Indirect Acquisition of Instrumental Gesture Based on Signal, Physical and Perceptual Information

Caroline Traube
Laboratoire d'acoustique musicale
Faculté de musique, Univ. de Montréal
200, avenue Vincent-d'Indy
Montréal (Qc), Canada H3C 3J7
caroline.traube@umontreal.ca

Philippe Depalle
Music Technology Area
Faculty of Music, McGill University
555 Sherbrooke Street West
Montreal (Qc), Canada H3A 1E3
depalle@music.mcgill.ca

Marcelo Wanderley
Music Technology Area
Faculty of Music, McGill University
555 Sherbrooke Street West
Montreal (Qc), Canada H3A 1E3
mwanderley@acm.org

ABSTRACT

In this paper, we describe a multi-level approach for the extraction of instrumental gesture parameters taken from the characteristics of the signal captured by a microphone and based on the knowledge of physical mechanisms taking place on the instrument. We also explore the relationships between some features of timbre and gesture parameters, taking as a starting point for the exploration the timbre descriptors commonly used by professional musicians when they verbally describe the sounds they produce with their instrument. Finally, we present how this multi-level approach can be applied to the study of the timbre space of the classical guitar.

Keywords

Signal analysis, indirect acquisition of instrumental gesture, guitar

1. INSTRUMENTAL GESTURE

When musicians play on a traditional musical instrument, they usually interact with a control surface made of keys (piano, clarinet), strings (violin), mouthpieces (trumpet), reeds (oboe), etc. In most cases, many years of motor skills development are necessary to control the instrument adequately in order to intentionally produce sounds of a given quality or timbre.

In this paper, we will call *instrumental gesture* the actual instrument manipulation and playing technique on an instrument [2]. We will consider here the *effective gesture*, defined as the purely functional level of the notion of gesture, i.e., the gesture necessary to mechanically produce the sound (like blowing in a flute, bowing on a string, pressing a key of a piano and so on). We will call *instrumental gesture parameters* the parameters characterizing the components of the instrumental gesture. They are, for example, the speed of an air jet, the location of a pluck along a string, and the pressure applied with a bow on a string. The variations of these parameters have an effect on the timbre and are usually clearly perceived by a trained listener such as a professional musician.

Considering the case of the guitar and referring to the typology established in [2] and [3], plucking is an *excitation* and *modification* gesture, while fingering is a *selection* and also a *modification* gesture since the choice of fingering on the neck of the guitar (string/fret combination) affects timbre as well.

2. RELATIONSHIP BETWEEN GESTURE AND TIMBRE

Musical expression has been traditionally related to expressive timing and dynamic deviations in performance [8]. Less attention has been given to the study of timbre and how it relates to musical expression. This is probably due to the difficulty of defining the features of timbre, which are related to the physical aspects of sound in very complex ways. On the other hand, pitch, duration and volume are perceptual phenomena that have fairly simple physical correlates. Here, we propose to limit the scope of the study to the aspects of timbre that musicians can clearly control, perceive and verbally describe.

2.1 Perceptual dimensions of an instrumental timbre space

Early studies on instrumental timbre were performed by David Wessel and John Grey in the late 1970's [9]. Based on similarity judgments, those studies used multidimensional scaling algorithms to reduce the number of dimensions in the timbre space. Timbral features such as brightness (associated with the spectral center of gravity), spectral irregularity (spectral flux) and transients density were identified. It is important to note that these axes were used to differentiate between different orchestral instruments—a macroscopic view of timbre—as opposed to differentiating between the possible palette of timbres in a single instrument—a microscopic view of timbre. This is precisely the viewpoint of our approach. In particular, we want to identify the dimensions of the timbre subspace corresponding to the classical guitar, the instrument chosen for investigation and validation in this study.

2.2 Source-mode of timbre perception

Handel proposed in 1995 an explanation for timbre perception, saying that the subjective identification of timbre could involve the observer's perception of the physical mechanisms and actions in the sound production. This is the *source-mode* of timbre perception, as opposed to the *interpretative mode* of timbre perception [10]. Other facts support this view such as the source-filter model of speech perception. It is also interesting to realize that mechanics and materials of vibrating systems are the bases for traditional Western musical instrument as well as World instrument classification systems (e.g. von Hornbostel & Sachs classification in aerophones, chordophones and membraphones).

