Estimation Of Saxophone Alternate Fingerings' Configurations

by

Marjan Rouhipour

B.Sc., Bahá'í Institute for Higher Education, 2009

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Fall 2015

APPROVAL

Name: Marjan Rouhipour

Degree: Master of Science

Title: Estimation Of Saxophone Alternate Fingerings' Configurations

Examining Committee: Chair: **Joseph Peters**
Professor

Tamara Smyth
Senior Supervisor
Department of Music,
University of California San Diego
Associate Professor

Ghassan Hamarneh
Supervisor
Computing Science,
Simon Fraser University
Professor

Arthur Kirkpatrick
Internal Examiner
Computing Science, Simon Fraser University
Associate Professor

Date Defended: September 4, 2015

Abstract

A method to estimate the saxophone fingerings' configurations is presented. This is not solely a problem of pitch detection since a sounding pitch can be produced by multiple fingering configurations. The widely used spectral envelope technique in similar problem of speech analysis as a vocal tract impulse response is not sufficient to solve the problem because a player can alter the envelope by controlling the reed vibration through embouchure and blowing pressure. Motivated by Saxophone playing technique, a theoretical model is proposed based on sub harmonics' and main harmonics' positions to identify fingerings by measuring the similarity of the input sound spectrum peaks' positions and models. The proposed method has achieved the accuracy of 73% on a dataset of *Bb* tenor saxophone recordings, the result has been improved to 87% with combination of pitch as a complementary feature.

Keywords: Saxophone fingering's configuration; sub harmonics; main harmonics; spectral envelope; source/filter model; transfer function .

To my parents

"Reason is powerless in the expression of Love."

— Mawlana Jalal-al-Din Rumi, Persian poet, *13th-century*

Acknowledgements

Foremost, I want to express my sincere gratitude toward my senior supervisor, Dr. Tamra Smyth, for her support, encouragement, and advice. Her expertise into the both areas of signal processing and music has been a great privilege during my study. Her enthusiasm and ideas in computer music have motivated me in my research, it was my honor to work under her supervision.

I would like to appreciate Joel Miller, the musician, and my senior supervisor who recorded and gathered the saxophone dataset for this research. My sincere thanks goes to my thesis committee, Dr. Ghassan Hamarneh, Dr. Arthur Kirkpatrick, and Dr. Joseph Peters for their feedback and insightful comments.

I would like to thank my parents and sisters for their love and encouragement to follow my dreams.

Contents

List of Tables

List of Figures

Glossary

cut off frequency Cut off frequency of a filter is a region of frequency response where the energy flow is being attenuated. An Array of open toneholes of a saxophone fingerings works as high pass filter and radiates high frequency, therefore the impedance spectrum is irregular and the peaks are weaker above the cut off frequency of the filter [5]. The sound spectrum also shows fall in harmonics above this frequency [5] . ix, 26, 28, 30, 36, 37, 47, 50

dynamic level Dynamic level refers to the volume of a note or sound. xiv, 17, 19, 23, 24, 43, 44, 47, 49–51

fingering A combination of open and close toneholes of a wind instrument such as saxophone is called a fingering configuration. Alternate fingerings refer to toneholes configurations that can produce the same sounding pitch. iii, ix–xiii, 2–13, 15–32, 35, 40–42, 45–51, 57, 59

impedance Acoustical impedance, Z, is the ratio of acoustic pressure, P, to acoustic volume flow V. Its International System Unit is $Pa \cdot s/m^3$. xi, xiii, 16, 42, 43, 45

overblowing Overblowing is a playing technique that uses lower pitch fingerings to produce higher sounding pitch through changing the input air flow inside the instrument's bore by alteration of input pressure and embouchure. xii, 11, 21–23, 27, 48, 51

register key suppression of the register key of saxophone by left hand palm key automatically opens a small register hole on the neck or bore of saxophone. For notes between $D5$ to $G\sharp5$, the hole opens on the bore and above $G\sharp5$ the hole opens on the neck. ix, xi–xiv, 11, 12, 21–23, 25, 31–34, 37, 38, 40–46, 48

sub harmonic Low amplitude harmonics between main harmonics of a sound spectrum which are associated with the resonances of the underlying fingering's transfer function of a saxophone are called sub harmonics. iii, xii–xiv, 11, 21–26, 28, 31–36, 38–40, 42–46, 49–51

transfer function In this study, transfer function is a filter that is described the characteristics of an instrument. When the saxophone is excited by a signal, the outputs which is tapped at the mouthpiece and the bell are called transfer functions of saxophone. xi, xii, 9, 11, 14–20, 22, 26–28, 48, 50, 51

Chapter 1

Introduction

The advancement of computer technology has provided musicians with an opportunity to extend musical instruments beyond their traditional use. Tracking the parameters of a musician performance allows him/her to interface with a computer while they are playing their acoustic instrument. Control parameters can then be remapped to synthesizers or processing algorithms that extend and/or enhance the instrument's produced sound. Though some interactive systems are capable of tracking certain parameters of a musician's control of their instrument, e.g. pitch, amplitude envelopes, and attack that are general to all instruments, challenges remain in tracking parameters specific to any one particular instrument, a task that often requires specific knowledge of that instrument. This research aims to estimate the tone hole configurations (fingerings) applied by a saxophonist, during a performance, using only a signal recording at the bell, with no other sensors. The solution is based on the study of saxophone acoustics, and a database of saxophone note recordings played with different fingerings and techniques. Estimation of saxophone fingering requires an understanding of playing techniques and acoustics. Since a specific pitch may be produced by a variety of fingering configurations and techniques involving blowing pressure and embouchure [4], fingering estimation is not solely a problem of pitch detection.

Separated from any particular instrument, the musicians' gesture is tracked through an interface and then remapped to control parameters of an interactive system as it is shown in Fig. 1.1. In case of a flute player gesture such as blowing pressure and thumb pressure can be identified through an interface, and then mapped to control parameter of a synthesizer to produce an interactive flanging effect [6]. This research focuses on gesture identification

to recognize the fingering configurations of saxophone and leave the rest, including mapping gesture to controlling parameters and designing an interactive systems for application of this research.

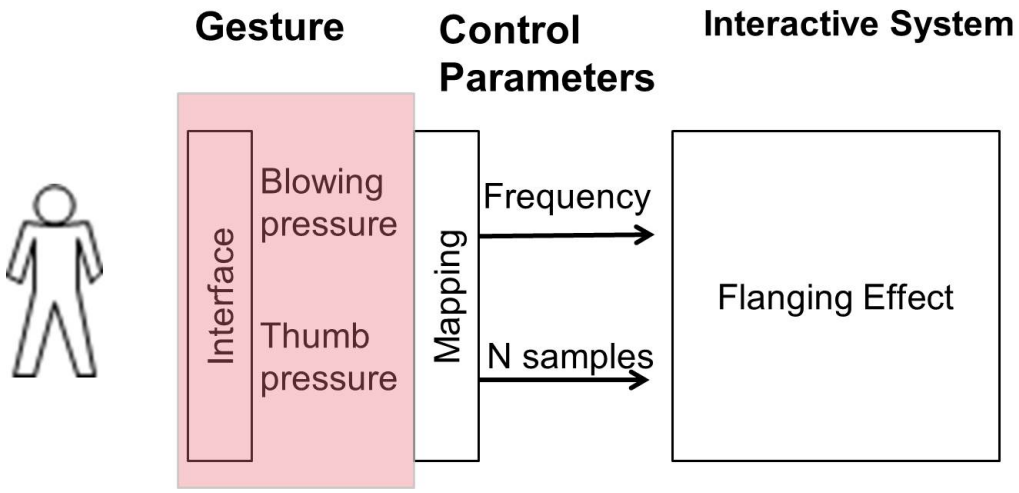

Figure 1.1: Connection between musicians and interactive systems. Musicians connect to an interactive systems through an interface, and then interactive system identify the performer's gesture and map it to parameters of interactive systems such as a synthesizer.

In related works, there are three main categories of interfaces that musicians use to interact with an interactive systems:

1. New Interface/Controllers
2. Extending Traditional Instrument
3. Signal Analysis/Synthesis

In the following sections, we will review these interfaces and bring successful examples of each to see how gesture can be tracked. We compare these interfaces and see why the signal analysis/synthesis is the most beneficial method for saxophonist, though it needs complex signal processing algorithms.

1.1 New Interface/Controllers

In this method, musicians' gestures are tracked by an electrical device which is equipped with multiple sensors. This electronic device or controller can have one of the following appearances:

- Traditional instrument's appearance
- Existing electronic device (such as tablets)
- New interface

These interfaces are popular for different reasons, familiar appearance, low cost, entertainment, and availability. Here, we are going to give examples of each.

Traditional Instrument Appearance

These interfaces have been inspired by the traditional instrument shape, therefore trained musicians benefit from the familiar interface. An example is the Vbow [7], virtual violin, to interface with a bowed string model. The Vbow's body has a violin shape with installed servomotors and encoders to track continual bow gestures and the length of virtual strings, then map the encoder's count to a violin's physical model parameters. The ability to continuously capture the kinematics of the bow by servomotor is a huge advantage for compositional purposes since the acoustic violin dose not have this capability by itself. Another example is the LMA flute [8] which has a flute-like interface with a mouthpiece attached to a pipe with multiple tone holes and key pads. A microphone inside the mouthpiece measures the blowing pressure and magnet-hall sensors on tone holes, and pads measure the output voltage which determines the fingering positions (open and closed tone holes), and then measured values are mapped to the input parameters of a flute synthesizer. The advanced LMA design even considers the key pad noise, and works as a controller in the absence of blowing pressure which provides extended controlling options, however some parameters such as the player's lip position and pressure are difficult to measure and therefore are not supported by LMA. The magnet-hall sensors can be used on the tone holes and key pads of an acoustical saxophone to track the fingerings' gestures. But wiring and sensor installation are not desirable. We will cover the augmented instrument separately in Extension of Traditional Instrument (Section 1.2), the second interface method.

Existing Interface

Usually a software program is developed for an existing device, such as tablets or smart phones, to produce music. The user interfaces are the built-in sensors and input devices such as keyboard, touch screen, microphone, temperature and light sensors, accelerometer

and gyroscope (for orientation and rotation). MadPad by SMULE is one of the interesting applications when users can record sounds in their everyday life by built-in microphones of their tablet or cellphone, then sounds are shown on a touch screen as grid table, and users can produce music by tapping on grid cells. Another type are virtual instruments such as a virtual piano [1] with a website interface and user plays it through the computer's keyboard. The new trend is making social music networks that connect musicians together. For example Singing Guitar applicationn by SMULE provides an online stage to connect guitar players with singers. A guitar player can invite a singer in this online social network to play and sing together. These interface devices are affordable and entertaining, and since the user interfaces are known, the gesture identification is straight forward, more creativity happens at mapping parameters to an interactive system. The advantages of these interfaces is assembled sensors and hardware which make the device ready to use. These electronic devices are great assets to musicians when they are programmed for tracking movements through an accelerometer and gyroscope sensors, then the tracking motions map to control parameters of a synthesizer. However, in case of Saxophone fingerings identification, these devices are big and difficult to carry while playing and they also are not accurate enough to track slight movement.

New Device

Some other devices go beyond conventional design to engage and entertain multiple players such as Jam-o-Drum [9] which has a tangible interface platform, inspired by board games to bring four players around a square screen table with four tangible spinners at each edge of the table. It provides a platform for game design to make an improvisational musical performance based on the play and rotation of each spinner. Besides tangible interfaces, wearable devices also are very popular for performance and voice synthesis. One of the successful wearable devices which also benefits speech and hearing impaired people is Gloves Talk [10] which synthesizes singing and speech by capturing hand gestures through a variety of input devices such as Cyberglove, a Contact Glove, a polhemus sensor, and a foot-pedal. The values from the input devices feed into a neural network to learn the input parameters of speech synthesizer. Enable Talk [2] is another wearable smart device which translates sign

[1] http://virtualpiano.net

[2] http://enabletalk.com

language to speech by tracking the fingerings and hand movements through contact and flex sensors attached to a glove. Wearable device can be used by dancers and musicians on stage where their hand movements can be tracked. A good example of this category of device is *Sound On Intuition* by Pieter-Jan Pieters that captures human body signals such as tapping fingers or foot movements by attaching devices and sensors to produce an effect on sound or music. For saxophone, cyber gloves seem useful to track the movement of fingerings, but wearing a glove limits finger movements, therefore it is also not easy to play the saxophone while wearing gloves.

While these electrical interfaces have enhanced human interaction with computers, musical interfaces have evolved at the fast pace of computers, often before their musical capabilities have been fully explored [11]. Moreover, musicians may lose their audience by playing an unfamiliar interface device. For example, the hand gestures of violinists make an impression on the audience, while using an unfamiliar interface to synthesize the violin may confuse and distract the audience [11]. Although these interfaces introduce new capabilities and new interaction interface between musicians and computer systems, lacking acoustical instruments' features and the learning curve of a new device are huge disadvantages. Musicians prefer to use their own instrument as an interface to interact with computer systems, since it is the most familiar interface for them and their audience. In the next two sections, we will see how the sensors and electrical devices, and signal processing can capture musicians' gestures while they play their acoustical instrument.

1.2 Extending Traditional Instrument (Augmented Instrument)

In contrast to sensor electrical instruments, gradual development of traditional musical instruments with computer and electronic technology gives musicians an opportunity to explore the playing techniques of an instrument and appraise the potential capability of their instrument to communicate with an interactive system [11]. This category is similar to the electrical instrument with the acoustical resemblance in Section 1.1 but the augmented conventional instruments is used for capturing the player's gesture instead of the electronic instrument. For this category, the choice of sensors is very important to be small enough

to not disturb the sound production of the instrument, be sensitive enough to capture the flow of the gestures and also have low latency for real-time application [6], while there is always a trade off between sensor sensitivity and price. A good example is McGill Air-Jet [6] to extend the traditional flute with a flanging effect synthesizer which passes the flute signal thorough the comb filter and combine the output of the comb filter with a delayed version of itself to produce flanging effect. A player can adjust the frequency sweep of the comb filter by changing the air pressure which is measured by pressure-based sensors at the mouthpiece and set the delay line length by pressure on the position of the right hand thumb through Force Sensitive Resistor (FSR).

Cameras are also used widely to capture the gestures and computer vision algorithms are applied for analysis. There is a trade off between sensitivity and cost, a cost-effective camera such as kinect has limited accuracy, resolution, and sensitivity, while the VICON camera with millimeters accuracy tracks all joins movements and is much more expensive. Besides the cost, the main draw back of cameras is a limitation of the players' actions while the musicians should be in front of camera. For saxophone fingerings' identification, more than one camera is needed to track all fingerings' gesture from different sides since tone holes' positions are at different position, though some fingerings which are facing the musicians or back of instrument are difficult to track by a camera, and therefore this method cannot be applied to saxophone fingering's identification. In some cases a low cost camera is used in combination with other sensors, for example an accelerometer attached to the violin bow to track direction and speed of the bow, but it cannot track the static and dynamics of the bow stroke. In this case an affordable camera is used to recover the accelerometer signal in [12]. In this case, the latency of the camera is a barrier for real-time application, however it is useful for any system that does not need real-time feedback. Although computer vision algorithms are advanced but sensitive cameras are very expensive to track the small movement of player in real-time.

The saxophone also can be equipped with the sensors to track the opening and closing tone holes using magnet-hall sensors attached to the keypads and tone holes. In case of Metasax [13], a tenor saxophone is equipped with FSR sensors on keypads to track the pressure and distance of the keypad with tone holes, each keypad triggers an effect like

a distortion, reverberation, and frequency modulation. An accelerometer IC chip on the bell also captures the position of the bell in three dimensional space, and multiple microphones are designed inside the saxophone for electro acoustic features such as multiplying channels. Clearly, augmented instruments are a prefect interface for musicians to interact with computer software for the long established practice of traditional instruments, but the wiring is undesirable for expensive instruments, needs technical experience, and restricts the players actions. For the drawbacks of augmented instruments, we are going to estimate the fingerings by the sound of an unaugmented saxophone recording at the bell using signal processing/analysis.

