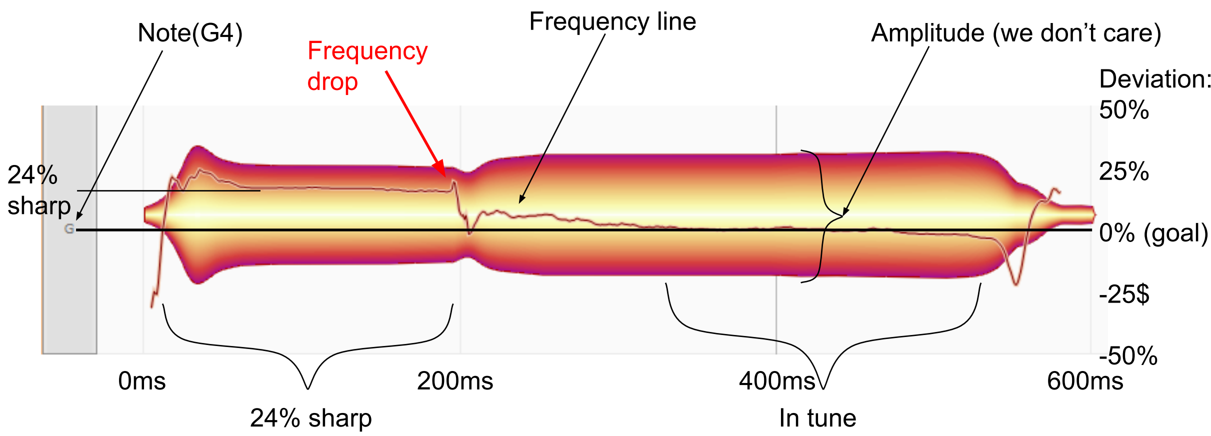


Fig. 12. Melodyne example

In the picture above, the colored block means amplitude, while the line means frequency. Since we are only analyzing the frequency, we do not care about the amplitude. The alphabet G at the left means that this is a G note. The sound is perfectly in tune if the frequency line is on the same height as the note letter G. The top ceiling means 50% sharp and the bottom floor means 50% flat. Deviation cannot be more than 50%, because that would enter the realm of another note. Thus, 50% is also called "halftone", which means that it is right in the middle of two neighboring notes. We aim to push the frequency line to the middle of the graph. The distance between each vertical grey line is 200ms. Melodyne can also automatically calculate the sound frequency of any designated interval by calculating the average of all frequency samples. In the diagram, the G note dropped from 24% sharp to in tune at around 200 ms.

Chapter 3: Design and Implementation

3.1 Requirements

3.1.1 Accuracy:

Being in tune is not a yes-or-no question but rather an open-ended question since different people have different perceptions of sound frequency. In general, there are three groups of people in terms of sound frequency perception.

a. Tone deaf people. Tone deaf is also known as amusia, which is a musical disorder that appears mainly as a defect in processing pitch. This group of people is the best audience because they are always happy no matter how out of tune a musician may be, thus we ignore this group of people. They represent roughly 4% of the entire population. (Peretz & Hyde, 2003, 362–367)
b. People with absolute pitch. Absolute pitch is not the ability to estimate a pitch value from the dimension of pitch-evoking frequency (30–5000 Hz), but to identify a pitch class category within the dimension of pitch class (e.g., C, C♯, D ... B, C). (Takeuchi & Hulse, 1993, 345–361) While it is commonly believed that absolute listeners can notice even the slightest frequency changes, there's an article suggesting that an absolute listener's sense of hearing is typically no keener than that of a non-absolute ("normal") listener. (Fujisaki & Kashino, 2002, 77-83) Those people are generally more picky with pitch accuracy than most people. We did not find any source about exactly to what extent people with absolute pitch can perceive frequency fluctuations during a music performance. However, absolute listeners represent only roughly 0.01% of the total population and 4% of the musician population. (Bachem, 1955, 1180) (Carden & Cline, n.d., 890-901) Considering the rarity of this group of people, although we do not fully understand them, we ignore them.
c. Normal people: Most of us are neither tone-deaf nor absolute pitch. Normal people represent 96% of the entire population and are the main target customers of our ATM. There is no widely agreed data showing to what extent most people distinguish frequency. According to research done by Loeffler, the smallest frequency difference that humans can distinguish is about 5% to 6%. (Loeffler, 2007-12-18, Page 6) However, Loeffler did not state in his essay how he obtained such data. According to renowned Professor of Music Elab Sobol, also my saxophone teacher, a good saxophonist should control his frequency deviation to below 10%. However, despite Sobol being a top saxophone player, his opinion still only represents his personal experience. Based on my empirical observation as a part-time saxophonist for 12 years, as long as a note's frequency deviation is less than 20%, most people will perceive that note as being in tune.

With the knowledge above, to make most people perceive our ATM as being in tune, we have 3 different standards all with some limited validity:

a. Loeffler standard: deviation < 5%
b. Elab Sobol standard: deviation < 10%
c. Our experience standard: deviation < 20%

For our ATM's accuracy, we aim to achieve the Loeffler standard so that almost everyone will perceive the tuned note as being in tune. If not possible, at least achieve the Elab Sobol standard.

3.1.2 Speed:

Music notes last a limited time. We need to correct the pitch in time. Most pieces of music have their standard speed designated by the composer. This standard speed is called Beat Per Minute (BPM). Different music genres have different BPMs. For saxophones, most pieces have a BPM of 60 to 120. The exact duration of each note can vary according to complicated music theories. But in this essay, we only use the most popular case of the Eighth Note at 4/4 time signature under 120 BPM as an example (Fig 13). Under the speed of 120 quarter notes per minute, an eighth note, which lasts half of the duration of a quarter note, lasts 250 ms. (Fig. 14)

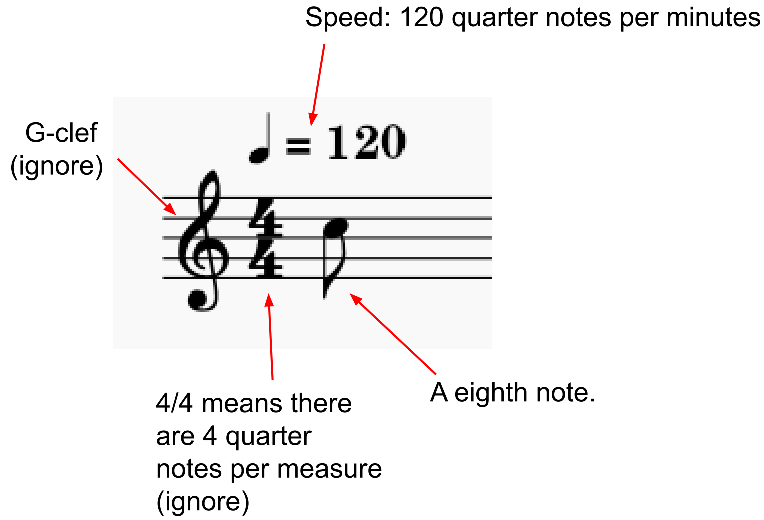


Fig. 13. An Eighth Note at 4/4 time signature under 120 BPM

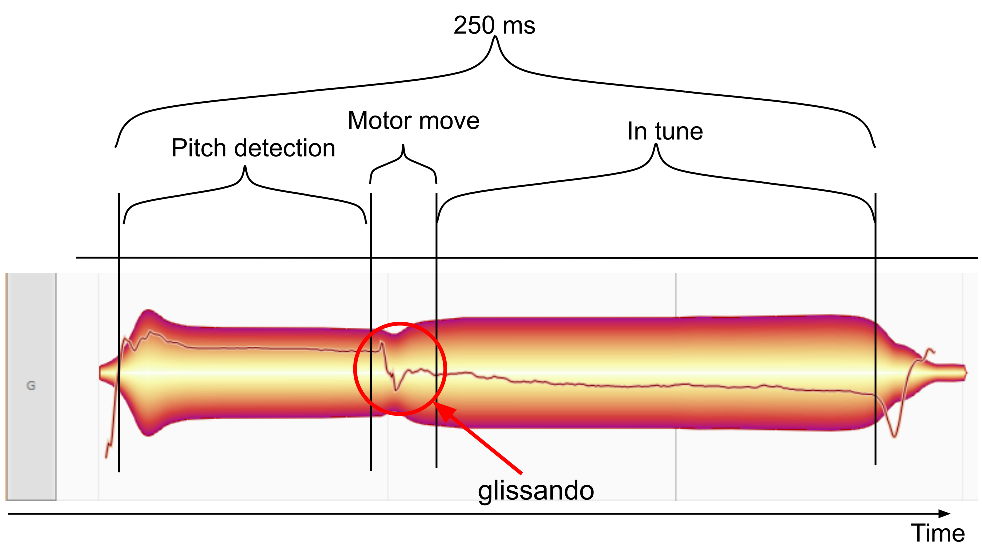


Fig. 14. Example tuning procedure of eighth note under 120 BPM

Now we know that the duration of the note that we need to tune is 250ms or longer, but we still need to explore the time portion of it that we can use for tuning without serious impact on the audience’s perception. More specifically, ATM tunes by first detecting the current pitch and then moving the position of the mouthpiece by a motor. Both pitch detection and motor movement take time. The motor movement also results in a glissando. A glissando means a continuous slide upwards or downwards between notes. This implies that the frequency of a glissando is always changing. Thus, we only have a limited time window for pitch detection and motor movement. The exact size of this time window depends on how long the in-tune section must be to make the audience perceive this note as in tune, despite the stall in the pitch detection phase and glissando in the motor movement phase.

Luckily, such an experiment has already been conducted by Jean-Pierre Rossi. Although Rossi's research was on the pronunciation of natural language words, his discoveries are comprehensive and should apply to human perception of all sounds, including music. Rossi expressed the conclusion in the ‘‘2/3 rule,’’ (Rossi, 1971, 1-33) which can be stated as follows:

> For dynamic tones in a vowel, the pitch perceived corresponds to a point between the second and the third third of the vowel. For example, if a linear glissando between 100 and 200 Hz lasts 150 ms, this rule predicts that the pitch perceived is somewhere in the F0 values of the glissando in the time interval 100–150 ms, i.e., between 166.6 and 200 Hz. (Rossi, 1978, 11-40)

Thus, for a 250ms note to be perceived as in tune, at least the latter 1/3 of it must be in tune. This led to a must-guaranteed in-tune time of 83.3ms, leaving the first 166.7ms for pitch detection and motor movement combined.

3.1.3 Ergometry:

Having something that continuously moves in and out of the mouth does not sound like a pleasant experience. Yet we still want to make our ATMs as comfortable as possible, under the premise that accuracy and speed are not compromised. Ergometry is hard to analyze quantitatively, thus we have plans to let a few saxophone players try our mouthpiece and ask them about their opinion about it.

In conclusion, we aim to make an Automatic Tuning Mouthpiece that can achieve an accuracy of a maximum 5% deviation within 166.7ms, without incurring too much discomfort.

3.2 Mechanical Part

Because of mechanical problems and our lack of knowledge in mechanical engineering, the ATM project took more time than usual. Mechanical limitations also greatly shaped the design and development of the embedded system.

3.2.1 Initial setbacks

When we first came up with the idea of the Automatic Tuning Mouthpiece in 2021, we realized that this was more than a purely embedded project. Considering the expected mechanical problems, in which we lack professional knowledge, we decided to start early in part-time. When we started producing static prototypes in 2022, we had many different ideas about mechanical design. We can only verify those ideas by making prototypes and trying them. Initial experimental prototypes are the slider design and the top-open design as shown in the following two pictures.

Fig. 15. Slider design verification prototype

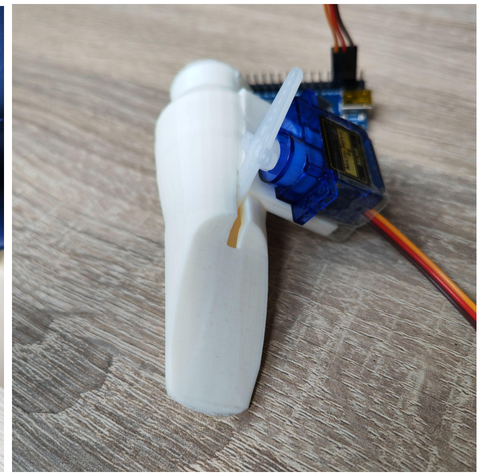


Fig. 16. Top-open design prototype

The slider design prototype (Fig. 15) was produced by Stereolithography(SLA) 3D printing. Stereolithography is a 3D printing technology that utilizes a laser to selectively solidify liquid photopolymer resin layer by layer, creating precise and detailed three-dimensional objects. It demonstrated a tremendous influence on the saxophone's timbre (frequency spectrum in scientific languages), however only a minimal influence on sound frequency. As a result, the slider design was soon scrapped. The top-open design (Fig. 16) was produced by Polylactic Acid(PLA) 3D printing. PLA 3D printing involves using thermoplastic filament derived from renewable resources and is a budget way to create three-dimensional objects, although with inferior details compared with SLA 3D printing. At first, it cannot make a sound because of too much air leaking through the top hole. We tried to solve this issue by blocking part of the hole with a rubber pad. However, the servo has a very unpredictable impact on the pitch. The relationship between opening size and sound frequency seems to be random. We believe that the reason behind this unpredictability can only be explained by aerodynamic simulations, in which we lack the necessary tools and knowledge. Thus, the top-open design also proved unsuccessful.

Despite setbacks, however, we did gain the following crucial pieces of knowledge that proved to be essential for our future designs:

1. The mouthpiece must be completely sealed. Air leaks must be prevented.
2. Material matters. Different materials have different sound characteristics. SLA is preferred over PLA.
3. Precision matters. Prototypes with higher precision reduce mechanical inaccuracies and thus make the ATM easier to play.
4. We need a powerful motor. Airtight means sealings, which means friction. Our motor must overcome this unavoidable friction
5. Doing things to the tip of the mouthpiece causes unpredictable responses. To generate predictable responses, we should work on the butt of the mouthpiece.

3.2.2 Retraction design

In the previous chapter, we talked about how the mouthpiece position impacts sound frequency. Now we need to design a mouthpiece that can easily change its position by motor. But first, we made a prototype (Fig. 17) to guarantee that we are on the right track:

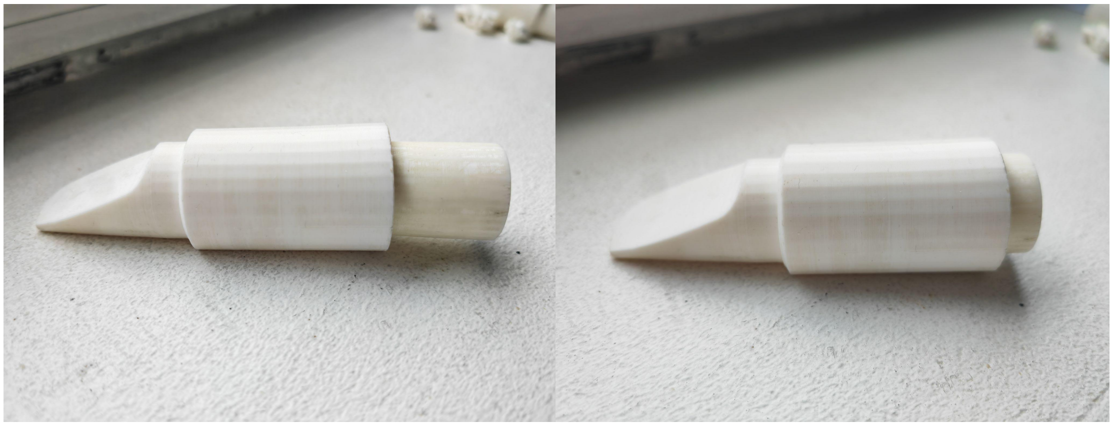


Fig. 17. Retraction design prototype. Left: extended. Right: retracted.