Considering the evident relationships between physical model constituents, instrumental gesture and perceptual attributes, we believe that the gestural information can be accessed via the identification of the parameters of a physical model. As it has been done for speech vowels, we propose to define an articulatory timbre space for individual musical instruments and to determine the relationships between this articulatory space and a perceptual timbre space (as defined by Grey in [9]).

3. INDIRECT ACQUISITION OF INSTRUMENTAL GESTURE

3.1 Direct vs indirect acquisition

There are different ways to capture the characteristics of instrumental gesture [5]:

- through *direct acquisition* of physical variables with sensors on the instrument or on the performer,
- through *indirect acquisition* of performance parameters from the analysis of the acoustic signal (namely from a recording).

In recent years we have seen an important development of technologies related to sensors and gestural interfaces. For example, many musical instruments can be augmented with devices that can monitor the performer's actions (choice of keys, pressure applied to a mouthpiece, etc.) and turn it into MIDI information. Direct acquisition is clearly a simpler way to capture the physical features of a gesture but it is potentially invasive and may ignore the interdependency of the various variables. For example, sensors on a clarinet could detect the air jet speed and the fingering but would not account for the coupling between the excitation and the resonator. As opposed to direct acquisition, indirect acquisition is based on the assumption that the performance parameters can be extracted from the signal analysis of the sound being produced by an instrument. The main difficulty of this task is to determine in the signal the specific acoustic signature of a particular performance parameter that has a perceivable influence on the sound.

3.2 From acoustic signal to instrumental gesture information

Most traditional musical instruments are stable during a performance, i.e., the acoustical properties of the instrument do not change over the time of the performance and an energy continuum needs to exist between the gesture and the perceived sound [3]. It is also interesting to note that in the case of most traditional acoustic instruments, the gestural interface is also part of the sound production unit. For instance, the reed, keys and holes of a clarinet are the elements the musician interacts with, but they are also responsible for the sound production [15].

Figure 1 schematizes the exchange of information between the three elements of a performance process: the performer, the instrument and the listener. Note that a musician is at the same time a performer and a listener.

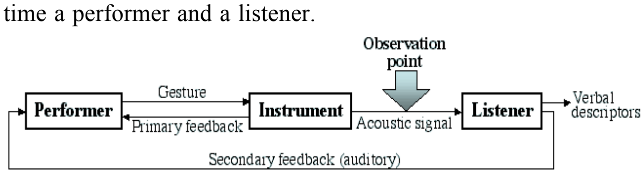

Figure 1. Interactions between the performer, the instrument and the listener.

The performer applies a gesture on the instrument, which in turn reacts to the gesture by producing a sound and by providing the performer with primary feedback, which can be visual, auditory (clarinet key noise, for instance) and tactile-kinesthetic [15]. The listener perceives sounds and attaches labels to them. Expert performers/listeners are generally able to discriminate and intuitively describe a large variety of sounds produced by their instruments.

In the approach that we propose, the observation point in the performance process loop is the acoustic signal, from which we extract structural information that allows us to get to the gestural information. To generate the data, musicians are recorded playing tones with specific gestures, varying one gesture parameter at a time. Figure 2 illustrates the procedure that we propose to access instrumental gesture information from the acoustic signal.

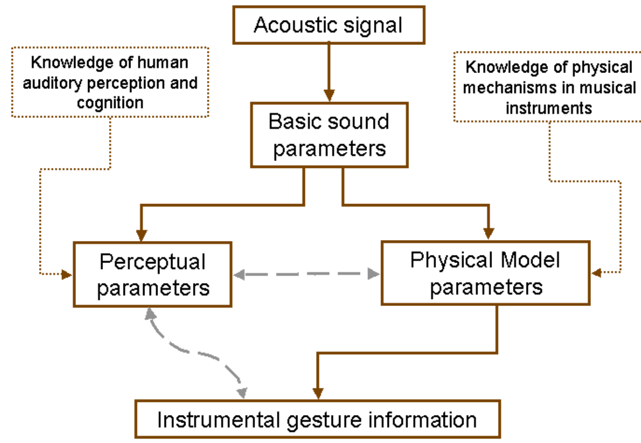

Figure 2. From acoustic signal to instrumental gesture information.