1.3 Signal Analysis/Synthesis

We have seen that augmented instruments provide extended capabilities for musicians, however the wiring and sensor installation is not desirable, even using a camera limits the movements of the player. The third method of gesture identification is processing the output sound of the instrument recording by a microphone. This approach is affordable since it only needs a microphone which is the case for a live performance, but the analysis of the signal and estimation of the gestures is more difficult. In the case of the saxophone, usually a microphone is attached to the bell which also gives the musicians the ability to move around. With woodwind instruments the volume flow of the reed and reed pulse include the player's gestures such as blowing pressure, lip vibration, and lip pressure on the reed. In a study on clarinets [14], an inverse model of the clarinet is introduced by physical modeling to extract the reed pulse which includes gesture information. In another research [15] the player's embouchure gesture extracted by applying the inverse model of the mouthpiece on the mouthpiece volume flow. The physical parameters of the model include effective stiffness of the reed, the effective reed surface, and the distance of the reed and mouthpiece and player's lip pressure on the reed, and are directly related to embouchure adjustment.

A very similar problem to fingering identification is in the area of speech/synthesis, which also can be seen as a system identification problem, whereby vocal tract frequency response is estimated from the spectral envelope of speech by a forward source-filter model

in which the filter represents the transfer functions of vocal tract shape [16, 17]. We will discuss in Chapter 2 that the source-filter model is not practical for the saxophone since the backward air propagation toward the mouthpiece contributes to the sound production, while massive glottal valve backward air propagation is negligible.

Other researchers have tried to estimate the gestures by comparing a database of synthesized sound with input sound. A model is usually built and the sound are synthesized by applying possible range of input parameters which associates to the player gestures. In a research on speech [16], a database of synthesized sounds is created by a cylindrical physical model of the vocal tract with a possible range of radius of cylinders. The most similar or likely model for an input sound and consequently associate vocal tract shape is selected. However, the higher resolutions for radius and size of dataset are barriers for real-time application. Another study uses the physical modeling on the saxophone for each fingerings configuration of the saxophone attached to reed model [3] [4]. The reed model parameter for each fingering is estimated by convex optimization on sound files of fingerings, then these parameters are used to synthesize the sound for each fingerings. The similarity between an input sound and a synthesized sound file is measured by Euclid's distance of waveforms. The result was promising since it works well on a dataset of recording data on low frequency fingerings configuration. However, comparison of waveform is computationally expensive. Our research covers wider frequency range (three octaves) and mainly focus on spectral features of sound.

1.4 Contribution

In this section, we have reviewed how the players' gestures are estimated based on three methods: new electronic device, augmented traditional instruments, and audio signal processing/synthesis of sound. Since the approach of this research is for a saxophone player we are not interested in the design of an electronic device. Augmented instruments are good solutions, where multiple sensors and/or camera tracking analysis the players' gestures. We have reviewed some augmented instruments such as MetaSax where the sensors track pressure on the holes, open and close tone holes, and even the movement of saxophone body.

However, wiring and tracing malfunction sensors needs technical knowledge, and usually limits players' actions. We have decided to use the audio signal processing approach to estimate fingerings configurations of saxophone since a microphone is usually attached at the bell in live performances. We continue this discussion in Chapter 2 with more details about fingerings of the saxophone, and the relation between pitch and fingerings, alternate fingerings, and playing techniques. We will review the similar problem of the voice vocal tract by feed-forward source-filter and discuss the saxophone source-filter model, and how it differs from speech source-filter model.

In Chapter 3, a theoretical model is introduced to solve the fingering identification problem. We will discuss the extendability of the model and the parameter of spectrum analysis, such as window type and size. Then the peak detection method and preprocessing and post processing steps are explained. In Chapter 4, the conclusion and future work are presented.

Chapter 2

Problem Statement

The aim of this research is to estimate the configuration of open and closed tone holes, applied to a saxophone by analysing its produced sound recorded at the bell. This is not solely a more general problem of pitch detection, since multiple fingerings can produce the same sounding pitch by a variety of playing techniques, e.g. alternate fingerings and bugling (overblowing) with/without the register key which fingerings is chosen by the player will depend on the musical/technical demands of a musical process such as faster playing and switching between notes. Consider the note $F\sharp 5$ that can be produced with four fingerings, as shown in Fig. 2.1: the spectra of these fingerings (in Fig. 2.2) show high amplitude peaks that are evenly spaced by the fundamental frequency, 300 Hz, corresponding to the sounding frequency. Though the sounding frequency and pitch are the same, the spectra are otherwise noticeably different. The low amplitude peaks between main peaks are called sub harmonics and related to the underlying fingering's transfer function (which will be discussed in more detail later in this chapter). In this chapter, the focus is the spectral envelope and source/filter model, and how the saxophone source/filter model is different and influences the spectrum. At the end we talk about overblowing technique and sub harmonics.

We introduce a notation to distinguish between fingering and pitch, which is shown in Table 2.1. A combination of pitch and name is shown by *pitchName/fingering*, for example $C5$ with three fingerings are shown as: $C5/C5$, $C5/sideC$, and overblown $C5/C4$. When the fingering is complex, it is put in parenthesis, for example overblown $Bb5/(Bb4/Bis)$ key means $Bb4$ fingering using Bis is overblown to produce the note $Bb5$.

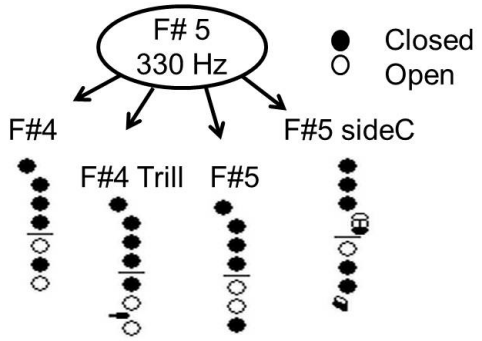

Figure 2.1: Alternate fingerings of $F\sharp 5$: the first two fingerings,$F\sharp 4$ and $F\sharp 4$ Trill, need the register key to play higher octave sounding pitch $F\sharp 5$ [1][2].

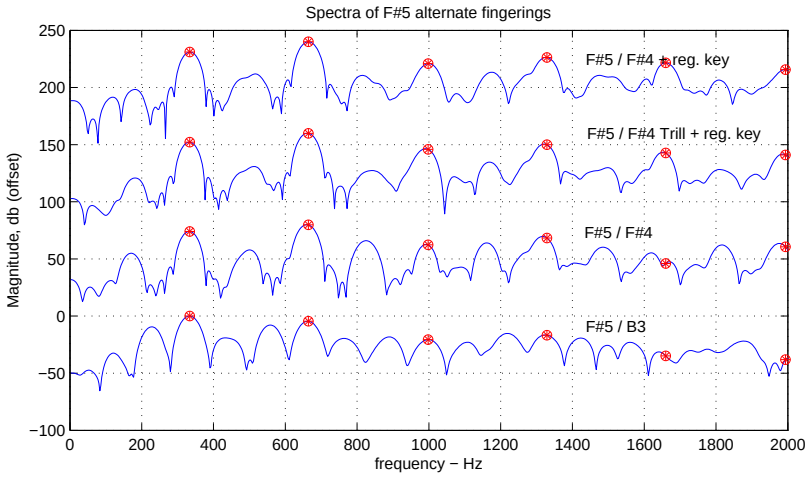

Figure 2.2: Spectra of $F\sharp 5$ alternate fingerings. Evenly spaced high peaks, red marks, demonstrate the pitch $F\sharp 5$ (330 Hz).

Description	Notation	Example
pitch	pitch name	$C6$
fingering	X fingering	$C5$ fingering
Pitch with fingering	pitchName/fingering	$C6/C5$
Pitch ($f_1 Hz$) with fingering ($f_2 Hz$)	pitchName/fingering (f_1/f_2)	$C6/C5$ ($466/233 Hz$)
Pitch ($f_1 Hz$) with fingering	pitchName/fingering (f_1)	$C6/C5$ ($466 Hz$)

Table 2.1: The pitch and fingering notation.

2.1 Spectral Envelope

Since pitch is not solely useful to solve the fingerings estimation problem, we study similar problems and their approaches. As mentioned in Chapter 1, the problem of the saxophone is very similar to the estimation of speech parameters, such as vocal tract shape, which has been researched extensively across disciplines even outside of computer music. Speech is often simulated as a source/filter model, where the source represents the input pressure signal from vibrating vocal cords that periodically open and close an aperture to the vocal tract. The vocal tract, in turn, filters the typically broadband source, giving it spectral characteristics, such as those that allow for distinction of vowel sounds. The spectra of vowel sounds /a/ and /e/ with the same sounding frequency (150 Hz) have different resonances and envelopes (Fig. 2.3) which are result of two different vocal tract shapes.

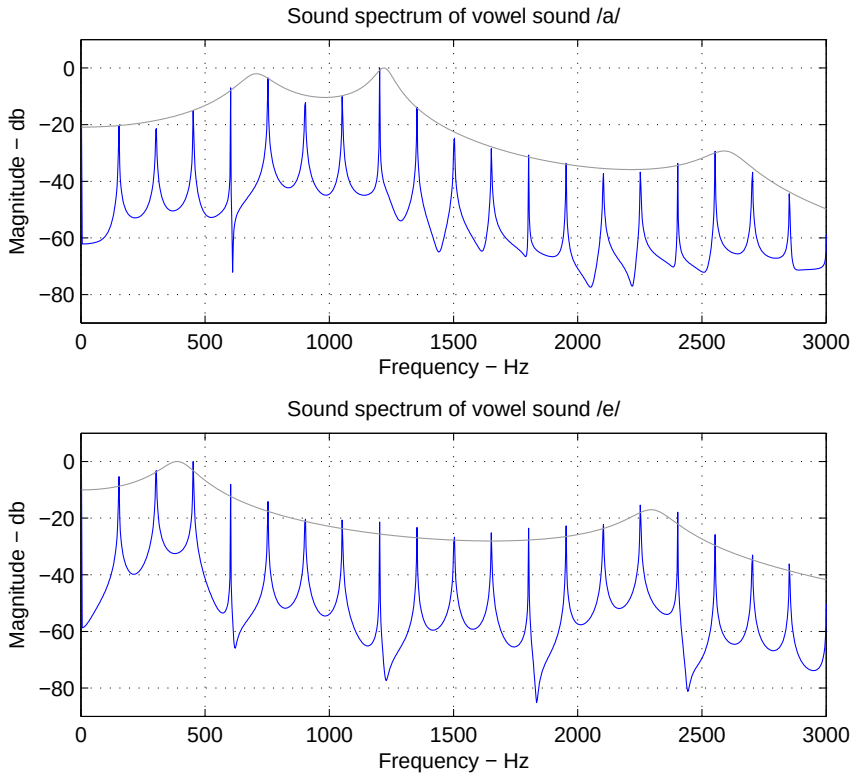

Figure 2.3: Comparison between two vowel sounds spectra and their envelopes: the vowel sound /a/ (top) and the vowel sound /e/ (bottom) have different spectral resonances but both have the same sounding frequency, 150 Hz.

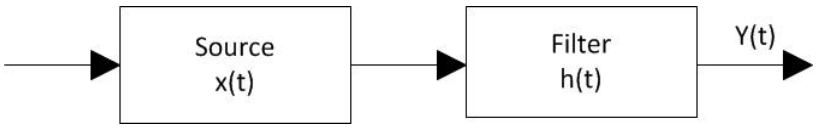

Figure 2.4: Simple Source Filter Model.

In a simple source-filter model, the source $x(t)$ consists of a periodic pulse train that excites a filter (vocal tract) having impulse response $h(t)$, producing a signal that is transmitted as speech $y(t)$. This process is depicted in the time domain in Fig 2.4, and is defined by the convolution of source and filter [18, 19] :

$$y(t) = (x * h)(t), \tag{2.1}$$

This may also be viewed in the frequency domain as a product between frequency responses of the source and filter spectra, defined as:

$$Y(\omega) = X(\omega)H(\omega). \tag{2.2}$$

From this, we see that the spectrum of the final sound, $Y(\omega)$, has contributions of both source $X(\omega)$ and filter $H(\omega)$. But typically, in speech, because the spectrum of the source is broadband and flat (as shown in Fig. 2.5), characteristics in $Y(\omega)$ are seen to be mostly due to $H(\omega)$. For example, a synthesized vowel sound $/a/$ having spectrum $Y(\omega)$, on the bottom of Fig. 2.5, is composed of the glottal pulse spectrum, $X(\omega)$, at 150 Hz (top), and the transfer function of the vocal tract, $H(\omega)$, (middle) which impacts the resonances seen in the spectral envelope, $Y(\omega)$, (bottom), [20].

Because the characteristic of envelope in the output $Y(\omega)$ are so dependent on $H(\omega)$, researchers have estimated the filter $H(\omega)$, from which the corresponding vocal tract shape may also be estimated, by analyzing the spectral envelope and resonance in the spectrum of the output sound. Usually, the vocal tract simulates as a n^{th} order all-pole filter, with poles associated with vocal tract resonances [18, 19, 21, 22, 23, 24]. To adjust the order and parameter of the all-pole filter, Linear Predictive (LP) and Auto Regression (AR) have widely been used. Some studies have utilized optimization techniques to explore the parameters [18, 19]. Another study introduced a piecewise physical model instead of the all-pole filter and compared the spectrum resonances of a sound with a database of synthesized sounds

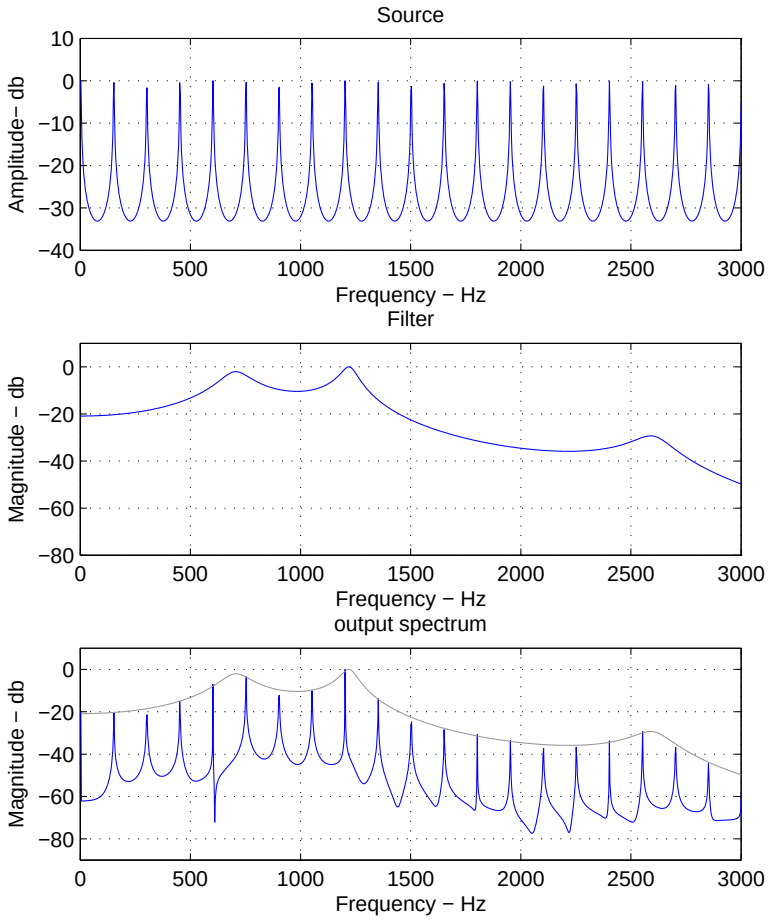

Figure 2.5: Source/filter model of a synthesized vowel sound /a/: pulse train of source (top), the transfer function of glottal shape for vowel sound /a/ with three resonances (middle), spectrum of vowel sound /a/.

[16]. Although estimation of the filter from spectral envelope is useful for speech, it is less helpful for identifying the filter in a saxophone signal.

The sound produced by a saxophone may also be modeled as a source/filter. The opening and closing reed produce a train of pulses the same way as the opening and closing glottal valve. And in the same way that the vocal tract dictates the filter $H(\omega)$, so does a particular fingering produce a filter $H(\omega)$. However, the source/filter model assumes a weak coupling between the source and the filter which is not true for the saxophone, since its reed is less massive and more strongly influenced by the resonances of the bore and bell.