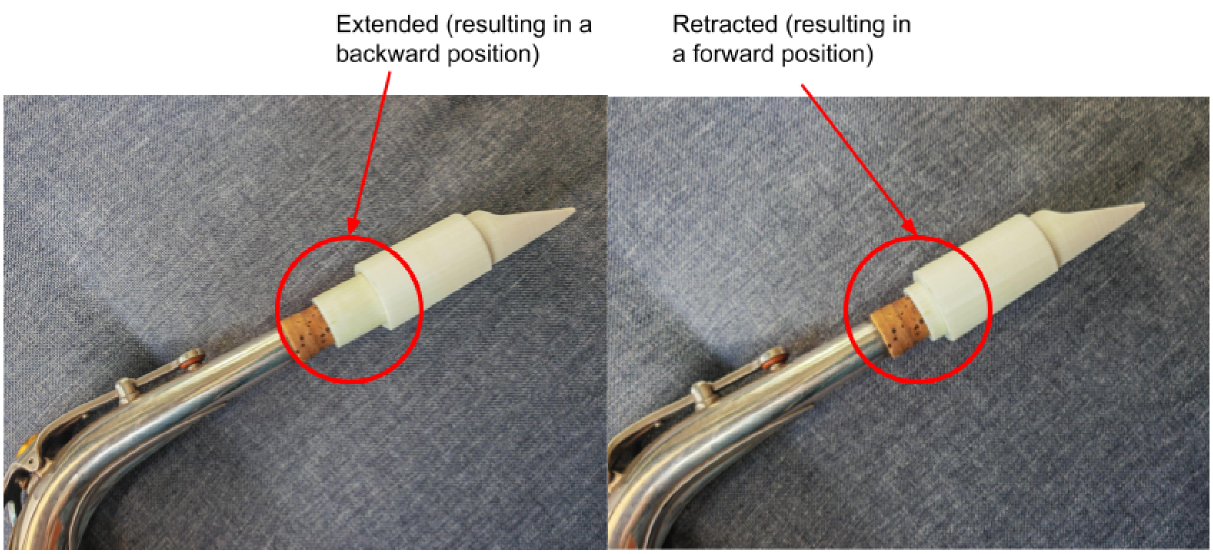


Fig. 18. Retraction design installed on the saxophone

Although we can only move its position by hand, it did prove that the mouthpiece position has a linear influence on sound frequency. Next, we need to add a transmission and a motor to make it a complete prototype. To solve the airtight problem, we installed an O-ring on the inner barrel (Fig. 19):

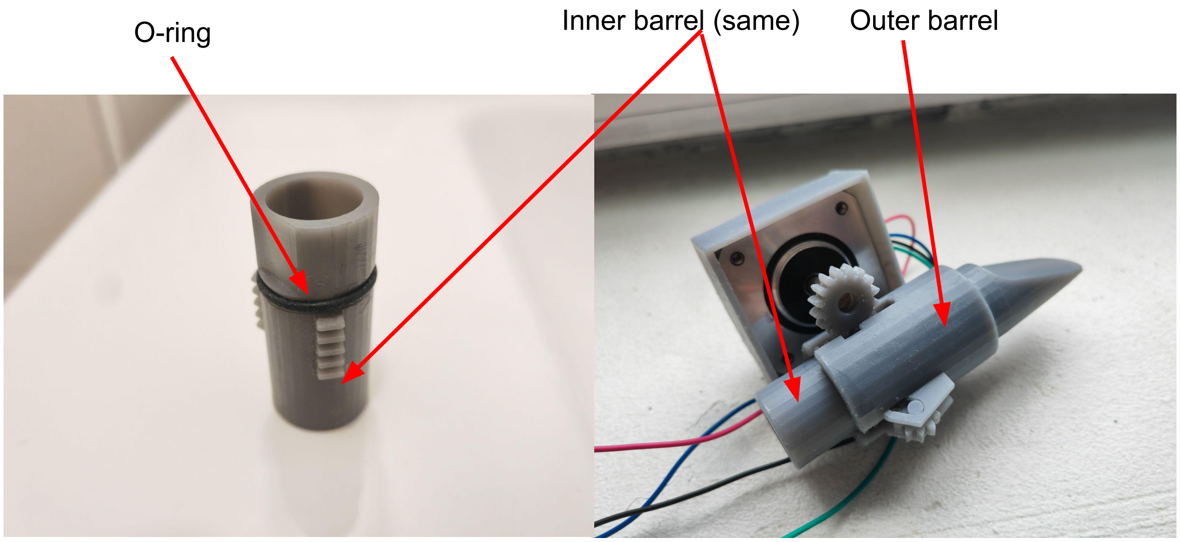


Fig. 19. Complete prototype (Gear transmission)

We tried it on our saxophone. To our disappointment, it only produces a whistling sound. After some research, we realized that we had neglected an unforeseen critical element: Vibration transmission.

3.2.3 Vibration transmission

Vibration needs mechanical contact to transmit. The transmission gears lift the inner barrel, making the vibration of the outer barrel unable to transmit to the inner barrel. Because the saxophone body is only connected to the inner barrel, the outer barrel becomes independent from the rest of the saxophone and thus works as a whistle.
To fix the vibration transmission problem, we designed another prototype, using saxophone corkwood as transmission support and cork grease as the lubricant.

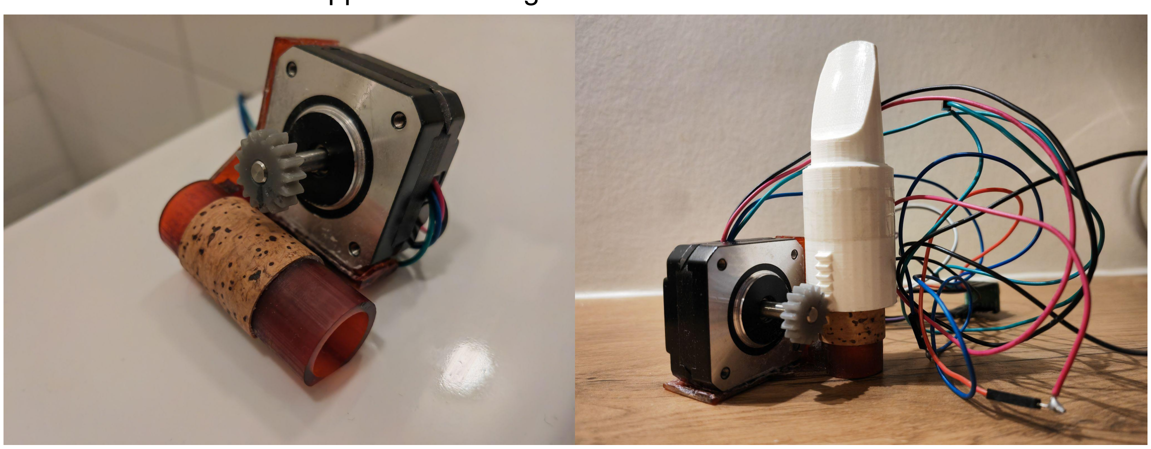


Fig. 20. Final prototype

Luckily, this prototype (Fig. 20) worked as expected. Since we had spent plenty of time on the mechanical development, and this prototype is, although not perfect, good enough, we decided to end the mechanical development here and concentrate on the embedded systems from now on. The reed implemented on the final prototype is Forestone Black Bamboo M. Here's a diagram that concludes all of our important prototypes:

Name	First prototype. (Fig. 15)	Top open prototype. (Fig. 16)	Retraction prototype. (Fig. 17)	First complete prototype. (Fig. 19)	Final prototype. (Fig. 20)
Material	SLA	PLA	PLA	SLA	SLA+PLA
Mechanical transmission	no	Top open, Servo motor	Retraction	Retraction with gear support, 42mm stepper motor	Retraction, 42mm stepper motor
Sealing	no	Rubber pad	no	O-ring	Cork wood
Vibration transmission	yes	yes	yes	no	yes
lubrication	Not needed	Not needed	no	Silicon lubricant	Cork grease

3.3 Embedded Part

The embedded part is made of the following four components (all pictures taken from Amazon.com):

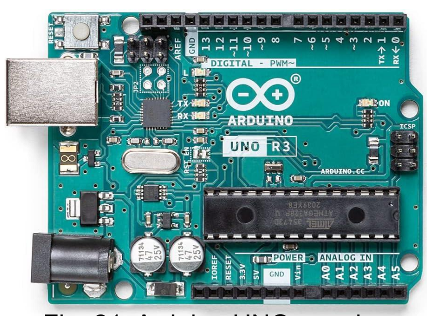


Fig. 21. Arduino UNO, used as the controller

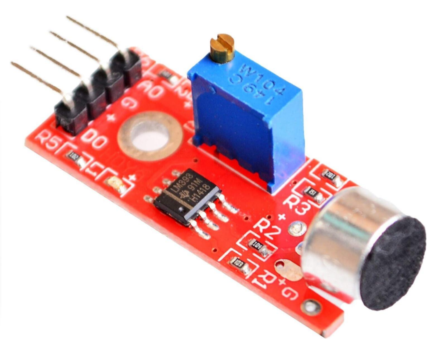


Fig. 22. KY-037 microphone.

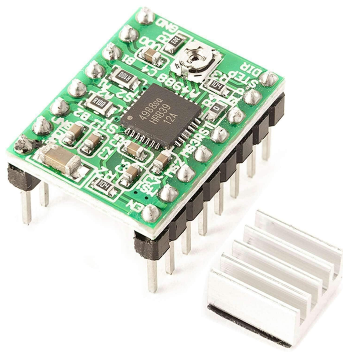


Fig. 23. A4988 stepper motor driver

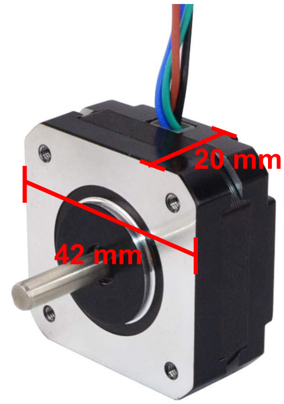


Fig. 24. 42X20 mm stepper motor

3.3.1 Repeated open loop

We installed a 42X20 millimeter stepper motor(Fig. 24) to ensure that we had enough power to overcome the friction. A stepper motor driver module(Fig. 23) is installed to supply 1A current at 12 volts to the stepper motor. We have the following control loop. The microcontroller (Fig. 21) needs to do two things: pitch detection and motor movement calculation. One KY-037 microphone(Fig. 22) was installed to record sound from the saxophone.

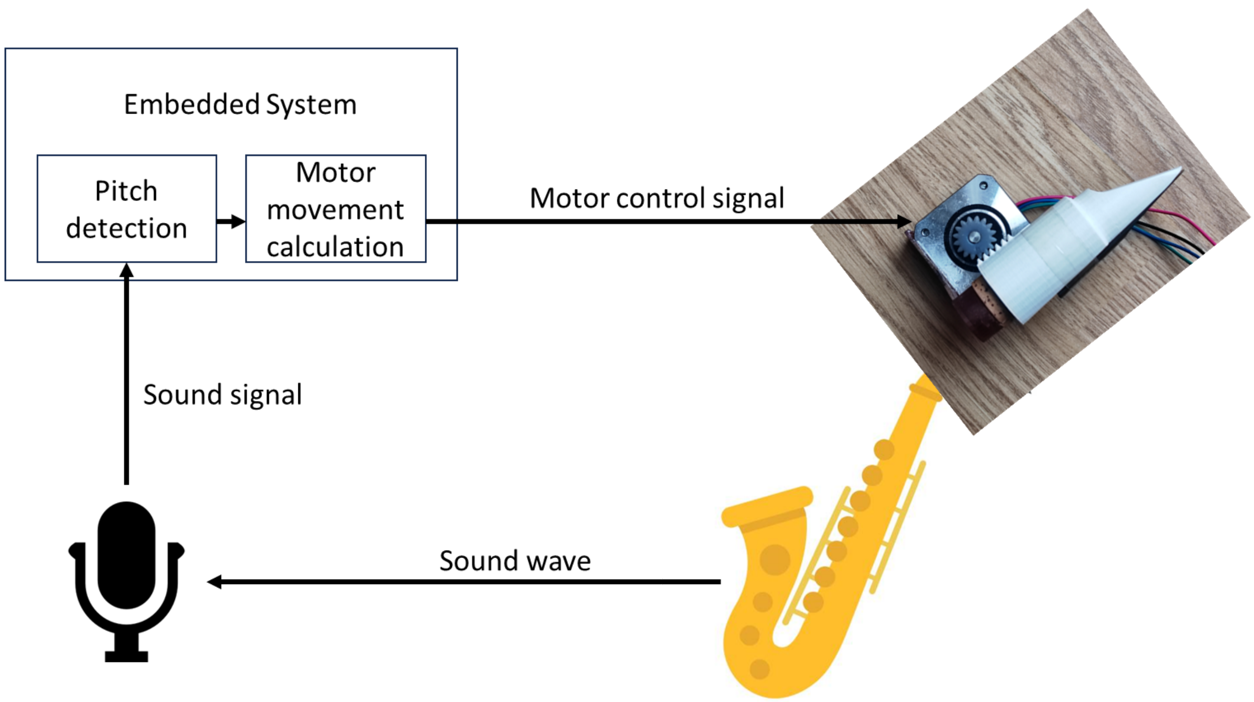


Fig. 25. ATM control loop

3.3.2 Pitch detection

Two elements of the FFT are sampling frequency and sample size. A higher sampling frequency enables us to detect higher frequencies, but the frequency resolution decreases as the sampling frequency increases. A bigger sample size increases frequency resolution, but costs more measuring time. Ideally, we want a very high sampling frequency with a big sample size so that we can detect frequency quickly and accurately. However, our Arduino UNO hardware limits our sample size to a maximum of 128. Consequently, we have to limit the sampling frequency to maintain an acceptable frequency resolution.

To quantitatively analyze the relationship between sampling frequency and pitch detection accuracy, we devise the following experiment:

1. Play an accurate reference frequency through a digital tuner.
2. Record the sound with ATM's microphone (KY-037), and conduct FFT to find the fundamental frequency.
3. Repeat step two 100 times to collect 100 data.
4. Find the maximum deviation from the average of these 100 data.

Here is the distribution of 100 frequency data of reference 440Hz under the setup of 2048Hz sampling frequency and 128 sound samples:

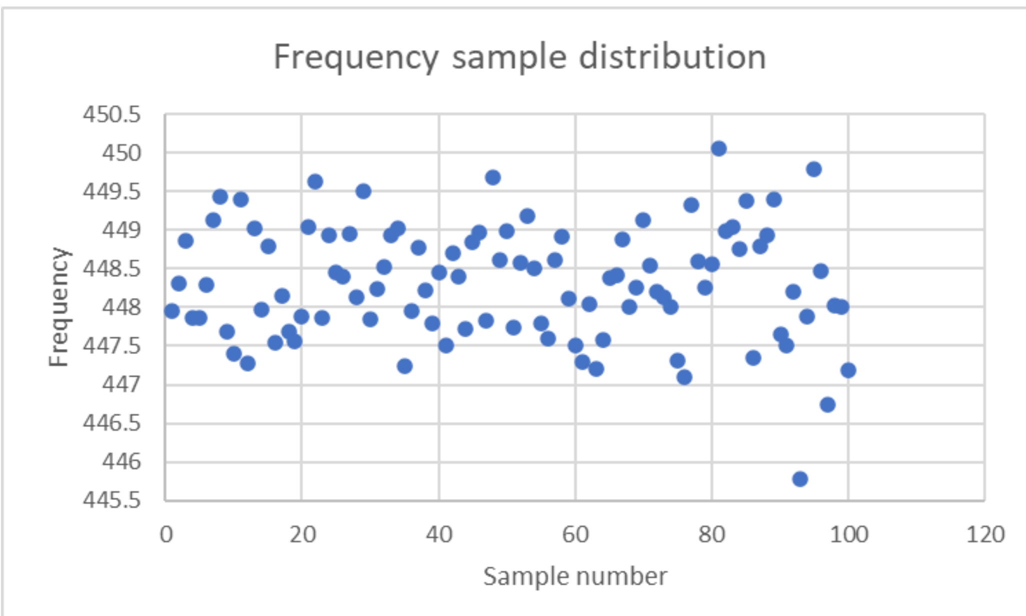


Fig. 26. 100 frequency data of 128 samples of 2048 sampling frequency

From the data of the diagram Fig. 26, we calculated the average to be 448.3Hz, and the maximum deviation to be 2.52Hz. The frequency difference from A4 = 440Hz to its closest note G#4 = 415Hz is 25Hz. To reach the less than 5% deviation goal, the maximum frequency deviation is 25Hz * 5% = 1.25Hz.

To see different sampling frequencies' impact on speed and accuracy, we picked four different sampling frequencies: 8192Hz, 4096Hz, 2048Hz, and 1024Hz. Pitch detection time is composed of the recording time and the FFT time. The recording time has a time complexity of

O(n), while the FFT has a time complexity of O(n×log(n)). Thus, in theory, pitch detection time should follow a logarithmic relationship with the sampling frequency. To determine different sample sizes' impact on speed and accuracy, we picked three different sample sizes: 128, 64, and 32 samples. Multiply four different sampling frequencies and three different sample sizes, we have a total of 12 different combinations of parameters. We record the speed and accuracy of these 12 combinations in the following tables:

Pitch detection time table

Sampling frequency / Sample numbers	8192Hz	4096Hz	2048Hz	1024Hz
128	96ms	96.8ms	127.6ms	191.4ms
64	44.8ms	45.8ms	61.2ms	92.8ms
32	22ms	21.2ms	29.4ms	44.8ms

Max deviation frequency table

Sampling frequency / Sample numbers	8192Hz	4096Hz	2048Hz	1024Hz
128	5.98Hz	5.22Hz	2.52Hz	1.17Hz
64	9.38Hz	5.39Hz	3.56Hz	2.62Hz
32	117.8Hz	30.22Hz	33.56Hz	4.6Hz

Max deviation percentage table

Sampling frequency / Sample numbers	8192Hz	4096Hz	2048Hz	1024Hz
128	23.92%	20.88%	10%	4.7%
64	37.52%	21.56%	14.2%	10.4%
32	471.2%	120.88%	134.24%	18.4%

If we want to guarantee less than 5% deviation accuracy, the only possible option will be 128 samples of 1024Hz. However, as shown in the Pitch Detection Time Table, such a combination will lead to an execution time of 191.4ms, which does not meet our speed requirement of

166ms. Thus we looked into the standard deviation to see if we could find something more with a relaxed standard.

We also noticed in the Pitch Detection Time Table that increasing the sampling frequency from 4096 Hz to 8192 Hz did not improve the speed. We decided to measure the maximum FFT sampling speed of the Arduino UNO board. According to our measurement, sampling 12800 data took 3.104 seconds, resulting in an actual sampling frequency of 4156 Hz. Because of the limitation of the hardware, the 8192 Hz sampling frequency option cannot be achieved, and will only result in worse frequency resolutions without any gain.