In the first stage of the analysis of the data, basic sound parameters are extracted from the acoustic signal, through time- and frequency-domain analysis. These *low-level parameters* include the short-time energy (related to the dynamic profile of the signal), fundamental frequency (related to the sound melodic profile), spectral envelope, amplitudes, frequencies and phases of sound partials, and power spectral density [16]. Using the knowledge of physical mechanisms taking place in musical instruments, physical model parameters are derived from the basic sound parameters. These parameters generally give direct access to the instrumental gesture parameters.

Finally, in order to understand the effect on timbre of the variation of the instrumental gesture parameters, we also use perceptual measures, which we could call *high-level parameters*, as opposed to the low-level parameters defined earlier. These perceptual measures are also derived from basic sound parameters and in particular from the amplitudes of the spectral components. They include widely used measures such as the spectral centroid, spectral irregularity, odd/even harmonic ratio, low/high harmonic ratio, and log-rise time [10]. These parameters are interesting to examine because they are correlated to perceptual attributes such as brightness, metallic quality and nasality. A strong correlation can generally be found between perceptual attributes and instrumental gesture parameters. For example, plucking a string closer to the bridge increases brightness. Modifying the angle of the air jet on the mouthpiece edge of a transverse flute affects brightness as well.

Although this study addresses issues related to the general problem of timbre recognition, the approach that we propose here for the analysis of instrumental timbre differs from the phenomenological approach taken in many timbre recognition systems described in the literature (in [6] for example). Timbre recognition systems implementing neural networks or using principal component analysis require a learning stage, meaning that a timbre can only be identified and labeled by the system after being compared to other typical examples of that timbre. Therefore they do not make explicit the relationships between the physical phenomena, the performer's actions and the obtained timbre. Here, we rather propose to develop analysis tools that use the knowledge that we have about the physical phenomenon taking place in the musical instrument and its effect on the acoustic signal, leading to an analytical model of the interaction between the performer and the instrument (cf. Figure 1).

4. APPLICATION TO EXPLORING THE TIMBRE SPACE OF THE CLASSICAL GUITAR

In order to validate the proposed approach for the analysis and understanding of the timbre of a musical instrument and its relationships with the physical phenomena and the performer's gesture, we will present how the approach is applied to the study of the timbre space of the classical guitar. The dimension of that timbre space that we want to start with is the one corresponding to brightness.

The guitar is an instrument that gives the player great control over the timbre. Different plucking techniques involve varying instrumental gesture parameters such as (a) the finger position along the string, (b) the inclination between the finger and the string, (c) the inclination between the hand and the string and (d) the degree of relaxation of the plucking finger. In [12], the author reports three analysis techniques that were used to investigate these four instrumental gesture parameters. Among these analysis techniques, Principal Component Analysis is used to verify that each of the instrumental gesture parameters induces significant changes in the cepstral envelope. However, it is not clear that this methodology can constitute an indirect acquisition system because the four sets of guitar tones were analyzed separately.

In the approach we propose, we rather want to make explicit the correspondences between a perceptual timbre space and a gestural timbre space of the instrument.

4.1 Timbre descriptors used by guitar players

As a starting point for the exploration of the timbre space, we want to inquire about the timbre descriptors commonly used by professional musicians. We asked 22 guitarists to define 10 adjectives they commonly use to describe the timbre nuances they can produce on their instrument. We asked the participants to intuitively describe the timbre itself ("How does it sound?") and to describe the gesture associated with it ("How do you make it?"). The compilation of these data lead to an inventory of over 60 adjectives. Dark, bright, chocolatey, transparent, muddy, wooly, glassy, buttery, and metallic are just a few of those adjectives used by guitarists to describe the brightness, the color, the shape and the texture of their sounds.