The source and filter modeling the voice (shown in Fig. 2.4) are loosely coupled since the backward air propagation through glottal valve has less significant influence on massive vocal cords, while the saxophone reed is less massive and moves more freely by feedback air propagation in the saxophone body. Moreover, autonomous vibration of the reed without locking to resonances of the bore is possible in some more advanced techniques which alters the spectral envelope resonances. The enhanced source/filter model of saxophone is shown in Fig. 2.6 where $h_b(t)$ and $h_m(t)$ are impulse responses of transfer functions with input at the mouthpiece and output at the mouthpiece and bell, respectively. In this model, the source output, p_r, depends on the output of the filter at the mouthpiece, $y_m(t)$. Hence, the spectral envelope depends on reed vibration and pressure, transfer functions at the bell and mouthpiece.

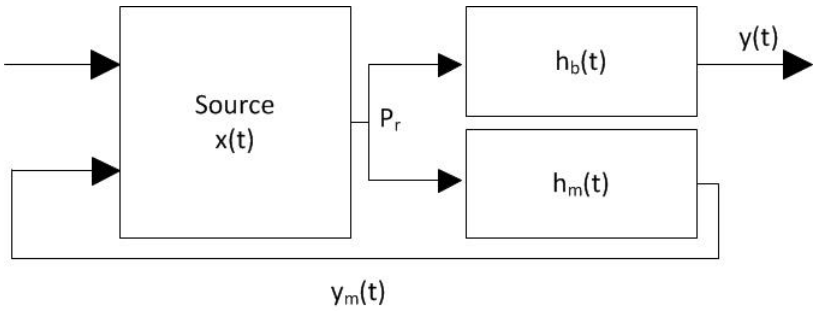

Figure 2.6: Saxophone Source Filter Model. $h_m(t)$ is impulse response of the input impedance, $h_b(t)$ refers to impulse response of transfer function at the bell [3]

.

The spectral envelope also depends on the blowing pressure and reed vibration which are controlled by the player to produce different dynamic levels sounds: *mezzo forte* (moderately loud), *fortissimo* (very hard), and *pianissimo* (very soft). Fig. 2.7 shows spectra of $C\sharp 5/C\sharp$ (247 Hz) for three dynamic levels, each has its own resonance and envelope. Although all of these dynamic levels share a resonance at their fifth harmonics, each one has their own resonances.

By studying the spectral envelope of some fingerings, we have found that some fingerings have similar transfer functions of $h_b(t)$ and $h_m(t)$ and may not be recognizable. For example the transfer function of $Bb5$ for Bis and $SideBb$ fingerings, which are measured at the mouthpiece, are exactly the same in Fig. 2.8. This similarity results in similar spectral

envelope for $Bb5/Bis$ and $Bb5/SideBb$ at mf dynamic level, which are depicted in Fig. 2.9. The evenly spaced harmonics of both notes correspond to the pitch of $Bb5$ ($208\ Hz$) and line up with transfer functions' resonances in Fig. 2.8. The resonance at fifth, ninth and thirteenth harmonics of spectrum, Fig. 2.9, (marked as red circles) of both fingerings are associated with sharp and high resonance of their transfer functions at those frequencies. As a result, fingerings with similar transfer function can have similar spectral envelope and therefore the spectral envelope is not enough to distinguish these cases, and better features should be considered.

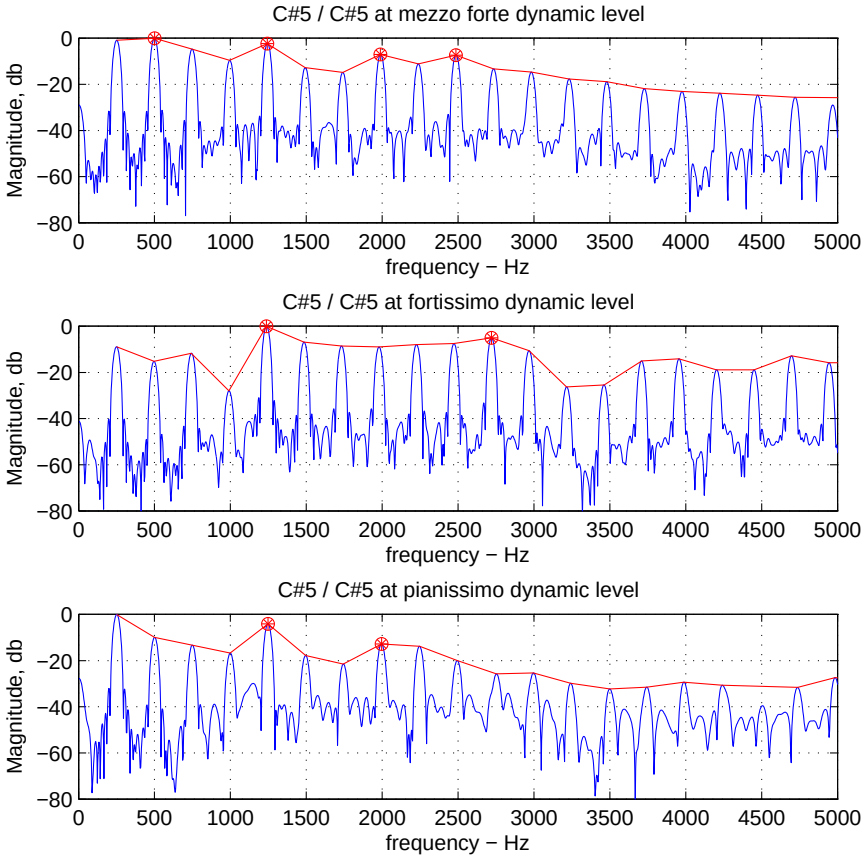

Figure 2.7: Sound spectra of $C\sharp 5$ for three dynamic levels: mezzo-forte (top), fortissimo (middle), pianissimo (bottom).

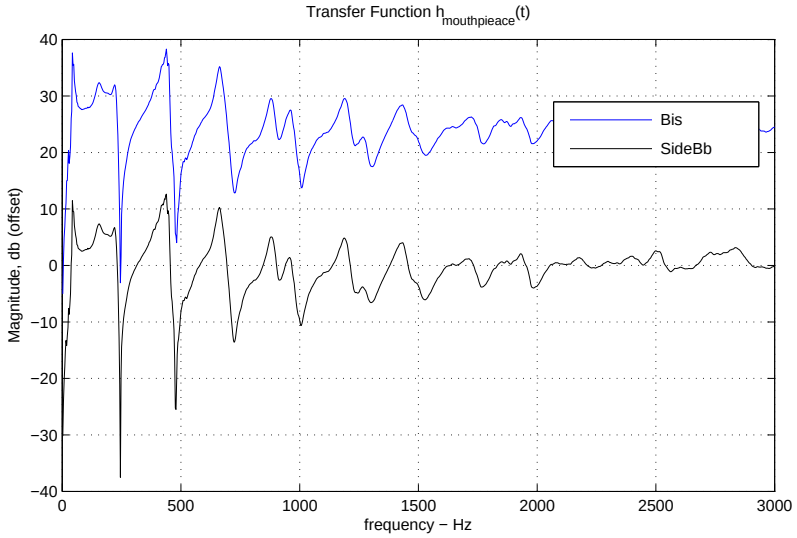

Figure 2.8: transfer functions of $Bb5$ with Bis and $SideBb$ fingerings which are measured at the mouthpiece [4].

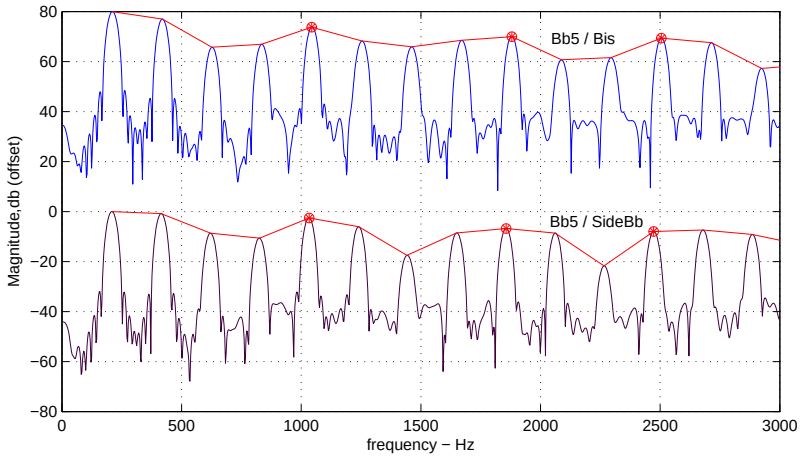

Figure 2.9: Sound spectra of $Bb5$ alternate fingerings . Both spectra have evenly spaced peaks associate with the $Bb5$ (208 Hz), and the resonance at the fifth, ninth and thirteenth harmonics which reflect the transfer functions' resonance at those positions.

On the other hand, an alternate fingering may produce a $Y(\omega)$ with different spectral envelopes. For example $C5$, with a sounding pitch of 233 Hz, with two alternate fingerings, $C5$ and *sideC*, which are played at the *mezzo forte* dynamic level, are shown in Fig. 2.10. Evenly spaced peaks of $C5$ fingering(top) and *sideC* fingering(bottom) line up and correspond to the pitch of $C5$ (233 Hz). Although the pitch is the same, $C5/C5$ (top) spectrum has a resonance around the third peak, which indicates the resonance of the transfer function of $C5$ fingering at that frequency region, while the *sideC* does not have such resonance in its transfer function as shown in Fig. 2.11.

As an initial approach we considered using spectral envelope as the first feature to filter some fingering candidates. Since the dynamic level is continuous metric, labeling data for each dynamic level is not easy task and thus classification method that requires labeling data is not applicable. We used K-mean clustering technique to partitions the examples to K classes with an average center. The K classes are initialized by a random example, and a new example is assigned to closest cluster based on distance of its spectral envelope from the cluster mean (e.g. center), then the cluster mean is updated. In prediction step, the input is assigned to the closest cluster center. The machine learning approach requires large dataset of recordings at different dynamic levels, since the accuracy of the final model or clusters depends on training data. We are looking for features that identify the fingerings with lowest overlap and not do not depend on huge data set for training. For this reason we are going to study the acoustic and playing techniques of saxophone to introduce a theoretical model that is easier to generalize for all players,and for all dynamic level while machine learning is so dependent on training data.

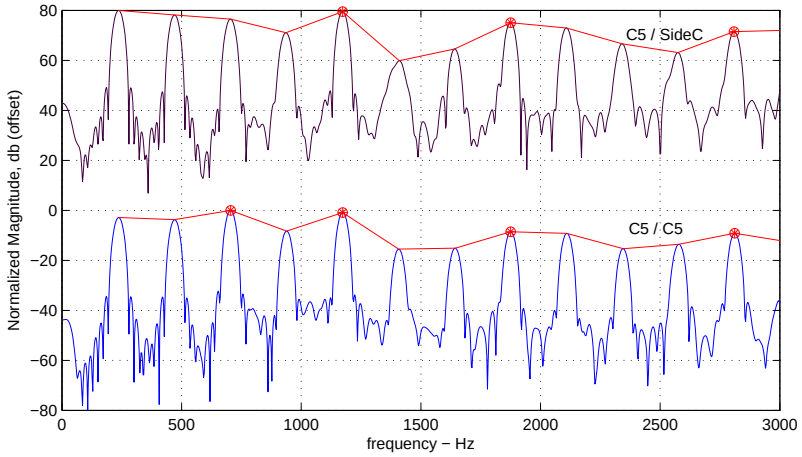

Figure 2.10: Sound spectra of $C5$ alternate fingerings: $C5/C5$ fingering (top) and $C5/sideC$ fingering(bottom). Both spectra have evenly spaced peaks correspond to the $C5$ (233 Hz), but the third peak of *C5/C5* (top) has a higher amplitude which corresponds to the transfer function resonance at that frequency region.

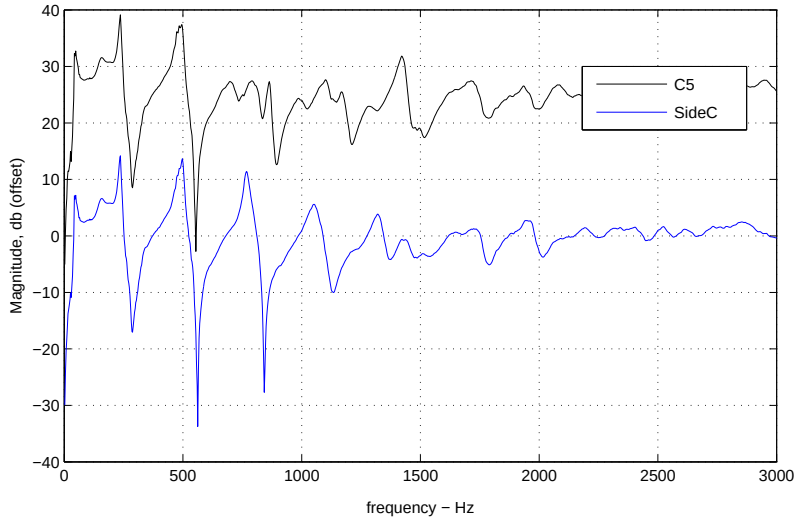

Figure 2.11: transfer functions of $C5$ with $C5$ and $SideC$ fingerings at the mouthpiece [4].

2.2 Overblowing

overblowing technique use low pitch fingerings and changes the input air flow of the bore through alteration of blowing pressure and embouchure [4]. For instance, the note $F5$ with a sounding pitch of 311 Hz can be played by the overblowing of two lower pitch fingerings: $Bb3$ (104 Hz) and $F4$ (156 Hz). In Fig. 2.12 every harmonics of the $F5$ sound spectrum, on the top, lines up with every third peak of $Bb3$ and every second peak of $F4$ sound spectrum which are marked as red circles, on the bottom.

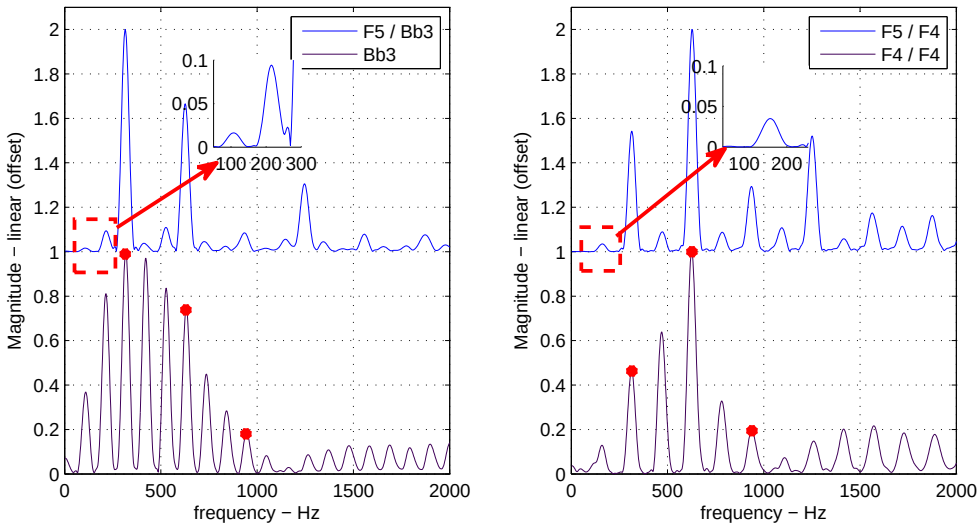

Figure 2.12: Sound spectra of overblown $F5/Bb3$ (left) and overblown $F5/F4$ (right). Every third harmonics of $Bb3$ (left bottom) and every second harmonics of $F4$ (right bottom) line up with every harmonics of $F5$ spectrum (top). The zoom in plots show the sub harmonics of lower fingerings that are suppressed by overblowing.

2.2.1 Effect Of The Register Key

To ease overblowing, the register key automatically opens a hole on the neck or body of the saxophone and decreases the amplitude of the fundamental in the transfer function and produces a higher sounding pitch. The register key opens a hole on the body for notes between $D5$ (262 Hz) to $G\sharp5$ (370 Hz) and on the neck for $A5$(391 Hz) to $F6$ (622 Hz). Fig. 2.13 demonstrates the spectrum of an alternate fingering of $F5$ with a sounding pitch of 311 Hz that uses the $F4$ fingerings (156 Hz) with the register key opening a hole on the body. Usually, the register key supports overblowing, however the register key is part of the

$D\sharp 6$ (523 Hz) to $F6$ fingerings which means these fingerings do not count as overblowing notes using the register key.