The standard deviation frequency table

Sampling frequency / Sample numbers	8192Hz	4096Hz	2048Hz	1024Hz
128	1.71Hz	1.05Hz	0.74Hz	0.46Hz
64	4.83Hz	1.38Hz	2.88Hz	0.56Hz
32	52.45Hz	14.3Hz	7.89Hz	1.34Hz

As with data from Fig. 26, we can generate the following box graph (Fig. 27), which shows that the data from Fig. 26 follows a normal distribution.

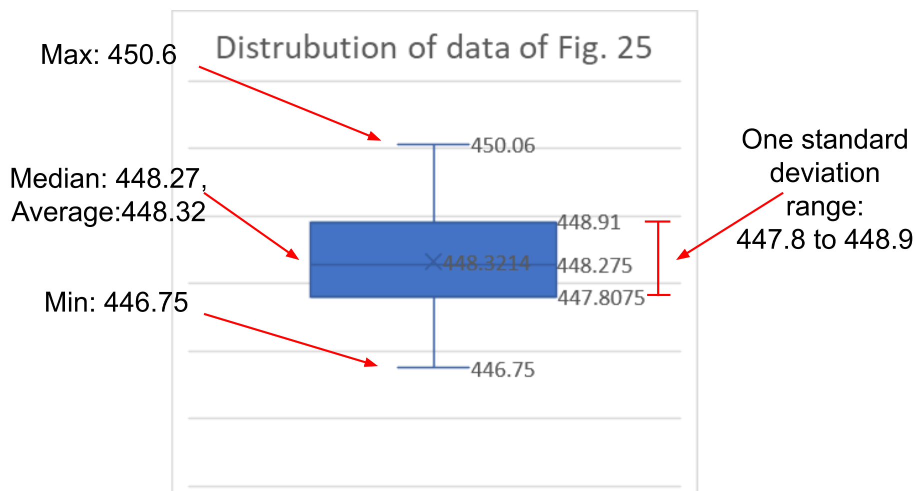


Fig. 27. Data distribution box graph

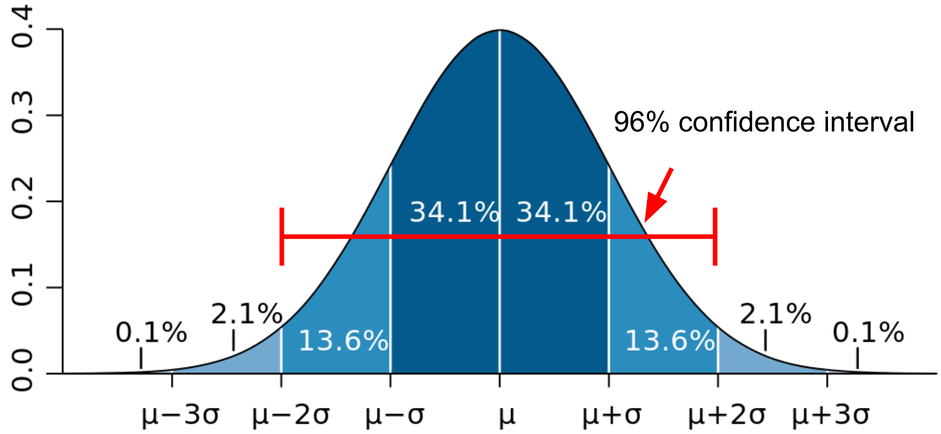


Fig. 28. Normal distribution diagram. Taken from Wikipedia.

To calculate the 96 confidence interval of 128 samples of 1024 Hz, we simply add 2 standard divisions as shown in the diagram above. We get 0.46Hz + 0.46HZ = 0.92Hz. To get the deviation percentage, we calculate 0.92Hz/(440Hz-415Hz) = 3.68%. Repeat this process for all elements of the table, and we will get the 96% confidence table:

The 96% confidence table

Sampling frequency / Sample numbers	8192Hz	4096Hz	2048Hz	1024Hz
128	13.68%	8.4%	5.92%	3.68%
64	38.64%	11.04%	23.04%	4.48%
32	419.6%	114.4%	63.12%	10.72%

According to the 96% confidence table, to achieve less than 5% deviation accuracy at 96 percent (2 standard deviations) of the time, we have 2 options: 128 samples of 1024Hz and 64 samples of 1024Hz. The 64 samples and 1024Hz combination also have an execution time of 92.8ms, successfully meeting our speed requirement of 166ms.

As a result, we have 2 reasonable options:

1. 128 samples of 1024Hz. Guarantees 5% accuracy. However, it needs more execution time.
2. 64 samples of 1024Hz. Guarantees that a cycle will take less than 250 ms, which ensures 120 BPM speed. However, there will always be around 4% chance that the pitch detected will deviate more than 5%.

We eventually decided to adopt option one for its guaranteed accuracy. Although the 192ms execution time will slow down our maximum speed from 120 BPM to 108 BPM, 108 BPM is still considered pretty fast music. More importantly, the primary purpose of the ATM project is to tune the saxophone, not speed. When we cannot have both, we value accuracy more than speed. Also, this experiment is done in an ideal environment where it is quiet and the microphone is close to the reference source, which is a perfect sinusoidal signal. In reality, according to our experience with the saxophone, we know that the saxophone sound is more complex and the room can be noisy, so the accuracy will likely be lower than the ideal one.

From Table I, we also see that the saxophone frequency ranges from 138.59 Hz to 880 Hz. According to the Nyquist frequency theorem, the sampling frequency must be two times the maximum possible frequency. With a sampling frequency of 1024 Hz, the highest note that we can sample is B4 (493.88 Hz). (See Table I) Thus, the current parameter setup reduces the ATM's upper range limit. Still, we consider this drawback to be worthwhile. Our goal is to make a mouthpiece that can automatically tune the saxophone in real time. As our first prototype, the ATM does not need to be capable of tuning all notes on a saxophone. Being able to tune some notes can still demonstrate the viability of ATM.

Thus, despite drawbacks in response time and maximum detection range, we still eventually decided to adopt the parameter of 128 samples of 1024Hz sampling frequency.

3.3.3 Bias

Because our hardware Arduino UNO does not read the negative part of the microphone signal, we need to add a bias. (*Arduino Frequency Analyzer*, 2017) We measured our detected frequency compared to the reference frequency and obtained the diagram below. Data for this diagram is in Appendix 1.

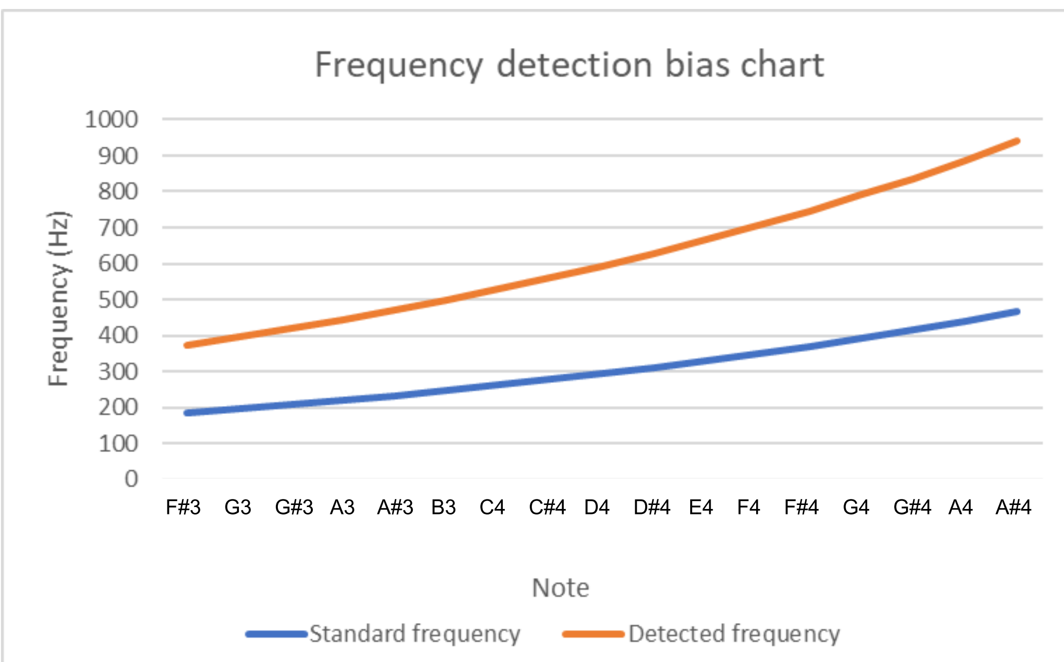


Fig. 29. Frequency detection dia chart

As we see (Fig. 29), the bias is not exactly linear. A safe solution is to make a lookup table. Designing an approximation mathematical function may also solve the problem, however, as we will soon see in the next section, we need a lookup table to compensate for mechanical limitations. If we will need to use a lookup table anyway, it is more convenient to just combine these two tables into one. Thus, the lookup table solution is adopted.

3.3.4 Sound intensity threshold

To not pick up noise in the background, or other distant sources, we need to set a threshold for the lowest sound intensity that can activate the ATM. The analog output of KY-037 can be controlled by the position of the screw (Fig. 30), so we set the screw to a position and measured a reasonable sound intensity threshold, and coded it in the software. However, KY-037 is not stable and its output current strength can change for many reasons, including temperature and pressure. We hard-coded the signal strength threshold into the software and adjusted the KY-037 microphone's analog signal strength with a screwdriver when the threshold was too high or too low. The user of the ATM should adjust the screw until only the sound from the musical instrument is heard and the background noise does not pass the threshold.

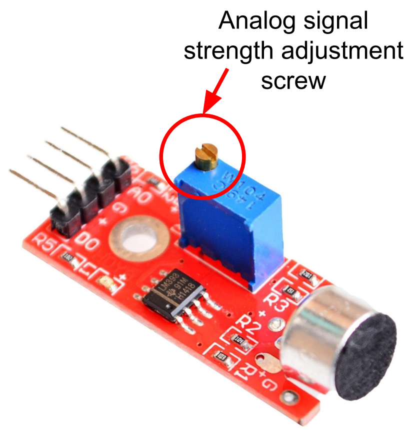


Fig. 30. KY-037 microphone.

Of course, the saxophone player is not always playing. When the saxophonist stops playing, the ATM will inevitably pick up sound from other musicians. To deal with such a case, we added a switch to the ATM so the saxophonist can manually turn off the ATM while not playing.

3.3.5 Motor movement calculation

Our stepper motor is 200 steps per rotation, which means a precision of 1.8 degrees. The gear radius and the number of teeth of the gear will transform this inaccuracy in degrees into inaccuracy in linear distances. With the current 16-teeth gear with a diameter of 20mm, one step of our motor corresponds to a linear movement by:

$$\frac{20\text{mm} \times \pi}{200 \text{ steps}} = 0.314\text{mm/step}$$

In Fig. 8, we see the relationship between the mouthpiece position and sound frequency. This relationship is not linear, but pretty close to linear, especially in a short range. Thus, we decided to treat it as linear and calculate an approximate relationship. By moving the mouthpiece from 14mm to 5mm, we increased the frequency from 430.5Hz to 443Hz.

$$\frac{(443\text{Hz}-430.5\text{Hz})}{(14\text{mm}-5\text{mm})} = 1.39 \text{ Hz/mm}$$

$$0.314 \text{ mm/step} \times 1.39\text{Hz/mm} = 0.436 \text{ Hz/step}$$

$$\frac{0.436 \text{ Hz/step}}{440 \text{ Hz} - 415 \text{ Hz}} = 1.4\% \text{ per step}$$

Thus, the approximate theoretical maximum pitch accuracy according to our motor is 1.4%, which is significantly lower than our allowed deviation of 5%. Thus, we concluded that our 200-step motor is accurate enough. Our stepper motor has a standard rotation speed of 2ms/step. Because rotating the motor means moving the mouthpiece into or out of the user's mouth. Moving a short distance might be alright, but moving too big a distance in a short time does not sound pleasant. Thus we decided that, in each cycle, the motor can only move a maximum of 20 steps. Multiply it by 2ms/step, we get 40ms. Add it with the frequency detection time of 192ms, and we get the total maximum processing time of 232ms. In most cases, the motor should only need to move a few steps. We estimate the average processing time to be around 200 ms.

Our mouthpiece parts were all 3D-printed. 3D printing has inaccuracies. Also, we did not consider the thickness of the multiple layers of glue that hold the motor when we drew the sketch. As a result, the gear on the motor and the gear rack on the mouthpiece do not perfectly fit into each other, leaving a small gap in between. To be more specific, moving one step on the motor will not move the mouthpiece at all. To move the mouthpiece, the motor needs to rotate at least 2 steps.

We decided to compensate for this mechanical inaccuracy with software. Combined with the frequency detection bias, we made a received frequency to the motor movement table. This table was first made heuristically, and improved by experiments. See this table in Appendix 2. With this lookup table, we can now instantly translate all possible recorded frequencies to motor movements.

3.3.6 Maximum movement range and reset

The gear rack on the ATM has only a limited length, thus we must limit the movement range of the gear. The total length of the gear rack is 12.6 mm (Fig. 31).

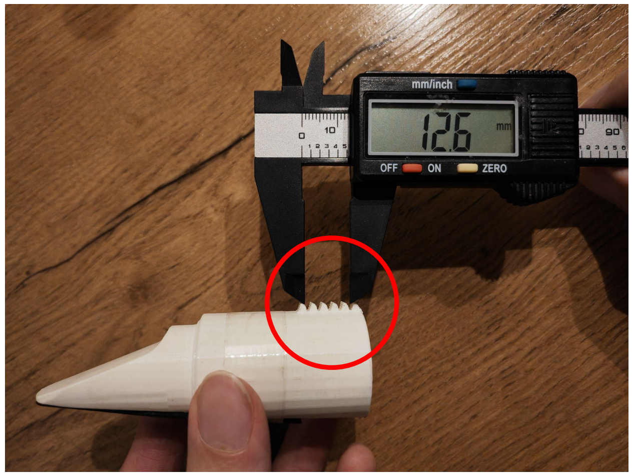


Fig. 31. Gear rack length

In Section 3.3.5, we calculated that the estimated linear movement of each step is 0.314 mm. With a total linear length of 12.6 mm, the motor can rotate a total of 40 steps.

We equally allocated these 40 steps in the forward and backward directions. The backward limit is recorded as position -20, the forward limit is recorded as position +20, and the reset position is 0. See Fig. 32 below.

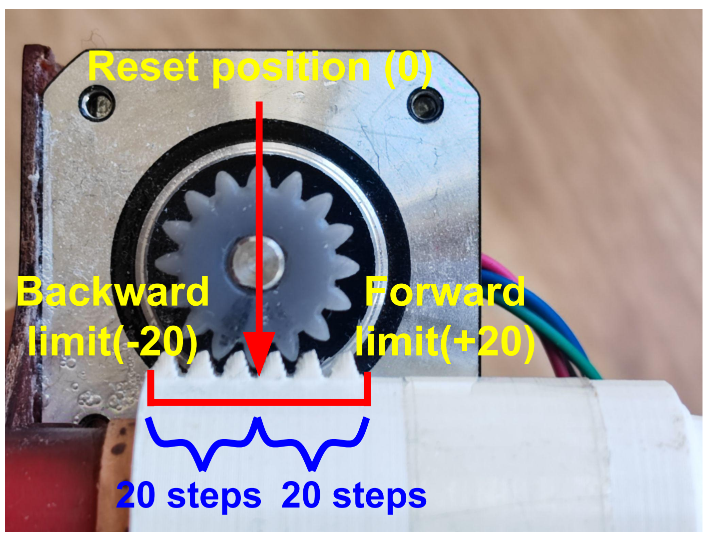


Fig. 32. Maximum movement range and Reset (0)

Because the maximum movement of the ATM is limited, its effect on sound frequency is also limited. Thus, we try to calculate the achievable maximum correctable frequency deviation under this movement range limitation. In Section 3.3.5, we calculated that each step would change the sound frequency by 1.4%. Rotating 20 steps will thus lead to a frequency change of 28%. As a result, we estimated that for A4, as long as the player's frequency deviation is within the 28% range, our ATM should be able to correct it. This estimated range based on A4 is very rough though, and the real maximum correction range of the notes can only be measured in later experiments.

Take together all our calculations and estimations during the development phase, we believe that we have made an Automatic Tuning Mouthpiece that can achieve a tuning accuracy of a maximum 4.7% deviation within 232ms, as long as the notes' original frequency deviation is under 28%.

3.4 Early experiments & software update

During our initial experiments, we noticed a critical problem. Our pitch detection algorithm finds the fundamental frequency, which is the lowest spike on the frequency diagram. This approach is fine when there is only one instrument playing, however, with multiple instruments, our ATM will not know which frequency is coming from the saxophone that it is controlling.

3.4.1 Inadequacy of the fundamental frequency

In the following example, the red spectrum comes from the saxophone, and the blue spectrum comes from a clarinet. Most human ears can easily differentiate the sound of a saxophone from the sound of a clarinet by their different timbres like most human eyes can differentiate red from blue. However, the ATM's simple frequency detection algorithm has no idea of timbre and thus cannot differentiate the saxophone sound from the clarinet sound. If the ATM continues to regard the lowest spike as the fundamental frequency of the saxophone note, in the case of Fig. 33, the ATM will mistakenly pick up the sound from the clarinet. In reality, there will be multiple different instruments (piano, guitar, bass, other saxophones, etc.) playing together even in a small band, making the fundamental frequency-based pitch detection inadequate. We need our ATM to pick up the frequency of the sound from the saxophone that it is controlling.

