Toward Real-Time Estimation of Tonehole Configuration

Tamara Smyth
Department of Music, University of California, San Diego
trsmyth@ucsd.edu

ABSTRACT

This work presents a strategy for developing an estimator of tonehole configuration or "fingering" applied by a player during performance, using only the signal recorded at the bell. Because of a player can use alternate fingerings and overblowing to produce a given frequency, detecting the sounding pitch does not produce a unique result. An estimator is developed using both 1) instrument transfer function as derived from acoustic measurements of the instrument configured with each of all possible fingerings, and 2) features extracted from the produced sound—indirect correlations with the transfer function magnitudes. Candidate fingerings are held in sorted stacks, one for each feature considered and a final decision is made based on a fingering's position within the stack, along with that stacks weighting. Several recordings of a professional saxophonist playing notes using all fingerings are considered, and features discussed.

1. INTRODUCTION

Much of the work presented here is motivated by a discussion with a professional tenor saxophonist[1] who frequently employs extended techniques in addition to his well-disciplined virtuosic playing:

> I have sometimes been frustrated by the limited control of the saxophone, particularly in lower registers where controlling the sounding pitch is done mainly by applying a particular fingering.

Miller continues to say that, as a result, transitions between notes at low frequencies is slow, and sliding between notes is nearly impossible. Control, from his point of view, starts to get interesting when playing in higher registers:

> Though it's more difficult to play up there, I feel as though I'm playing a more responsive instrument, one that is more ideal, one that approaches the human voice.

This work presents preliminary research toward the end objective of identifying a saxophone fingering, that is, a configuration of open/closed toneholes, as shown in Figure 1, during real-time performance. In so doing, a performer would be able to apply or map estimated fingerings to parameters of a synthesis model, perhaps even a physical model of his/her very own instrument, allowing for both the benefits of 1) better nuanced control in altissimo playing (as described in the above quote), and 2) improved sound such as that available in the lower saxophone register or an altogether different processed sound. For this to be of use to most saxophonists, identification would require only appending a microphone at the bell—a sensor with which they are usually accustomed. Any additional sensor or device might not withstand the rigor of playing, or might impede the player's technique, ultimately interfering with expression on the the instrument.

Though a pitch estimator can get an idea of fingering, it doesn't consider the whole story. In extended techniques, many notes are produced by overblowing or *bugling*, resulting in several possible alternate fingerings that can be used to produce a given note. Furthermore, since resonant frequencies of the saxophone are not precisely harmonic, that is, they are not strictly integer multiples of a fundamental but rather are stretched with increased frequency, overblowing on a particular fingering can produce a note that may be sharper than expected—sometimes by as much as a semitone or more. Though the player can adjust the tuning with embouchure, it might happen after the attack which could be too late depending on the desired latency.

The problem of fingering estimation is, therefore, a system identification problem, akin to that of extracting the glottal pulse from recorded speech [2], the inverse problem for a trumpet physical model [3], or of estimating the clarinet reed pulse from instrument performance [4]. In the former case for speech, it is common to use Mel-frequency cepstral coefficients (MFCC), linear predictive coding (LPC) [5], or more distinctly and recently, convex optimization [2] to separate a source-filter model. In the case of the saxophone (and indeed the clarinet) however, the reed has a much smaller mass than the vocal folds, and its vibration is more effected by the internal state of traveling waves in the bore. This, along with the fact that it generates a more significant reflection than a fleshy biological valve, makes source-filter estimation methods less appropriate—that is, the "filter" for woodwind reed instruments is not well described by an all-pole representation of the produced sound's spectral envelope. That approximation is already tenuous for speech; it is even more remote for blown closed cane reeds.