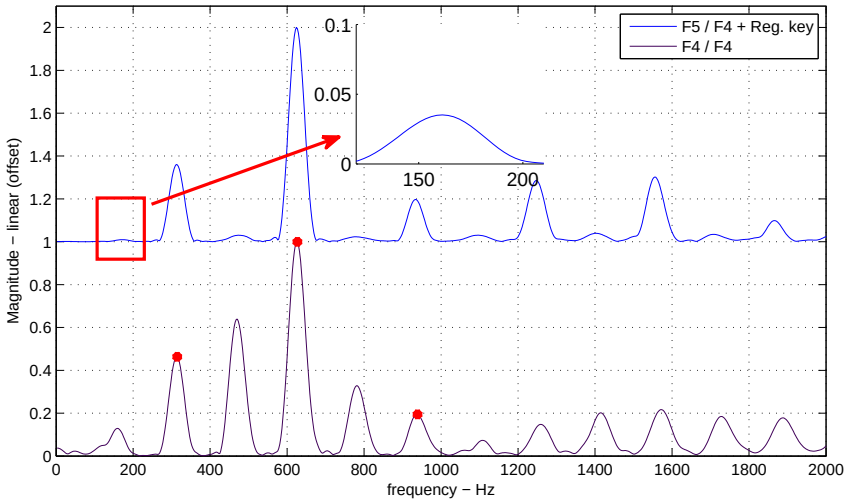

Figure 2.13: Spectrum of overblown $F5/F4$ with the register key. Every peaks of this spectrum correspond to a peak of underlying $F4$ fingering's transfer function at those frequency regions, while high amplitude peaks associate with the sounding pitch of $F5$ (311 Hz) as a result of suppressing even harmonics by the register key. The zoom in plot shows the track of lower octave fingering $F4$ which are suppressed by the register key.

2.2.2 Sub Harmonic

Low amplitude peaks between the main harmonics in the overblown spectra of $F5$, shown on the top of Fig. 2.12 and 2.13, are called sub harmonics which align to the peaks of the underlying fingerings' transfer functions of $F4$ and $Bb3$, the spectrum of these fingerings shown at the bottom of the figures. Generally, all overblown notes have traces of their fingerings transfer function's peaks during the attack and sustain portions of the note.

Comparison between overblown $F5/F4$ with/without the register key shows that the sub harmonics tend to have lower amplitudes (Fig. 2.12 and 2.13) when the register key is activated, it alters the fundamental of the transfer function, and suppresses the sub harmonics to make overblowing easier. The highest of sub harmonics will be studied in more details in Section 3.4.2.

Though the envelope and pitch cannot demonstrate a unique feature of fingerings, sub harmonics' and main harmonics' positions are more or less stable during the attack and sustain portions of a sound. Similar study on flute fingering identification has spotted the sub harmonics to estimate the underlying fingerings [25]. The feature space is defined as the average power spectrum energy of continuous frames of a sound, over small sub band centered at $i.F0/4$ and width of $F0/8$ where $i = 1, 2, 3, 4, 5, 6$, and $F0$ is pitch. The accuracy of 100% has been achieved with applying Principal Component Analysis and Support Vector Machine. However, this method works only at one dynamic level (here *mezzo forte*), for only one and two octave overblown notes, and pitch assumes to be known. An extension of this research estimates the octave and non octave related underlying fingering for overblown notes of the flute at *mezzo forte* and *fortissimo* dynamic levels [26]. The spectral Energy of smaller sub bands are considered to support non octave related underlying fingerings with pitch. The error changes between 1.3% to 13.3%, while more fingerings configurations for a pitch increases error. In both methods for flute fingerings configurations identification, the pitch is known and the features are extracted from he envelope while we have seen spectral envelope changes for different dynamic levels and therefore it limits the study on different dynamic levels. Saxophone fingerings configurations are more complicated than flute since overblowing is possible with and without the register key with the same fingering. We are going to propose a theoretical model for saxophone fingering which support all dynamic levels in next Chapter.

Chapter 3

Estimation Of Saxophone Fingerings By Sub Harmonics

In this chapter we discuss the proposed method for identification of saxophone fingerings. First a theoretical model is proposed, then the distance measurements is introduced. Finally the effect of the register key on the amplitude of sub harmonics will be discussed and experiments will be presented.

3.1 Theoretical Model

In Section 2, it is shown that the sub harmonics of the overblown spectrum relate to the underlying fingerings configurations. For example, the spectrum of overblown note $F5/F4$ ($311\ Hz$), Fig. 2.12, has tracks of $F4$ fingering peaks, called sub harmonics. These sub harmonics' and main peaks' positions are more or less consistent for a note duration and for different dynamic levels, whereas the spectral envelope is not stable during the attack and is altered by the blowing pressure, for example the spectral envelope of $C\sharp5/C\sharp5$ is different for three dynamic levels, depicted in Fig. 2.7. The consistency of peak positions, the frequency at which the peaks lie, over the spectral envelope leads to introducing the peak position model to solve the saxophone fingerings identification problem.

We have also discussed in Chapter 1 why pitch is not solely useful, since multiple saxophone fingering configurations can produce the same sounding pitch. A good example

of such pitch detection is introduced in [27], which directly uses the peak position and amplitude to estimate the likelihood of a sounding pitch at $f\ Hz$, as follow:

$$\mathcal{L}(f) = \sum_{i=1}^{K} a_i t_i n_i, \tag{3.1}$$

where K is the number of spectral peaks, a_i is the amplitude of the peak, t_i is the frequency ratio between the peak position and frequency of $f\ Hz$, and n_i is the harmonic number for the fundamental frequency f. Similarly, saxophone fingering identification can use peak position to estimate fingerings. In the following we are going to introduce a theoretical model based on peak position, then a likelihood function is introduced to map an input sound spectrum to the model.

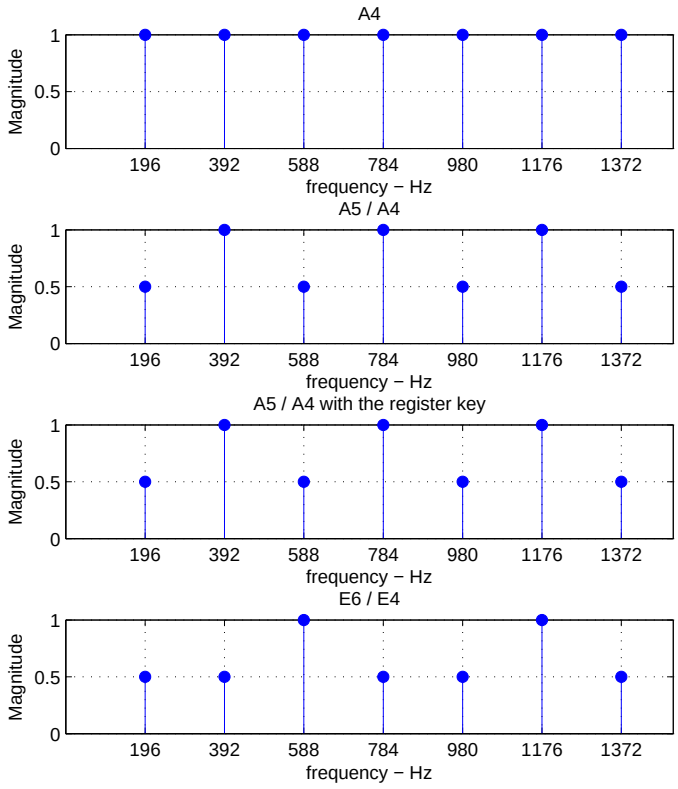

Figure 3.1: Peak position models of $A4$ fingering with three different sounding pitches: $A4/A4$ (196 Hz)(top), $A5/A4$ (391 Hz) with and without the register key (middle), and $E6/F4$ (588 Hz) (bottom). In this Figure, the high amplitude spikes refer to the sounding pitch, and the low amplitude spikes represent sub harmonics.

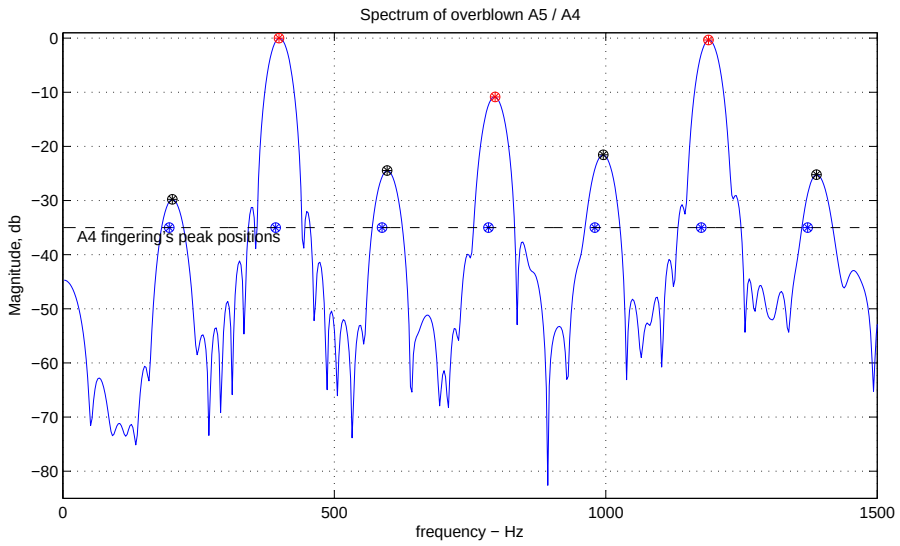

Figure 3.2: The spectrum of $A5/A4$ (391 Hz): The black circles show the sub harmonics, the red circles are the main harmonics, and the blue circles on dashed lines are the peak position of models with $A4$ fingering.

When an underlying fingering is used to produce a sounding pitch, the subharmonic of the underlying fingerings and main harmonics appear in the sound spectrum, and therefore each combination of fingering and pitch model contains the main and sub harmonics of the underlying fingering. In our modeling, the position of the fundamental harmonic is set by the position of the first peak in the sound spectrum and the rest of harmonics' positions are adjusted in relation with it. In Fig. 3.1, four models of $A4$ (196 Hz) fingering are shown where high amplitude peaks are associated with the sounding pitch, and the low amplitude peaks are sub harmonics of $A4$ fingering. The fact that all peaks line up under the peaks of the $A4$ model (top) helps to estimate the underlying $A4$ fingering. A set of 62 models based on known possible combination of fingerings configurations and sounding pitch is introduced, listed in Appendix A.

The peaks in the transfer function of saxophone are not ideally harmonic, and above the cut off frequency of tenor saxophone (1800 Hz) they are more irregular and weaker, and so do the sound spectral peaks that coincide with the peaks of transfer function [5]. Fig. 3.2 shows how peaks drift away from ideal positions in frequency for overblown $A5/A4$ (391 Hz). We consider the range of study around the cut off frequency, 1500 Hz, to have evidence of sub harmonics and main harmonics. The irregularity of harmonics is usually

negligible (except for two octave overblown notes) under 1500 Hz, except for two octave overblown notes, when the player locks to the higher resonance of transfer function. We consider ideal harmonic positions for modeling and try to modify model for two octave overblown notes. However, for better result we recommend to have exact location of peaks based on resonance of fingerings'transfer functions.

When a player uses overblowing technique and locks to a higher resonance of the instrument, the note is not always going to be in-tune because the peaks of transfer function are not ideally harmonic. This effect influences two octave overblown notes the most and some of the examples in our dataset are half a semitone off-tune. Fig. 3.3 shows the spectrum of overblown $E6/E4$ on the top and the spectrum of $E4/E4$ on the bottom as a reference point that plays at the first resonance of E4 fingering transfer function. The first main harmonic of $E6/E4$ lines up with the fourth resonance of $E4/E4$ spectrum , and as a result the peaks locks to that resonance in transfer function and show a shift from ideal position. We expect an expert player adjusts the note and play in-tune, but playing an overblown note is very difficult. To compensate harmonic shift we consider two models for overblown notes. The gray dashed line in Fig. 3.3 represents the second model, where each peak of the note $E4$ is shifted by a factor of $2^{1/48}$ or by a half a semitone.

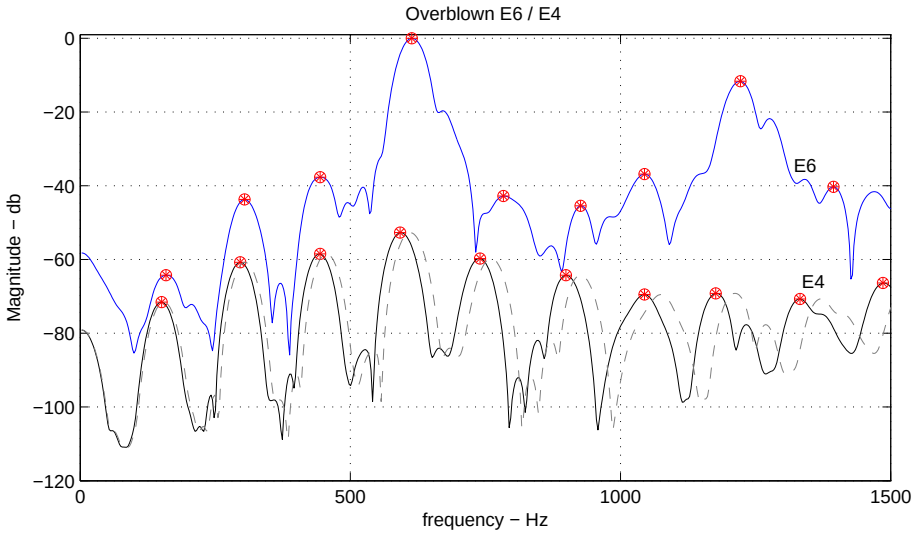

Figure 3.3: Harmonics shift for two octave overblown $E6/E4$

We have seen some peaks that are not harmonically associate to the fundamental harmonic may appear in the spectrum over the cut off frequency range. These peaks correspond to a resonance in transfer function over the cut off frequency. When one or two of such partials constantly appear in the sound spectrum of a fingering, the model has been tweaked and those partials added to the model. So, we talked about two modification of model by considering two models for overblown notes and considering peaks over cut off frequency. We will test the the original model in Section 3.5.1, and the modified models in Section 3.5.2.

This modeling is based on peak position, and cannot distinguish between some models. For example $Bb5/Bis$ and $Bb5$/1&4 fingerings have the same peak position models, only peaks at the main harmonics. Similarly, overblown $Bb6$/(Bb5/Bis) and $Bb6$/(Bb5/1&4) have the same model, peaks at sounding pitch of $B6$ (415 Hz) and sub harmonics at the sounding pitch of $Bb5$ (208 Hz). Identifying these cases will be left for future research and here they are treated the same. In next Section (Section 3.2) we propose a distance algorithm based on peak position to tackle saxophone fingering identification problem.

3.2 Fingering Estimation

Having a database of theoretical models, a distance function in Algo. 1 measures the distance between the sound spectrum and a model, $H_{locs}(i)$ based on the sum of logarithmic distance of the sound spectrum peaks from the closest peaks of each model (inner For loop). Since the musical scale is logarithmic the distance defines as logarithmic distance between two frequency positions to show how close or far they are. The lower the sum of distances is, the more likely it is the right model, thus the fingering of the lowest distance model is selected as the underlying fingering. For example, the spectrum of overblown $A5/A4$ (391 Hz) is shown in Fig. 3.2, where the red filled circles are the main harmonics and the black filled circle are sub harmonics. The most likely models, listed in Table .3.1, all share $A4$ fingering and therefore, $A4$ fingering is chosen as the underlying fingering.

In some cases when two fingerings with exactly one octave frequency difference appear on the top of the most likely fingerings, the higher frequency fingering is selected. For example Fig. 3.4 shows the sound spectrum of $B4/B4$ (220 Hz) with red circles as peaks.