When playing the guitar, the location along the string where the plucking is performed strongly affects the resulting timbre. If the plucking point is closer to the bridge, the sound is brighter, sharper, more percussive. If the plucking point is closer to the middle of the string or the soundhole, the resulting sound is warmer, mellower, duller, as expressed by expert performers/listeners.

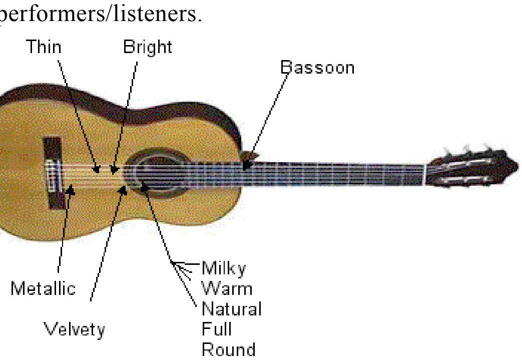

Figure 3. Timbre descriptors and corresponding plucking locations along the string according to guitarist Zane Remenda (participant in guitar timbre study).

So, for the case of the guitar, we find that a dimension of the gestural timbre space (the plucking position) clearly corresponds to a dimension of the perceptual timbre space (the brightness). As illustrated on Figure 4, we should be able to check this correspondence by calculating the spectral centroid of the spectrum, which is a measure on the acoustic signal that has been shown to strongly correlate with perceived brightness [10].

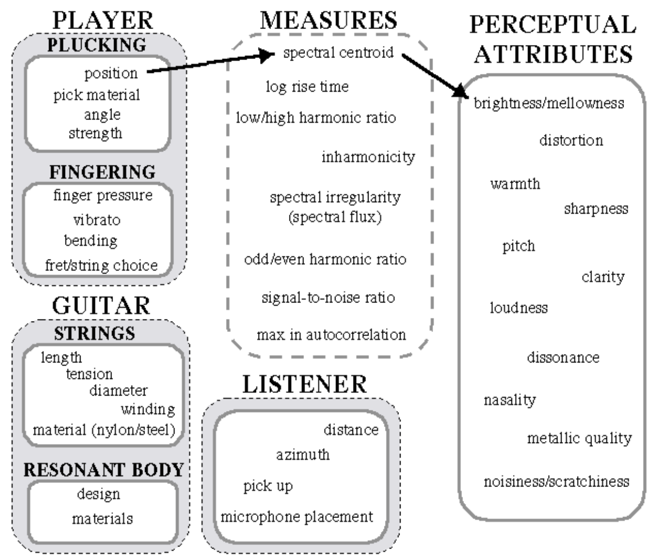

Figure 4. Factors influencing the timbre of guitar tones.

Figure 4 also inventories the other factors influencing the timbre of tones played on a guitar. Besides instrumental gesture parameters (characterizing the plucking and the fingering), the materials and the physical features of the instrument itself affect the timbre as well, in the sense that they constrain the palette of timbre nuances that can be achieved by the performer. Finally, the listening conditions also have an impact, due to the particular radiation pattern from the instrument, the characteristics of the microphone (in the case of a recording) and the acoustics of the room.

4.2 Acoustic signature of plucking position

The next step in our approach is to learn about the physical phenomenon taking place in the instrument.

Plucking a string sends an acceleration impulse along the string in both directions. Those impulses are reflected at the ends of the string (the bridge on one side and the nut or the finger on the other side). All those impulses combine to form a standing wave on the string. The resultant motion consists of two bends, one moving clockwise and the other counter-clockwise around a parallelogram [7]. In the ideal cases, the output from the string (force at the bridge) lacks those harmonics that have a node at the plucking point. The amplitude C_n of the nth mode of the displacement of an ideal vibrating string of length l, with an initial vertical displacement h is given by:

$$C_n(h,R)=\frac{2h}{n^2\pi^2R(1-R)}\sin(n\pi R) \qquad (1)$$

where R is the relative plucking position, defined as the fraction of string length from the point where the string was plucked to the bridge [13].