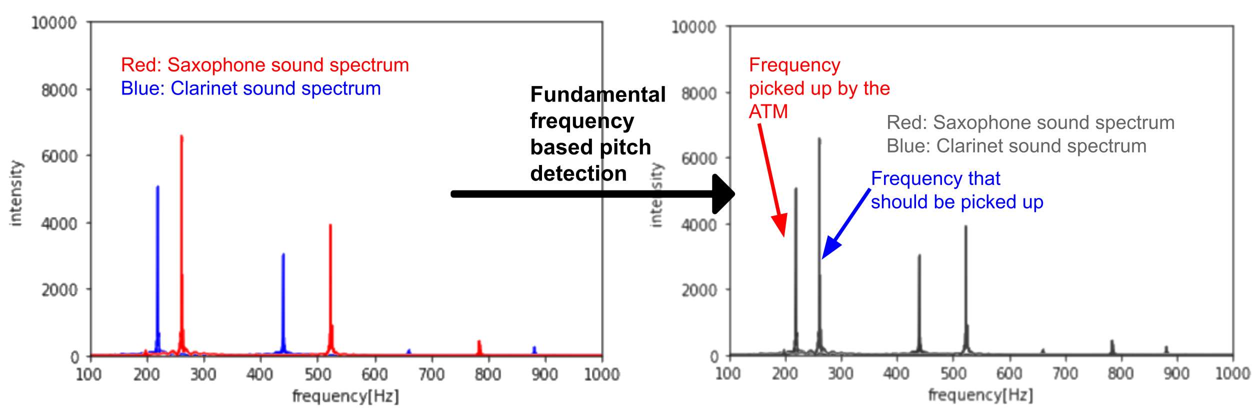


Fig. 33. ATM's incorrect perception of saxophone and clarinet sound

3.4.2 Dominant frequency

To fix this problem, we need to first introduce the dominant frequency. In a frequency spectrum, the dominant frequency refers to the frequency with the highest intensity, or in other words, the loudest. The dominant frequency can be any spike, including the fundamental frequency spike. In Fig. 6, both frequency spectrums' fundamental frequencies are also their respective dominant frequencies. In Fig. 34 below, the dominant frequency is 440 Hz.

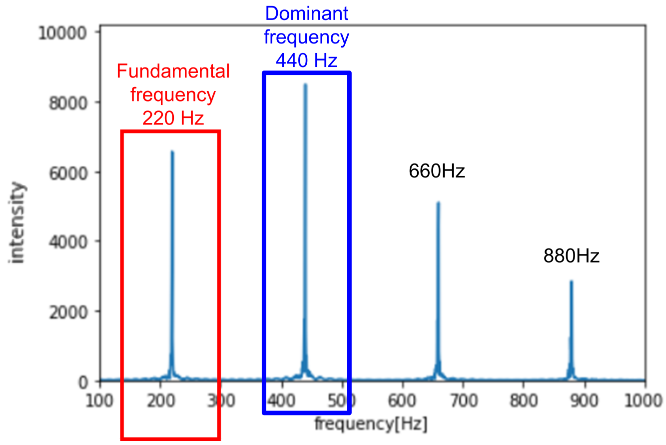


Fig. 34. Fundamental frequency (red) and dominant frequency (blue)

The ATM's microphone is around 10 centimeters away from the bell (Fig. 2) of the saxophone. Using an online sound volume detection service, we measured the sound volume 10 centimeters away from the bell to be an average of 107 dB during a normal performance. Here is the link to the online sound volume measurement service:
https://youlean.co/online-loudness-meter/

All sound volume measurements in dB are conducted through this website with the laptop's integrated microphone. Notice that this does not mean that the ATM is immune to any noise that is below 107 dB, because the saxophone's sound volume changes depending on different music pieces, and the 107 dB volume only represents the average. We personally can play saxophone with a sound of from 59 dB up to 122 dB. The ATM's resistance to background noise largely depends on the loudness of the saxophone player.

Now our new ATM picks up sound from the saxophone, but another problem occurs. We talked about in Section 2.5 that tuning a musical instrument is tuning its fundamental frequency. Now we don't have the fundamental frequency but the dominant frequency instead. However, we also mentioned that all non-fundamental frequencies, which certainly include the dominant frequency, have mathematical relationships to the fundamental frequency. We explored this relationship to find a way to reconstruct the fundamental frequency using the dominant frequency.

However, as musicians, we know that actually, we don't even need to reconstruct the fundamental frequency, because the fundamental frequency is always there to be heard, even when it is missing. (Schnupp et al., 2011) Musicians can determine whether a note is too sharp or too flat without even knowing that note is. To explain why this is possible, we need to talk about some music theories.

3.4.3 Dominant frequency based pitch detection

A solid music theory background is indispensable for the development of the ATM but is also pretty much irrelevant to the embedded systems. To be less distractive, we will only briefly introduce the necessary music knowledge and jump to the result. For a more detailed explanation, see Appendix 4.

Sound from musical instruments is itself a harmony, which is composed of the fundamental frequency and higher frequencies. Higher frequencies of a note are in multiples of the fundamental one, and most of them are magically also notes (Fig. 35). With some more knowledge in music theory, we know that among the first six multiples of the fundamental frequency, only the fifth one is not a note. Our current ATM can only cover at most the third multiple because of the mechanical limitations, thus circumventing the “small problem” in the fifth multiple (Fig. 36).

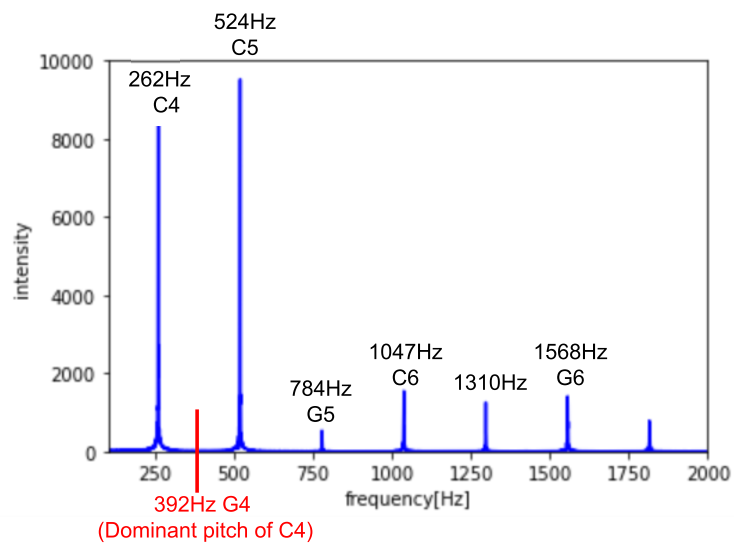


Fig. 35. Frequency spectrum of C4

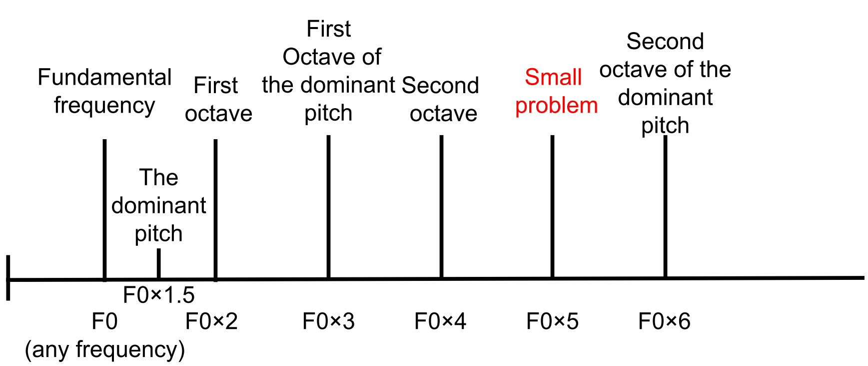


Fig. 36. First six spikes for any note

Because all multiples of the fundamental frequency (inside the ATM range) are also notes, we can tune any multiple of the fundamental frequency instead. Because the dominant frequency is always a multiple of the fundamental frequency, tuning the dominant frequency will have the same effect as tuning the fundamental frequency. Thus We conclude that **for our current ATM, it is okay to use the dominant frequency as the fundamental frequency without incurring any extra problem**.

Chapter 4: Experiment and Evaluations

Our ATM now listens to the dominant frequency, and we believe that our ATM is now ready for tests. We want to see how good our ATM is in different kinds of music in different performance environments. But before playing any music pieces, we want first to measure the note range and correction range of all notes.

4.1 Note range

Note range refers to the collection of all music notes playable. In Table I from Section 2.2, we introduced the normal note range for a saxophone, which is from C#3 (138.59 Hz) to A5 (880 Hz). During the development of the ATM, we sacrificed some note ranges for better pitch detection accuracy. The software thus limited the ATM's note range to from C#3 (138.59 Hz) to B4 (493.88 Hz).

Also, present saxophone mouthpieces are made of hard rubber and those high-end ones are still being produced by artisans by hand. We lack experience in traditional saxophone mouthpiece design, and our PLA material is far inferior to hard rubber in terms of vibration. As a result, it is extremely hard to play very high or very low notes on the ATM. Our ATM can only handle notes between G3 (196 Hz) and A4 (440 Hz).

For a note to be playable on the ATM, it has to meet the following three conditions:

1. The note is originally within the range of the saxophone.
2. The note is detectable by the microcontroller (below 512 Hz).
3. The note can be played by the saxophonist with a reasonable amount of effort, despite mechanical imperfections.

We overlap the range of these three requirements in Table I, with different colors, and obtain Fig. 37.

	Octave 0	Octave 1	Octave 2	Octave 3	Octave 4	Octave 5	Octave 6	Octave 7	Octave 8
C	16.35	32.70	65.41	130.81	261.63	523.25	1046.50	2093.00	4186.01
C#	17.32	34.65	69.30	138.59	277.18	554.37	1108.73	2217.46	4434.92
D	18.35	36.71	73.42	146.83	293.66	587.33	1174.66	2349.32	4698.64
D#	19.45	38.89	77.78	155.56	311.13	622.25	1244.51	2489.02	4978.03
E	20.60	41.20	82.41	164.81	329.63	659.26	1318.51	2637.02	5274.04
F	21.83	43.65	87.31	174.61	349.23	698.46	1396.91	2793.83	5587.65
F#	23.12	46.25	92.50	185.00	369.99	739.99	1479.98	2959.96	5919.91
G	24.50	49.00	98.00	196.00	392.00	783.99	1567.98	3135.96	6271.93
G#	25.96	51.91	103.83	207.65	415.30	830.61	1661.22	3322.44	6644.88
A	27.50	55.00	110.00	220.00	440.00	880.00	1760.00	3520.00	7040.00
A#	29.14	58.27	116.54	233.08	466.16	932.33	1864.66	3729.31	7458.62
B	30.87	61.74	123.47	246.94	493.88	987.77	1975.53	3951.07	7902.13

Fig. 37. ATM final range

As we see in Fig. 37, the range that satisfied all three requirements is from G3 (196 Hz) to A4 (440 Hz). For later experiments, we will limit all notes in this range. The blue limit as a result of low sampling frequency can be improved by using a more powerful microcontroller. The red limit as a result of unrefined mechanical mouthpiece designs can be improved by getting help from a traditional mouthpiece craftsman. Because of the limited time we have on this project, we are satisfied with our current note range. We then proceed to measure the correction range on each note.

4.2 Correction range

In Section 3.3.6, we calculated the estimated maximum correction range to be 28%. Now we measure the exact correction range of all notes by experiment. We measure by using a special program that will move the mouthpiece position first to the -20 position, wait 5 seconds, then move to the +20 position (40 steps of movement), wait another 5 seconds, and repeat forever. During the 5-second waiting time, we can play and measure the frequency of a note. We measured all the playable notes one by one and made the figure below. (Fig. 38) Frequencies of notes are labeled in dashed lines, and the correction ranges of the ATM on different notes are colored orange. Data from Fig.38 is in Appendix 3.

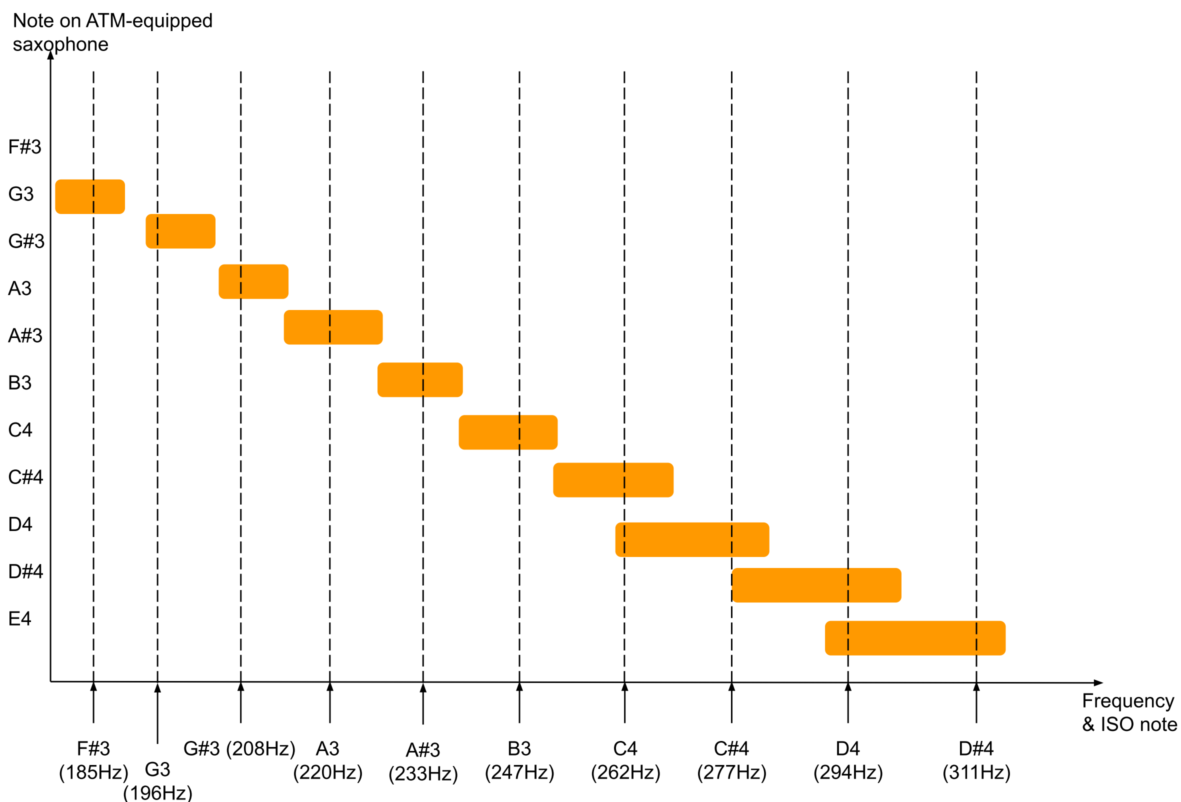


Fig 38(1). ATM correction range of all playable notes

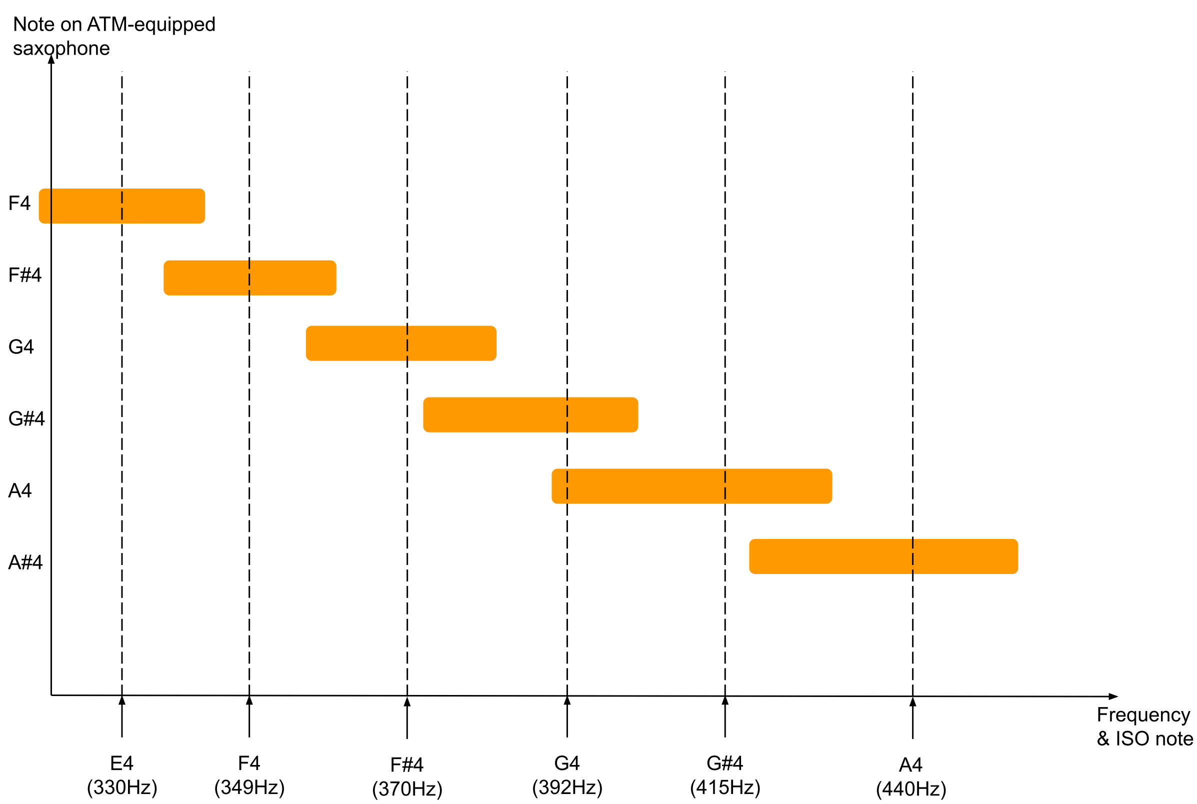


Fig 38(2). ATM correction range of all playable notes

We see that all dashed lines are covered, which means that our ATM can correct all notes, though to a different degree. Some correction ranges even cover two notes. A problem that we have noticed is that the ISO standard notes covered (labeled at the button) are all one note lower than the desired note of the ATM-equipped saxophone.

