GESTURALLY-CONTROLLED DIGITAL AUDIO EFFECTS

Marcelo M. Wanderley and Philippe Depalle

Faculty of Music - McGill University
555, Sherbrooke Street West
H3A 1E3 - Montreal - Quebec - Canada
mwanderley@acm.org, depalle@music.mcgill.ca

ABSTRACT

This paper presents a detailed analysis of the acoustic effects of the movements of single-reed instrument performers for specific recording conditions. These effects are shown to be mostly resulting from the difference between the time of arrival of the direct sound and that of the first reflection, creating a sort of *phasing* or *flanging* effect. Contrary to the case of commercial flangers – where delay values are set by a LFO (low frequency oscillator) waveform – the amount of delay in a recording of an acoustic instrument is a function of the position of the instrument with respect to the microphone. We show that for standard recordings of a clarinet, continuous delay variations from 2 to 5 ms are possible, producing a naturally controlled effect.

1. INTRODUCTION - ANCILLARY GESTURES

Musicians constantly perform movements – or *gestures* – that are not directly related to sound production [1]. These gestures have been called *expressive*, *accompanist*, *ancillary* or *non-obvious*. For the case of a woodwind instrument performer, these movements can consist of postural adjustments, upwards/downwards movements of the instrument, and circular patterns, among others [2]. Although there is no clear consensus on the origin of these gestures, it seems obvious that they are present in skilled performer's technique [3] [4], and are dependent on several factors, therefore presenting different movement levels [5].

2. CLARINET PERFORMER'S ANCILLARY GESTURES

The detailed study of several clarinet player's ancillary gestures is presented in [5], where the first author used an Optotrak 3D Infrared tracker system[1] to measure clarinetists' movements playing several solo and chamber pieces, both classical and contemporary.

The pieces were performed with the player standing and seated, and with different expressive characteristics: a) expressive, b) standard and, c) with the player consciously trying not to move the instrument. An example of typical expressive gestures can be seen in figure 1, showing the vertical movement of the clarinet bell for a subject performing Poulenc's first clarinet sonata.

[1] In collaboration with the Free University of Amsterdam and the NICI, Nijmegen, the Netherlands.

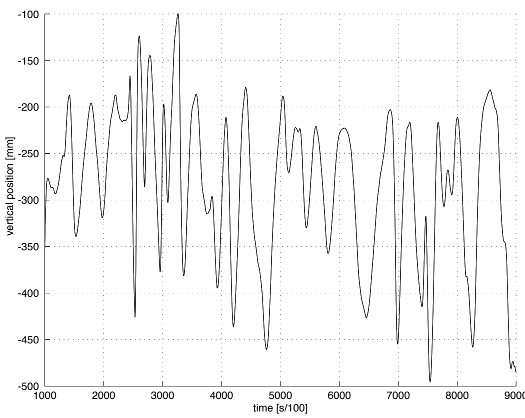

Figure 1: *Subject performing Poulenc's Clarinet Sonata, first movement (excerpt). Vertical bell position over time.*

One can notice from the analysis of figure 1 that movements of the instrument are constantly produced throughout the performance, with a maximum amplitude range of the vertical movement of the instrument's bell of 40 centimeters.

Figure 2 shows different gestures and postures of a performer in a series of still images[2] taken from a video of an expressive performance of a contemporary piece for solo clarinet. Note the various postures and the different angles of the instrument with respect to the performer.

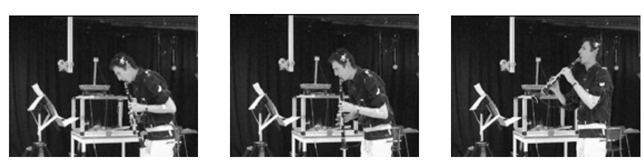

Figure 2: *Three photographs showing a subject performing (expressively) an excerpt of a contemporary piece –* Domaines, cahier A, *by Pierre Boulez.*

Furthermore it can be shown that a performer will tend to reproduce the same movements when playing a piece several times [6]. This therefore indicates that these expressive movements are an integral part of the performance, not simply a visual effect or produced randomly.