Both $B4$ (220 Hz) and $B3$ (110 Hz) fingerings peak positions, which are marked as blue circles, match with the spectral peaks. However, the $B4$ fingering is the right fingering otherwise the partials of $B3$ fingering should have increased the distance of the $B4$ fingering, as reported in Table 3.2.

function Distance_func (spectrum, H_{locs}, N_{locs})
%spectrum : spectrum of a sound file
%H_{locs} : a list of peak positions' models
%N_{model} : number of models in H_{locs}
  y_{locs} := findpeaks(spectrum);
  N_{pks} := length(y_{locs});
  For i:=1 **to** N_{model} **do**
    Set $Diff[1...N_{pks}]$ to zero;
    For j:=1 **to** N_{pks} **do**
      $Diff(j) = min(abs(1400 \cdot log_2(\frac{y_{locs}(j)}{H_{locs}(i)})))$
    end
    $distance(i) := \sum_{j=1}^{N_{pks}} Diff(j)$
  end
end function

Algorithm 1: This **Distance function** measures the distance between an input sound spectrum and all models. After extraction of sound spectrum peaks y_{locs}, the outer for loop goes over all models, H_{locs}, and inner for loop calculates the distance between peak positions of i_{th} model,$H_{locs}(i)$, and input sound spectrum peaks,y_{locs}.

Pitch / Fingering	Distance
$E6$ / $A4$	604.271234
$A5$ / $A4$ +reg. key	604.271234
$A5$ / $A4$	604.271234
$A4$ / $A4$	604.271234

Table 3.1: The distance of the models from $A5/A4$ sound spectrum: the first column represents a model of pitch and fingering, and the second column lists the distance of sound spectrum from each model.

Pitch / Fingering	Distance
$B5$ / $B4$ (reg. key)	38.268165
$B5$ / $B3$ (overblown)	38.268165
$F\sharp 5$ / $B3$ (overblown)	38.268165
$B4$ / $B4$	38.268165
$B3$ / $B3$	38.268165

Table 3.2: The lowest distance model from the sound spectrum of $B4$: The first column shows the pitch name and underlying fingering, and the second column lists the distance of the spectrum from each model.

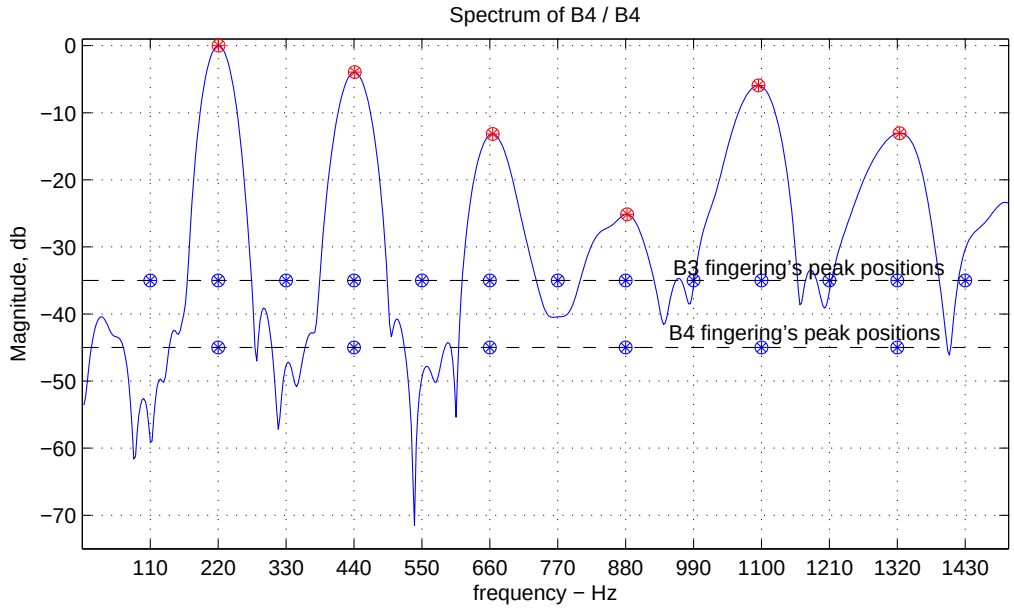

Figure 3.4: Spectrum of $B4/B4$ (220 Hz): the red circles are selected partials and the blue circles are $B3$ (top) and $B4$ (bottom) fingering models.

The time complexity of Algo. 1 depends on measuring distance and extracting spectral peaks. The time complexity of distance measurement is $O(1)$ since the number of models and spectral peaks are limited to 62 and the maximum number of harmonics which is considered up to 1500 Hz for cut off frequency of saxophone and therefore limited [5]. The sum over distances also depends only on the number of spectral peaks and therefore is done in $O(1)$. The feature extraction or peak detection depends on the calculation of spectral components, its resolution, and peak detection algorithm. The Fast Fourier Transform for spectral calculation is known to have computing complexity of $O(n \log n)$ where n is data size or frame size. Since the goal is not developing an on-line system, we have used Matlab *findpeaks* for peak detection, though it can also implement in many other ways. In

a real-time application there is always a trade off between accuracy of peak detection and performance, the decision is made based on application of this research. As result, the time complexity of the algorithm is an upper bound of computing complexity of Fast Fourier Transform, $O(nlogn)$, and peak detection algorithm.

3.3 Peak Detection

In this section, the spectral peak detection is studied in more details since the accuracy of spectral peaks directly influences the distance metric in Alg. 1 and fingering estimation. First, the parameter of the spectrum, such as scale, window type, window size, and start point of segmentation are studied. Then, the smoothing function is discussed to remove the noise and side lobes from the spectrum. Finally, Matlab *findpeaks* function with 100Hz as minimum peak distance is applied, which is about the lowest frequency that a tenor saxophone plays, and consequently the lowest distance between harmonics. At the end, why a constant threshold cannot be applied and peaks are refined based on a local threshold will be discussed.

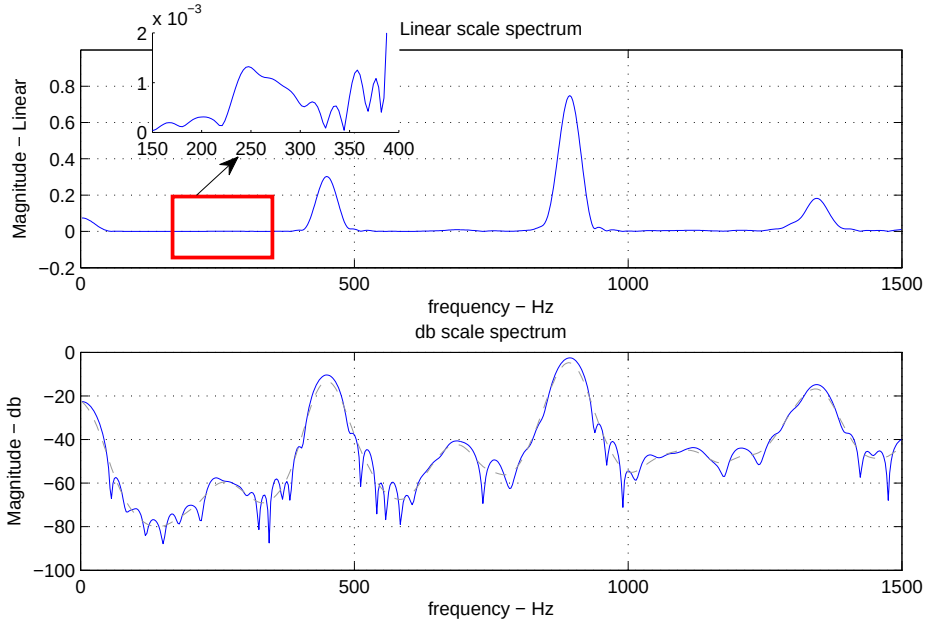

Figure 3.5: Comparison between spectral sub harmonics in linear and db scale: the spectrum of $C6$ (466 Hz) with the register key and $C5$ fingering is shown in linear scale (top) and db scale (bottom)

3.3.1 Scale

Comparing between the linear and logarithmic scale shows that the low amplitude sub harmonics are very close to the noise floor in the linear scale and difficult to detect, while the db scale exposes sub harmonics. For example, the sub harmonics of the note $C6/C5$ (466 Hz) with the register key are very low and hide in the normalized linear spectrum in Fig. 3.5 (top), only zooming into the signal shows high energy around the first sub harmonic at 233 Hz, while in the logarithmic scale spectrum (bottom) the sub harmonics are high enough to be detected. As a result, the analysis is done in logarithmic scale.

3.3.2 Window Type

Hanning and Hamming windowing are very popular in music signal analysis to handle the discontinuity around the window edges to decrease side lobes. In the fingering estimation scenario, sub harmonics can be at a very low level, and may be buried under the side lobes of high main harmonics. For this reason, the window type with rolling side lobes is a better choice. Moreover, narrower main lobes increase the accuracy of peak positions which is very important to detect the right model based on peak position in Algo. 1.

Applying the Fourier Transform on a simple sinusoid vibrating at 800 Hz, the Hanning window shows a wide main lobes and roll off side lobes, shown in Fig 3.6 (left), while the Hamming window has a narrower main peak with flat side lobes (right). The Fig 3.7 shows the spectrum of a windowed signal of the note $C6/C5$ with the register key where the sub harmonics buried in the side lobes for Hamming windowed signal (right), while attenuation of side lobes avoids this problem when the Hanning window is applied (left). Comparison of the sound spectrum peaks (gray line) with windowed signal spectrum (blue line) shows both window types preserve peas' positions very well and therefore the Hanning window with rolling side lobes is a better option.

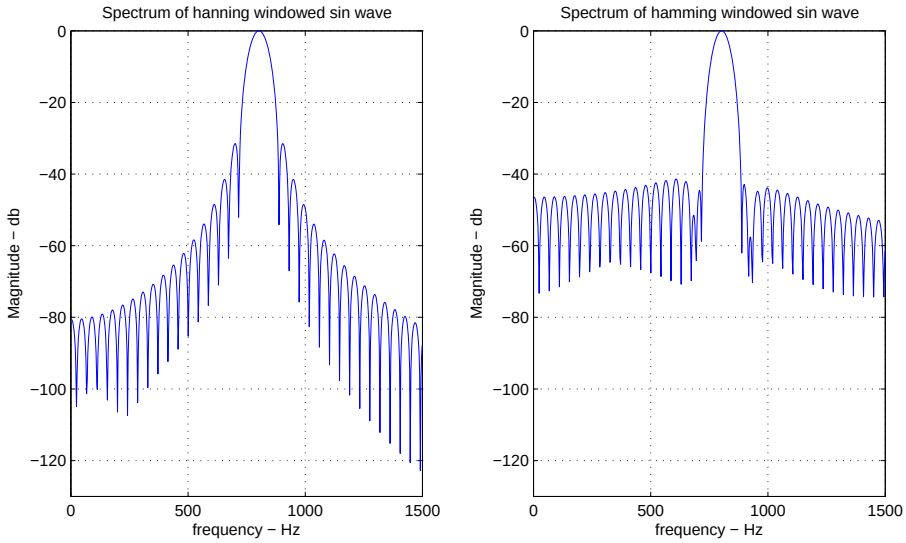

Figure 3.6: Comparison between Hanning and Hamming window on sin wave: spectrum of a Hanning windowed signal has wider peaks with roll of side lobes (left), whereas the spectrum of Hamming window has narrower peaks with higher amplitude and flat side lobes (right).

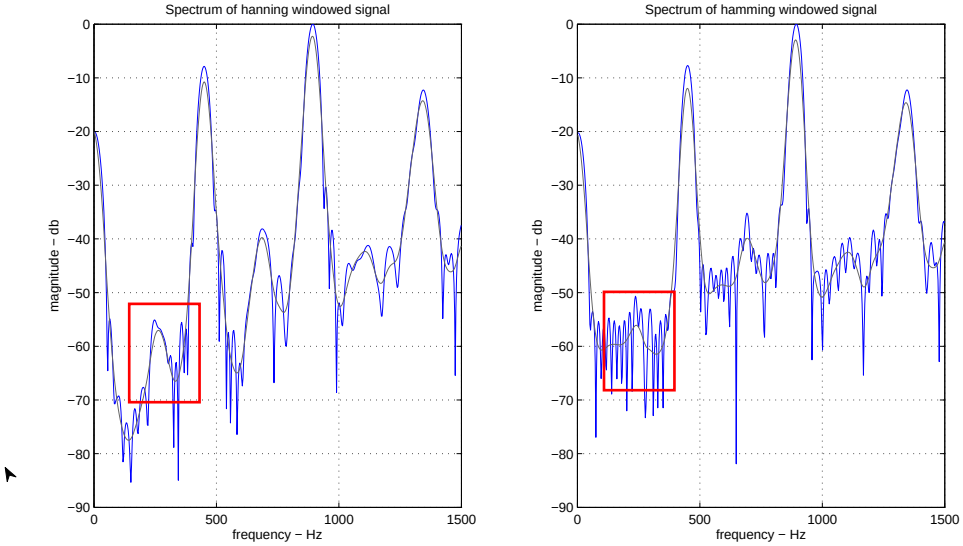

Figure 3.7: Comparison between spectral sub harmonics in linear and db scale: the spectrum of $C6/C5$ (466 Hz) with the register key is shown in linear scale (top) and db scale (bottom)

3.3.3 Frame Size

In the signal processing, a signal breaks into short segments which length is a trade off between time and frequency resolutions, while the longer segment gives more resolution in frequency domain, and the smaller segment provides more time resolution. Since the peak positions are extracted form the spectrum, the frequency domain resolution is more important and the larger frame size of 1024 (23 ms) and 2048(46 ms) samples with a sampling rate of 44100 Hz are considered. The spectrum of a signal of 23 ms and 46 ms are shown in Fig. 3.8, where the top is the spectrum of 46 ms with more frequency components details and therefore noisier than the two separate continuous 23 ms segments (bottom). The second spectrum of size 23 ms (bottom right) does not show the sub harmonics, while the first frame on the left (bottom left) exposes sub harmonics very clearly, which means sub harmonic fade out in time. So, the window size should be large enough to cover part of the region where sub harmonics appear strongly.

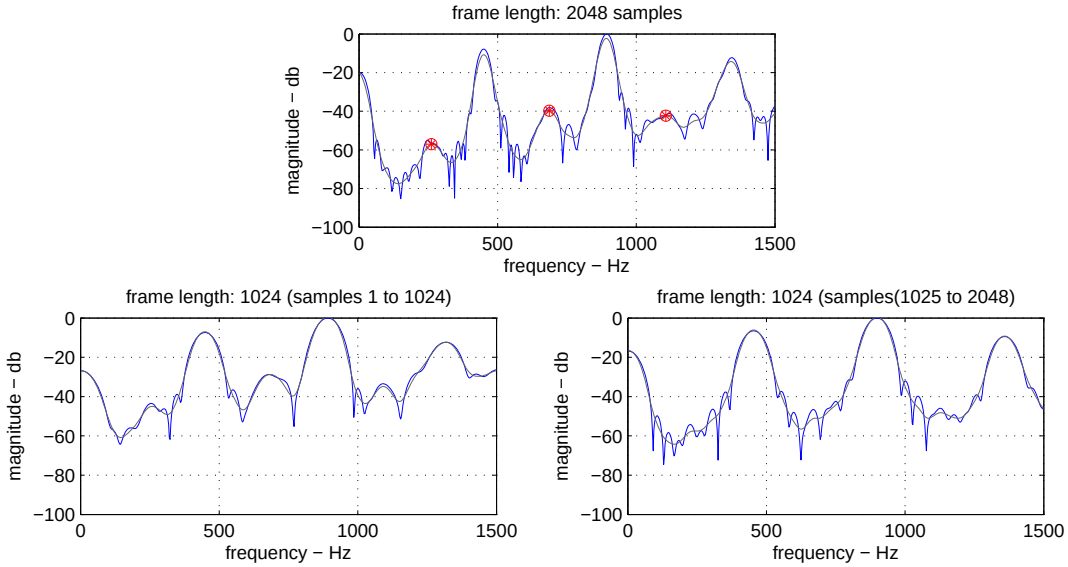

Figure 3.8: This figure shows the effect of window size on the note $C6/C5$ (466 Hz) with the register key. The top left spectrum is the result of 2048 window size from the signal, bottom spectra breaks the large window to two window size of 1024 samples. It is clear that the bigger spectrum has more resolution and shows the peak position more accurately.