The location of the plucking point along a string has an effect on the spectrum of the sound that is similar to the effect of a comb filter. In fact, in a simple digital physical model of a plucked-string instrument, the resonant modes would translate into an all-pole structure, while the initial conditions (a triangular-shaped initial displacement for the string and a zero-velocity at all points) would result in a FIR comb filter structure. The delay of this comb filter is related to the time the wave needs to travel from the plucking point to the fixed end of the string (the bridge or the nut) and back. Therefore, the comb filter delay can be expressed as the product of the relative plucking position *R* and the fundamental period T_o.

The comb filtering effect is illustrated on Figure 5 showing the magnitude spectrum of a guitar tone plucked at 12 cm from the bridge on a 58 cm open A-string. The relative plucking position *R* is approximately 1/5 (12 cm / 58 cm = 1 / 4.83). If it was exactly 1/5, all harmonics with indices that are multiples of 5 would be completely missing.

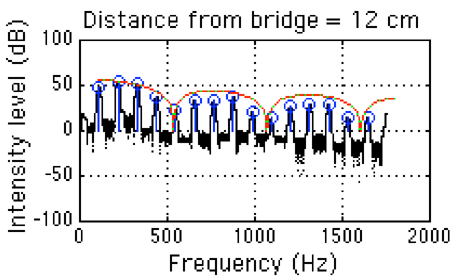

Figure 5. Magnitude spectrum of a guitar tone plucked at 12 cm from the bridge on a 58 cm open A-string.

4.3 Perceptual effect of plucking position

In order to understand the effect on timbre of the variation of parameters related to the performer's actions, we derive perceptual measures from the basic sound parameters.

As guitarists intuitively associate increasing brightness with decreasing plucking distance from the bridge, we assume that it is possible to check this correspondence by calculating the spectral centroid *SC* of the spectrum:

$$SC = \frac{\sum_{n=1}^{N} f_n C_n^2}{\sum_{n=1}^{N} C_n^2} \qquad (2)$$

where C_n is the magnitude of the *n*th spectral component and f_n its frequency [10]. Figure 6 displays the plots of the theoretical spectra for various plucking distances, calculated from the theoretical expression of the amplitude of the velocity modes (proportional to $n\ C_n$). We can visually notice that the center of gravity of the spectrum would decrease as the plucking distance from the bridge increases.

This trend is in fact confirmed by the plot displayed on Figure 7, showing the spectral centroid of the theoretical spectra (shown on Figure 6) as a function of plucking distance from the bridge. Also shown on Figure 7 is the spectral centroid curve from the spectra of recorded guitar tones played with different plucking distances. The real data curve follows the same trend as the theoretical curve although the spectral centroid is generally lower.

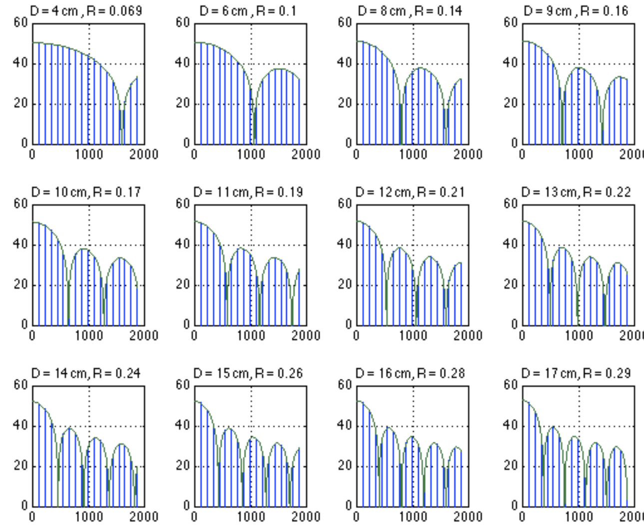

Figure 6. Variation of theoretical spectral envelope (magnitude in dB vs frequency in Hz) with plucking position.

