

# Evaluation of the Technical Level of Saxophone Performers by Considering the Evolution of Spectral Parameters of the Sound

Matthias Robine and Mathieu Lagrange

SCRIME – LaBRI, Université Bordeaux 1  
351 cours de la Libération, F-33405 Talence cedex, France  
firstname.name@labri.fr

## Abstract

We introduce in this paper a new method to evaluate the technical level of a musical performer, by considering only the evolutions of the spectral parameters during one tone.

The proposed protocol may be considered as front end for music pedagogy related softwares that intend to provide feedback to the performer. Although this study only considers alto saxophone recordings, the evaluation protocol intends to be as generic as possible and may surely be considered for wider range of classical instruments from winds to bowed strings.

**Keywords:** music education, performer skills evaluation, sinusoidal modeling.

## 1. Introduction

Several parameters could be extracted from a musical performance. The works of Langner & al [1] or Scheirer [2] explain how to differentiate piano performances using velocity and loudness parameters for example. Studies presented by Stamatatos & al [3, 4] use differences found in piano performances to recognize performers. We propose here a method to evaluate the technical level of a musical performer by analyzing non expressive performances, as scales. Our results are based on the analysis of alto saxophone performances, however the same approach can be used with other instruments.

Before us, Fuks [5] explains how the exhaled air of the performer can influence the saxophone performance, and Haas [6] propose with the SALTO system to reproduce the physic influence of the saxophone instrument on the performance. Here we do not want to consider the physic behavior of the instrument, or what is influenced by the physiology of the performer. Since the spectral envelop is strongly dependent to the physics of the couple instrument / instrumentalist, this kind of observations can not be considered.

On contrary, the long-term evolution of the spectral parameters over time reflects the ability of the performer to control its sound production. Even if this ability is only one

aspect of saxophone technique, it appears to be strongly correlated to the overall technical level in an academic context. Moreover, we will show in this paper that considering this evolution is relevant to evaluate performers over wide range, from beginners to experts.

We think this work could be useful in music education. A software that can automatically evaluate the technical level of a performer would be a good feedback of his progress, especially when the teacher is not present. This kind of software will surely be welcome in music schools since music teachers we met during the recordings were very exciting with this idea. Other projects are already on the same way, as IMUTUS [7, 8], the Piano Tutor [9], and the I-MAESTRO [10] projects.

After presenting the sinusoidal model for sound analysis in Section 2, we explain in Section 3 the experiment protocol we used to record the 30 alto saxophonists playing long tones, exercises usually played by instrumentalists as scales. We also detail how and why the music exercises have been chosen, and how the database has been built from the recordings.

Conclusions of our study are based on the analysis of this database, using metrics to evaluate the musical performance. These metrics proposed in Section 4 are defined to correspond to the perceptive criteria of technical level commonly used by music teachers to evaluate the quality of the produced sound. The results presented in Section 5 show how these metrics are well-suited to evaluate the technical level of a performer.

## 2. Sinusoidal Model

Additive synthesis is the original spectrum modeling technique. It is rooted in Fourier's theorem, which states that any periodic function can be modeled as a sum of sinusoids at various amplitudes and harmonic frequencies. For stationary pseudo-periodic sounds such as saxophone tones, these amplitudes and frequencies continuously evolve slowly with time, controlling a set of pseudo-sinusoidal oscillators commonly called *partials*. This representation is used in many analysis / synthesis programs such as AudioSculpt [11], SMS [12], or InSpect [13].

Formally, a partial is composed of three vectors that are respectively the time series of the evolution of the frequency,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.  
© 2006 University of Victoria

linear amplitude, and phase of the partial over time:

$$P_k = \{F_k(m), A_k(m), \Phi_k(m)\}, \forall m \in [b_k, \dots, b_k + l_k - 1]$$

where  $P_k$  is the partial number  $k$ , of length  $l_k$ , and that appeared at frame index  $b_k$ .