Since many studies [28, 29, 30, 31] have found sustain and attack portions of a sound as the most informative parts, we look for sub harmonics in those regions. Based on our observation on spectrograms of saxophone recording sounds, the sub harmonics appear with different strengths during attack and early in the sustain part. The spectrogram of the overblown note $C6/C4$ (466/233 Hz) fingering in Fig. 3.9 shows the sub harmonics of $C4$ fingering during the attack period and the beginning of the steady state in a black rectangle. So, we try to start segmentation from the beginning of the attack and choose a large window size of 46 ms to cover the attack and the early part of sustain. For the purpose of this research, we found a simple threshold of 0.008 on signal amplitude can be helpful, but when the sound is soft, or with hard attack or very short, we adjust this threshold manually. In an online application, this threshold cannot be helpful and a more complicated method such as spectral fluctuation with consideration of background noise should be considered.

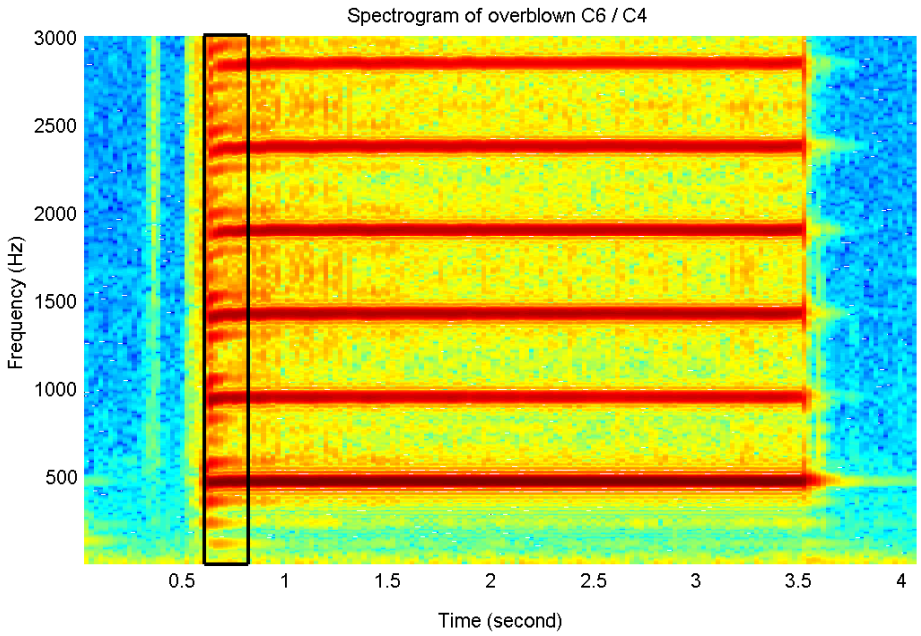

Figure 3.9: Spectrogram of overblown $C6/C4$: This figure shows the existence of sub harmonics of $C4$ fingering during attack period inside a black rectangle.

3.3.4 Preprocessing: Smoothing Spectrum

Smoothing a signal is a common peak detection preprocessing step, which involves removing noise from the spectrum. Here, two common smoothing functions are studied: a low pass filter and a sliding window, in which the low pass filter uses previous samples and the sliding window uses previous and future samples to smooth the spectrum. Another way of seeing the smoothing is finding the trend of the data and it means a tight and fit spectral envelope to show the sub harmonics peaks as well as the main harmonics. We study the Linear Predictive Coding (LPC) to extract the spectral envelope and show that even the high order LPC cannot represent the small sub harmonics and therefore is not useful for this research.

Smoothing functions reduce the noise in a signal, in our case it reduces the noise and removes the small side lobes in the spectrum. The low amplitude sub harmonics are very close to the noise floor, and very difficult to detect, smoothing functions help to detect the right peaks (main harmonics and sub harmonics) instead of the side lobes. Knowing that the low pass filter removes the noise and fluctuation from a time domain, the same idea can be applied to smooth the spectrum with a specific cut off frequency with a recursive function over the spectrum defined as: $y[i] := a * x[i] + (1 - a) * y[i - 1]$. We uses the implementation of butterworth filter of Matlab with a low pass parameter and a cut off frequency of 1323 Hz. Low pass should be used very carefully since it can have the effect of shifting peaks slightly.

A Sliding window is another type of smoothing, in which a window passes through the signal and each point of the signal centered at the middle of the window is replaced by the average of point by point multiplication of the window with a segment of the signal [32]. The rectangle window, or unweighted window is the simplest type, but usually one pass cannot smooth the spectrum very well, two and three passes of a rectangle window over a signal results in a triangle and a pseudo-Gaussian window with $n * w - n + 1$ smooth-width where w is the length of rectangle window and n is the number of passes [32]. The odd window length, w, preserve peaks' positions and therefore we only use odd window size for

sliding window [32].

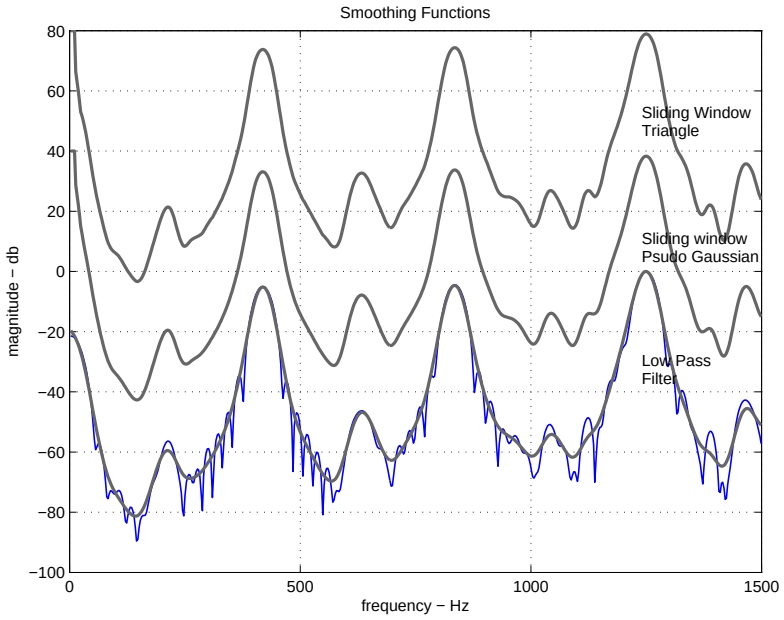

Figure 3.10: Smoothing functions: comparison between low pass filter, sliding triangle window, and sliding pseudo-Gaussian window on the spectrum of $Bb6/(Bb5/Bis)$ with the register key

A comparison between the low pass filter and the sliding window is shown in Fig. 3.10 for the spectrum of $Bb6/(Bb5/Bis)$ with the register key, since one pass of the slide window is not usually helpful we consider two and three passes (triangle and pseudo-Gaussian types). The low pass filter with a cut off frequency around 1323 Hz smooths the spectrum very well, while showing off the peaks. The triangle and pseudo-Gaussian created by a rectangular sliding window of size 7, works as good as a low pass filter. However, the sliding window with an odd window size does not change the peak positions [32], while the low pass may. The pseudo-Gaussian and low pass filter with different cut off frequency and window size will be compared in Section 3.5.

The spectral envelope is another way to find the resonances and peaks in a spectrum. Linear Predictive Coding (LPC) is a method to extract the spectral envelope, by finding the trend of a signal. It tries to find the coefficient of $A = [1\ A(1)\ A(2)\ ...]$ of estimated signal X as $Xp(n) = -A(2) * X(n-1)\ -\ A(3) * X(n-2)\ -\ ...\ -\ A(N+1) * X(n-N)$ to

reduce the error between the original signal, X, and the estimated version of it, Xp. Fig. 3.11 shows the spectral envelope on top of the spectrum with orders of 46 and 100, using the Matlab *lpc* function. Even the high order of 100 cannot detect the low amplitude sub harmonics, though it works well on the main harmonics. As a result, LPC ignores the sub harmonics and only shows the resonance of the spectrum at very high amplitude harmonics and thus is not helpful for this problem.

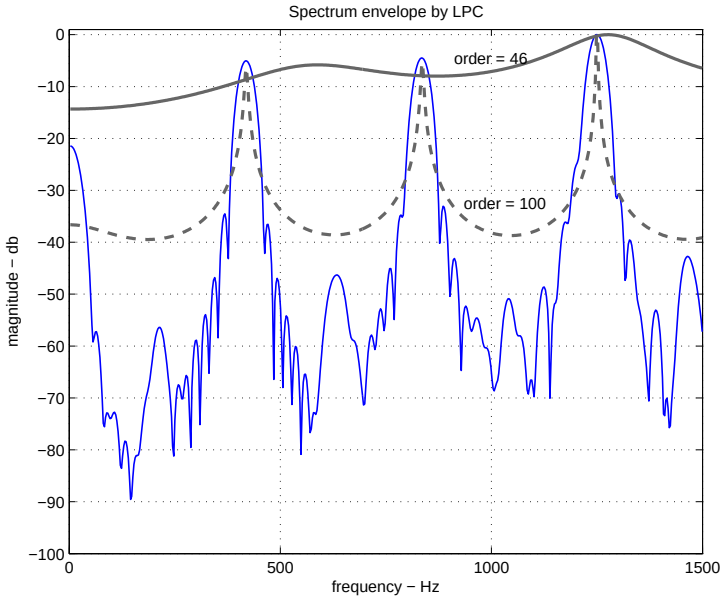

Figure 3.11: LPC envelope on the spectrum of $Bb6/(Bb5/Bis)$ with the register key

3.3.5 Post Processing: Threshold

After smoothing the spectrum, and applying the peak detection, there may still be peaks from the side lobes or the noise floor in the peak list. These peaks are lower in amplitude and closer to the noise floor. A post processing step is necessary to refine the peak lists by a simple threshold, however a constant threshold does not work. The logarithmic scale spectrum in Fig. 3.12 shows increasing trend of the noise floor with a dashed line like a high pass filter with respect to frequency and therefore a constant threshold for peak detection ignores sub harmonics at low frequencies. A high constant threshold around -50 db ignores some of the sub harmonics, while a lower threshold at -75 db is too low for the higher frequency range. The solution is considering the local threshold from the adjacent valleys for each peak as shown by double arrows in Fig. 3.12 for the first sub harmonic. Since the spectrum is noisy and the valley maybe very low in the original spectrum, the valley is measured on the smooth spectrum curve. We found the threshold of 3 db as the minimum height of a peak works well to remove the irrelevant peaks.

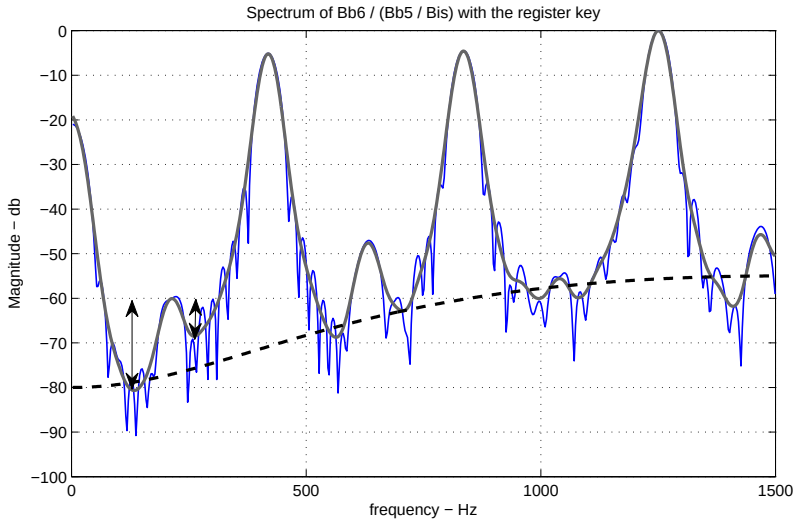

Figure 3.12: The high pass trend of spectrum shows by dashed line, and therefore the constant threshold ignores the low frequency sub harmonics. Therefore, a local threshold validates a peak's height by measuring its amplitude from neighboring valleys which is shown by double arrow in the figure.

3.4 The Register Key Detection

The theoretical model introduced in Section 3.1 has a similar model for a note with and without the register key. For instance the peak position model for $F5/F4$ (311 Hz) with the register key is exactly the same as the overblown $F5/F4$ without the register key, shown in Fig. 3.1. The list of these fingerings with either overblown notes or the register key is listed in Table 3.3, where solid lines separate the underlying base fingering , and dashed lines separate the same fingering model with different pitch, such as $F5/F4$ and $F6/F4$ that share the $F4$ fingering but have different pitches. This table shows that even knowing the pitch leaves us with two or more options of overblown fingerings with and without the register key. For example knowing the pitch $F5$ and $F4$ fingering cannot help to distinguish the activation of the register key. Here we are going to study the effect of the register key on the spectrum and how it can influence the sub harmonics, and peak positions.

pitch Name / Fingering
$A5/A4$ + reg. key
$A5/A4$ (OB)
$E6/A4$ (OB) + reg. key
$G\sharp5/G\sharp4$ + reg. key
$G\sharp5/G\sharp4$ (OB)
$D\sharp6/G\sharp4$ (OB)
$D\sharp6/G\sharp4$ (OB) + reg. key
$G5/G4$ + reg. key
$G5/G4$(OB)
$D6/G4$(OB) + reg. key
$D6/G4$(OB)
$F\sharp5/F\sharp4$ + reg. key
$F\sharp5/F\sharp4$ (OB)
$C\sharp6/F\sharp4$ (OB)
$F5/F4$ + reg. key
$F5/F4$ (OB)
$F6/F4$ (OB)
$E5/E4$ + reg. key
$E5/E4$ (OB)
$E6/E4$ (OB) + reg. key
$E6/E4$ (OB)
$Eb5/Eb4$ + reg. key
$Eb5/Eb4$ (OB)
$D5/D4$ + reg. key
$D5/D4$(OB)
$D6/D4$ (OB)
$A5/D4$ (OB)
$Bb6/Bb5/Bis$ + reg. key
$F6/Bb5/Bis$ + (OB) + reg. key

Table 3.3: This table shows the overblown fingerings with/without the register key for the same pitch, where*reg. key* means overblown with the register key and *(OB)* refers to overblown note without the register key.

3.4.1 Function Of The Register Key

To play the tenor saxophone in the higher frequency range, which is limited by the saxophone's bore length, the register key is designed to open a register hole on the body or the neck which works as a short circuit to air flow in the low frequency range, consequently weakening the first peak of the input impedance (shown in Fig. 3.13) [5]. The register hole is small enough to seal the air flow in the higher frequency range, and leave the second and higher peaks of input impedance unaltered , making it easier to play at the second peak and higher sounding pitch [5]. As a result, sufficient higher frequency waves travel to the end of the bore and the low frequency waves reflect back at the first open hole [5]. However, the register holes cannot filter all low frequencies and therefore low amplitudes peaks, called sub harmonics, which are associated with the first input impedance peak, appear in the spectrum [5]. For example $A5/A4$ with the register key (open the register key on the neck) has sub harmonics of $A4$ fingering as shown in Fig. 3.14 with black marks. The $A4$ fingering, plus the register key, opens a hole on the neck and the low frequency components leak at that point.

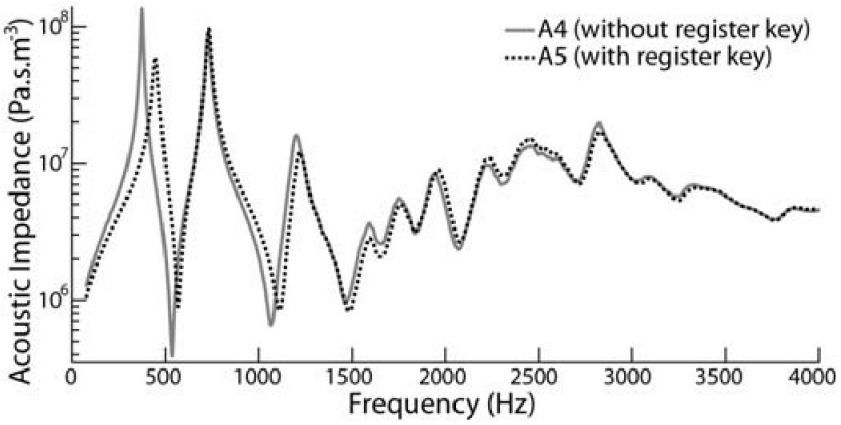

Figure 3.13: The effect of the register key on the input impedance of $A4$ and $A5$ fingerings for a soprano saxophone [5]. The register key weakens and shifts the first peak in the input impedance of $A4$ fingering and make it easier to play at the second peak for the frequency range $A5$.

