Improving instrumental sound synthesis by modeling the effects of performer gesture

Marcelo M. Wanderley, Philippe Depalle* and Olivier Warusfel

Ircam - Centre Georges Pompidou - 1, Place Igor Stravinsky - 75004 - Paris - France

mwanderley@acm.org, phd@ircam.fr, warusfel@ircam.fr

Abstract

In this paper we study the effects of performers' gesture on the sound produced by an acoustic instrument as well as the modeling of these effects. More particularly, we propose to focus on the effects of ancillary gestures – those not primarily intended to produce the sound, but nevertheless omnipresent in professional instrumentalists' technique. We show that even in the case of non-expressive performance of wind instruments, performers' ancillary gestures are recognized in the resultant sound by means of strong modulation of sound partial amplitudes. We claim that these modulations account for a naturalness that is usually lacking in current synthesis techniques.

1   Introduction

Control of sound synthesis methods may greatly benefit from the massive corpus of experience developed through centuries of improvement in acoustic instruments' playing techniques. *Expert gestures*, or the fine control achieved by highly-skilled performers, account for a degree of expressiveness that is hardly matched in live computer music. Moreover, apart from performers' *instrumental gestures* [1], the analysis of instrumentalists' performances reveals that *non-obvious –* or *ancillary* – gestures occur to a great extent [2]. These gestures are particularly important, for instance, in the case of wind instruments such as the clarinet, since any movement of the instrument results in sound source(s) displacements.

2   Ancillary gestures

Although there is no clear consensus on why these gestures are performed, it seems obvious that they are present in highly-skilled clarinettists' technique [3].

Figure 1: (Left) *Shot from a performance of Jean-Guy Boisvert playing "Eschroadepipel".* (Right) *Another shot less than one second afterwards.*

For instance, let us consider a video recording of clarinettist J. G. Boisvert playing a piece of Zack Settel, *Eschroadepipel*, performed at Pollack Hall, McGill University on May 18, 1999. Changes in posture and instrument position can be noticed in the two shots of figure 1.

3   Database Sample Analysis

In order to explore the effects of ancillary gestures in the resulting sound for a given acoustical condition (such as microphone position and room response), we have analyzed clarinet samples from Ircam's Studio-on-Line project[1], a sound database of orchestral instruments.
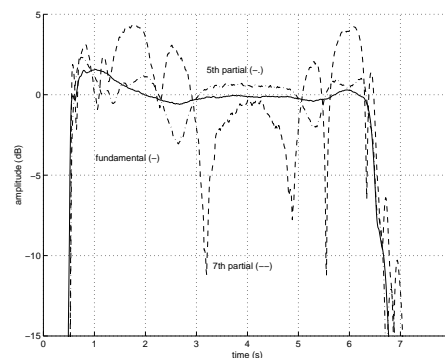

Figure 2: *F3, fortissimo (Player: Pierre Dutrieu).*

Figure 2 represents the normalized amplitude versus time of three partials of an F3 (F0 = 174Hz) played *fortissimo*, an example from that database. An omnidirectional microphone is placed 2 meters away from the instrument, at 2 meters height, and the player is seated. For the sake of clarity, only partials 1, 5 and 7 have been displayed.

In order to explain the origin of these modulations, one must consider the manner in which a clarinet radiates. Since it radiates from multiple tone holes – whose configuration depends on the note being played – interferences between these sources may cause amplitude cancellations. Nevertheless, figure 3 shows that even

---

*Also with Faculty of Music, Université de Montréal, Québec, Canada, e-mail: depallep@ERE.UMontreal.CA

for the case of a D3 (F0 = 147Hz) (all side holes closed) modulations are still present.
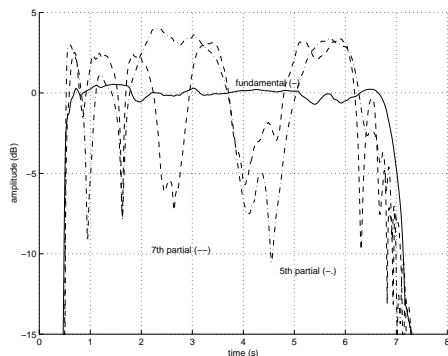