To evaluate the technical level of a performer, its performance is recorded following a protocol detailed in the next section. From these recordings, the partials are extracted using tracking algorithms [14]. Since the protocol proposed in this article is intended to evaluate the technical level of wind and bowed string instrument performer, the frequencies of the partials that compose the analyzed tones are in harmonic relation and the evolution of the parameters of these partials are correlated [15]. Consequently, only the fundamental partial is considered for the computation of the metrics proposed in Section 4.

### 3. Experiment Protocol

To evaluate the technical level of saxophonists, we ask them to play long tones, exercises they use frequently to warm up, as scales. These exercises are commonly used in music education to improve and evaluate the technical level of a performer, either by the teacher or by the instrumentalist himself. It consists in controlling music parameters as nuance, pitch and vibrato.

Recordings took place in the music conservatory of Bordeaux and in the music school of Talence, France. More than 30 alto saxophonists have been recorded, from beginners to teachers including high technical level students. They played long tones without directive about duration on 6 different notes: low B, low F, C, high G, high D, and high A altissimo. For each note, they executed 5 exercises: first a straight tone with nuance *piano*, a straight tone *mezzo forte* and a straight tone *forte*, respectively corresponding to a sound with low, medium and high amplitude. Then they played a long tone *crescendo / decrescendo*, from *piano* to *forte* then *forte* to *piano*, corresponding to an amplitude evolving linearly from silence to a high value, then to silence again. They ended the exercises by a long tone with vibrato. An example of these exercises with the note C is given by Figure 1.

The sound files were recorded using a microphone SONY ECM-MS907 linked to a standard PC sound card. The chosen format was PCM sampled at 44100 kHz, and quantized on 16 bits. A database containing about 900 files (5 long tones per saxophonist) has been built from the recordings. The fundamental partial has been extracted for each file using common sinusoidal techniques referenced in Section 2.

While comparing the performances of several saxophonists, an important factor to consider is the multiplier coefficient of amplitude from *piano* straight tone to *forte* straight tone, noted  $\alpha$ . Its value depends on the control of the air pressure. A *piano* tone is much more difficult to perform at a very low amplitude. The technical effort to differentiate nuances can therefore affect the results, but increases the  $\alpha$



**Figure 1.** Musical exercises performed by saxophonists during recordings. Here is an example with the note C. It consists in first playing a straight tone piano, then a straight tone mezzo forte and a straight tone forte. Then a long tone crescendo / decrescendo must be played before ending with a long tone with vibrato. There is no directive about duration.

coefficient. In the results presented in Section 5,  $\alpha$  is computed as the ratio between the sum of the amplitudes of all the partials extracted from the *forte* tone versus the amplitudes of all the partials extracted from the *piano* tone.

## 4. Evaluation Metrics

For each exercise presented in the last section, we introduce the metrics that are computed to evaluate the quality of the performance. These metrics consider the evolution of the frequency and the amplitude parameters of the partial  $P_k$  as defined in Section 2. For the sake of clarity, the index  $k$  will be removed in the remaining of the presentation.

### 4.1. The Weighted Deviation

When performing a straight tone, the instrumentalist is requested to produce a sound with constant frequency and amplitude. It is therefore natural to consider the standard deviation to evaluate the quality of its performance.

$$d(X) = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (X(i) - \bar{X})^2} \quad (1)$$

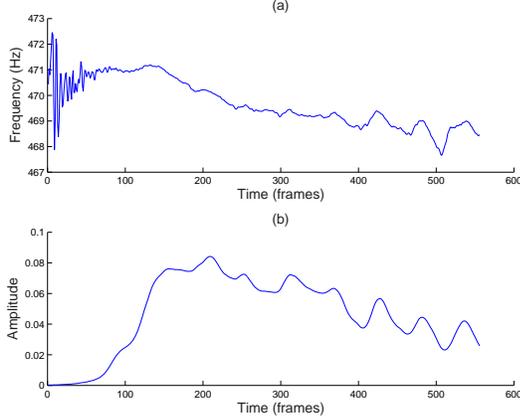
where  $X$  is the vector under consideration of length  $N$ , and the mean of  $X$  is:

$$\bar{X} = \frac{1}{N} \sum_{i=0}^{N-1} X(i) \quad (2)$$

However, if the amplitude is very high, a slight deviation of the frequency parameter will be perceptively important. On contrary, if the amplitude is very low, a major deviation of the frequency parameter will not be perceptible. To perform this kind of “perceptual” weighting, we consider a standard deviation weighted by the amplitude vector  $A$ :

$$\text{wd}(X) = \sqrt{\frac{1}{N \bar{A}} \sum_{i=0}^{N-1} A(i) (X(i) - \bar{X})^2} \quad (3)$$