This problem occurs because the ATM is significantly longer than traditional mouthpieces, even at the most forward position. A longer tube means a lower frequency. This can also be categorized as part of the mechanical problems. To fix this problem, we should either redesign a shorter ATM or just instruct the saxophonist to always press the key pattern of the note that is higher than the note he wants to play. For example, with the ATM installed, if the saxophonist plays the note G4, the note will actually be F#4. If the player wants to play G4, he should play G#4 instead. To play the desired note, the saxophonist using the current ATM has to press the key pattern of the higher note.

When we were exploring the relationship between the mouthpiece's position and the sound frequency two years ago, we realized that it is very hard to find a reliable mathematical relationship. This realization eventually guided us to adopt the lookup table design in the final prototype. Now we have measured data from the completed ATM, we can revisit our speculations two years ago and try to find a mathematical relationship between mouthpiece position and sound frequency again.

We have plotted ATM correction range data from Fig. 38 in Fig. 39 below. The black line, which is the ISO standard frequency, is always between the red and blue lines. This positional relationship has the same implication that all dashed lines in Fig. 38 are contained by the yellow bars.

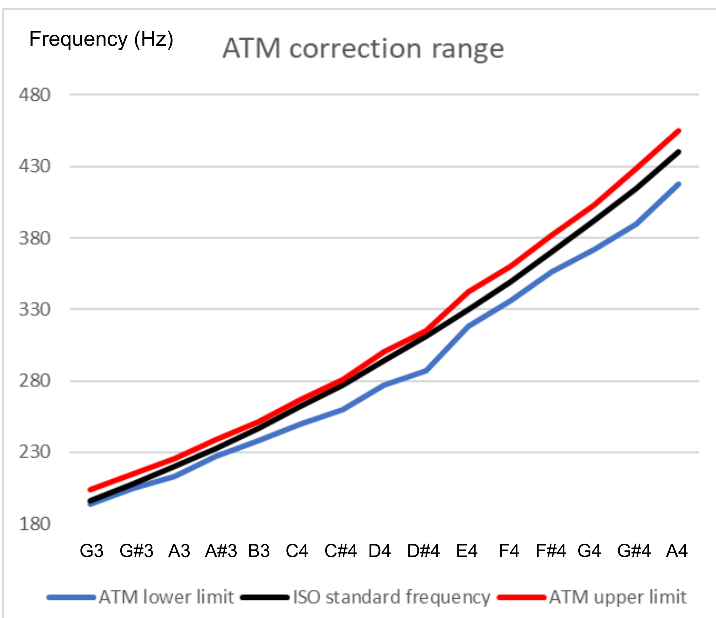


Fig. 39. ATM correction range and ISO standard frequency

We explore more about the relationship among these three lines. We decided to calculate the relative position of the ISO standard frequency between the ATM limits. This ISO percentage calculation considers the difference between the ATM limits as 100% and calculates the ISO standard frequency's position at a percentage. The ISO percentage is calculated with the following formula:

$$\text{ISO percentage} = \frac{\text{ISO standard frequency - ATM lower limit}}{\text{ATM upper limit - ATM lower limit}}\ \%$$

For example, in Fig. 40 below, we calculate the ISO percentage of G4 to be 64.5 %.

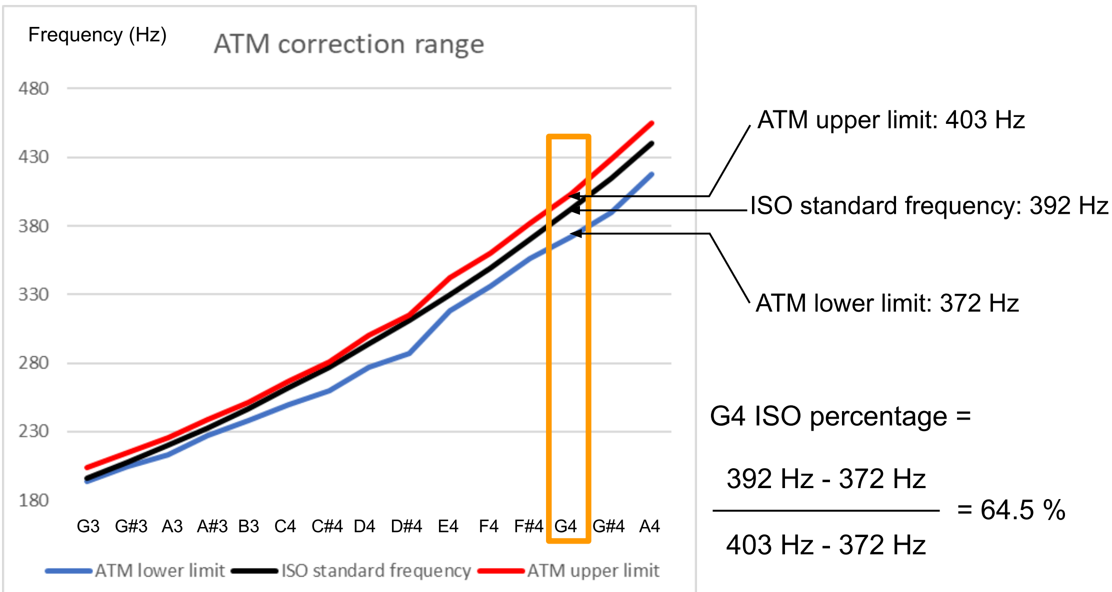


Fig. 40. ISO standard frequency position example

We repeated the same process in Fig. 41 for all notes and got Fig. 45 below.

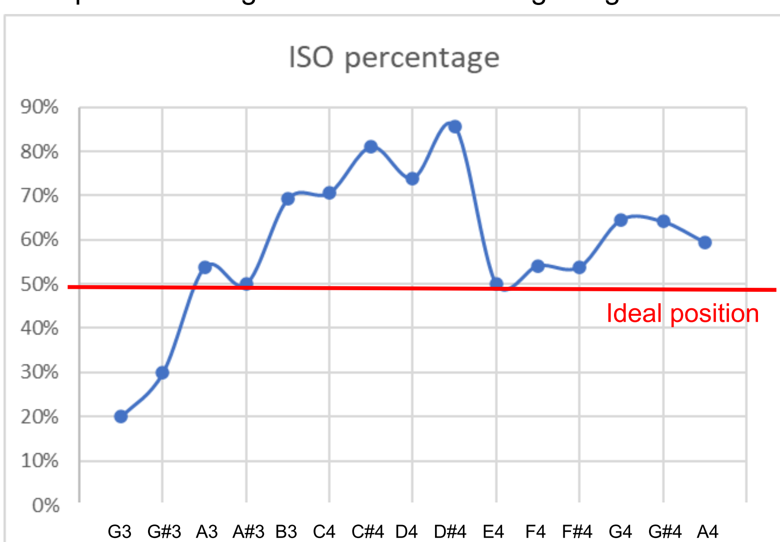


Fig. 41. ISO percentage of all notes

We see that the allocation of all notes' ISO percentages does show some trends. However, these trends are too obscure and cannot be translated into a reliable mathematical correlation. Still, we feel that some tendencies do exist and might be discovered in the future by someone with a more solid math background. For now, using the lookup table is the best approach.

4.3 Accuracy and speed: single note analysis

Now we know that our ATM can correct all notes from G3 to A4. Then we want to see how fast and how accurate our ATM corrects saxophone sound frequency. Since it is not possible to incorporate sound in an essay, we can only provide pictures of sound waveforms instead (See Fig. 12). We play the saxophone with the ATM installed, record the sound, and analyze it. All soundtracks in this chapter are recorded by a TASCAM DR-05 recorder and analyzed by Meloyne 5.3.1. TASCAM DR-05 is a stereo recorder, which means that it has two microphones and will record two channels simultaneously to simulate the perception of two human ears. For simplicity in frequency analysis, we only use the soundtrack from the right channel (Fig. 42).

Fig. 42. TASCAM DR-05
Taken from amazon.com

Whenever there is a pitch inaccuracy detected, the ATM intervenes. In different situations, the ATM intervention may lead to different scenarios. After thorough experiments, we conclude that there are in total 4 different scenarios:

1. Ideal correction
2. Extreme correction
3. Inaccurate correction
4. Human versus computer

4.3.1 Ideal correction:

Ideally, ATM corrects the pitch perfectly in one cycle. A cycle includes pitch detection and motor movement, which can take a maximum of 232 ms, which includes 192 ms frequency detection time, and a maximum of 20 ms motor movement time. (Section 3.3.6) In the ideal correction case (Fig. 43), the microphone picks up a frequency of +21% out of tune. This frequency detection process takes 192ms. Then, according to the lookup table, the ATM should move 9 steps inward. The motor movement time is thus 18 ms. After 18 ms, the frequency was correctly adjusted and stabilized at -1%.

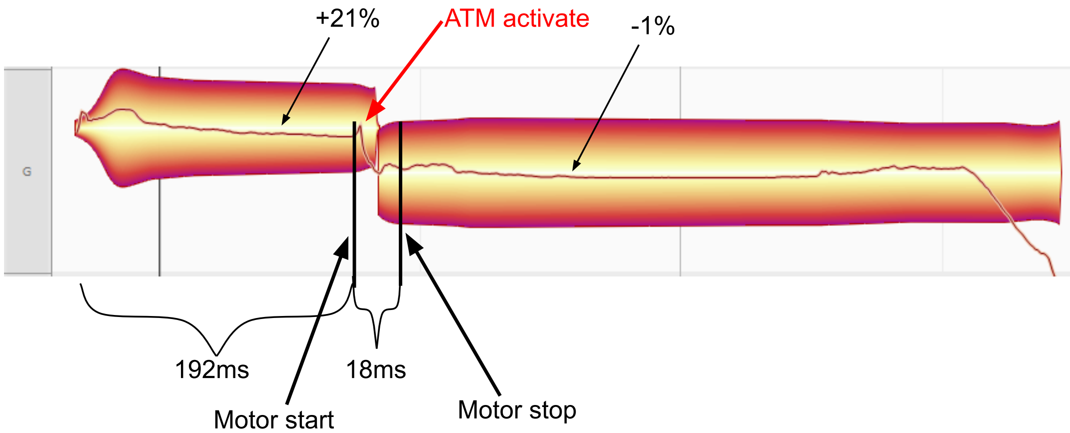


Fig. 43. Ideal correction

4.3.2 Extreme correction:

Sometimes, the player might be out of tune so much that a very big motor movement is needed. In the case shown in Fig. 44, 16 steps outward (which correspond to roughly 5 mm) proved to be slightly too much for the player's mouth to sustain. To be more specific, when the stepper motor rotates, the mouthpiece will move either into or out of the player’s mouth. Some slight movements are alright, but a quick and forceful thrust into the mouth is not a pleasant experience. The player loses his control over the embouchure and will have a shaky mouth. This leads to a duration of very unstable frequency changes until the player regains control over his mouth. The exact time that a player needs to regain mouth control depends on his skill and preparation. Skill is intuitive in that more skilled players generally have better control over the embouchure.

Mental preparation can also be helpful. When we first started testing, even movement of 11 steps may break the embouchure. After some practice, our mouth muscles learned that the mouthpiece might move. With both mental and mouth muscle preparation for the ATM to suddenly move into or out of the mouth, the same player can maintain his embouchure control during motor movement as big as 20 steps. As such, we decided that the maximum motor movement allowed per cycle is 20 steps. This is also shown in our detected frequency to motor movement lookup table in Appendix 2.

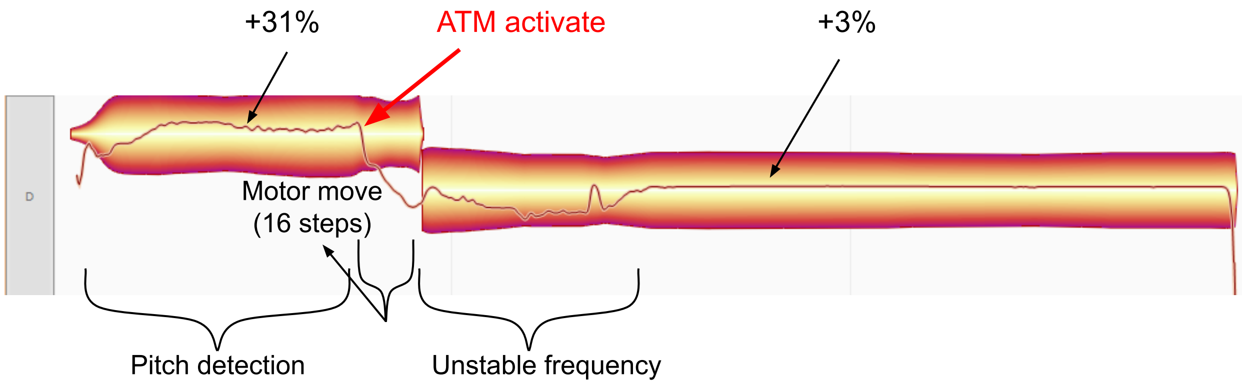


Fig. 44. Extreme correction resulting in shaky mouth

Sometimes, the player may be out of tune even more so that even moving 20 steps cannot tune him. In such a case (Fig. 45), our ATM will do the correction in 2 cycles. Cycle 1 corrects the deviation from +43% to +14%. A deviation of +14% is still too high, so Cycle 2 corrects it from +14% to +4%. At last, we get +4% deviation, which is below our 5% goal, thus the frequency stabilizes.

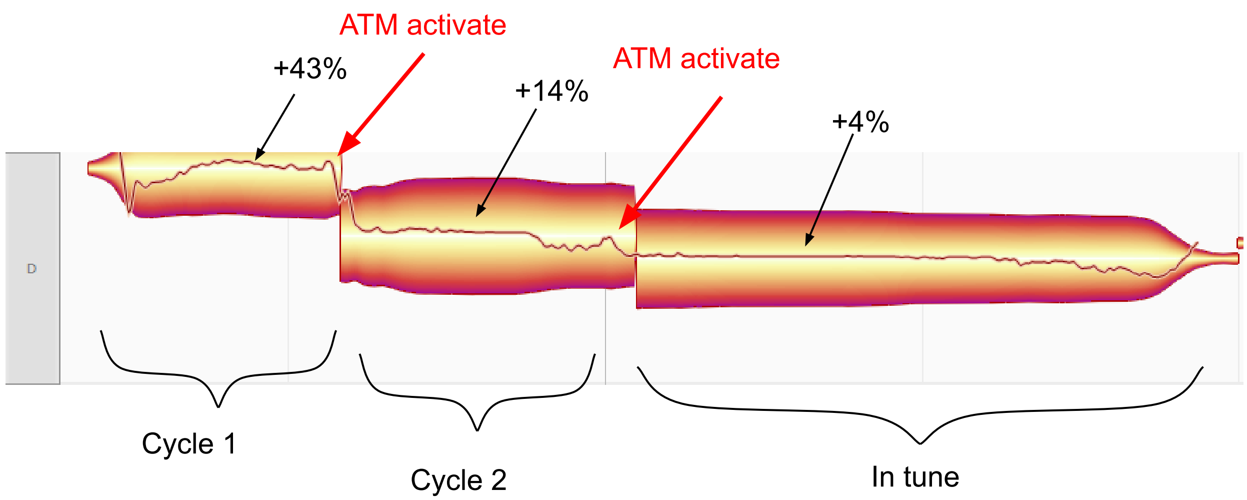


FIg. 45. 2-cycle response for extreme pitch deviation

Extreme correction is not ideal, however such cases do not happen frequently. Considering the rarity of the extreme correction case, an extra correction cycle is acceptable.

4.3.3 Inaccurate correction:

Although we tried to achieve perfect correction in one shot by applying a lookup table, unfortunately, we underestimated the inconsistency of human mouth muscles. When the ATM moves, not all its movement will be applied to the length increment of the tube. Some of the movement will be absorbed by mouth muscles. Different people's mouth muscles have different strengths, and even the same person may apply different mouth strengths at different times. Thus, an accurate detected-frequency-motor-movement lookup table does not exist. ATM might need to correct the frequency by multiple cycles if the player's mouth is either too loose or too tight. In such cases, our ATM is still effective in that it can greatly reduce pitch deviation, just not as perfectly as in the ideal case. Here are 2 diagrams of the multiple-cycle correction scenario.