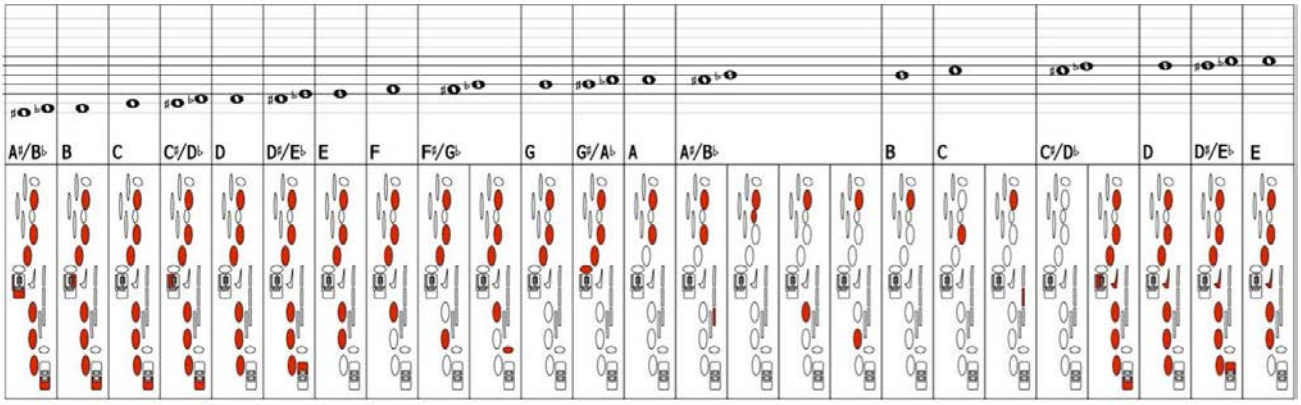

Figure 1. The fingerings or tonehole configurations, for the tenor saxophone.

Furthermore, the dynamics of a played note can significantly alter the spectrum of the produced sound, vastly changing the spectral envelope. In quieter notes there is less (if any) beating of the reed and the input pressure to the bore is relatively pure in frequency content. As a result, the resonances of the instrument are less likely to be excited, and there is less contribution of the instrument bore and bell which characterize a particular fingering.

As described in Section 2, we expand upon a previously described measurement technique for obtaining the instrument transfer function (the *filter*) at the bore base (mouthpiece), H_M, and the transfer function outside the bell, H_B, corresponding to all regular fingering used on the tenor saxophone [6]. We then attempt to explain salient features of a sound spectrum, incorporating known characteristics of a particular fingering.

Figure 2. Joel Miller, saxophonist, applies fingering during a measurement session.

2. OBTAINING FINGERING TRANSFER FUNCTIONS

2.1 Saxophone Waveguide Model

The transfer function of the saxophone bore and bell make be approximated in one-dimension with a bi-directional delayline accounting for the acoustic propagation delay in a conical bore, as well as filter elements $\lambda_N(z)$ and $R_M(z)$ accounting for the propagation loss, and reflection at the mouthpiece, respectively, and elements $R_B(z)$ and $T_B(z)$ describing the reflection and transmission functions of the bell, the non-cylindrical/non-conical section at the end of the instrument [7]. This leads to the following instrument transfer functions as measured at the mouthpiece

$$H_M(z) = \frac{Y_M(z)}{X(z)} = \frac{1 + R_I(z)}{1 - R_M(z)R_I(z)} \tag{1}$$

and the bell,

$$H_B(z) = \frac{Y_B(z)}{X(z)} = \frac{T_I(z)}{1 - R_M(z)R_I(z)} \tag{2}$$

where $X(z)$ is the pressure input into the bore, the product of volume flow and the characteristic of the bore, $Y_M(z)$ is the transfer function of the pressure at the bore base (downstream from the reed), $Y_B(z)$ is the transfer function of the pressure recorded outside, and on axis with, the bell, R_I is the round-trip instrument reflection function (from reed to bell then back to reed) given by

$$R_I(z) = R_B(z)\lambda_N^2(z)z^{-2N}, \tag{3}$$

and $T_I(z)$ is the one-way transmission (from reed to bell) given by

$$T_I(z) = T_B(z)\lambda_N z^{-N}. \tag{4}$$

If $R_B(z)$ and $T_B(z)$ are permitted to have "long-memory" acoustic information, the model given by (1) and (2) can be made to include tonehole configurations by lumping open tonehole radiation and scattering into $R_B(z)$ and $T_B(z)$. Here, however, we apply an existing measurement technique in [6] for obtaining $R_I(z)$ and T_I from measurement, and we apply the technique for all possible fingering in the range of the tenor saxophone (see Figure 2).

2.2 Measurement Setup

Since the spectral characteristic of any particular fingering is governed by its transfer function, measurement of the horn is required for each of the possible fingering within the playable range of the tenor saxophone.