[2] with total duration of 1 second.

An example can be seen in figure 3, where a second performer plays the same piece three times. Note the striking consistency on both the spatial movements and their timing[3].

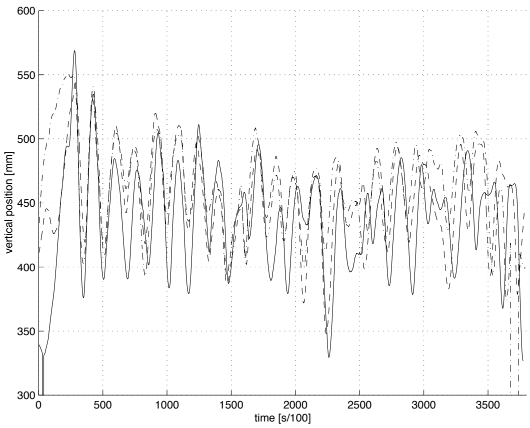

Figure 3: *Three performances of Brahms Clarinet Sonata, first movement (excerpt), by another performer. Vertical bell position over time.*

3. ACOUSTICAL EFFECTS OF PERFORMER MOVEMENTS

It is interesting to note that performer movements – for the case of woodwind instruments[4] – may influence the sound produced and recorded under close microphone conditions.

For instance, considering the case of a clarinet, for standard recording conditions [7], movements of the instrument will cause significant amplitude modulations (and even cancellations) of sinusoidal sound partials due to the displacement of the sound source (the open holes) with respect to the microphone [8]. In the same reference, we have presented a detailed report of the analysis of several clarinet samples recorded in various acoustically controlled conditions, including an anechoic chamber. This was to investigate and evaluate the effects of ancillary performer gestures on the timbre of the instrument. We have shown that the influence of ancillary gestures mostly results from the reflection off the floor, as compared to variations in the mouthpiece, directivity effects, or speed of performer movements.

The floor reflection, which is, in this case, the first reflection of the room reverberation, interferes with the direct sound of the clarinet. This effect can be represented by a simple model consisting of two delay lines each one including a variable delay ρ (expressed in samples), and a variable gain g. The first (characterised by ρ_1, g_1) represents the propagation of the direct sound, while the second one (ρ_2, g_2, with $\rho_1 < \rho_2$) represents the propagation of the sound that reflects off the floor. For a fixed position of the clarinet, the transfer function $H(z)$ of this model (cf. figure 4) can be written as:

$$H(z) = g_1 z^{-\rho_1} + g_2 z^{-\rho_2} \tag{1}$$

[3]Extra comparisons of performances by other musicians and the similarities and differences between the performances of different musicians are presented in [6].

[4]and any other instrument for which sound sources move with performer gestures, such as strings, brass, etc...

that factorises into:

$$H(z) = g_1 z^{-\rho_1} H_c(z) \tag{2}$$

where H_c is a comb filter

$$H_c(z) = 1 + \alpha z^{-D} \tag{3}$$

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

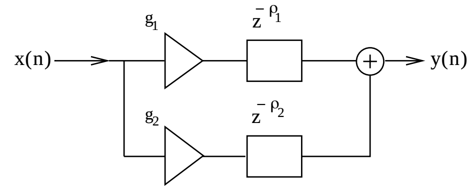

Figure 4: *Symbolic representation of the two-path acoustical propagation system.*

The magnitude of the frequency response of such a system exhibits an interleaved structure of evenly spaced soft peaks at frequencies $f_p(k) = k\frac{1}{D}$ (k being an integer), and sharp dips at frequencies $f_d(k) = (k - \frac{1}{2})\frac{1}{D}$. As an example, assuming that the amplitudes of the direct sound and of the first reflection are equal (i.e. $\alpha = 1$), and that the delay difference $D = \rho_2 - \rho_1 = 106$ samples (i.e. 2.4 ms at a sampling rate of 44,100 Hz, which represents a distance difference of 0.792 m), one can plot the frequency response shown below where zeroes are distributed on odd harmonic locations of $f_d(1) = 208$ Hz, while poles lie on harmonic locations of $f_p(1) = 416$ Hz.