This weighting operation is also useful to minimize the influence of sinusoidal modeling errors. Due to time / frequency resolution problems, a partial extracted with common partials tracking algorithms is often initiated with a very low amplitude and a noisy frequency evolution before the attack, see Figure 2.



**Figure 2.** Frequency and amplitude vectors of a partial corresponding to the first harmonic extracted using common sinusoidal analysis techniques. Before the attack, the frequency evolution is blurred due to the low amplitude.

This unwanted part could be automatically discarded by considering an amplitude threshold. However, the attack is important to evaluate the performance of an instrumentalist and could be damaged by such a removal. By considering the amplitude weighted version of the standard deviation, we can safely consider the entire evolution of the parameters of the partial.

#### 4.2. Sliding Computation

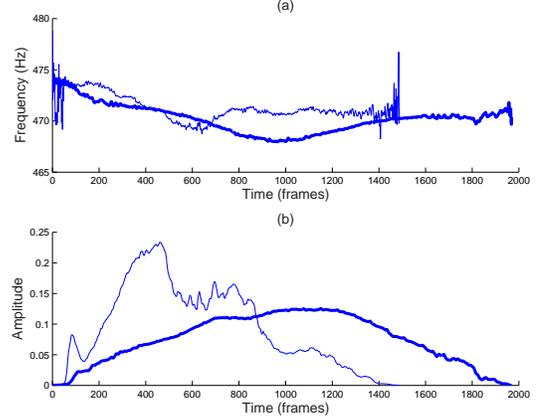
As presented in Section 3, no particular directive about the duration of the tone has been given to performers. Thus, the length of the partial may be different for each instrumentalist. To compare the deviations of multiple performers on a same time interval, we consider a sliding computation of the weighted deviation:

$$\text{swd}(X) = \frac{1}{K} \sum_{i=0}^{K-1} \text{wd}(X[i\Delta, \dots, i\Delta + 2\Delta]) \quad (4)$$

where  $\Delta$  is the hop size and  $K = \lfloor N/\Delta \rfloor$ .

This sliding computation is also useful to consider a mean value computed on a local basis which leads to a less biased estimation of the deviation.

The choice of the window length is therefore critical. If the length is too small, we will consider very local deviations which are probably non perceptible. On the other hand, if the window is too long, the mean value will be very biased and we will consider global variations. Although these slow variations are not perceived as annoying, they will be penalized. For example in Figure 3, the evolution of the parameters plotted in double solid line would be penalized, however it reflects a good control of the exercise. In the experiments reported in Section 5, we use a window length of 80 ms.



**Figure 3.** Frequency and amplitude vectors of the partials corresponding to the first harmonic of a long tone *crescendo* / *decrescendo* played by two performers. In double solid line, the performer is an expert and in solid line, the performer is a mid-level student.

#### 4.3. Metrics for the Straight Tones

The instrumentalist performing a straight tone is asked to start at a given frequency and amplitude and ideally these parameters should remain constant until the end of the tone. The sliding and weighted deviation can then be considered directly. Since the pitch and the loudness differ between different exercises, we apply a normalization to obtain the following metrics:

$$d_f(P) = \frac{1}{F} \text{swd}(F) \quad (5)$$

$$d_a(P) = \frac{1}{A} \text{swd}(A) \quad (6)$$

#### 4.4. Metric for the Long Tones *crescendo* / *decrescendo*

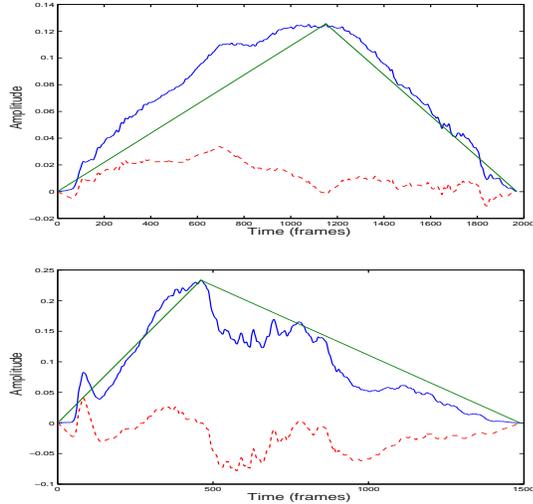
When the instrumentalist performs a long tone *crescendo* / *decrescendo*, the amplitude should start from an amplitude close to 0, linearly increases to reach a maximum value  $M$  at index  $m$ , and linearly decreases to reach the amplitude close to 0. From the evolution of the amplitude of a partial  $A$ , we can compute the piecewise linear evolution  $L$  as follows:

$$L(i) = \begin{cases} s_1(i - b) + A(b) & \text{if } i < m \\ s_2(l - b - i) + A(b + l) & \text{otherwise} \end{cases} \quad (7)$$

where  $b$  and  $l$  are respectively the beginning index and the length of the partial  $P$ . The coefficients  $s_1$  and  $s_2$  are respectively the slopes of the linear increase and decrease:

$$s_1 = \frac{M - A(b)}{m - b}$$

$$s_2 = \frac{M - A(b + l)}{l - m + b}$$



**Figure 4.** Amplitude vector  $A$  and piecewise linear vector  $L$  of a partial for two long tones crescendo / decrescendo . The difference between the two vectors is plotted with a dashed line. On top, the performer is an expert and at bottom, the performer has a mid-level.

Two examples of the difference between  $A$  and its piecewise linear version  $L$  are shown in Figure 4.

As a metric, we consider the sliding weighted deviation of the difference between the amplitude of the partial  $A$  and a piecewise linear evolution ( $L$ ). Since the objective of the exercise is to reach a high amplitude from a low amplitude, we propose to weight the deviation as follows:

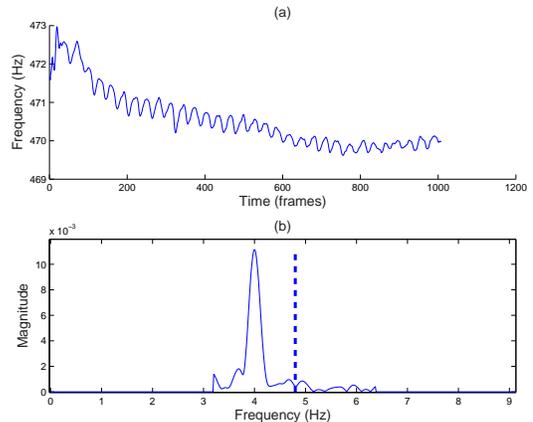
$$d_{<>}(P) = \frac{1}{(M - \min(A))} \text{swd}(A - L) \quad (8)$$

#### 4.5. Metrics for the Vibrato Tones

When performing a vibrato tone, the frequency should be modulated in a sinusoidal manner. The evolution of the frequency during the vibrato is plotted on Figure 5. As the classical saxophone vibrato is commonly taught using 4 vibrations by quarter note with 72 beats per minute, we fix that the frequency of the sinusoidal modulation should be close to 4.8 Hertz. The amplitude of the vibrato should remain constant for all the tone duration. We therefore consider these two criteria to evaluate the performance of an instrumentalist in the case of a vibrato tone.

We estimate the evolution of the frequency and the amplitude by performing a sliding time spectral analysis of the frequency vector  $F$ . For each spectral analysis, we consider a time interval equivalent to four vibrato periods at 4.8 Hertz, a Hanning window and a zero-padded fast Fourier transform of 4096 points.