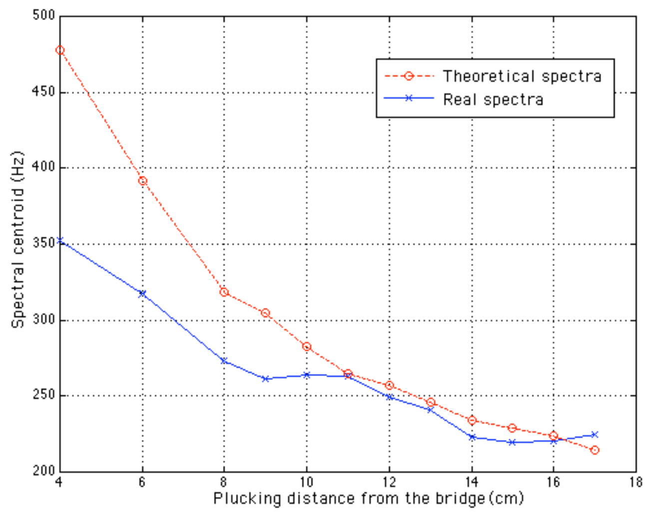

Figure 7. Variation of spectral centroid with plucking distance from the bridge.

4.4 Indirect acquisition of plucking position

In order to derive the plucking position from the recording of guitar tones, we propose a signal processing method that extracts the location of the zeros in the spectral envelope starting from a FFT-analysis and a measure derived from the autocorrelation. This work adds on to other methods proposed previously and reported in [1], [13] and [14].

The autocorrelation function can be very useful to estimate the fundamental frequency of a periodic signal, since it should show a maximum at a lag corresponding to the fundamental period. Figure 8 displays the plots of the autocorrelation function calculated for 12 recorded guitar tones plucked at various distances from the bridge on an open A-string (fundamental frequency = 110 Hz).

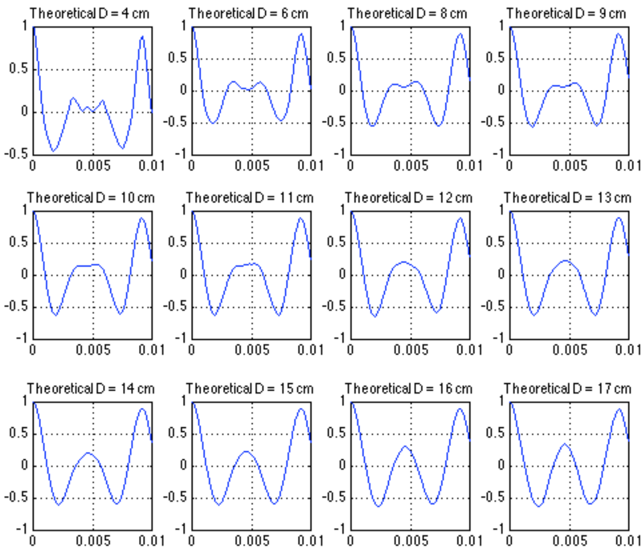

Figure 8. Autocorrelation graphs for 12 guitar tones plucked at distances from the bridge ranging from 4 cm to 17 cm.

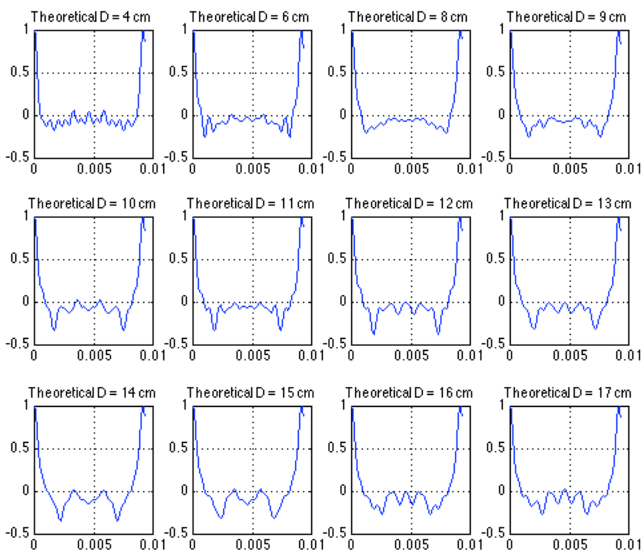

Figure 9. Log-correlation graphs for 12 guitar tones plucked at distances from the bridge ranging from 4 cm to 17 cm.