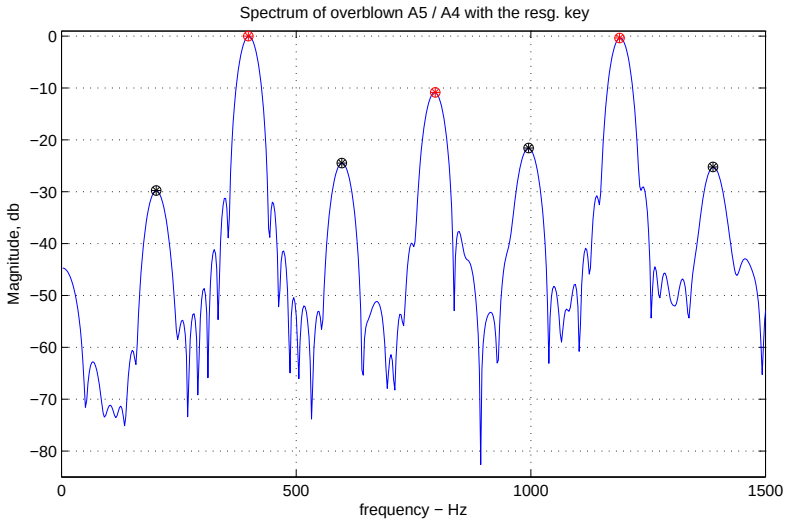

Figure 3.14: sub harmonic: sub harmonics of note the $A5/A4$ (391 Hz) and the register key are shown by by black starts and main harmonics with red stars.

3.4.2 Sub Harmonic Amplitude

Studying the effect of the register key on amplitude of the first input impedance peak (shown in Fig. 3.13 in section 3.15), the amplitude of the sub harmonics, especially the fundamental sub harmonic, expect to suppress more compared with overblown notes without the register key. For example, sub harmonics and the fundamental sub harmonic of $G5/G4$ with the register key have lower amplitudes compared to sub harmonics of $G5/G4$ without the register key, shown in Fig. 3.15 on the top with red marks for the sub harmonics. On the other hand, our study on sub harmonic amplitude, in Chapter 2, shows that the amplitude depends on many other parameters, such as blowing pressure, air flow, and dynamic level. For example the spectra on the bottom of Fig. 3.15 shows the higher sub harmonics for the note $G5/G4$ with the register key at *mezzo forte* dynamic level compared to overblown $G5/G4$ without the register key (top right), even the sub harmonic fundamental height is almost the same for the $G5/G4$ with the register key (bottom) and without the register key (top right). This hypothesis of higher amplitude for overblown notes without the register key doses not supported by our dataset, while we think it may be valid and should be studied on a dataset with controlling blowing pressure, dynamic level, and attack.

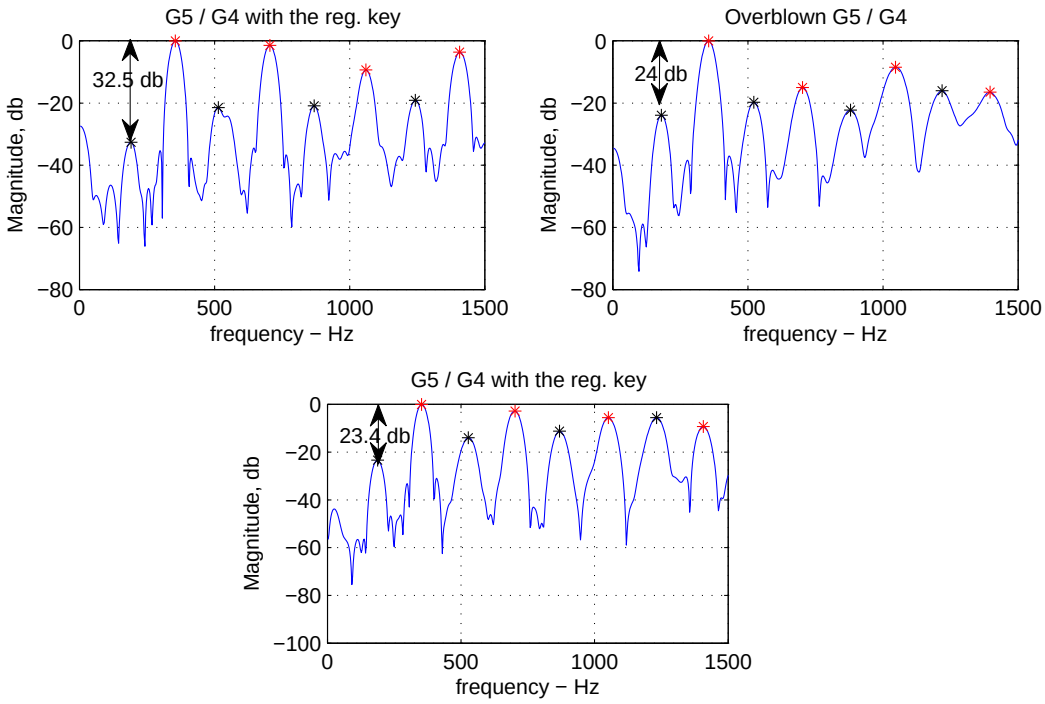

Figure 3.15: Comparison between the sub harmonic amplitude of $G5/G4$ with and without the register key which is played at *mezze forte* dynamic level. The top figure shows the sub harmonics of $G5/G4$ with the register key (top left) are lower than the sub harmonics of overblown $G5/G4$ (top right), where the sub harmonics are marked as black stars and the main harmonics by red. However, another sound file spectrum of $G5/G4$ with the register key are shown with high amplitude sub harmonics on the bottom.

3.4.3 Sub Harmonic Fluctuation

Listening to the sound file, the attack part of some overblown notes without the register key are not as clear as overblown ones with the register key. Our hypothesis is that the sub harmonics' amplitude of overblown notes without the register key drop faster. The changes in peaks' height is studied under spectral Flux (SF) to measure the changes in spectral energy over time, which is defined as Euclid's square distance between two continuous frames, as follows [33]:

$$SF(i) = \frac{2}{N} \sqrt{\sum_{k=0}^{N/2} ((X(i,k) - X(i-1,k))^2}, \tag{3.2}$$

where $X(i,k)$ is the magnitude of the central frequency bin, k, calculated by Fourier Transfer of the i^{th} frame , $N/2$ associates with the highest frequency of interest, which is 1500 Hz [33]. In this test, two frames are selected, the first frame from the very beginning, when

normalized time domain signal is over 0.08, and the second frame from the steady state, after 46.4 msec. Contrary to our hypothesis, the spectral flux is usually higher for overblown notes with the register key when the pitch is the same and the only difference is the activation of the register key, reported in Table 3.4 (for a complete list of Spectral Flux for all fingerings refer to Table A). However, in most cases the spectral flux is very close for overblown notes with/without the register key, and hard to conclude. Moreover, spectral flux includes changes in spectral envelope, main harmonics' amplitude besides sub harmonics' amplitude. As a contrary to our hypothesis, Fig. 3.16 shows the spectra of $A5/A4$ with/without register key during attack, solid line, and steady state, gray dashed line. It shows that the fall in sub harmonics is less for overblown $A5/A4$ without the register key, while some sound files support our hypothesis of faster fall of amplitude without the register key. We think it may be hard to play at the higher resonance of input impedance without using the register key and that is why sub harmonics sometimes drop slower for these overblown notes. Since the peak's amplitude depends on many factor other than resonance of the saxophone bore such as blowing pressure and control of the reed by the player, the sub harmonics' amplitude should be studied on a dataset with controlling parameters to have a fair comparison.

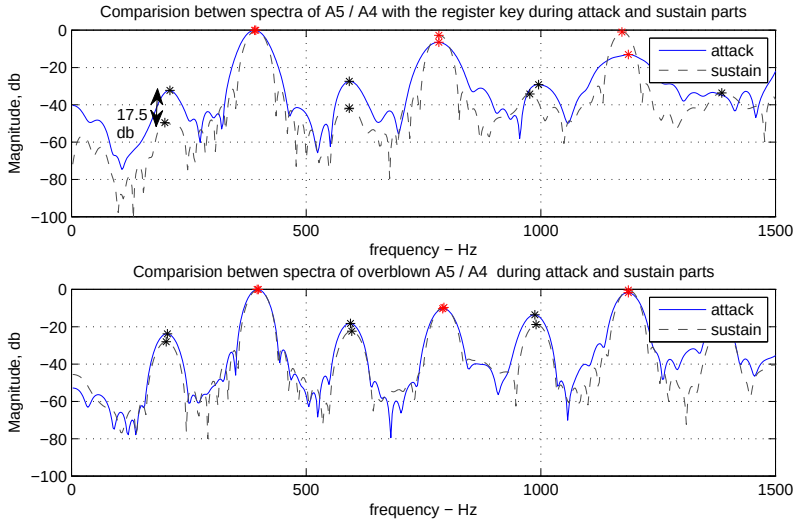

Figure 3.16: Comparison between sub harmonics' amplitude during attack, solid black line, and sustain part, dashed gray line, for overblown notes $A5/A4$ with/without the register key.

pitch Name / Fingering	**Spectral Flux(90-1500 Hz)**
$A5/A4$ + reg. key	**345.035949**
$A5/A4$ (OB)	327.237785
$G\sharp5/G\sharp4$ + reg. key	**447.012359**
$G\sharp5/G\sharp4$ (OB)	395.615029
$D\sharp6/G\sharp4$ (OB) + reg. key	**427.882864**
$D\sharp6/G\sharp4$ (OB)	222.960760
$G5/G4$ + reg. key	4**55.615059**
$G5/G4$(OB)	442.078846
$D6/G4$(OB) + reg. key	**284.022165**
$D6/G4$(OB)	281.326763
$F\sharp5/F\sharp4$ + reg. key	**384.229877**
$F\sharp5/F\sharp4$ (OB)	378.602346
$F5/F4$ + reg. key	**390.543537**
$F5/F4$ (OB)	302.907116
$E5/E4$ + reg. key	333.998636
$E5/E4$ (OB)	**336.603008**
$E6/E4$ (OB) + reg. key	**275.552258**
$E6/E4$ (OB)	262.238699
$Eb5/Eb4$ + reg. key	**316.565615**
$Eb5/Eb4$ (OB)	234.640366
$D5/D4$ + reg. key	299.533917
$D5/D4$(OB)	**237.721536**

Table 3.4: This table shows the Spectral Flux for overblown notes with/without the register key.

3.5 Experiment

In this section, the results of Alg. 1 is reported. First the influence of the smoothing function on the Alg. 1 is going to be studied, after that the modified models based on adding sub harmonics and harmonics' shift are examined. Finally, pitch is going to limit the number of estimated fingering configurations' candidates to improve the accuracy. In each experiment, we are going to compare accuracy, octave error and semitone error, where semitone error is difficult to improve, and we see only pitch can remove all of these semitone errors. Octave errors happen because there is not enough sign of the sub harmonics or in the worst case the peak detection does not work well.

Our dataset contains 62 combinations of fingerings and sounding pitch for a tenor saxophone, which is reported in Appendix A, with a total number of 295 examples, where each combination has 5 sound files except a few with high frequency have 4 files. The recording

was done by an expert player, Joe Miller, in two sessions, the first session includes one sample of recording for each combination of pitch and fingering at Mezzo forte dynamic level, and the second session covers the rest of dataset at different dynamic level. This recordings have variety of length from one to five seconds. The configuration for this experiment is window size of 46 ms (equal to 2048 samples with sampling rate of 44100 Hz), Hanning window type, and 0.008 threshold of signal energy for the start point of the segmentation, which have been discussed earlier in this chapter.

3.5.1 Experiment 1: Smoothing Function

In this experiment, the smoothing functions are going to be studied with Alg. 1 which only considers the peak positions of fingerings. The accuracy of fingerings estimation by using a low pas filter and a pseudo-Gaussian are reported in Table 3.5 with different configurations, where the low pass with a cut off 1323 Hz and a pseudo-Gaussian with a window size of 31 have the highest accuracy, around 73%. The error includes some mismatching that could be octave related or a semitone, and both smoothing functions have error, reported in Table. 3.5. We also consider combination of these methods by applying a pseudo-Gaussian and then a low pass filter, however the accuracy decreases by 4%. Either a low pass filter or pseudo-Gaussian method can be used for smoothing if the right cut off and window size are selected.

Smoothing Function	Accuracy	Octave Error	A Semitone Error
Low Pass (cut off = 1323 Hz)	**73.9**%	05.1%	7.8%
Low Pass (cut off = 1200 Hz)	**73.9**%	06.8%	7.8%
Low Pass (cut off = 1102.5 Hz)	70.5%	08.1%	8.5%
Pseudo-Gaussian (window=37)	71.9%	06.1%	8.8%
Pseudo-Gaussian (window=31)	**72.9**%	05.8%	8.1%
Pseudo-Gaussian (window=25)	69.8%	09.5%	7.8%
Pseudo-Gaussian (window=19)	67.5%	10.8%	8.1%
Pseudo-Gaussian (window=13)	62.4%	15.6%	8.1%
Pseudo-Gaussian (window=7)	60.3%	16.3%	7.1%
Pseudo-Gaussian and Low Pass	69.8%	09.2%	7.8%

Table 3.5: Fingering estimation accuracy by using a low pass and pseudo-Gaussian Smoothing function for peak detection. The last row is a combination of pseudo-Gaussian sliding window with window size of 31 and low pass filter with cut off frequency of 1323 Hz.

3.5.2 Experiment 2: Model Modification

In Section 3.1 we have discussed some model modification based on shift in transfer function peaks and peaks above cut off frequency of the instrument. Altered models include: $B6$, $C6(C5/SideC)$ with the register key, $D5/SideD$, $Bb5/1\&4$. It is also mentioned considering two models for two octave overblowings. So, we consider the original model and the shifted version by a half semitone for $C6/F\sharp4$, $D6/G4$ and the register key, $E6/E4$ with/without the register key, $F6/(Bb5/Bis)$ and the register key, $E6/A4$. These modifications improved the result from 73% to almost 76% for both the low pass and sliding window smoothing functions, reported in Table 3.6. Clearly the semitone error has not changed, while the octave error has decreased to a third. We think a better estimation of peak position from the resonance of the transfer function can improve the accuracy.

Smoothing Function	Accuracy	Octave Error	A Semitone Error
Low Pass (cut off = 1323 Hz)	75.9%	2.0%	8.8%
Pseudo-Gaussian Sliding window (Window Size = 31)	76.6%	2.4%	8.8%

Table 3.6: Fingering estimation accuracy by using a low pass and pseudo-Gaussian Smoothing function for peak detection, and model modification.

3.5.3 Experiment 3: Pitch

The segmentation starts very soon in time, before even the pitch is known during an attack. We studied whether if the estimation of fingering postpones to the steady state, the accuracy improves over 10 percent. But pitch detection is not easy and is very complicated in real-time application, for this experiment we used MIR toolbox, to estimate the pitch and label dataset. This experiment uses the modification models and Alg. 1 for distance measurements based on only peak positions. The candidates from Experiment in Section 3.5.2 are sorted and eliminated by pitch. Removing the candidate with wrong pitch increased accuracy over 10 percent to 87% and removed all semitone errors, Table 3.7.

Smoothing Function	Accuracy	Octave Error	A Semitone Error
Low Pass (cut off = 1200 Hz)	87.5%	5.9%	0.00 %
Pseudo-Gaussian Sliding window	86.1%	6.1%	0.00%

Table 3.7: Fingering estimation by distance method and pitch

Chapter 4

Conclusion

4.1 Conclusion

We have studied the saxophone fingering identification and proposed a theoretical model based on harmonics' and sub harmonics' frequencies as exclusive identifiers of these alternate gestures. It is shown that, pitch and spectral envelope alone cannot distinguished alternate tonehole configurations. A sounding pitch can be produced by multiple fingerings, and is therefore not a unique feature. Likewise, the combination of sounding pitch and the fingerings does not have a unique spectral envelope unto itself, since many playing parameters at the mouth (e.g. blowing pressure, embouchure) contribute. For example, a player can change the resonance of the spectral envelope by his/her embouchure and blowing pressure to the reed. We have also discussed why source/filter model of the speech is not practical for saxophone since the backward air propagation in the saxophone body impacts the reed vibration, and consequently alters the spectral envelope resonances; the backward air propagation in vocal tracts, on the other hand, have negligible impact on relatively massive vocal cords, and therefore the characteristics of the envelope mainly derives from the impulse response of the vocal tract or filter. Moreover, it has been observed that some fingerings, which are playing at the same dynamic level, have similar resonances and spectral envelopes. Therefore spectral envelope is not a unique identifier for some alternate fingerings.