It's well known that if the input to an LTI system is an impulse, the output is the impulse response of the system. There are problems, however, in using an impulse as the test signal—an impulse having sufficient energy to excite

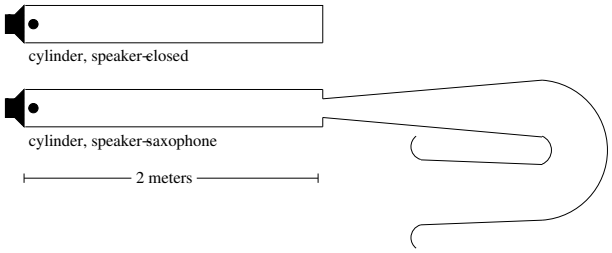

Figure 3. The measurement system consisting of a 2-meter tube with a speaker and co-located microphone at one end. The tube is measured first closed (top), then with a saxophone appended (bottom) to produce the measurement's impulse response under both terminating conditions.

the system above the noise floor will likely produce distortion (and nonlinearities) in the driver. An alternative is to use a sine swept linearly, or logarithmically (as was here), over the frequency range of interest. By smearing the impulse in time, more energy can be applied to the system without introducing distortion.

Estimating the round-trip instrument reflection $\hat{R}_I(z)$ is done by first taking a measurement of the tube with a closed termination as in Figure 3 (top). Following the steps described in [8] allows for estimation of speaker transmission, the reflection off the speaker, and the propagation loss of the measurement tube. All these filter elements related to the measurement system are necessary before further estimating round-trip reflection and one-way transmission from a second measurement, taken with the saxophone appended to the measurement tube as in Figure 3 (bottom). following the steps in [6], the measurement system can be expressed algebraically before isolating for R_I and T_I.

3. ESTIMATOR

Once measurement and postprocessing is complete for each fingering, a stack $\mathbf{S}$ is produced containing candidate magnitudes $G_B(\theta) = |H_B(\theta;\omega)|$ for tonehole configuration θ. Each $G_B(\theta)$ may be consulted by the estimator as described below before making an informed decision as to which fingering θ is most likely to have produced the sound spectrum recorded at the bell $Y_B(\omega)$.

In developing an estimation strategy, a stack of magnitude transfer functions $\mathbf{S}_\mu$ is created and sorted according to the strength by which $G_B(\theta)$ possesses the feature described by μ. The final candiate fingering θ is selected based on the position of each $G_B(\theta)$ in the stack, as well as the weighting of feature μ.

In the following, the features are described and illustrated with examples of how recorded data $Y_B(\omega)$ might fare.

3.1 Selection of Initial Candidates Based on Frequency

Initial selection of stack $\mathbf{S}_{f_0}$ is done based on an estimation of the fundamental frequency f_0 of the sound recorded at the bell $Y_B(\omega)$—consistently its lowest resonant peak for the saxophone. Possible candidate fingerings θ are selected and sorted based on the whether there is a peak in $G_B(\theta)$ that is in alignment, within a certain threshold to allow for flexible tuning, with f_0 (see Figure 4). Though it's possible to do this theoretically, i.e. fingering candidates θ could be reduced to those for which f_0 is an integer multiple of the pitch frequency of $H_L(\theta;\omega)$, but since $G_B(\theta)$ is not strictly harmonic, alignment with actual transfer functions obtained from measurement greatly improves accuracy.

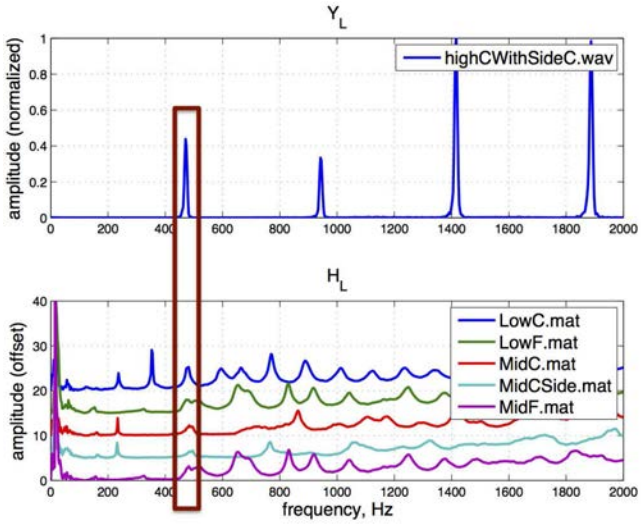

Figure 4. An initial selection of candidates is made by comparing the fundamental of $Y_B(\omega)$ with all measured fingerings $H_B(\omega;\theta)$ and finding fingerings with an aligned partial (within a threshold of tolerance).