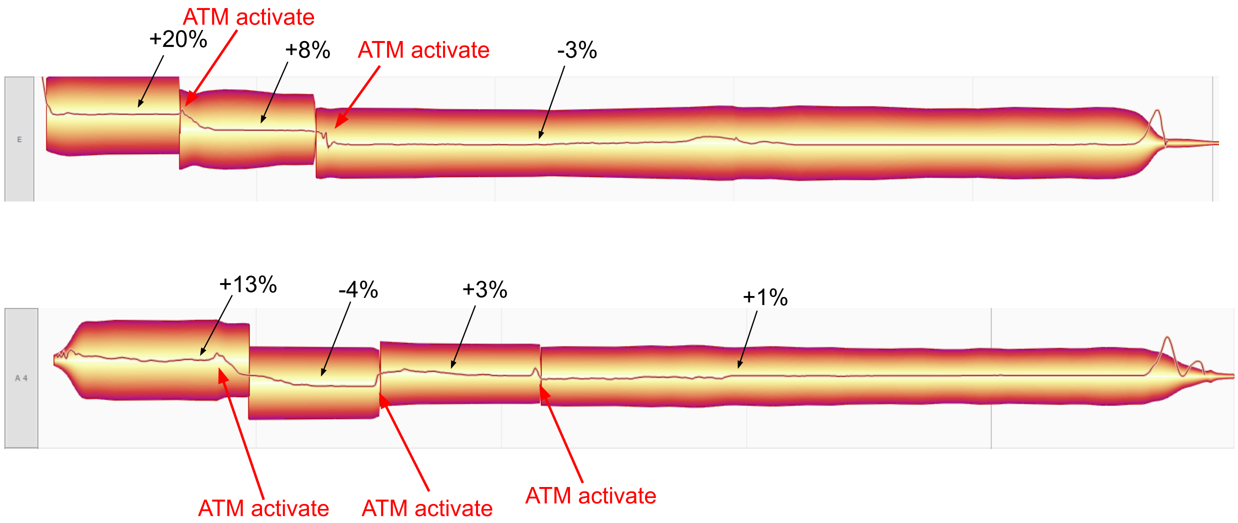


Fig. 46. Multiple cycle correction diagrams

In the first case, the note is out of tune by +20%. The ATM tries to increase the note's frequency, with an amount of movement from the lookup table. Ideally, the ATM should bring the frequency down to below 5%. The result of +8% is caused by the saxophonist's mouth being too loose.

In the second case, when the ATM tries to correct the +13% out of tune, the player's mouth is more firm than the lookup table's anticipation. As a result, the motor moves too much, resulting in -4%. Because our ATM's pitch detection has an inaccuracy of up to 4.7%, the -4% is probably perceived as something lower, thus activating the ATM again to correct the frequency back to +3%. For the same reason, the ATM was activated yet again to finally correct the frequency to +1%.

When the player's mouth is too loose, the ATM's movement tends to be insufficient; when the player's mouth is too firm, the ATM's movement tends to be too much. Our lookup table only represents the needed amount of motor movement under an average mouth firmness. To reduce such inaccuracy, we have come up with an idea to make the ATM learn the player's mouth firmness. Depending on the correction result of the former note, the ATM can measure the saxophone player's mouthpiece firmness, and adjust the lookup table accordingly.

4.3.4 Human versus computer:

Through years of practice, some saxophone players gradually develop an instinct to change their embouchure to make their notes in tune. This is traditionally how saxophonists develop better pitch accuracy. However, now with the presence of the ATM, such natural instinct might pose a problem. In the scenario below, both the player and the ATM noticed a pitch deviation and both tried to correct it. The result is that the frequency was corrected twice, leading to an overshoot (Fig. 47). For this problem, there is no obvious solution. Maybe saxophone players using the ATM need to develop another instinct that tells their mouth muscles to do nothing even when the ears notice that he is out of tune and wait for the ATM to do the job.

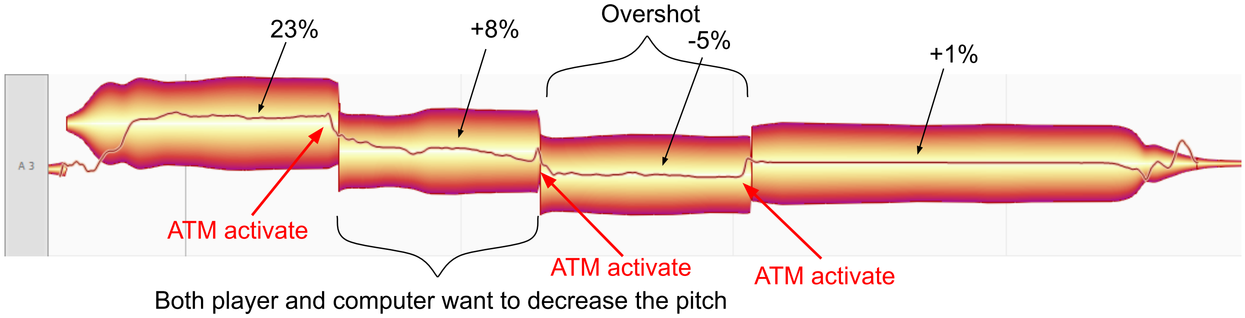


Fig. 47. Human versus computer

To fix this issue, we have an idea to give our ATM an Expert Mode, which reduces motor movement. This mode is for intermediate saxophonists who are capable of applying some adjustments using their mouth muscles. The ATM thus does not need to move that much.

4.3.5 correction time analysis:

In theory, each of our correction cycles takes from 194 ms to 232 ms. In reality, the saxophone player's mouth firmness can influence the speed of pitch rise or drop. After analyzing all data from section 4.3.1 to section 4.3.4, we estimate the approximated response time of the ATM to be roughly from 200 ms to 240 ms. This is the time from the emergence of the out-of-tune frequency to the point when the frequency stabilizes on the correct note. This measured correction time is only a few milliseconds longer than the theoretical response time. For human ears, a few milliseconds' delay is not noticeable. Thus, we conclude that the actual correction time is the same as the theoretical correction time.

4.4 Music piece analysis

We want to see how our ATM behaves in more realistic situations. We choose a relatively slow and easy piece of music: Wiegenlied (otherwise known as Lullaby) by Johannes Brahms. We have transcribed part of it for the saxophone, and the music score for the part that we will use for analysis looks like this:

Fig. 48. Part of Lullaby by by Johannes Brahms

At 80 BPM, the shortest note in the piece of the music score above lasts 375 ms, which satisfied the ATM's requirement of 232 ms. We performed this piece of music on the saxophone with the ATM and with a traditional mouthpiece(Theo Wanne 6). We compared the soundtracks and agreed that our ATM had significantly augmented the player's pitch accuracy. Below is the frequency diagram of the ATM working in a quiet environment. Pitch inaccuracies have been added. Pitches that have been corrected are annotated in red.

ATM frequency diagram

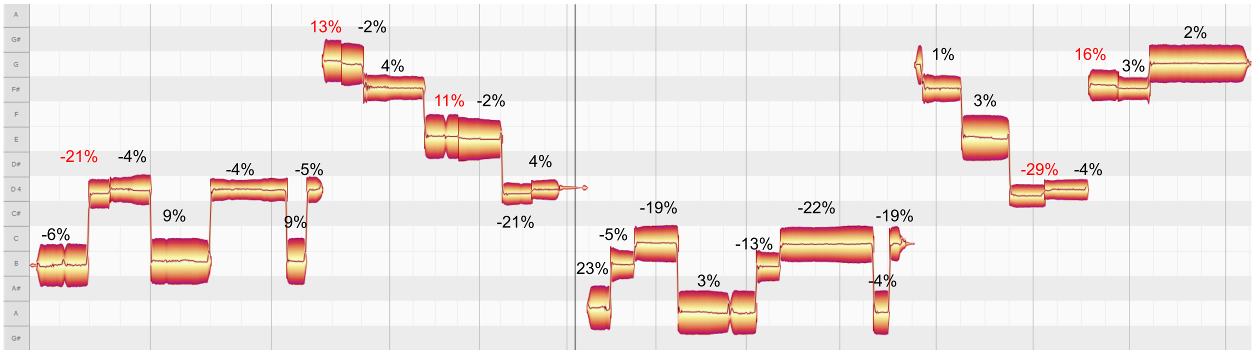


Fig. 49. ATM Frequency diagram, ideal environment

Different notes last different durations. Notes with longer durations generally have a bigger impact on the audience's perception of the overall intonation of the entire music piece. Thus, we decided to give all notes in Fig. 48 a weight corresponding to duration. The weights of some "heavy" notes are highlighted with bright colors to show the importance of those notes (Fig 50).

Fig. 50. Part of Lullaby by by Johannes Brahms, weight of all notes added

Combining the frequency data from Fig. 49 and the weight data from Fig. 50, we get the following chart:

Note	Weight	Frequency	Note	Weight	Frequency	Note	Weight	Frequency
B3	1	-6%	F#4	3	4%	B3	1	-13%
B3	1	-6%	E4	1	11%	C4	4	-22%
D4	3	-4%	E4	2	-2%	A3	1	-4%
B3	1	9%	D4	2	4%	C4	1	-19%
B3	2	9%	A3	1	23%	F#4	1	1%
D4	4	-4%	B3	1	-5%	E4	1	3%
B3	1	9%	C4	2	-19%	D4	2	-4%
D4	1	-5%	A3	2	3%	F#4	2	3%
G4	2	-2%	A3	1	3%	G4	4	2%

TABLE II. Pitch inaccuracy and weight

We calculate the overall pitch deviation by calculating the standard deviation of weighted pitch inaccuracy. Using data from Table II, we can calculate the standard deviation of the weighted frequency deviation to be 6.57%. From now on, when we mention the inaccuracy of a piece of music, we refer to the standard deviation of the weighted frequency deviation.

We got a higher number than our theoretical maximum deviation of 5%. This is the result of motor step loss. Step loss is caused by hindered movement, which is caused by the player's mouth being too firm. The stepper motor has no output signal. The microcontroller controls the stepper motor to rotate by sending command signals. The microcontroller does not know whether its command has been successfully executed by the motor. When the stepper motor tries to move but cannot, the microcontroller still records a movement. For example, after a few steps losses, the microcontroller might think that the mouthpiece is in position +10 while the mouthpiece is actually at -5%. Sometimes, when a lot of step loss happens, the ATM might mistakenly think that it has reached its movement limit and refuse to move. (circled in red in Table II) To solve this problem, we need to update mouthpiece position data by measurement and send it to the microcontroller. We believe that installing a sliding rheostat structure can solve the problem. The resistance of the sliding rheostat changes corresponding to the position of the mouthpiece. The microcontroller supplies a constant voltage to the sliding rheostat and measures the current, which can be used to calculate the mouthpiece position. If there is no step loss, we remove the data of C4 as circled in red, and get a better standard deviation of 4.91%.

To know how good our ATM is, we need to compare it with a traditional mouthpiece. We played the same part of Lullaby on a traditional mouthpiece (Model: Theo Wanne Durga 6). Repeat the same analysis demonstrated in Fig. 37 and Table II, we get a standard deviation of 10.55%.

We also did not forget to evaluate the negative impact of ATM's inferior build material and mechanical designs. We turned off our ATM and played on it again to see what result we would get. Repeat the same calculation, and we get a standard deviation of 11.42%. This result seems to be slightly worse than the traditional mouthpiece result of 10.55%, however, as saxophone players, we know that human inconsistency is way bigger than the slight difference between 11.42% and 10.55%. The mood, tiredness, and level of attention of the saxophone player can all impact his pitch accuracy. Thus, it is safe to say that although our ATM has a poor timbre and is hard to play, it is not inferior compared with traditional mouthpieces in terms of pitch accuracy, even when powered off.

In conclusion, according to the standard deviation analysis, turning on our ATM reduced the pitch deviation of a part of Brahms' Lullaby from 11.42% to 6.57%. Compared with traditional mouthpieces' pitch deviation of 10.55%, ATM also shows a superior pitch deviation of only 6.57%.

4.5 Realistic Environment Analysis

The recordings of Section 4.3 and Section 4.4 were all recorded in a quiet living room. In reality, saxophonists may need to perform in environments less ideal for pitch detection. There are primarily three kinds of noises: random background noise, wind pressure, and sound from other musical instruments.

4.5.1 Random background Noise

Random background noise means that the noise does not contain any obvious frequency. This type of noise is present in places like bars and markets. We simulate a random background noise by playing a market background noise from:
https://www.youtube.com/watch?v=vE_-CnEEYC8&ab_channel=Mr.RainandThunder-relaxingASMRsounds
The noise is played by the laptop's integrated speaker, and the ATM's microphone is roughly half a meter away from the laptop' speaker. Using the online sound volume measure service, we measured that the average and maximum generated noise volume at half a meter is 87 dB and 93 dB.

Before playing the saxophone, we adjust the microphone threshold screw until it is unaffected by the background noises (See Fig. 21). This means that the ATM will activate only if its perceived sound volume is higher than 87 dB.

We then conducted the same evaluation of section 4.2 and obtained a standard deviation of 4.71%. This result is actually better than the recording in the ideal environment (6.57%), mainly because our mouth is less firm this time, thus no step loss happened. This result also demonstrates that moderate background noise will not decrease pitch detection accuracy and thus will not affect the performance of our ATM. Considering the fact that the average sound volume of the saxophone (107 dB) is significantly higher than this noise volume (87 dB), it is expected that our ATM is immune to noises of such level. As long as we adjust the microphone screw to make the ATM activation threshold higher than the noise value, noise should never be a problem.

4.5.2 Wind pressure

Wind is another factor with a great impact on the microphone reading. The KY-037 microphone has a thin wind-proof layer on its surface to reduce wind pressure noises, so moderate wind should not affect it in theory, but we still devised an experiment to confirm this.

Because there is no power socket outside, we simulated wind using a hair dryer. We used the hair dryer to continuously blow cold wind toward the microphone of the ATM while playing. The hairdryer's sound volume is 82 dB, thus should not affect the ATM, as long as we adjust the microphone screw to make the ATM activation threshold higher than 82 dB. So, with the microphone threshold properly adjusted, we consider the wind to be the only variable in this experiment. With the same experiment procedure as section 4.5.1, we got a standard deviation of 5.66%. This result proves that moderate wind has indeed negligible impact on our ATM.

4.5.3 Another musical instrument

We want to determine the ATM's performance while there is another saxophone playing nearby. We want to see how far away the other saxophone must stay in order not to interfere with the ATM's pitch detection. The sound of another saxophone is simulated with the laptop. The result is as expected: the ATM picks up the loudest sound's frequency. To make our ATM not affected by other instruments, we need to play louder than all other instruments (measured from the position of the ATM's microphone). For example, consider the saxophone plays at 90 dB and the drum plays at 120 dB. We know that doubling the distance decreases the sound volume by 6 dB. The microphone's distance to the saxophone is roughly 10 centimeters. To guarantee that ATM picks up the saxophone's sound, the drum's distance to the microphone must be:

$$10\text{ cm} \times 2^{\frac{(120\text{ DB} - 90\text{ DB})}{6\text{ DB}}} = 3.2\text{ m}$$

Thus, in this hypothetical scenario, in order not to interfere with the ATM's pitch detection, the drum should be at least 3.2 meters away from the saxophone. In live performances, the distance between band members is limited by the stage size. The typical sizes of stages measure between 6 m x 4 m and 10 m x 8 m. Creating a 3.2 m distance between the drum and the saxophone should not be a problem.

4.6 User experience

Until now, the ATM has only been played by us, which is the developer of it. We want to see what most saxophone players think of the ATM. We found five participants. Among them, one is an intermediate-level saxophone player while the other four are beginners. Participant 1 Lou Xinhao is an intermediate saxophone player with around five years of experience in classic saxophone music. Before playing, we explained to him how the ATM works, and he was aware that the ATM might move forward and backward in his mouth.

4.6.1 Objective analysis (intermediate)

When Lou first tried on the ATM, the supposed automatic pitch correction had no effect at all. We were confused, and repeated the experiment, and soon realized that a lot of step loss was happening. In fact, the ATM was pushed to the most forward position and then barely moved at all.

In classical saxophone teaching, players are supposed to have a very firm mouth. Lou has a very firm mouth. Many intermediate players, including Lou, also tend to push the mouthpiece forward with teeth. Lou forcefully pushed the ATM to the most forward position and kept the ATM there with a very firm mouth. So firm that all movement from the stepper move was blocked. Here’s the frequency diagram of a note picked from Lou’s second try on the ATM (Fig. 51).

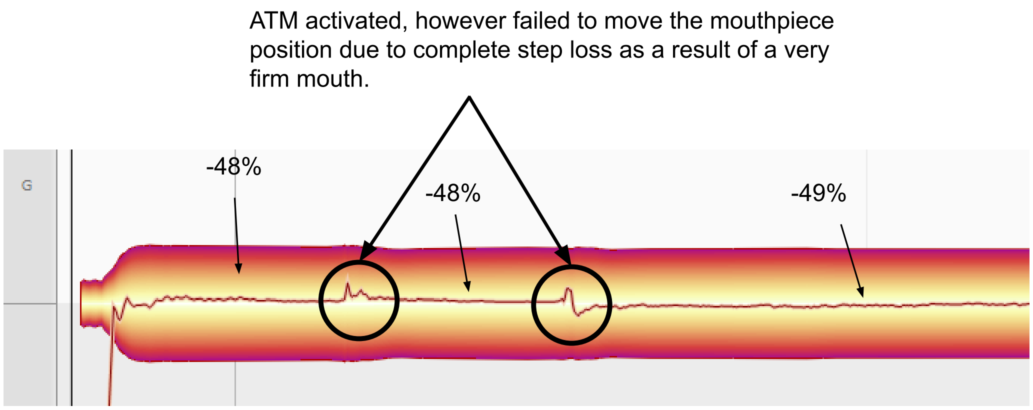


Fig. 51. Failed correction due to step loss

After realizing this problem, Lou was instructed to try to keep his mouth as loose as possible. Lou was hesitant because, for his many years of saxophone experience, he had always kept a firm mouth in order to control his intonation. When asked to loosen his mouth, Lou expressed his worry that he might not be able to keep a steady intonation in such a situation. We asked him to forget about the intonation control and just play with a loose mouth anyway.