Figure 3: *D3, fortissimo (Player: Pierre Dutrieu).*

Moreover (Cf. figure 4), the analysis of the simultaneous internal microphone recording of the same note shows no similar modulation (amplitude variations do not exceed ± 0.5 dB). This accounts for a certain stability of the *mouthpiece* [4].
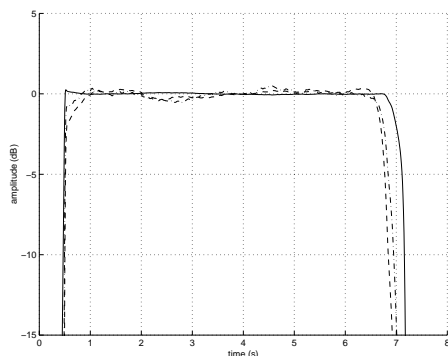

Figure 4: *D3, fortissimo, internal microphone (Player: Pierre Dutrieu).*

4 Further investigation

In order to ascertain whether the amplitude modulations are mainly produced by room reverberation, we first performed recording tests using a clarinet kept immobilized by a mechanical apparatus.
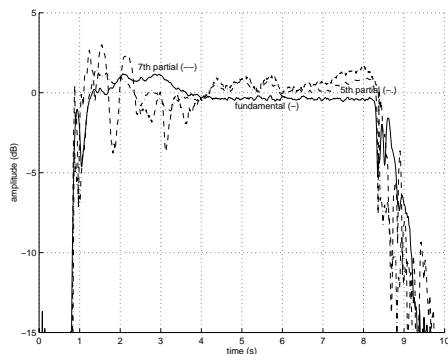

Figure 5: *D3, fortissimo, highly reverberant auditorium, clarinet immobilized (Player: Peter Hoffmann).*

The clarinet was played by a human performer and recorded in a highly reverberant auditorium. Recording conditions were kept similar to those used in the previously described sound data base.

There exist some modulation (Cf. figure 5) starting after the attack of the note up to 4 seconds. Afterwards, however, the amplitudes of partials are rather constant. Thus, it appears that movement is necessary to produce the amplitude modulations observed in figure 2.

Three main hypotheses have been considered in order to explain the effects induced by ancillary gestures: the instrument directivity, the effect of early room reflections and the speed of the movement.

4.1 Instrument directivity

Clarinet's directivity patterns are almost uniform at low frequencies (≤ 700 Hz), but may be very complicated depending on the number of open side holes [5] [6]. This would account for a reduced role of the directivity in the amplitude modulations when all the holes are closed, as for a D3 sample, for instance. Since strong modulations occur at low frequencies and also when all the holes are closed, the directivity of the instrument cannot be the main factor that generates this effect.
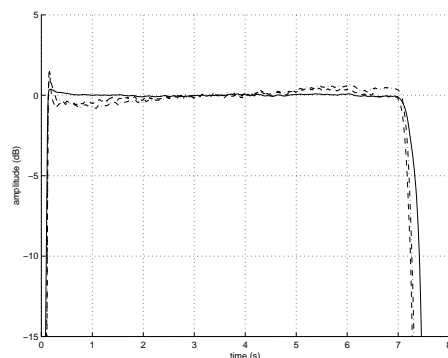

Figure 6: *D3, fortissimo, anechoic chamber, standard performance (Player: Joseph Butch Rovan).*

To further demonstrate that directivity does not play a major role when gestures are of a small amplitude, we have recorded a clarinet in an anechoic chamber. The clarinet was played in three styles : expressive performance (quasi-jazz); non-expressive performance (standard) and performance with the instrument kept completely immobilized by a mechanical apparatus. The analysis of the recordings shows that modulations exist only for the case of large amplitude movements (Cf. figures 6 and 7).

4.2 Early room reflections

Next, these recordings were repeated in the same conditions, except that a wood floor was placed underneath the performer and the microphone[2]. It results in deepened modulations (Cf. figure 8) which prove that the

---

[2]The player was asked to repeat the same movements (bell up), at the beginning and at the end of the note. Nevertheless, during the session, he could not guarantee making exactly the same movement at precisely the same time during the note.
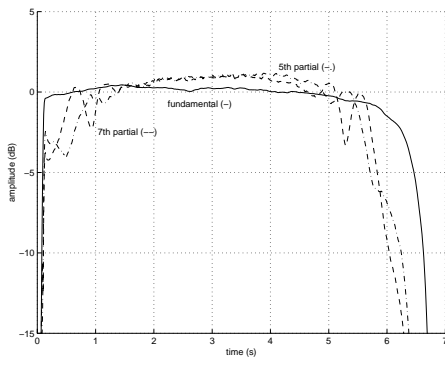
Figure 7: *D3, fortissimo, anechoic chamber, expressive performance (Player: Joseph Butch Rovan).*

first reflection induced by the floor has a great influence on the effect.
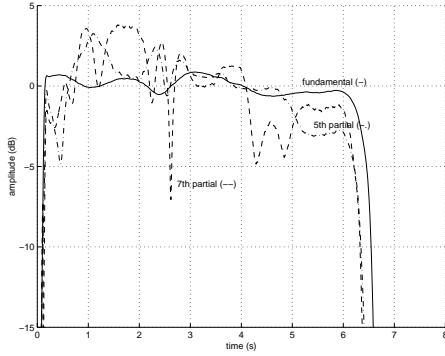