We have also seen machine learning technique requires a huge data set of different playing techniques and dynamic levels for training phase, while our data set is small and does

not cover such variety of recordings. Moreover, machine learning does not explain the reason behind weakness of features to recognize fingerings, for example it does not reveal the parameters that alter the spectral envelope and how similarity in spectral envelope relates to transfer functions of fingerings, while study of saxophone source/filter model explains variation of spectral envelope. Instead of complicated machine learning techniques, a simple theoretical model has been proposed by studying the acoustics of saxophone to identify fingerings without need of a huge dataset.

Motivated by the acoustic of saxophone and its playing techniques, we have proposed a dynamic level-independent model for each combination of a sounding pitch and fingerings' configuration, based on expected frequency positions of main harmonics and sub harmonics. A logarithmic distance metric has been introduced to measure the similarity between the input sound spectrum and theoretical models, and the fingering of the closest model is selected as the best match. However, the accuracy of this estimation depends on precise peak detection from the input sound spectrum. We have identified the attack and early sustain region of a note as the most informative portion for tracking sub harmonics, and therefore, a Fourier Transform is applied on a big window size of 46 msec to cover both the attack and early steady state of a note. Evaluation of window types on 46 msec samples has shown that a Hanning window with higher rolling off side lobes is a better option than a Hamming window, as low amplitude sub harmonics and main harmonics can stick out from the noise level and side lobes. Further study will include studying other window types with higher rolling off side lobes and narrower main lobes, which contributes to the accuracy of feature extraction.

We have evaluated three smoothing functions to apply on a db scale spectrum to remove fluctuation and increase the accuracy of feature extraction. It is found that both the low pass filter with a cut off frequency of $1300Hz$ and the pseudo-Gaussian sliding window size of 31 provide competitive accuracy for fingering estimation - around 73%. We also studied Linear predictive coding (LPC) with a high order of 100, but it could not show the sub harmonics resonances, and was only useful for extraction of spectral envelope. The frequency range is studied under the 1500 Hz which is over the cut off frequency of the instrument. One important selection of such selection is getting more evidence of sub harmonics and

main harmonics. However, the resonance of transfer function are not ideally harmonic, and so do the spectral peaks, but such irregularity is negligible at least for non overblown notes under 1500 Hz. The lowest distance between the peaks of the spectrum is set to $100Hz$, which is close to the lowest note on the tenor saxophone. We presented the post processing step as a local threshold from neighboring valleys on the smoothing spectrum, since a constant threshold does not work on the high pass trend of spectrum under $1500Hz$. We think there may be a relation between such a high pass trend and tonehole configurations. We leave this point for further study.

The error of peak position model includes 5% octave error, 8% semitone error, and 14% mismatched fingering. A modified peak position model is presented, which considers the semi semi-tone off tune of by two octaves overblowing notes and shifting the peaks, as well as some recurrent peaks which do not belong to either main harmonics nor sub harmonics. The modified version improves the octave error by 3% percent. In future studies, this experiment should be applied on a bigger data set, since we believe an expert player can avoid off-tune notes and adding recurrent peaks need more evidence at different dynamic levels. Besides the position, we try the pitch as a complementary feature, which resolves all semitone errors and some mismatched fingerings estimation, as well as achieving an accuracy of 87%. However, the pitch is a costly feature and postpones the estimation to the sustain part of a sound, and it sometimes cannot be accurate in on line applications. When few seconds latency of pitch detection does not matter, the only concern would be an accuracy of pitch detection in a noisy environment. Finding complementary features that are reliable during attack is encouraged for future work.

The next step is finding new features that can distinguish fingering with the same model and spectral envelope. For this study, a dataset with controlling parameters such as blowing pressure, dynamic level, and attack time is necessary. As a start point for feature extraction, we suggest looking at how the main harmonics and sub harmonics form and how their amplitude fluctuates during attack at each dynamic level. The noise robustness of the features should also be evaluated. Although camera limits action of a player, it worth studying computer vision as a complementary method to track fingerings and improve accuracy. The other interesting future work is how likely it is to move from one fingering to another which

involves studying the spectral and temporal features during transition between two notes.

Bibliography

[1] Fingering scheme for saxophone, . URL http://www.wfg.woodwind.org/sax/sax_fing.html. x, xi, xiv, 11, 55, 57

[2] Alternate fingering chart for saxophone, . URL http://www.wfg.woodwind.org/sax/sax_alt_2.html. x, xi, 11, 57

[3] Cheng-i Wang, Tamara Smyth, and Zachary C Lipton. Estimation of saxophone control parameters by convex optimization. 2014. xi, 8, 15

[4] Tamara Smyth and Marjan Rouhipour. Saxophone modelling and system identification. In *Proceedings of Meetings on Acoustics*, volume 19, page 035010. Acoustical Society of America, 2013. xi, xii, 1, 8, 17, 19, 20

[5] Jer-Ming Chen, John Smith, and Joe Wolfe. Saxophone acoustics: introducing a compendium of impedance and sound spectra. *Acoustics Australia*, 37(1-19), 2009. xiii, xv, 25, 29, 41

[6] Andrey R Da Silva, Marcelo M Wanderley, and Gary Scavone. On the use of flute air jet as a musical control variable. In *Proceedings of the conference on New interfaces for musical expression*, pages 105–108. National University of Singapore, 2005. 1, 6

[7] Charles Nichols. The vbow: development of a virtual violin bow haptic human-computer interface. In *Proceedings of the conference on New interfaces for musical expression*, pages 1–4. National University of Singapore, 2002. 3

[8] Sølvi Ystad and Thierry Voinier. A virtually real flute. *Computer Music Journal*, 25 (2):13–24, 2001. 3

[9] Tina Blaine and Tim Perkis. The Jam-O-Drum interactive music system: a study in interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 165–173. ACM, 2000. 4

[10] S Sidney Fels and Geoffrey E Hinton. Glove-Talk II-A neural-network interface which maps gestures to parallel formant speech synthesizer controls. *Neural Networks, IEEE Transactions on*, 8(5):977–984, 1997. 4

[11] Atau Tanaka. Sensor based musical instruments and interactive. *The Oxford handbook of computer music*, page 233, 2009. 5

[12] Erwin Schoonderwaldt, Nicolas Rasamimanana, and Frédéric Bevilacqua. Combining accelerometer and video camera: Reconstruction of bow velocity profiles. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 200–203. IRCAM Centre Pompidou, 2006. 6

[13] Matthew Burtner. The metasaxophone: concept, implementation, and mapping strategies for a new computer music instrument. *Organised Sound*, 7(02):201–213, 2002. 6

[14] Tamara Smyth and Jonathan Abel. *Estimating the reed pulse from clarinet recordings*. Ann Arbor, MI: MPublishing, University of Michigan Library, 2009. 7

[15] Vasileios Chatziioannou and Maarten van Walstijn. Estimation of clarinet reed parameters by inverse modelling. *Acta Acustica united with Acustica*, 98(4):629–639, 2012. 7

[16] Adam P Kestian and Tamara Smyth. Real-time estimation of the vocal tract shape for musical control. In *Proceedings of the 7th Sound and Music Computing Conference, Barcelona, Spain*, pages 206–211, 2010. 8, 14

[17] Peter Ladefoged, Richard Harshman, Louis Goldstein, and Lloyd Rice. Generating vocal tract shapes from formant frequencies. *The Journal of the Acoustical Society of America*, 64(4):1027–1035, 1978. 8

[18] Pamornpol Jinachitra and JO Smith. Joint estimation of glottal source and vocal tract for vocal synthesis using kalman smoothing and em algorithm. In *Applications of Signal Processing to Audio and Acoustics, 2005. IEEE workshop on*, pages 327–330, 2005. 13

[19] Hui-Ling Lu and Julius O Smith. Joint estimation of vocal tract filter and glottal source waveform via convex optimization. In *Applications of Signal Processing to Audio and Acoustics, IEEE workshop on*, pages 79–82, 1999. 13

[20] Julius O. Smith. *Introduction to Digital Filters with Audio Applications*. http://-ccrma.stanford.edu/~jos/filters/, accessed (date accessed). online book. 13

[21] Perry Cook. Identification of control parameters in an articulatory vocal tract model, with applications to the synthesis of singing. Master's thesis, Stanford University, Stanford, California, 12/1990 1990. URL https://ccrma.stanford.edu/files/papers/stanm68.pdf. 13

[22] Olaf Schleusing, Tomi Kinnunen, Brad Story, and J-M Vesin. Joint source-filter optimization for accurate vocal tract estimation using differential evolution. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(8):1560–1572, 2013. 13

[23] Arantza Del Pozo and Steve Young. The linear transformation of lf glottal waveforms for voice conversion. In *Interspeech*, pages 1457–1460, 2008. 13

[24] Soumya Bouabana and Shinji Maeda. Multi-pulse lpc modeling of articulatory movements. *Speech Communication*, 24(3):227–248, 1998. 13

[25] Corey Kereliuk, Bertrand Scherrer, Vincent Verfaille, Philippe Depalle, and Marcelo M Wanderley. Indirect acquisition of fingerings of harmonic notes on the flute. In *33rd Int. Computer Music Conf*, volume 1, pages 263–6, 2007. 22

[26] Vincent Verfaille, Philippe Depalle, and Marcelo M Wanderley. Detecting overblown flute fingerings from the residual noise spectrum. *The Journal of the Acoustical Society of America*, 127(1):534–541, 2010. 22

[27] Miller S. Puckette, Theodore Apel, and Zicarelli David D. Real-time audio analysis tools for Pd and MSP. 1998. 24

[28] Gerard R Charbonneau. Timbre and the perceptual effects of three types of data reduction. *Computer Music Journal*, 5(2):10–19, 1981. 34

[29] Melville Clark Jr, Paul T Robertson, and David Luce. A preliminary experiment on the perceptual basis for musical instrument families. *Journal of the Audio Engineering Society*, 12(3):199–203, 1964. 34

[30] John M Grey and James A Moorer. Perceptual evaluations of synthesized musical instrument tones. *The Journal of the Acoustical Society of America*, 62(2):454–462, 1977. 34

[31] Sébastien Schiesser and Caroline Traube. On making and playing an electronically-augmented saxophone. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 308–313. IRCAM Centre Pompidou, 2006. 34

[32] T OHaver. Smoothing. URL http://terpconnect.umd.edu/~toh/spectrum/Smoothing.html. 35, 36

[33] Tao Li, Mitsunori Ogihara, and George Tzanetakis. *Music data mining*. CRC Press, 2011. 43

Appendix A

Fingerings' Configurations

Table. A.1 includes the sounding pitch with their underlying fingerings, where the key names are standard and based on Fig. A.1 [1].

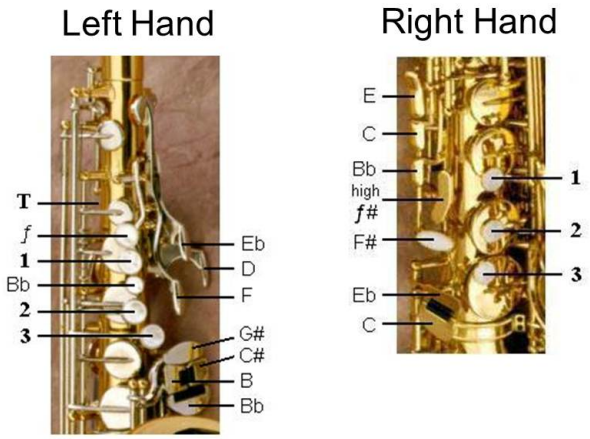

Figure A.1: Saxophone keypads: the figure shows the key pad names of alto saxophone, while this naming is general to most of saxophones [1].

Note name / Fingering	Overblown / Reg. key	SF(90-1500 Hz)
$F6/F6*$	reg. key	220.734151
$F6/Fork$		237.185748
$F6/(Bb5/Bis)$	OB + reg. key	221.002097
$F6/F4$	OB	544.719137
$E6/E6*$	reg. key	273.092278
$E6/E4$	OB	262.238699
$E6/A4$	OB	212.057694

Note name / Fingering	Overblown / Reg. key	SF(90-1500 Hz)
$E6/E4$	OB + reg. key	275.552258
$D\sharp6/D\sharp6*$	reg. key	237.765522
$D\sharp6/G\sharp4$	OB	222.960760
$D\sharp6/G\sharp4$	OB + reg. key	427.882864
$D6/D6*$	reg. key	195.776197
$D6/D4$	OB	371.993134
$D6/G4$	OB	281.326763
$D6/G4$	OB + reg. key	284.022165
$C\sharp6/C\sharp5$	OB + reg. key	201.159374
$C\sharp6/C\sharp4$	OB	511.762391
$C\sharp6/F\sharp4$	OB	272.708053
$C6/C5$	OB + reg. key	171.181494
$C6/C4$	OB	415.149039
$C6/(C5/SideC)$	OB + reg. key	345.056799
$B5/B4$	OB + reg. key	460.711805
$B5/B3$	OB	383.094181
$Bb5/(Bb4/1\&4)$	OB + reg. key	299.740158
$Bb5/(Bb4/Bis)$	OB + reg. key	231.922015
$Bb5/Bb3$	OB +	384.172054
$Bb5/(Bb4/SideBb)$	OB + reg. key	417.402547
$A5/A4$	OB + reg. key	345.035949
$A5/D4$	OB	409.351663
$A5/A4$	OB	327.237785
$G\sharp5/G\sharp4$	OB + reg. key	447.012359
$G\sharp5/G\sharp4$	OB	395.615029
$G\sharp5/C\sharp4$	OB	291.002713
$G5/G4$	OB + reg. key	455.615059
$G5/C4$	OB	346.905340
$G5/G4$	OB	442.078846
$F\sharp5/F\sharp4$	OB + reg. key	384.229877
$F\sharp5/F\sharp Trill$	reg. key	396.271387
$F\sharp5/B3$	OB	326.833495
$F\sharp5/F\sharp4$	OB	378.602346
$F5/F4$	OB + reg. key	390.543537
$F5/Bb3$	OB	271.648993

Note name / Fingering	Overblown / Reg. key	SF(90-1500 Hz)
$F5/F4$	OB	302.907116
$E5/E4$	OB + reg. key	333.998636
$E5/E4$	OB	336.603008
$Eb5/Eb4$	OB + reg. key	316.565615
$Eb5/Eb4$	OB	234.640366
$Eb5/SideEb$		229.648690
$D5/D4$	OB + reg. key	299.533917
$D5/D4$	OB	237.721536
$D5/SideD$		229.318745
$C\sharp5/C\sharp5$		291.934798
$C\sharp5/C\sharp4$	OB	320.596828
$C5/C5$		231.614134
$C5/C4$	OB	261.167369
$C5/SideC$		351.762861
$B4/B4$		286.850065
$B4/B3$	OB	242.322996
$Bb4/Bis$		298.460517
$Bb4/1\&4$		251.396721
$Bb4/Bb3$	OB	228.044438
$Bb4/SideBb$		319.757570

Table A.1: The table shows the list of pitch name and their underlying fingerings as *pitch-Name/fingering*. The pitch from $D6$ to $F6$ with star marks use the register key with their own underlying fingerings not one octave below. The second row specifies the activation of the register key or applying overblown technique. The last column shows the spectral flux between 90 to 1500 Hz in db scale. The actual fingerings on saxophone and fingerings configurations can be found at [1] [2]

Appendix B

Tenor Saxophone Recordings

Creator:
The audio files were played by Joel Miller, gathered by effort of Dr. Tamara Smyth, and organized and labeled by Marjan Rouhipour.

Description:
The audio files are dataset of tenor saxophone recordings which are used in this research. There are four to five recordings for each combination of pitch and alternate fingering. The file with zero index belong to the first recording session and usually in mezzo forte dynamic level, the rest of files belong to the second recording session. The dynamic level of each file is labeled by mf for *mezzo* forte, ff for *fortissimo* and pp for *pianissimo*.

Filename:
audio.zip