3.2 Presence (and Absence) of Subharmonics

As shown in Figure 5, it is often the case that, if a note is overblown, the magnitude of $Y_B(\omega)$, will have peaks present below the fundamental frequency f_0, called *subharmonics*. If this occurs, the task of estimation, and the creation of stack $\mathbf{S}_h$ for subharmonics h, is facilitated considerably. In the presence of subharmonics, the note is certainly overblown and certain candidates θ can be omitted altogether in the formation of $\mathbf{S}_h$.

Furthermore, the subharmonics will typically correspond to the resonant peaks of $G_B(\theta)$, and so can be used in sorting $\mathbf{S}_h$. As shown in Figure 5 (left), an example of middle C played with a low C fingering produces subharmonics at the octave below. In this case, an estimation of the frequency of the subharmonic clearly shows that the sounding note is the second harmonic of the fingering for low C. It is, of course, often the case that more than one subharmonic is produced. In Figure 5 (middle), there are 3 subharmonics, clearly making the fundamental f_0, corresponding to note high C, the 4th harmonic of $G_B(\theta)$, for θ being low C (2 octaves below). It is often the case, however, that the magnitudes of the subharmonics are so slight they might not be detected, or their inharmonicity makes it difficult to simply detect a pitch for determining θ. Consider, for example, the magnitude of the second subharmonic in Figure 5 (right)—it is so low in amplitude that it risks not being noticed by a peak detector. Other examples, not shown here, have shown 3 subharmonics with the amplitude of the sec-