At a given frame  $i$ , the magnitude and the location of the maximal value of the power spectra respectively estimate



**Figure 5.** Frequency vector  $F$  of a vibrato tone. At bottom, the spectrum of the vector  $F$  is plotted in solid line and the vertical dashed line is located at 4.8 Hz. The difference between the frequency location of the maximal value of the spectrum and this frequency is one of the metrics considered for the vibrato tones.

the amplitudes  $VA(i)$  and the frequencies  $VF(i)$  of the vibrato of the partial  $P$ . We seek for this maximal value in the following frequency region: [3.2, 6.4] Hz.

The first metric  $d_{vf}$  for vibrated tones is defined as the difference between the mean value of  $VF$  and the reference frequency 4.8 Hz, see Figure 5. The second one,  $d_{va}$ , is defined as the standard deviation of the amplitude of the vibrato over time:

$$d_{vf}(P) = \|\overline{4.8 - VF}\| \quad (9)$$

$$d_{va}(P) = d(VA) \quad (10)$$

## 5. Results

For each sound, the metrics presented in the last section are computed from the evolution over time of the parameters of the fundamental partial. For convenience, the values computed using these metrics are converted in marks.

### 5.1. Conversion from Metrics to Marks

The technique of an instrumentalist is principally evaluated according to the best performers in his class or music school. It is what explains that technical marks are here dependent on the best performances. Indeed, this dependence respects the technical difficulties of the instrument. Even for an expert saxophonist, playing a low B *piano* is very difficult, because of the physics of the instrument. Relative evaluation, instead of absolute one, allows to evaluate the performance without being influenced by the instrument itself. We have chosen the *confirmed* class as mark reference (mark 100). It groups high level students and teachers, and contains 7 elements. Although the *experts* class could be a better reference due to the better marks obtained by its elements, it does not contain enough elements (3).

Amplitude results						
	$\alpha$	p	mf	f	<>	vibrato
experts (3)	17	55 (8)	105 (40)	<b>122</b> (67)	126 (24)	114 (150)
confirmed (7)	11	100 (22)	100 (33)	<b>100</b> (10)	100 (28)	100 (98)
mid (6)	7	45 (12)	89 (22)	<b>88</b> (36)	80 (21)	41 (42)
elementary (8)	4	47 (22)	65 (22)	<b>76</b> (19)	44 (7)	100 (90)
beginners (6)	4	32 (16)	49 (22)	<b>65</b> (39)	47 (17)	- (-)

Frequency results						
	$\alpha$	p	mf	f	<>	vibrato
experts (3)	17	106 (49)	115 (54)	100 (47)	169 (68)	92 (18)
confirmed (7)	11	100 (37)	100 (36)	100 (32)	100 (24)	100 (26)
mid (6)	7	54 (19)	71 (19)	74 ( <b>9</b> )	75 (12)	61 (34)
elementary (8)	4	50 (15)	59 (19)	60 ( <b>10</b> )	65 (15)	45 ( <b>7</b> )
beginners (6)	4	42 ( <b>11</b> )	69 (23)	64 (28)	73 (27)	- (-)

**Table 1.** Results for low note F. 5 level classes of performers are represented (with the number of performers by class within parentheses), where the confirmed class is the reference 100 to give marks to individual performances. The results are marks given by class with standard deviation within parentheses. We can notice that the level classes are homogeneous, with reasonable standard deviations, and that the technical marks correspond to the supposed technical level, illustrated for example by the values of the amplitude results for the straight tone forte.

We distinguish amplitude results and frequency results. For the amplitude results we use the metrics defined in Section 4:  $d_\alpha$ ,  $d_{<>}$  and  $d_{va}$  to compute the technical marks for respectively the straight tones, the long tone *crescendo* / *de-crescendo* and the vibrato tone. The marks given as frequency results are computed using metrics  $d_f$ , again  $d_f$ , and  $d_{vf}$  respectively for straight tones, long tone *crescendo* / *de-crescendo* and vibrato tone.