When Lou started playing with a loose mouth for the first time in his saxophone career, he was out of tune by -22% as expected — however, the ATM immediately (around 210 ms) corrected his sound to the right frequency (Fig. 52).

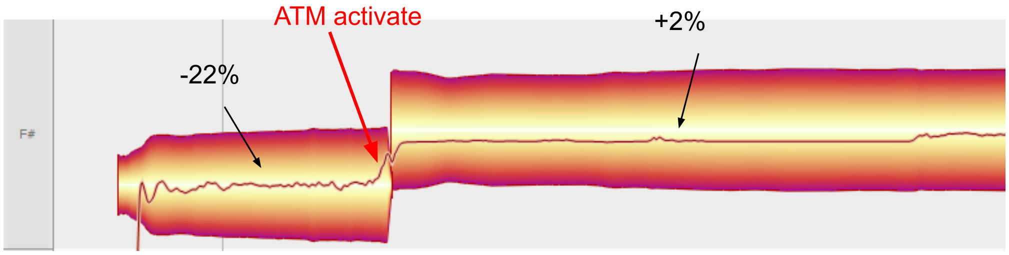


Fig. 52. Successful correction

Lou was surprised but kept playing different notes. Then he was even more surprised when he saw that the ATM corrected all his out-of-tune notes. (Fig. 53 & Fig. 54) He seemed to be amazed and continued to play with the ATM for quite some time.

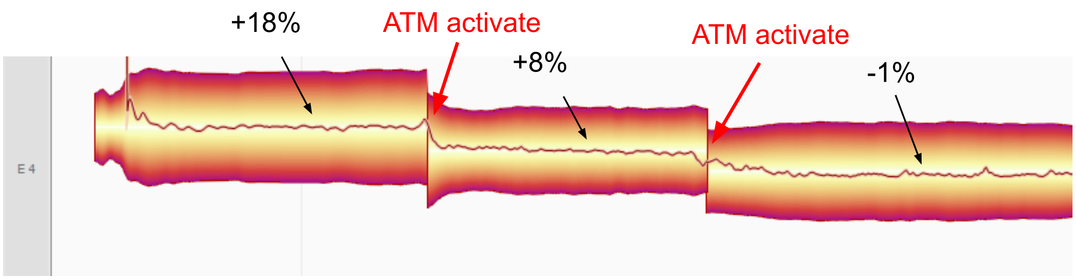


Fig. 53. Successful inaccurate correction of a sharp note

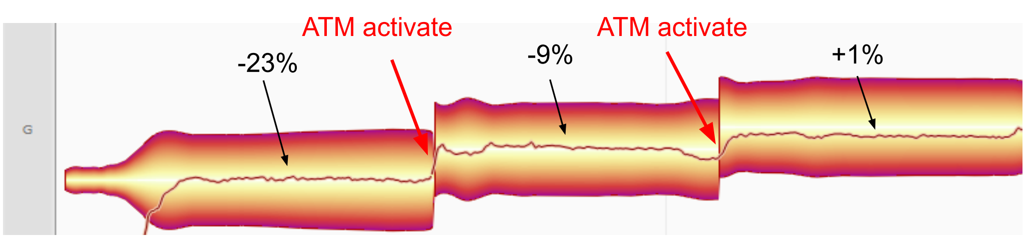


Fig. 54. Successful inaccurate correction of a flat note

Something we noticed is that the ATM corrected most of Lou's notes in two instead of one cycle. This is reasonable because the lookup table is made based on our mouth firmness. It seems that Lou's mouth actually became looser than ours after being asked to do so, thus needing more motor movement than what is specified by our current lookup table.

4.6.2 Subjective analysis (intermediate)

After Lou had spent around 20 minutes with the ATM, we started to ask him about his feelings and opinions about it. Lou said that the ATM has provided him with a completely new experience. He said that it feels weird that his notes are getting tuned automatically, and the player will need some time to get used to it. He also mentioned that the human-versus-computer problem is significant. The player needs to learn to not do anything when they realize that they are out of tune and wait for the ATM to respond instead. But Lou's foremost concern is that classical music saxophone players generally bite harder than we had expected, and we might need an even more powerful motor to prevent step loss, or simply tell our customers to not bite that hard. Lou believed that saxophonists need to learn how to cooperate with the ATM. This process does not take long. In Lou's case, with our instruction, he learned the correct way to use the ATM in 15 minutes.

As for Ergometry, Lou claimed that it is actually more comfortable than he had expected. He said that having something that moves a little bit in the mouth did not incur any discomfort. We thus speculate that our maximum movement limitation is sufficient. Overall, Lou believes that the ATM is a very cool invention and can provide a never-heard-of experience for saxophone players, and gave it a 7/10 rating.

4.6.3 Beginner experience analysis

All four beginners have no prior saxophone experience. We gave each of them a 10-minute basic introduction and training about how to play the saxophone with a traditional mouthpiece. Because of the ATM's material and mechanical design, it is very hard to play compared with traditional mouthpieces. As a result, beginners can only play for a few minutes before they are tired out and we can only collect limited data about beginner's experience on the ATM. In this section, we only will talk about the general conclusion of the experience of four beginner saxophone players. For the detailed interviews, please check Appendix 5.

Among the four beginner participants, two of them successfully played stable notes. Notes of these two participants are corrected by the ATM to the standard frequency (Fig. 55). The third participant could play notes, but his notes were very unstable (Fig 56). The ATM needs a stable frequency of at least 192 ms duration to function. The last participant could not make a sound out of the ATM-equipped saxophone because of the ATM's compromised building material and mechanical design.

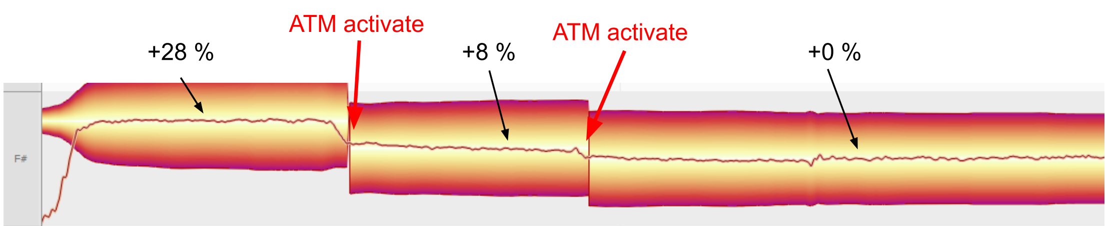


Fig. 55. A stable note from beginner participants

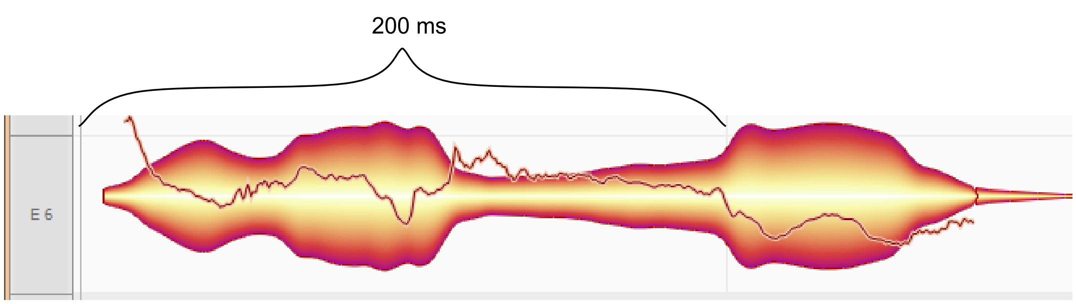


Fig. 56. Unreadable frequency from the third beginner participant

Among all five participants, the ATM succeeded in helping three of them but failed in the other two. The biggest problem is still the mechanical design, which is crucial for the success of the ATM as a product. This result also shows that the ATM still requires the saxophone player to have some basic saxophone-playing capability. A person who cannot play a traditional saxophone still cannot play an ATM-equipped saxophone. But for saxophone beginners with some but very limited capability, the ATM can be tremendously helpful. Saxophone players will need an average of roughly 15 minutes (with our instruction) to learn how to use the ATM properly. By lowering the difficulty level of playing the saxophone in tune, the ATM can also encourage more people to learn the saxophone.

4.7 Future works

We have many ideas that might improve the ATM's performance. However, as we mentioned in the beginning of Section 3.2, the ATM project already took more time than usual and we do not have more time to implement and test all these ideas. For now, we are happy with our current ATM; in the future, when we have time, or anyone wants to resume our work, this section can be the starting point.

4.7.1 Better microcontroller

The current pitch detection bottleneck is the speed of the CPU and RAM of the microcontroller. Better hardware would help. Our Arduino UNO's RAM size limited our maximum sample number to 128. If we can get a microcontroller with a bigger RAM, then we would be able to save more samples and get more accurate pitch detection results. Arduino UNO is equipped with a very weak ATmega328 processor, which only achieved a 4156 Hz maximum sampling frequency in our case. (despite the ATmega handbook claiming a maximum 9600Hz sampling frequency) Ideally, we want to achieve a 48 kHz sampling frequency, which is today's audio engineering's standard sampling frequency.

4.7.2 Expert mode

When we first started this project, we thought that the ATM was only for amateurs or at most intermediates. However, now we realize that the ATM can be helpful even for pro players. Pro players have different needs. While amateur players need extreme pitch correction on many notes, pro players will only need subtle correction on a few notes. Pro players may also employ skills like quarter tone and vibrato, which our current ATM cannot deal with. If we have more time, we can develop a new "expert mode" for our ATM to meet the needs of pro saxophone players. In the expert mode, major note deviations and deviations below a certain duration of time are considered intentional and thus ignored. The ATM should only correct subtle pitch deviations during a long note.

4.7.3 Behavior learning

Different players have different levels of mouth firmness. The ATM can learn its master's mouth firmness and change the lookup table accordingly. The current lookup table is based on our mouth firmness, which does not necessarily represent the average mouth firmness of all saxophone players. The mouth will "absorb" part of the ATM movement. A loose mouth will "absorb" more movement than a firm mouth. Thus, to achieve that same amount of adjustment, for a player with a more firm mouth, the ATM should move less, and for a player with a more loose mouth, the ATM should move more. We can give our ATM a learning mode that can record data from the inaccurate correction case (Section 5.1.3), determine whether the player's mouth is firmer or looser than the preset, and change the lookup table accordingly.

4.7.4 Solve step loss issue

Because our 42 mm stepper motor's (Fig. 24) limited power, step loss is unavoidable. Using a bigger motor will need bigger space and more electricity, which is a big compromise to the already wacky mechanical design and thus should be avoided. The problem is actually not step loss but unrecorded step loss. We need to give the microcontroller a way to know the current position of the ATM. An easy way to measure distance is to use a sliding rheostat. The sliding rheostat moves in parallel with the ATM, and its resistance changes as the ATM's position changes. The microcontroller can measure the current through the sliding rheostat under a constant voltage to calculate the position of the ATM. With the sliding rheostat distance measure mechanism, the microcontroller will not need to remember the movement distance, but can rather calculate the distance, and step loss will no longer be a problem.

4.7.5 Vibration detection

There exists a kind of guitar tuner that detects frequency by vibration:

Fig. 57. Vibration detect guitar tuner.
PIcture taken from amazon.com

However, it only applies to a flat surface. This is logical because guitars have a lot of flat surfaces, but the saxophone has no flat surface. However, modifications can be made to make it suitable for curved surfaces of the saxophone. If we can measure frequency by vibration, then we can insulate all sounds from other instruments and background noises forever. We will also need to design new frequency detection algorithms based on vibration. Such an algorithm might be simpler than our current algorithm and can thus potentially improve the execution speed. The major difficulty is how to make a device design for a flat surface suitable for a curved surface.

Chapter 5: Conclusion

The Saxophone is a musical instrument that tends to be out of tune. Today's pre-performance saxophone tuning can tune only one note, while all other notes can only be tuned during the performance by mouth. A saxophone player needs years of practice to master tuning by mouth. We want to find a new way to help saxophonists tune their instruments in real time during performances. We did this by inventing the Automatic Tuning saxophone Mouthpiece (ATM). The ATM comprises an adjustable mouthpiece, a stepper motor, a microphone, and a microcontroller. The microcontroller calculates the player's note's frequency using signals from the microphone and sends corresponding control signals to the stepper motor to adjust the mouthpiece position to correct the player's note's frequency.

We set our goal to create an ATM with less than 5% pitch deviation and less than 166 ms response time. We built a prototype that can in theory achieve a 4.7% deviation, but as a compromise and sacrifice, the worst response time was increased to 232 ms. In actual experiments, our prototype achieved less than a 6.57% deviation, and each correction cycle takes roughly from 200 ms to 240 ms. Because of mechanical imperfections, the ATM will also limit the playable range of the saxophone. We cannot say that we have achieved comprehensive success, but we can say that we have proven that the Automatic Tuning Saxophone Mouthpiece is a viable idea. The ATM's performance is compromised by a slow processor, insufficient RAM, and mechanical problems. Nevertheless, our current ATM is still good enough as a prototype, and it successfully demonstrated that it is possible to tune a saxophone in real time during the performance.

The biggest constraint on the development of the ATM is the mechanical issue. We know that traditional mouthpieces are made of hard rubber, but we only have PLA and SLA material. These materials proved to be far inferior acoustically. We also lacked professional modeling skills and had to learn 3D modeling from the beginning. We also overestimated the accuracy of 3D printing and wasted some time on remodeling with 3D printing inaccuracy in consideration. With hindsight, we believe that we made a very wise decision two years ago to start early. We gave ourselves plenty of time to deal with unexpected obstacles encountered during the development, and successfully solve all mechanical issues to some degree.

Knowledge of music theory also proved to be crucial for the development of the ATM. Through experiments, we realized the inadequacy of the fundamental frequency-based pitch detection. Thanks to our solid music theory knowledge, we discovered and justified that it is okay to use the dominant frequency instead.

As for ergometry, the discomfort caused by the ATM is less severe than we had expected, especially after we limited the maximum motor movement amount in a single cycle. According to our user experience experiments, an average saxophone player will need around 15 minutes to learn how to use the ATM properly.

There are still many possible improvements that we are not able to implement due to limited time (section 4.7). One of our user experience participants commented “The current ATM is limited, but its future is infinite”. When the time is appropriate, our work on the ATM can always be resumed to transform it from a prototype to a product.

And, despite our best efforts, it is still very hard to illustrate the sound effect of our ATM through words and diagrams. If you want to see and hear how exactly our ATM behaves during a performance, watch this video:

https://www.youtube.com/watch?v=-isqbL49Cvc&ab_channel=IDAlreadyExisted

Acknowledgement

The ATM prototype will never be completed without the help from many people. I would like to thank people from TU Delft Demo and Camlab for their support in 3D printing. I would like to thank all my five user experience participants: Lou Xinhao, Zhao Wenjie, Zhao Zhouyangguang, Gao Xinyu, and Zhao Yikai. But foremost, I would like to thank my supervisor Professor Koen Langendoen, who not only allowed me to create the ATM as my thesis project but also kept giving me advice, instructions, and encouragement during the ATM’s 2-year-long development.

References

Arduino Frequency Analyzer. (2017, November 9). forum.arduino. Retrieved January 14, 2024, from https://forum.arduino.cc/t/arduino-frequency-analyzer/598041/7

Artificial Tunes on Tumblr. (n.d.). Retrieved January 11, 2024, from https://artificialtunes.tumblr.com/

Bachem, A. (1955). Absolute Pitch. *The Journal of the Acoustical Society of America*, *27*(6).

Benward, B., & Saker, M. N. (2003). *Music in Theory and Practice*. McGraw-Hill.

Carden, J., & Cline, T. (n.d.). Absolute pitch: Myths, evidence and relevance to music education and performance. *Psychology of Music*, *47*(6). 2019

Celemony. (n.d.). Celemony | Tomorrow's audio today. Retrieved January 14, 2024, from https://www.celemony.com/en/start

Double harmonic scale. (n.d.). Wikipedia. Retrieved January 11, 2024, from https://en.wikipedia.org/wiki/Double_harmonic_scale

Fujisaki, W., & Kashino, M. (2002). The basic hearing abilities of absolute pitch possessors. *Acoustical Science and Technology*, *23*(2), 77-83. https://www.jstage.jst.go.jp/article/ast/23/2/23_2_77/_article

ISO 9001:2015(en), Quality management systems — Requirements. (n.d.). ISO. Retrieved January 11, 2024, from https://www.iso.org/obp/ui/#iso:std:iso:16:ed-1:v1:en

Loeffler, B. D. (2007-12-18). Instrument Timbres and Pitch Estimation in Polyphonic Music. In (p. 6). Department of Electrical and Computer Engineering, Georgia Tech.

Peretz, I., & Hyde, K. L. (2003). *What is specific to music processing? Insights from congenital amusia*. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/12907232/

Rossi, M. (1971). Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. In (pp. 1-33). Phonetica.

Rossi, M. (1978). La perception des glissandos descendants dans les contours prosodiques (The Perception of Falling Glissandos in Prosodic Contours). In (pp. 11-40). Phonetica.

The saxophone - popular in all genres • lebrass. (n.d.). leBrass Wind instruments. Retrieved January 10, 2024, from https://www.lebrass.com/saxophone/

Scala pentatonica. (n.d.). Wikipedia. Retrieved January 11, 2024, from https://it.wikipedia.org/wiki/Scala_pentatonica#Scala_cinese

Schnupp, J., Nelken, I., & King, A. (2011). *Auditory Neuroscience: Making Sense of Sound*. MIT Press.