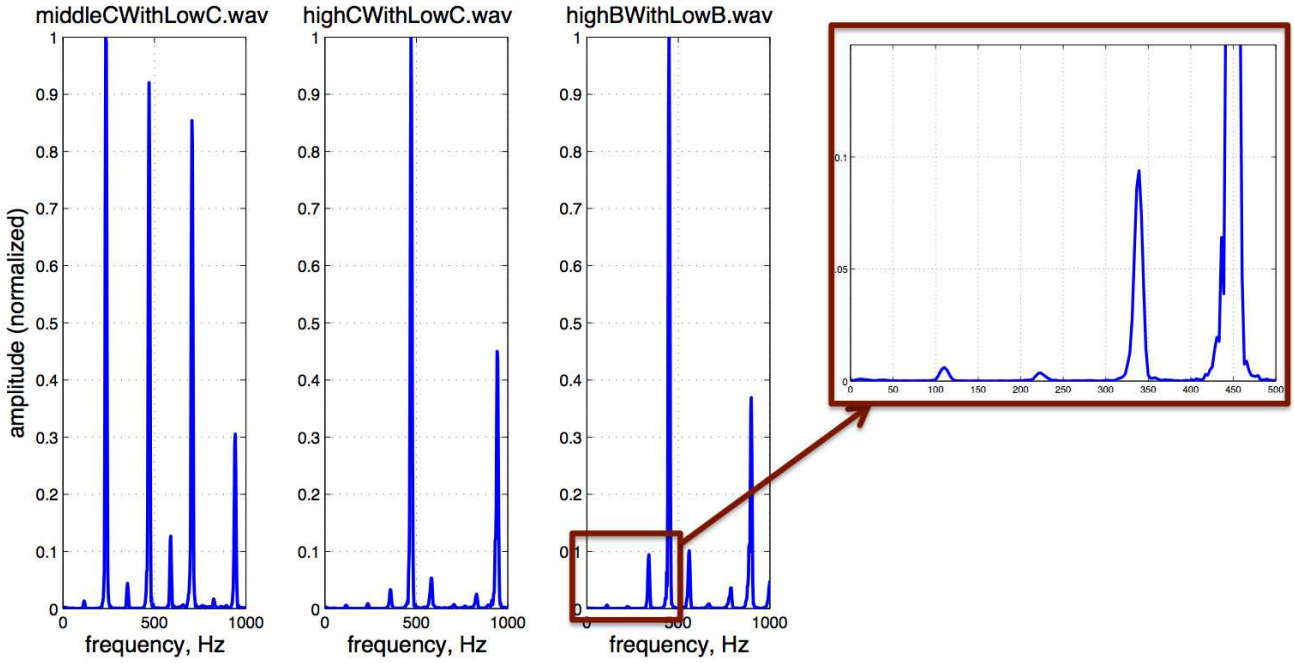

Figure 5. Subharmonics for middle C played with low C fingering, high C played with low C fingering, and high B with low B fingering. Though the subharmonics are rarely actual "harmonics", i.e. they are typically not evenly spaced, their presence greatly simplifies the estimation task.

ond harmonic being significantly more pronounced than the first and/or third—possible suggesting to a peak detector that the sounding frequency f_0 is actually the second, rather than the fourth harmonic corresponding to fingering θ (producing an error of an octave).

It is preferable, therefore, to use a salience measure between the subharmonics of $Y_B(\omega)$ and $G_B(\omega;\theta)$ when sorting $\mathbf{S}$—again showing how the existing measured transfer functions can inform, and provide greater accuracy to, the estimator.

In addition to the presence of subharmonics, their *absence* can be similarly revealing. With the current data set of recordings, subharmonics have been observed in $Y_B(\omega)$ for all cases where f_0 is two or more octaves above the sounding frequency of $G_B(\omega;\theta)$. Though it's too early to say whether this is a definitive feature, a stack is, nevertheless, created and sorted based on the absence of subharmonics. If no subharmonics are detected in $Y_B(\omega)$, the stack is reordered giving less priority to candidate fingerings for which f_0 would the the fourth (or greater) harmonic of $G_B(\omega;\theta)$. Though this stack is created, because of the uncertainty of the feature, it is not as strongly considered in the final estimation.

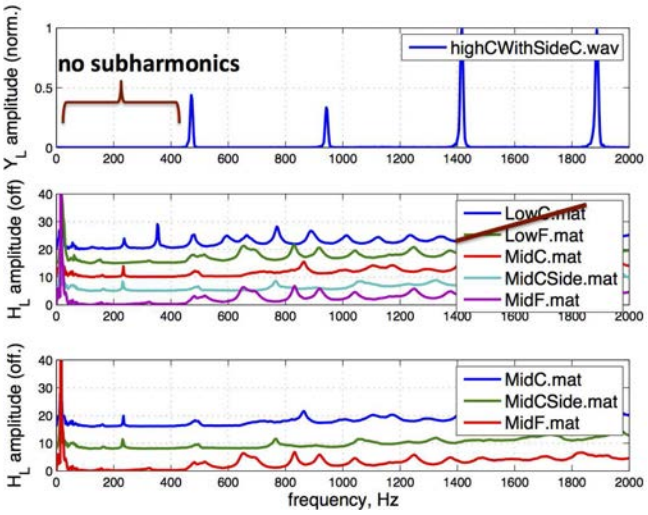

Figure 6. Absence of subharmonics in $Y_B(\omega)$ (top) reduces the likelihood of a candidate fingering having a $G_B(\theta)$ (middle and lower) pitch frequency two octaves below the sounding frequency. Though the figure suggests removal of candidates low C and low F, the stack is actually sorted giving these candidates less priority.

3.3 Gains in $Y_B(\omega)$ Spectral Envelope

The natural state of harmonics in the spectrum produced by a vibrating reed attached to a cylinder is for them to decrease with frequency. It follows, therefore, that gains (peaks in the spectral envelope) in the sounding note that occur above the fundamental frequency are explained by resonant peaks in the instrument. It should perhaps be emphasized, however, that the spectral envelope of the sounding note bears very little resemblance to the magnitude of the instrument transfer function (why the use of LPC for estimation of $H_B(z)$ is not accurate). For this reason, the magnitude $G_B(\omega;\theta)$ cannot be used to directly to estimate the fingering from $Y_B(\omega$. It can, nevertheless, be of tremendous use.