As expected, the graphs show a maximum around 1/110 = 0.009 seconds, the fundamental lag of the autocorrelation. One can also see that the autocorrelation takes on different shapes for different plucking positions but the information about the comb filter delay can not be extracted in an obvious way, directly from these graphs.

To increase the negative contribution of low amplitude harmonics (around the valleys in the comb filter spectral envelope), the log of the squared Fourier coefficients C_n are used to calculate a modified autocorrelation function, that we propose to name *log-correlation* and express as follows:

$$\Gamma(\tau)=\sum_{n=1}^{N}\log\left(C_n^2\right)\cos\left(\frac{2\pi n}{T_o}\tau\right) \qquad (3)$$

Figure 9 displays the log-correlation graphs for the same 12 recorded guitar tones (as for Figure 8). Those plots reveal an interesting pattern: the minimum appears around the location of the lag corresponding to the relative plucking position. We can conclude that the relative plucking position can be approximated by the ratio $R = \tau_{min} / \tau_o$, where τ_{min} is the lag corresponding to the global minimum in the first half of the log-correlation period, and τ_o is the lag corresponding to the fundamental period T_o, as shown on Figure 10.

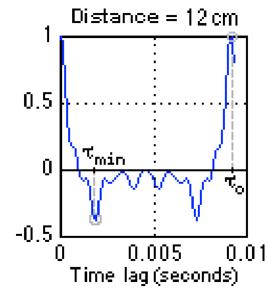

Figure 10. Log-correlation for a guitar tone plucked 12 cm from the bridge on 58 cm open A-string.

Figure 11 summarizes the estimation results for the data set of 12 tones. Except for a significant error for the first distance (at 4 cm from the bridge), the estimation is accurate for all other distances (within 1 cm of error). At 4 cm, $R = 4 / 58 = 1 / 14.5$ and the error probably comes from the fact that the spectrum contains only one "zero" over the frequency range that is considered (up to the 15th harmonic).

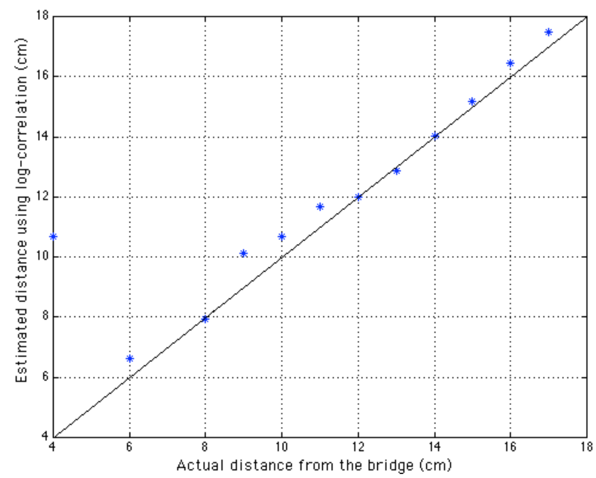

Figure 11. Plucking point estimation with log-correlation (estimated distance vs actual distance from the bridge)

5. CONCLUSION

In this paper, we have proposed a multi-level approach for the extraction of instrumental gesture parameters from the characteristics of the signal captured by a microphone and based on the knowledge of physical mechanisms taking place on the instrument. Starting from the timbre descriptors commonly used by professional musicians when they verbally describe the sounds they produce with their instrument, we explore the relationships between some features of timbre and gesture parameters. Finally, we presented how this multi-level approach can be applied to the study of the timbre space of the classical guitar. More specifically, we have confirmed the relationship between perceived brightness and decreasing plucking distance from the bridge (intuitively expressed by guitarists) and we have presented a way to extract the plucking position from the signal, which is related to the delay of a comb filter in the physical modeling of the instrument.