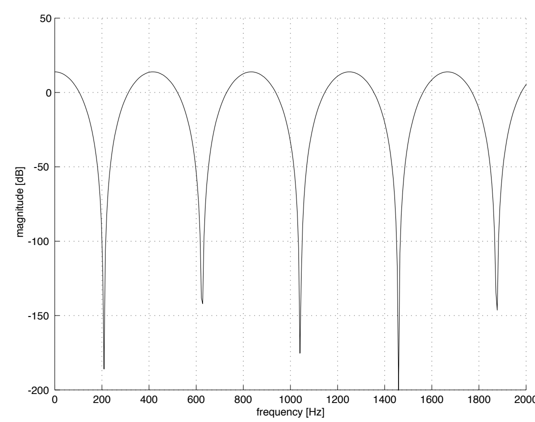

Figure 5: *Frequency response of the two-path system for a delay difference of 2.4 ms and* $\alpha = 1$.

There are several factors that influence the specific values of the two gains g, and the two delays ρ. For a refined model, one has to take into account the radiation of the sound from the clarinet, the distances travelled by the two waves, the losses due to the propagation through the air, the acoustic absorption when reflecting off the floor, and the characteristics of the receiver (the microphone). The performer controls two parameters (the note and the orientation angle of the clarinet) that modify each of these aforementioned factors as follows:

- The radiation in the air depends on the frequency (the directivity patterns are almost omnidirectional at low frequencies but become more complicated at higher frequencies depending on the configuration of opened/closed clarinet holes). It also depends on the angle, as the radiation is far from being isotropic, except at low frequencies[5]. An example can be seen in figure 6, where the directivity pattern at frequency $f = 2,352$Hz (the eight partial of a D4) is ploted three times, each one for a different angle of the instrument. Note the complex structure of the pattern that demonstrates the clarinet to be fairly unidirectional at this frequency. Therefore, for different angles, the radiated amplitude will take very different values.
- The distance depends mainly on the angle of the instrument, but also on which hole principally radiates the sound, since most of the radiation occurs in the first few open holes [9].
- The propagation losses depend on the frequency of the wave (the note), and on the distance.
- The absorption when reflecting off the floor depends on the angle, and on the frequency.

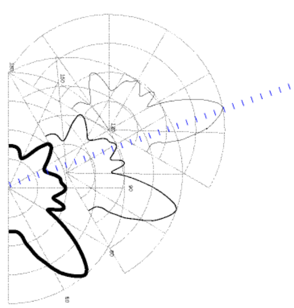

Figure 6: *Directivity pattern at the frequency $f = 2,352$Hz ploted for three clarinet angles with respect to the mouthpiece.*

A precise representation of the effect should therefore take all of these factors into account in order to control the model. In this study, we use a phenomenological model which implicitly takes these factors into account by using measured gains and delays. These factors are going to be implemented in further explicit models.

4. GAIN AND DELAY MEASUREMENTS

The gain and delay parameters of the model are determined using experimental measurements through the estimation of concert hall's impulse responses.

To achieve this goal, we have used standard techniques for impulse response estimations. We present the experimental details regarding the auditorium and recording techniques that ensure that the relevant part of the impulse response is consistent with our approach.

[5]However, we have to quote that according to our knowledge, no directivity patterns are available for near field conditions. The information available in the scientific literature addresses the far field case.

4.1. *Espace de Projection* - Acoustical details

The concert hall used for the project is the *Espace de Projection*, located at IRCAM, which allows for the modification of its acoustic configuration through the choice of reflective, diffusive or absorbent panels in the walls and ceiling, and the specification of its total volume.

In the present case, the auditorium has a volume V of approximately 3,430 m^3, a 60-dB reverberation time T at 1 kHz varying from 1.258 seconds (totally absorbent) to 3.018 seconds (totally reflecting), depending on the configuration of the acoustic panels in the walls and ceiling [10].