Because gains are a result of resonances in the instrument, overblown notes typically have a steep decay in the

spectral envelope—harmonics of the vibrating reed are not being supported by resonances of the instrument beyond the fundamental. But a decaying spectrum in $Y_B(\omega)$ cannot necessarily be used to identify an overblown note; a note played at a soft dynamic, one where the input pressure $X(\omega)$ is nearly sinusoidal, will similarly exhibit a decay in energy above the fundamental frequency.

This suggests, therefore, that gains in the spectral peaks of $Y_B(\omega)$ can be used as a feature that is explained by closer observation of $G_B(\omega;\theta)$. As shown in Figure 7, the frequencies of harmonic peaks in $Y_B(\omega)$ are first determined, producing vector f_h; the ratio of the amplitudes at these frequencies produces a ratio vector R which may be used to determine significant gains at frequencies f_h. The gain in $G_B(\omega;\theta)$ at frequencies f_h is then observed (as shown by the black dots in Figure 7 (left)), and the ratios similarly taken to sort a stack $\mathbf{S}_g$, based on a closest match.

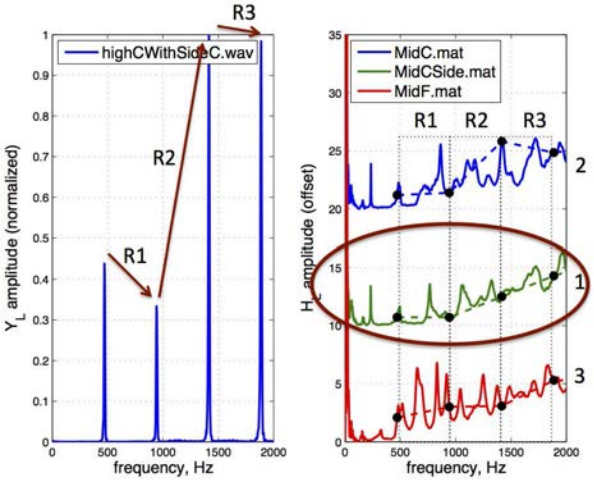

Figure 7. Frequencies of peaks in $Y_B(\omega)$ are looked up under the curve $G_B(\omega;\theta)$ (black dots on dotted line in right-most figure). The slope of amplitudes at peaks in $Y_B(\omega)$, given by R1, R2, R3, etc., are used to determine gains above the fundamental frequency (formant peaks in the spectral envelope). The similarity between gains in R and gains under the curve in $G_B(\omega;\theta)$ is used to sort $\mathbf{S}_g$.

3.4 Frequency-Centered Energy Above the Noise Floor (Subbands)

It is sometimes the case that the magnitude spectrum of a recording $Y_B(\omega)$ will have neither gains nor subharmonics, making it difficult to estimated accurate based on the features previously discussed.

Subbands are defined as regions centered about a frequency having increased energy above the noise floor. Subbands are related to subharmonics, and though subharmonics may not have been detected by a peak picker, it may be possible to determine the presence of subbands using a salience measure, such as a cross correlation of $Y_B(\omega)$ with each $H_B(\omega;\theta)$, up to the fundamental frequency f_0. Figure 8 shows and example of pitch high D played with a middle G fingering—no subharmonics are present, and

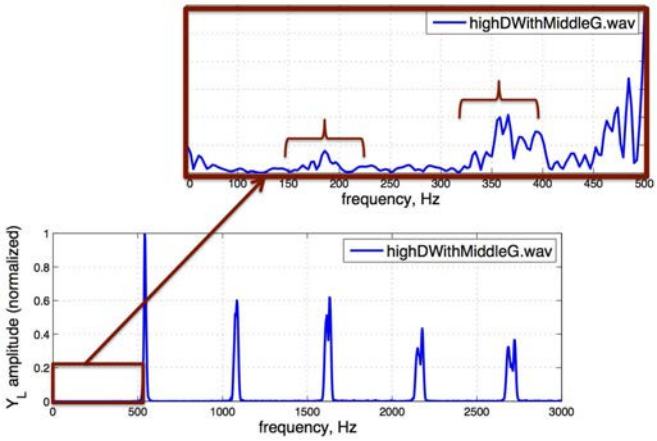

Figure 8. A magnitude spectrum of $Y_B(\omega)$ shows a decay in amplitude with frequency (i.e. there are no characteristic gains in the spectral envelope) and no subharmonics). Zooming in to the region below the fundamental frequency however, shows the existence of energy in clearly define frequency bands—called *subbands*. In this example, the increased energy centered around 350 Hz coincides with the first peak in $G_B(\omega;\theta)$ for the fingering middle G.