Since these values computed using the metrics introduced in the previous section are errors, we consider as marks the inverse of the values multiplied by 100. These marks are then divided by the mean of the marks obtained by instrumentalists of the *confirmed* class.

## 5.2. Presentation of Results

Saxophonists played long tones and only a few succeed with altissimo high note A. Table 1 shows results for the note F, where  $\alpha$  is the multiplier coefficient of amplitude from *pi-*

*ano* straight tone to *forte* straight tone. *p*, *mf*, and *f* correspond to the straight tones played respectively with low (*piano*), medium (*mezzo forte*) and high (*forte*) amplitude. The tone <> correspond to the long tone *crescendo* / *de-crescendo*, and *vibrato* to the long vibrated tone.

Saxophonists were clustered in five classes (*beginners*, *elementary*, *mid*, *confirmed*, *experts*) according to their academic level validated by school teachers. Marks obtained with the proposed metrics reflect fairly this ranking since level classes are homogeneous, with reasonable standard deviations.

For example, with the long tone *mezzo forte*, *experts* got 105 as amplitude mark, *confirmed* got 100, *mid* 89, *elementary* 65 and *beginners* 49. We can notice that levels under *confirmed* class have big difficulties with the constance of frequency for the *piano* and *mezzo forte* tones. The frequency result for the vibrato seems to be a good criterion to differentiate performers under *confirmed* class, but not over. The amplitude results for the vibrato do not exactly correspond to the supposed technical level of performers. Surely the metrics used to evaluate the quality of the vibrato could be improved in future work.

Results of Marion and Paul are presented in Tables 2 and 3. Marion is a confirmed performer of the music conservatory of Bordeaux, and Paul is a mid-level performer from the music school of Talence. We can infer technical information from marks they got. The results for Marion, given by Table 2, show for example that she respects better the amplitude constraints than the frequency ones. She must be careful with the pitch, especially with low note F and note C.

Paul must work to increase his  $\alpha$  coefficient for the extreme notes of the saxophone, since alto saxophonists can play notes from low B flat to high F sharp, without considering the altissimo notes. He only got a 2 for the  $\alpha$  of high D as shown in Table 3. The same problem appears with the frequency results of his vibrato, that decrease for high note D and low note B.

Thus, with few exercises and the metrics we propose in Section 4, it is possible to evaluate a performer according to confirmed performers, to identify his technical facilities or defaults. It is a good way to increase the technical progress of a performer.

## 6. Conclusion

We have proposed a protocol to evaluate the technical level of saxophone performers. We have shown that the evolution of the spectral parameters of the sound during the performance of only one tone can be considered to achieve such a task.

We introduced metrics that consider this evolution and appear to reflect important technical aspects of the performance. It allows us to automatically sort performers of the evaluation database with a strong correlation with the ranking given by professional saxophonist teachers.

Amplitude results						
note	$\alpha$	p	mf	f	<>	vibrato
low B	5	<b>126</b>	100	81	88	24
low F	4	95	98	112	109	95
C	9	<b>147</b>	108	59	83	150
G	4	117	81	102	58	85
high D	6	<b>141</b>	77	<b>126</b>	106	39

Frequency results						
note	$\alpha$	p	mf	f	<>	vibrato
low B	5	94	<b>65</b>	90	93	64
low F	4	<b>60</b>	73	71	92	51
C	9	104	<b>64</b>	<b>67</b>	<b>53</b>	65
G	4	114	130	98	98	47
high D	6	137	110	189	142	50

**Table 2. Results for Marion, a confirmed performer. The amplitude results of Marion are high, with a good  $\alpha$  coefficient. But she must improve the control of the pitch of notes, regarding to her low frequency results.**

This new protocol may be considered as a front end for music education related softwares that intend to provide feedback to the performer of a wide range of classical instruments from winds to bowed strings.

Additionally, the use of pitch estimation techniques instead of considering the fundamental partial in a sinusoidal model may lead to a better robustness. This issue will be considered for further researches, as the problem of giving a single technical mark to a performer by combining the proposed metrics.