Stigler, S. M. (n.d.). *Normal distribution*. Wikipedia. Retrieved January 14, 2024, from https://en.wikipedia.org/wiki/Normal_distribution

Takeuchi, A. H., & Hulse, S. H. (1993). *Absolute pitch*. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/8451339/

12 equal temperament. (n.d.). Wikipedia. Retrieved January 11, 2024, from https://en.wikipedia.org/wiki/12_equal_temperament

Appendix

Appendix 1: Bias table

Note	Standard frequency	Measured frequency	Delta
F#3	185	189	4
G3	196	200	4
G#3	208	211	3
A3	220	224	4
A#3	233	236	3
B3	247	251	4
C4	262	266	4
C#4	277	282	5
D4	294	299	5
D#4	311	316	5
E4	330	335	5
F4	349	355	6
F#4	370	377	7
G4	392	399	7
G#4	415	422	7
A4	440	448	8
A#4	466	474	8

Appendix 2: detected frequency to motor movement lookup table

In the detected frequency to motor movement lookup table below, red indicates frequencies considered as in tune, white indicates frequencies that need to be adjusted, grey indicates no adjustment needed, blue indicates the number of steps moving counterclockwise(out), and yellow indicates the number of steps moving clockwise(in).

189	
190	3
191	6
192	9
193	12
194	15
195	15
196	12
197	9
198	6
199	3
200	
201	3
202	6
203	9
204	12
205	15
206	15
207	12
208	9
209	6
210	3
211	
212	3
213	6
214	9
215	12
216	15
217	18
218	18
219	15
220	12
221	9
222	6
223	3
224	
225	3
226	6
227	9
228	12
229	15
230	18
231	15
232	12
233	9
234	6
235	3
236	
237	3
238	6
239	8
240	11
241	14
242	16
243	18
244	18
245	16
246	14
247	11
248	8
249	6
250	3
251	
252	3
253	6
254	8
255	11
256	14
257	16
258	18
259	18
260	16
261	14
262	11
263	8
264	6
265	3
266	
267	3
268	6
269	8
270	11
271	14
272	16
273	18
274	18
275	17
276	15
277	12
278	10
279	7
280	5
281	3
282	
283	3
284	5
285	7
286	10

287	12	320	8	353	4
288	15	321	10	354	2
289	17	322	12	355	
290	18	323	14	356	2
291	18	324	16	357	4
292	17	325	18	358	6
293	15	326	18	359	8
294	12	327	16	360	10
295	10	328	14	361	12
296	7	329	12	362	14
297	5	330	10	363	16
298	3	331	8	364	18
299		332	6	365	20
300	3	333	4	366	20
301	5	334	2	367	18
302	7	335		368	16
303	10	336	2	369	15
304	12	337	4	370	13
305	15	338	6	371	11
306	17	339	8	372	9
307	18	340	10	373	7
308	18	341	12	374	6
309	17	342	14	375	4
310	15	343	16	376	2
311	12	344	18	377	
312	10	345	20	378	2
313	7	346	18	379	4
314	5	347	16	380	6
315	3	348	14	381	8
316		349	12	382	10
317	2	350	10	383	12
318	4	351	8	384	14
319	6	352	6	385	16

386	18
387	20
388	20
389	18
390	16
391	15
392	13
393	11
394	9
395	7
396	6
397	4
398	2
399	
400	2
401	4
402	6
403	7
404	9
405	11
406	13
407	15
408	16
409	18
410	20
411	20
412	18
413	16
414	15
415	13
416	11
417	9
418	7
419	6
420	4
421	2
422	
423	2
424	4
425	5
426	7
427	9
428	11
429	12
430	14
431	15
432	17
433	18
434	20
435	20
436	18
437	16
438	15
439	14
440	13
441	11
442	10
443	9
444	7
445	5
446	4
447	2
448	
449	2
450	4
451	5
452	7
453	9
454	11
455	12
456	14
457	15
458	17
459	18
460	20
461	20
462	18
463	16
464	15
465	14
466	13
467	11
468	10
469	9
470	7
471	5
472	4
473	2
474	
475	2
476	4
477	5
478	7
479	9
480	11
481	12
482	14
483	15
484	17

Appendix 3: ATM correction range table

Note	ISO standard frequency	ATM low limit (position -20)	ATM high limit (position +20)
G3	196 Hz	181 Hz	191 Hz
G#3	208 Hz	194 Hz	204 Hz
A3	220 Hz	205 Hz	215 Hz
A#3	233 Hz	213 Hz	226 Hz
B3	247 Hz	227 Hz	239 Hz
C4	262 Hz	238 Hz	251 Hz
C#4	277 Hz	250 Hz	267 Hz
D4	294 Hz	260 Hz	281 Hz
D#4	311 Hz	277 Hz	300 Hz
E4	330 Hz	287 Hz	315 Hz
F4	349 Hz	318 Hz	342 Hz
F#4	370 Hz	336 Hz	360 Hz
G4	392 Hz	356 Hz	382 Hz
G#4	415 Hz	372 Hz	403 Hz
A4	440 Hz	390 Hz	429 Hz
A#4	466 Hz	418 Hz	455 Hz

Appendix 4: Dominant frequency based pitch detection (detailed)

Dominant pitch and octaves

The dominant pitch, also known as the dominant key, or simply the dominant, refers to the fifth degree of a scale. The musical definition of the dominant pitch may be obscure, but its scientific definition is clear: Every note has one and only one dominant pitch, which is of approximately 50% higher frequency, and is also a note in Table I. For example, the dominant pitch of A3 (220 Hz) is E4 (329.63 Hz). Notice that while the fundamental pitch and the fundamental frequency mean essentially the same thing, the dominant frequency and the dominant pitch are completely different concepts.

Fundamental frequency = Fundamental pitch
Dominant frequency: The frequency with the highest intensity
Dominant pitch: the frequency that is 50% higher than the current note's frequency

Every note also has an octave, which is another note with 100% higher frequency. For example, the octave of A3 (220 Hz) is A4 (440 Hz). The octave of A4 (440 Hz) is A5 (880 Hz). Because we get A5 after rising A3 for two octaves, we also call A5 as A3's second octave. The same rule applies to the third or higher octaves. Also, in the frequency spectrum of any note, the dominant pitch itself is never a spike, while the first two octaves of the dominant pitch are always spikes. See the frequency spectrum of C4 (262 Hz) below (Fig. 35).

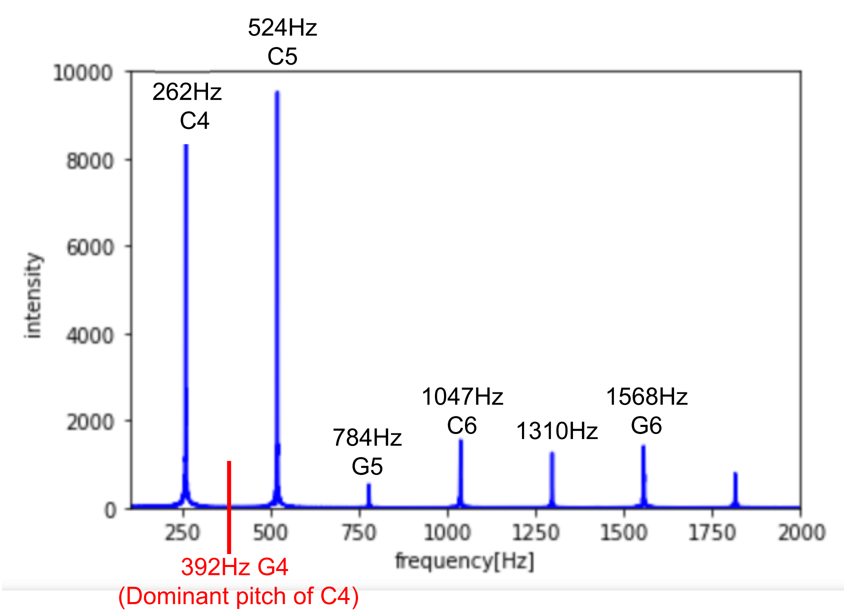


Fig. 35. Frequency spectrum of C4

We see that in Fig. 30, a fundamental frequency (first spike) of 262 Hz shows that this is a C4. The dominant pitch of C4 is G4 (392 Hz). The first octave of C4 is C5 (524 Hz), which is the second spike, which also happens to be the dominant frequency in this case. The third spike is G5 (784 Hz), which is the octave of the dominant pitch G4. The fourth spike is C6 (1047 Hz), which is the second octave of C4. The fifth spike has a frequency of 1310 Hz and is not related to any notes. The sixth spike G6 (1568 Hz) is the second octave of the dominant pitch. We see that, for any note, its frequency spectrum's first six spikes are also all notes, except the 5th one. (See Fig. 36) F0 denotes the fundamental frequency. It can be the frequency of any note in Table I.

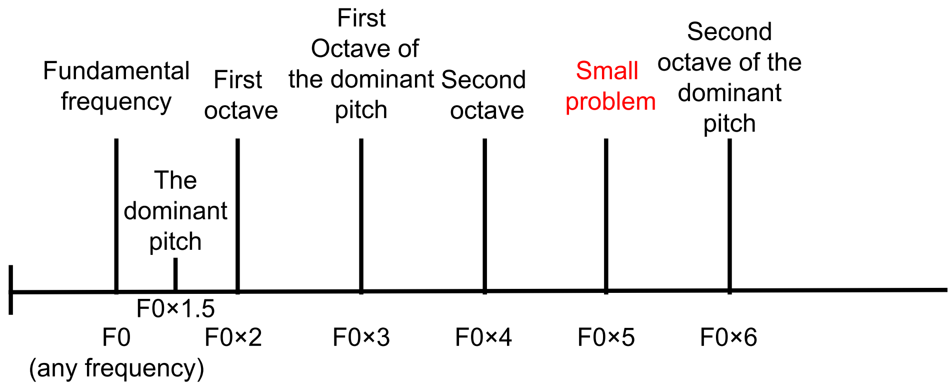


Fig. 36. First six spikes for any note

Dominant frequency over the fundamental frequency

We know that the dominant frequency is one of those six spikes, in which five are also notes. Now we just temporarily forget about the fifth note labeled as "small problem" to make everything easier. We know that if the fundamental frequency is out of tune, all its non-fundamental frequencies will also be out of tune by the same ratio. Then, we can tune any non-fundamental frequencies, and the fundamental frequency will also be tuned automatically. Since all non-fundamental frequencies are also frequencies of other notes, we can tune the other notes instead. For example, if the microcontroller receives a dominant frequency of 880 Hz, then there will be five possibilities:

1. It is the fundamental frequency. The note is A5 (880 Hz).
2. It is the first octave of the fundamental frequency. The note is A4 (440 Hz)
3. It is the first octave of the dominant pitch of the fundamental frequency. The note is D4 (293 Hz).
4. It is the second octave of the fundamental frequency. The note is A3 (220 Hz)
5. It is the second octave of the dominant pitch of the fundamental frequency. The note is D3 (147 Hz).

We do not know which note this 880 Hz points to. However, if the microcontroller receives a dominant frequency of 884 Hz, then it will be considered sharp in all five possibilities. No matter which one possibility is true, the ATM's response is the same: move outward to decrease the sound frequency. Thus, we do not need to know whether this note is an A5, or A4, or D4, or A3 or D3. The ATM can simply treat them as the same and give the same response.

One may ask, but why only the first six spikes? That is because of the frequency range limit of the saxophone. In Section 2.2, we mentioned that the maximum frequency of a saxophone under normal conditions is from 138.59 Hz to 880 Hz. Thus, we use a low pass filter to cut off all frequencies above 880 Hz. The highest multiple under 880 Hz that the saxophone can achieve is:

$$\frac{880 \text{ Hz}}{138.59 \text{ Hz}} = 6.35$$

Thus, we only need to check six spikes. The only remaining problem is the fifth note. But, according to our experience, the dominant frequency of a saxophone is almost always in the first three spikes. This does not guarantee that the fifth spike is not dominant frequency, however, such a possibility is low.

What's more, in Section 3.3.2, because of the limitations of our hardware, we have limited the upper limit of the detection range of our ATM to 512 Hz. Under such a frequency cap, the highest multiple that can be achieved is:

$$\frac{512 \text{ Hz}}{138.59 \text{ Hz}} = 3.69$$

Thus, our ATM can actually only detect only first three spikes, which are all notes, therefore circumventing the problem of the fifth spike. Thus, the ATM picking up the wrong sound problem is solved without any compromise, for now. However, if we want to develop a better ATM that covers the entire normal saxophone range, we would need to figure out a way to deal with the fifth spike.

Appendix 5: Beginner experience interview

Participant 1

Participant 1 has been playing the harmonica for many years. He learned how to make sound on a saxophone relatively quickly. With the ATM installed, he initially could not make a sound. After a few more minutes, he finally managed to play a stable long note, which was successfully corrected by the ATM. (Fig. 58) After that, participant 1 claimed that his mouth muscles were tried out, so we let him go. Later, he mentioned that one of the reasons that he chose the harmonic is that harmonica is always in tune. He mentioned that many people avoid learning woodwind instruments because they do not want to practice for a few years just to get in tune. ATM has the potential to encourage more people to learn woodwind instruments. He said that the current ATM might be limited, but its future is infinite. Despite our asking for a 1 to 10 rating, he insisted on rating the ATM 100/10. We eventually decided to accept his goodwill and recorded this as abnormal feedback.

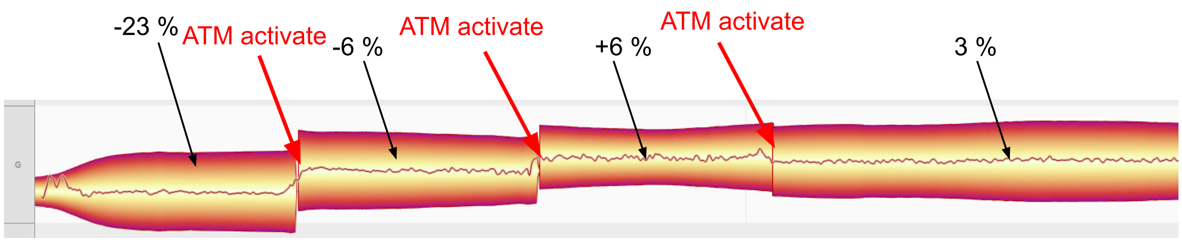


Fig. 58. The only stable note from participant 1

Participant 2

The second participant has some experience in playing Xiao, which is a kind of traditional Chinese wind instrument. He did not manage to play a stable note on the ATM. Here's one of his notes (Fig 55). As we see, the frequency is very unstable. The ATM needs a stable frequency of at least 192 ms to function. As a result, the ATM is not able to help Participant 2. He rated the ATM 999/10. We consider this as a rating of our friendship rather than the ATM.

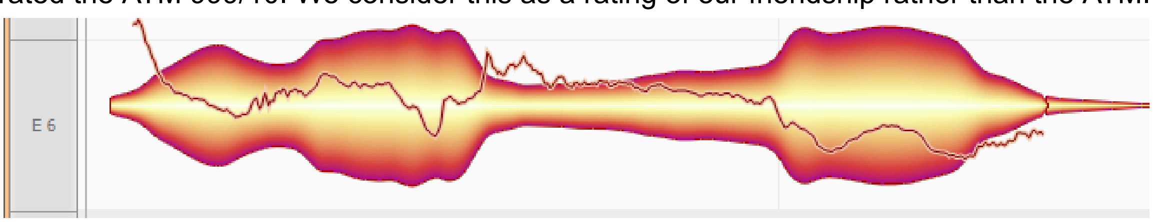


Fig. 55. Unreadable frequency from participant 2

Participant 3

The third participant is a very experienced flute and ukulele player, with extensive knowledge in music theories. She was interested in playing the saxophone, so we spent roughly 30 more minutes training her on how to play a traditional saxophone. Thus, she is slightly more experienced than the other three participants in saxophone. With the ATM installedd, she said that layering on the ATM requires more air than playing on a traditional mouthpiece. After only a few minutes, she managed to play many notes and even some simple melodies. Fig 56 shows one of her many stable notes that are further corrected by the ATM. She was surprised and amazed that she could play in tune and commented that the ATM could be the salvation of woodwind beginners. She rated that ATM 10/10.

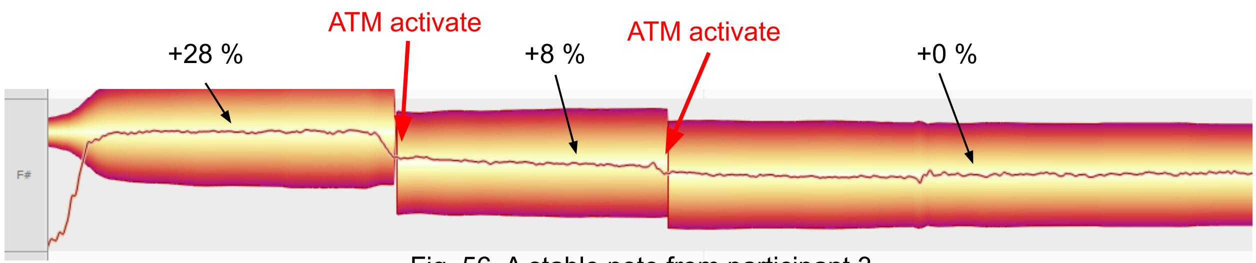


Fig. 56. A stable note from participant 3

Participant 4

The fourth participant has no experience with any musical instruments. He tried for 15 minutes, his mouth muscles were already very tired but still could not play a note on the saxophone with ATM implemented. We gave up. He still rated the ATM 9/10, because it looks cool.