Figure 8: *D3, fortissimo, anechoic chamber with wood floor, expressive performance (Player: Joseph Butch Rovan).*

4.3 Speed of movement

Finally we have experimented with the possible influence (such as a Doppler effect) of the gesture speed on the amplitude's modulations. For each note played, the performer was using three durations (8.5 s, 6.5 s and 5 s), while rising the clarinet's bell over the same distance. This results in three different speeds which we designate respectively as slow, normal and fast.
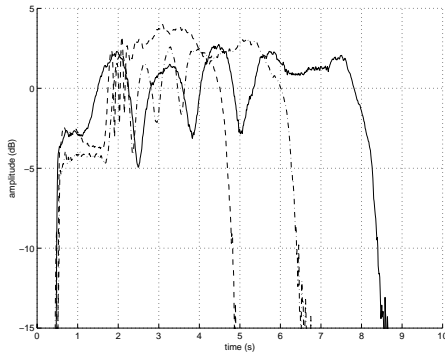

Figure 9: *D3, fortissimo, anechoic chamber with wood floor, 7th partial played at slow (solid line), normal (semi-dashed line) and fast (dashed line) speed (Player: Peter Hoffmann).*

Figure 9 presents the influence of the speed on the 7th partial of a D3. It shows that there is no significant effect of the gesture speed on the amplitude's modulations. As expected speed change results only in a simple time-stretching of the time evolution of the amplitude of the partials, the source speed being of a magnitude order (1m/s) that can be neglected when compared to the speed of sound.

5 Simulation

As the first reflection mostly influences the modulation effects, we decided to simulate it by using a simple system (Cf. figure 10) which models propagation of direct sound and of the first reflection. The transfer function of the system is ($0 \leq \rho_1 \leq \rho_2 \leq 1$):
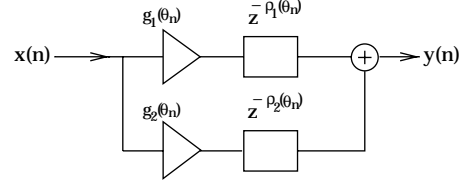

Figure 10: *Model of the first reflection effect for a close microphone (θ stands for the orientation angle of the clarinet.)*

$$H(z) = g_1 z^{-\rho_1}\left(1 + \frac{g_2}{g_1} z^{-(\rho_2 - \rho_1)}\right) \qquad (1)$$

where the second term of the right part of Eq. 1 represents an inverse comb filter $H_c(z)$:

$$H_c(z) = 1 + \alpha z^{-p} \qquad (2)$$

with $\alpha = \frac{g_2}{g_1}$ and $p = \rho_2 - \rho_1$.

Parameters of the model may vary according to the angle θ measured between a vertical line and the direction of the clarinet.
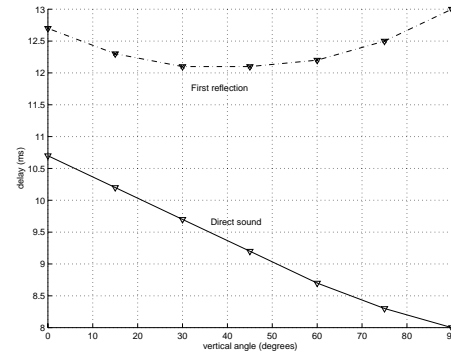

Figure 11: *Delay of direct sound and first reflection for the room response to a clarinet with all side holes closed.*