The search for other relationships between physical model constituents, instrumental gesture parameters and perceptual attributes would be worth being pursued. Inspired by Grey's timbre space study, a multidimensional scaling analysis of guitar tones could be useful to determine the dimensions of the subspace of guitar timbre nuances.

This works finds applications in the context of hybrid instruments, generating control parameters for physical model based synthesizers and automatic tablature generation.

6. APPENDIX

The recorded tones that are used in this study were played with a plastic pick, 0.88 millimeters in thickness and triangular shaped, on a plywood classical guitar strung with nylon and nylon-wrapped steel Alvarez strings. The intended plucking points were precisely measured and indicated on the string with a marker. The tones were recorded with a Shure KSM32 microphone in a sound-deadened room, onto digital audio tape at 44.1 kHz, 16 bits. The microphone was placed in front of the sound hole, approximately 25 cm away, which was far enough to capture a combination of waves coming from different parts of the string, in that way limiting the filtering effect of the pick-up point. A 4096-samples portion was extracted from the middle of the tone (after the attack) and the Fast Fourier Transform analysis was performed with zero-padding factor of 8 and parabolic interpolation. The magnitudes of the first 15 harmonics were used to calculate the log-correlation and the spectral centroid.

7. REFERENCES

[1] Bradley, K., M.-H. Cheng, and V.L. Stonick. Automated Analysis and Computationally Efficient Synthesis of Acoustic Guitar Strings and Body. in IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics. 1995.

[2] Cadoz, C. Instrumental Gesture and Musical Composition. in International Computer Music Conference. 1988: International Computer Music Association.

[3] Cadoz, C. and M.M. Wanderley, Gesture - Music, in Trends in Gestural Control of Music, M.M. Wanderley and M. Battier, Editors. 2000, Ircam - Centre Pompidou: Paris, France. p. 71-94.

[4] Delalande, F., La gestique de Gould: éléments pour une sémiologie du geste musical., in Glenn Gould, Pluriel, G. Guertin, Editor. 1988, Louise Courteau Editrice Inc. p. 83-111.

[5] Depalle, P., S. Tassart, and M. M. Wanderley, Instruments virtuels, in Résonance. 1997.

[6] Egozy, E. B., Deriving Musical Control Features from a Real-Time Timbre Analysis of the Clarinet. Master thesis, 1995. Department of Electrical Engineering and Computer Science. Massachussetts Institute of Technology.

[7] Fletcher, N. H. and T. D. Rossing. The Physics of Musical Instruments, 1998, 2d edition, New York, Springer-Verlag.

[8] Gabrielsson, A. The Performance of Music. In Diana Deutsch (ed) The Psychology of Music, 1999, San Diego, CA: Academic Press. Second Edition, pp. 501-602.

[9] Grey, J.M., Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America, 1977. 61(5): p. 1270-1277.

[10] Hajda, J.M., et al., Methodological issues on timbre research, in Perception and Cognition of Music, I. Deliège and J. Sloboda, Editors. 1996, Pychology Press, Oxford: Hove, UK. p. 253-306.

[11] Machover, T., Hyperinstruments - A progress Report 1987-1991. 1992, MIT Media Laboratory Massachussets Institute of Technology.

[12] Orio, N. The timbre space of the classical guitar and its relationship with the plucking techniques. in ICMC - International Computer Music Conference. 1999.

[13] Traube, C. and J.O. Smith. Estimating the plucking point on a guitar string. in Conference on Digital Audio Effects (DAFX00). 2000. Verona, Italy.

[14] Traube, C. and J.O. Smith. Extracting the fingering and the plucking points on a guitar string from a recording. in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01). 2001. Mohonk Mountain House, New Paltz, New York.

[15] Wanderley, M.M., Gestural Control of Music. Proceedings of the International Workshop on Human Supervision and Control in Engineering and Music. Kassel, Germany, pp. 101-130.

[16] Wanderley, M.M. and P. Depalle, Contrôle Gestuel de la Synthèse Sonore, in Interfaces Homme-Machine et Création Musicale. 1999, Hermes Science Publications. p. 145-163.