## References

- [1] Jörg Langner and Werner Goebel, "Visualizing Expressive Performance in Tempo-loudness Space," *Computer Music Journal*, vol. 27, no. 4, pp. 69–83, 2003.
- [2] Eric D. Scheirer, *Computational Auditory Scene Analysis*, chapter Using Musical Knowledge to Extract Expressive Performance Information from Audio Recordings, pp. 361–380, Lawrence Erlbaum, 1998.
- [3] Efstathios Stamatatos, "A Computational Model for Discriminating Music Performers," in *Proceedings of the MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, 2001, pp. 65–69.
- [4] Efstathios Stamatatos and Gerhard Widmer, "Music Performer Recognition Using an Ensemble of Simple Classifiers," in *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI)*, 2002, pp. 335–339.
- [5] Leonardo Fuks, "Prediction and Measurements of Exhaled Air Effects in the Pitch of Wind Instruments," in *Proceedings of the Institute of Acoustics*, 1997, vol. 19, pp. 373–378.
- [6] Joachim Haas, "SALTO - A Spectral Domain Saxophone Synthesizer," in *Proceedings of MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, 2001.

Amplitude results						
note	$\alpha$	p	mf	f	<>	vibrato
low B	<b>3</b>	75	89	86	<b>17</b>	29
low F	6	65	94	98	57	111
C	4	28	46	68	22	147
G	5	65	49	76	36	52
high D	<b>2</b>	49	65	89	<b>42</b>	260

Frequency results						
note	$\alpha$	p	mf	f	<>	vibrato
low B	<b>3</b>	88	83	81	82	86
low F	6	47	65	69	69	91
C	4	53	52	68	46	119
G	5	62	59	82	36	63
high D	<b>2</b>	<b>30</b>	47	74	61	<b>38</b>

**Table 3. Results for Paul, a mid-level performer. It appears that Paul must improve his control of pitch and loudness specially playing the lowest and the highest notes of the saxophone. For these notes (here low B and high D), his technical marks decrease and the  $\alpha$  coefficient is low.**

- [7] Erwin Schoonderwaldt, Kjetil Hansen, and Anders Askenfeld, "IMUTUS - an interactive system for learning to play a musical instrument," in *Proceedings of the International Conference of Interactive Computer Aided Learning (ICL)*, Auer, Ed., Villach, Austria, 2004, pp. 143–150.
- [8] Dominique Fober, Stéphane Letz, Yann Orlarey, Anders Askenfeld, Kjetil Hansen, and Erwin Schoonderwaldt, "IMUTUS - an Interactive Music Tuition System," in *Proceedings of the Sound and Music Computing conference (SMC)*, Paris, 2004, pp. 97–103.
- [9] Roger B. Dannenberg, Marta Sanchez, Annabelle Joseph, Robert Joseph, Ronald Saul, and Peter Capell, "Results from the Piano Tutor Project," in *Proceedings of the Fourth Biennial Arts and Technology Symposium*, Connecticut College, 1993, pp. 143–150.
- [10] "I-MAESTRO project," Online.  
URL: <http://www.i-maestro.org>.
- [11] IRCAM, Paris, *AudioSculpt User's Manual*, second edition, April 1996.
- [12] Xavier Serra, *Musical Signal Processing*, chapter Musical Sound Modeling with Sinusoids plus Noise, pp. 91–122, Studies on New Music Research. Swets & Zeitlinger, Lisse, the Netherlands, 1997.
- [13] Sylvain Marchand and Robert Strandh, "InSpect and ReSpect: Spectral Modeling, Analysis and Real-Time Synthesis Software Tools for Researchers and Composers," in *Proc. ICMC*, Beijing, China, October 1999, ICMA, pp. 341–344.
- [14] Robert J. McAulay and Thomas F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.
- [15] Mathieu Lagrange, "A New Dissimilarity Metric For The Clustering Of Partial Using The Common Variation Cue," in *Proc. ICMC*, Barcelona, Spain, September 2005, ICMA.