the spectral magnitude is decaying with frequency. Zooming into the the region below f_0, however, shows the existence of subbands, centered approximately around 175 Hz and 350 Hz. These peaks correlate most strongly with the peaks in $G_B(\omega;\theta)$ for middle G fingering, however a sorted stack $\mathbf{S}_b$ is created holding correlation values for all fingerings.. Had the region centered about 175 Hz been stronger to that centered about 350 Hz, the subbands would have better correlated with the low G fingering. This is a reasonable result given that the primary difference between the $G_B(\omega;\theta)$ for low G and middle G is that the latter has reduced amplitude at the first resonant peak (caused by the applied octave key) which makes it easier to overblow.

Once stacks $\mathbf{S}_{f_0}$, $\mathbf{S}_h$, $\mathbf{S}_g$, $\mathbf{S}_b$, are created and sorted for features pitch (fundamental frequency), subharmonics, spectral gains, and subbands, respectively, the final candidate may be chosen by assigning each fingering a score. The score is determined based on the position of each fingering θ within stacks $\mathcal{S}_{f_0,h,g,b}$ (the lower the position index, the lower the score and the greater the likelihood the fingering was used to produce the sound), weighted by the strength of each feature.

4. CONCLUSIONS

In this work, features of sound recorded at the saxophone bell, played with an applied fingering, are discussed in relation to instrument transfer functions derived from measurement, with measurements taken of the instrument configured with of all possible fingerings throughout its range. A databased of sound, having notes played with alternate fingerings and overblowing, is used to assess four (4) features that may be used to inform an estimator which makes a final decision on the most likely fingering used to produce a given sound. Candidate fingerings, represented by

transfer function magnitudes $G_B(\theta)$, are held in stacks, one for each feature considered, and sorted according to the salience of a feature in a particular fingering. Each fingering is given a score based on the index of $G_B(\theta)$ within a stack, as well as that stack's weighting.

It is likely, in future work and algorithm refinement, that the final feature measuring the salience of subbands could replace several of the other features. Focusing on this has become of higher priority since the spectral envelopes and gains vary so tremendously with dynamics, making estimation based on gains in the spectrum rather tenuous and unreliable. It is believed that perhaps looking at what seems to be absent might be as revealing as what is obviously present.

Though performance of the algorithm is quite successful with the current database and shows good promise, refinement is needed before brining it into a real-time performance situation, where increased noise floor, and possible bleed from other instruments, will introduce further difficulties.

Acknowledgments

Many thanks to Joel Miller who's fine musicianship, intuition and commitment has been instrumental in this work.

5. REFERENCES

[1] J. Miller, "Private correspondence."

[2] H.-L. Lu and J. O. Smith, "Joint estimation of vocal tract filter and glottal source waveform via convex optimization," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'99)*, New Paltz, NY, October 1999, pp. 79–92.

[3] T. Helie, C. Vergez, J. Levine, and X. Rodet, "Inversion of a physical model of a trumpet," in *Proceedings of the 38th IEEE Conference on Decision and Control*, vol. 3, Phoenix, Arizona, December 1999, pp. 2593–2598.

[4] T. Smyth and J. S. Abel, "Toward an estimation of the clarinet reed pulse from instrument performance," *Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. 4799–4810, June 2012.

[5] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, April 1975.

[6] T. Smyth and S. Cherla, "Saxophone by model and measurement," in *Proceedings of the 9th Sound and Music Computing Conference*, Copenhagen, Denmark, July 2012, p. 6 pages.

[7] J. O. Smith, "Physical audio signal processing for virtual musical instruments and audio effects," http://ccrma.stanford.edu/~jos/pasp/, December 2008, last viewed 8/24/2009.

[8] T. Smyth and J. Abel, "Estimating waveguide model elements from acoustic tube measurements," *Acta Acustica united with Acustica*, vol. 95, no. 6, pp. 1093–1103, 2009.