For our measurements, the room was configured with one-third reflective and two-thirds absorbent panels. Figure 7 shows the impulse response of the global system (clarinet, microphone, and auditorium) for the chosen configuration (cf. section 4.2).

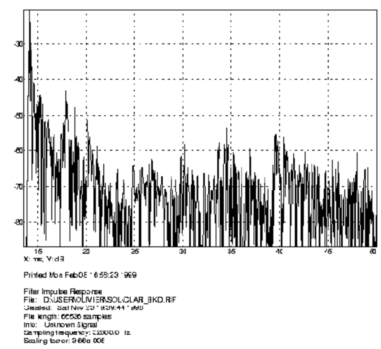

Figure 7: *Auditorium measurement - TR = 1.4760s @ 1kHz. First 50 ms.*

In order to verify the validity of the measurements, we have to be sure that the sound source behaves as specified in [11]. In this case, the sound pressure amplitude p_a of a sound captured by a microphone placed in the immediate vicinity of a sound source with strength S, frequency f, normalized directivity Δ is:

$$p_a(r) = (S\rho f)\frac{\Delta(r,\theta)}{2r} \tag{4}$$

where ρ is the density of air in the room.

Equation 4 is valid when both source and microphone are more than half a wavelength away from the walls, and when the distance between source and microphone, r, is much smaller than a critical distance r_c, given by:

$$r_c = 0.0565\Delta(r_c,\theta)\sqrt{(\frac{V}{T})} \tag{5}$$

where V is the room volume and T the 60-dB reverberation time.

Considering then a source with $\Delta(r,\theta)$ = 6.00 dB at 1 kHz and $T = 1.4760s$, r_c equals to 16.34 meters[6].

Since the clarinet recordings we have analyzed thus far have been realised with a microphone placed 2 meters away from the instrument and several meters away from the walls, we can reasonably consider that we are in the case described by equation 4.

[6]Obviously, a radiating source comparable to a musical instrument would not have such a high value of $\Delta(r,\theta)$, thus reducing the value of r_c.

4.2. Measurements

For these measurements, the sound was generated by using a loudspeaker connected to a clarinet tube, all side holes closed. The temporal response of the global system (clarinet, microphone and auditorium) was recorded for several clarinet orientation angles, as shown in figure 8.

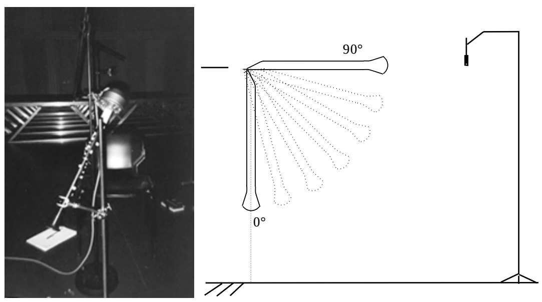

Figure 8: *Room response measurements with excitation provided by a loudspeaker connected to a clarinet tube.*

Table 1 shows the values for the gain and the delay of the direct sound and of the first reflection, measured by a microphone at 2 meters away from the mouthpiece, at a height of 2 meters.

Angle [degrees]	g_1 [dB]	g_2 [dB]	ρ_1 [ms]	ρ_2 [ms]
0	-49	-47	10.7	12.7
15	-48	-45	10.2	12.3
30	-47	-44	9.7	12.1
45	-46	-43.5	9.2	12.1
60	-43.5	-43.5	8.7	12.2
75	-42	-44.5	8.3	12.5
90	-39	-47	8.0	13.0

Table 1: *Measurements of gain and time delay for both the direct sound and the first reflection recorded with a microphone 2 meters away from the mouthpiece of the instrument.*

Figure 9 shows the delays obtained for the direct sound (ρ_1) and for the first reflection (ρ_2), under the conditions described above.