In order to provide the model with realistic gain and delay functions $g(\theta)$ and $\rho(\theta)$, impulse responses have been measured in the same auditorium where the original recordings were made. The clarinet was excited by a loudspeaker connected to the clarinet barrel, all side holes closed. From these impulse responses the arrival time of the direct sound and floor reflection together with their respective amplitudes could be obtained for each clarinet orientation. Figure 11 shows the delays obtained for direct sound and first reflection.
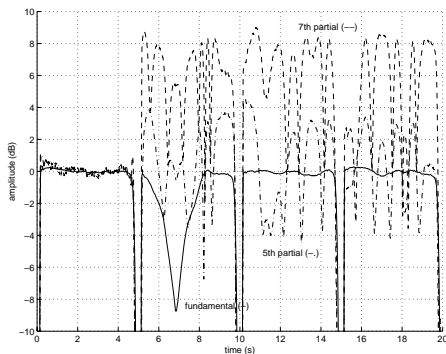
Figure 12: *Simulation using model $H(z)$. Original sample by Joseph Butch Rovan (0 to 5s) followed by three arbitrary movements ($[5-10]$, $[10-15]$, and $[15-20]$ seconds).*

A real-time implementation of the model has been implemented in the FTS environment[3]. Its input is a recording of an immobilized clarinet played in an anechoic chamber (figure 12, $[0-5]$ seconds). A slider allows the simulation of one-dimensional movements, thus changing accordingly the amplitude and delay of the direct sound and of the first reflection. Some experimental results are shown in figure 12, where three different slider movements were performed ($[5-10]$, $[10-15]$, and $[15-20]$ seconds).

5.1 Remarks

The difference between the measured delays of the direct sound and the floor reflection may also support some quantitative investigation in order to explain previous observations.

For instance, it may be observed on different figures presenting modulation effects on a D3 sample that the first partial is not significantly disturbed. Using the data shown in figure 11 one can notice that the delay difference varies from 2.0 ms to 5.0 ms for a clarinet angle of 0 to 90 degrees, respectively. This corresponds to a comb filter whose first zero ranges from 250 Hz to 100 Hz.

For a realistic movement, the clarinet angle will seldom reach or exceed 60 degrees [3]. Moreover, for a seated clarinettist the angle range is even further reduced. Hence, we can reasonably consider the zero range to be limited to values greater than 200 Hz. This means that the first partials of both D3 and F3 samples will not be affected, since this value is above their fundamental frequencies.

On the contrary, during the simulation presented in figure 12, the input angle could easily be spanned through the full range, thus creating the dip that appears on the first slider movement ($[5-10]$ seconds).

As a last remark, considering that the samples' recording conditions throughout this research comply with the standard clarinet recording procedures suggested in the literature (Cf. [4]), and also that a clarinet player will most likely produce ancillary gestures during a performance, it is reasonable to expect that, in these circumstances, modulations will be an integral part of the recorded sound. Consequently, the use of a model of such effects may naturally improve standard sound synthesis techniques.

6 Conclusions

We have presented a study of the influence of performers' ancillary gestures in the sound produced by a clarinet. These gestures, which have an undeniable visual impact during performances, are part and parcel of top instrumentalists' technique.

We have shown that ancillary gestures also affect sound production and may generate strong sound modulations which are perceived as beating or phasing-like effects. It seems to us that these modulations account for a naturalness that is usually lacking in current synthesis methods. Also, it could explain the extensive use of flanging and phasing devices in sound mixing techniques.

In order to identify the causes of these modulations, we have recorded an extensive set of clarinet sounds in different reproducible environments, such as a variable acoustic auditorium and an anechoic chamber. We have shown through analysis and measurements that these modulations are primarily caused by the influence of the performer's gestures coupled with the room's characteristics, especially the early reflections.

Finally, a real-time model simulating the effect of the floor reflection has been implemented, allowing the production of similar effects depending on an angle parameter which simulates the orientation of the instrument.

Aknowledgements

References

[1] C. Cadoz and M. Wanderley. Gesture - music. In M. Wanderley, M. Battier, and J. Rovan, editors, *Trends in Gestural Control of Music*. Ircam, 1999. To appear.

[2] F. Delalande. La Gestique de Glenn Gould. In *Glenn Gould Pluriel*, pages 84–111. Louise Courteau Éditrice, 1988.

[3] M. Wanderley. Non-Obvious Performer Gestures in Instrumental Music. In *Proc. of the III Gesture Workshop*, Gif-sur-Yvette, 1999. Springer Verlag. To appear.

[4] A. H. Benade. From Instrument to Ear in a Room: Direct or Via Recording. *J. Audio Eng. Soc.*, 33(4), April 1985.

[5] J. Meyer. *Acoustics and the Performance of Music*. Verlag das Musikinstrument, Frankfurt/Main, 1978.

[6] C. Lheureux. Simulation et mesure du rayonnement de différents instruments à vent à trous latéraux. Master's thesis, Ircam, 1997.

---

[3]http://www.ircam.fr/produits/logiciels/log-autres/