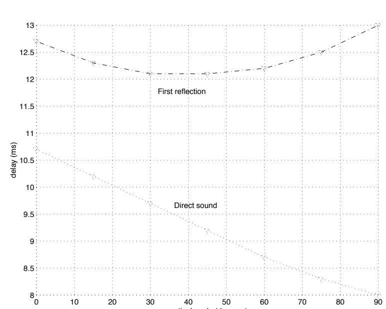

Figure 9: *Delay of direct sound (ρ_1) and first reflection (ρ_2) measured in the auditorium excited by the experimental device shown in figure 8.*

When moving the clarinet tube from the horizontal to the vertical position, the delay difference evolves from 2 ms to 5 ms, which generates a harmonic structure of zeroes in the spectrum, the fundamental frequency of which decreases from 250 Hz down to 100 Hz. For sound whose partial frequencies coincide with the positions of the zeroes of the system, a strong attenuation will be noticed. The same will also happen for the odd multiples of these frequencies.

Considering that the samples' recording conditions throughout this research comply with the standard clarinet recording procedures suggested in the literature (cf. [7]), and also that a clarinet player will most likely produce ancillary gestures during a performance (cf. [2] [5]), it is reasonable to expect that, in these circumstances, modulations are an integral part of the recorded sound.

5. REAL-TIME SIMULATION

A real-time implementation of the model presented in figure 4 has been performed in jMax, IRCAM's Linux/IRIX real-time synthesis and audio processing environment.

The sound input $x[n]$ to the model shown in figure 4 is a 5-second musical excerpt recorded in an anechoic chamber. We then simulate five different 5-second angular movements with a slider that controls the orientation angle. This angle is used for table look-up of gain and delay values for the direct sound and first reflection, as shown in table 1.

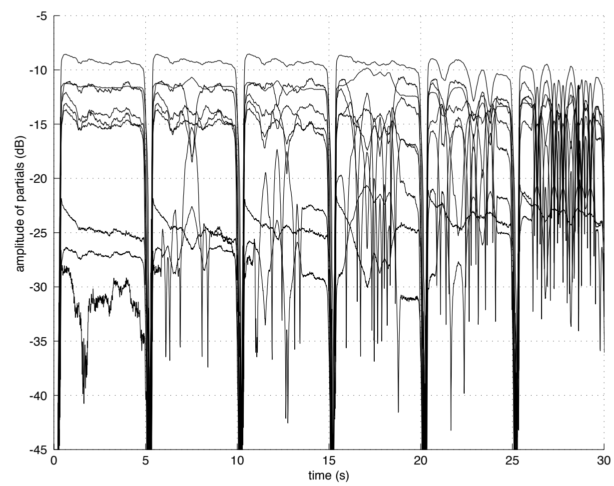

Figure 10: *Evolution of partials amplitudes for simulated motions applied to an original anechoic room sample ([0 - 5] seconds) – D3* ff *standard performance. Arbitrary movements with increasing amplitudes were performed at ($[5-10]$, $[10-15]$, $[15-20]$, $[20-25]$, and $[25-30]$ seconds).*

This results in a timbre modulation that sounds similar to a flanging effect. This effect, which is often used in recording studios, consists of adding to a signal a slightly delayed copy of itself. This constitutes a comb filter structure that is very similar to the two-path acoustical propagation system presented in section 3.

By changing the delay, one makes the dips sweep over the spectrum of the input signal, causing a very recognisable sound effect. Commercial flangers control the delay variations through the use of a LFO (low frequency oscillator) waveform and present typical delay values evolving between 1 and 10 ms [12]. As these periodic variations may be perceived to be repetitive, some authors have proposed improvements by adding random variations to the LFO waveform [13].

Considering the structural analogy presented above, it seems that a further improvement in the control of flanger effects is to modify its delay and gain (or depth) parameters by performer gestures that naturally occur during instrumental performances. These gestures imply variations that are neither too repetitive nor random, and are tightly related to musical events being performed.

The amplitude modulation effect on sound partials can also appear in other circumstances, such as a beating effect in instruments having several slightly detuned strings associated to the same note, as in the case of a piano. Conversely, a similar modulation effect can be produced by a fixed comb filter applied to a time-varying spectrum, as in the case of sound coloration in auditoriums [14]. The lack of such modulation in electronic sounds, in electric instuments, or in "sanitized" sounds recorded in absorbing rooms, is likely to explain the success of flanger devices in modern studio technology.

6. CONCLUSIONS

We have shown that performer's expressive gestures affect sound production and generate strong sound amplitude modulations that are perceived as a flanging effect with delay amounts continuously dependent on the position of the instrument with respect to a close recording microphone. Continuous delay variations from 2 to 5 ms were measured for a microphone 2 meters away from the instrument and applied in a first-order model of the effect used in real-time simulations.

It appears that this modulation accounts for a naturalness that is often lacking in current synthesis methods. The correlation between performer gestures and musical parameters such as tempo, articulation, etc... opens up new possibilities regarding the control of digital audio effects, for which ancillary movements may provide coherent relationships between this control and the musical interpretation.

7. ACKNOWLEDGEMENTS

Part of this research was performed while the first author was a doctoral candidate at IRCAM, thanks to funding from CNPq, the Brazilian Research Council.

The acoustical measurements were performed in collaboration with Gérard Betrand and Federico Cruz-Barney. Olivier Warusfel largely contributed to this research through his technical expertise on the recording setup and on the analysis of the acoustical model of the influence of performer's movements. The far field directivity pattern from figure 6 is courtesy of René Caussé. Thanks to Stéphan Tassart for various discussions and suggestions; to Erwin Schoonderwaldt for help with the measurements of movements at the Free University of Amsterdam and the NICI; to Peter Beek and Peter Desain for the use of the Optotrak system; and to all performers by their participation.

Thanks to Geoff Martin who proofread this article and provided several comments.

8. REFERENCES

[1] A. Gabrielsson. "Music Performance." In D. Deutsch, ed. *The Psychology of Music*, 2nd edition, pp. 501–602, 1999.

[2] M. M. Wanderley. "Non-Obvious Performer Gestures in Instrumental Music." In A. Braffort et al., eds. *Gesture Based Communication in Human-Computer Interaction*, Springer-Verlag, pp. 37–48, 1999.

[3] F. Delalande. "La gestique de Gould." In *Glenn Gould Pluriel*, Louise Courteau, editrice, pp. 85–111, 1988.

[4] J.-W. Davidson. "Visual Perception of Performance Manner in the Movements of Solo Musicians." *Psychology of Music*, vol. 21, pp. 103-113, 1993.

[5] M. M. Wanderley. "Quantitative Analysis of Non-Obvious Performer Gestures." In *Proc. of the IV Gesture Workshop*, London, April, 2001.

[6] M. M. Wanderley. *Performer-Instrument Interaction: Applications to Gestural Control of Sound Synthesis*. PhD Thesis, Université Pierre et Marie Curie - Paris VI, 2001.

[7] A. H. Benade. "From Instrument to Ear in a Room: Direct or via Recording." *J. Audio Engineering Society*, 33(4), 1985.

[8] M. M. Wanderley, P. Depalle and O. Warusfel. "Improving Instrumental Sound Synthesis by modeling the Effects of Performer Gesture." In *Proc. of the 1999 International Computer Music Conference - ICMC99*, pp. 418–421, 1999.

[9] N.-H. Fletcher and T.-D. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, 2nd edition, 1998.

[10] F. Cruz-Barney and O. Warusfel. "Prediction of the Spatial Information for the Control of Room Acoustics Auralization." in *Proc. of the 103rd Audio Eng. Soc. Convention*, New York, USA, 1997.

[11] H. Kuttruff. *Room Acoustics*. Elsevier Applied Science, 3rd edition, 1991.

[12] S. Lehman. "Flanging." In *Effects Explained*. Harmony Central. 1996.
http://www.harmony-central.com/Effects/Articles/Flanging/

[13] P. Fernandez-Cid and F. J. Casajus-Quiros. "Enhanced Quality and Variety for Chorus/Flange Units." In *Proc. of the 1st COST G-6 Workshop on Digital Audio Effects (DAFx98)*, Barcelona, November 19-21, 1998.

[14] T. Halmrast. "Musical Timbre Combfilter-Colloration from Reflections." in *Proc. of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, NTNU, Trondheim, December 9-11, 1999.