

Transmodal Feedback as a New Perspective for Audio-visual Effects

Christian Jacquemin
LIMSI & Université Paris 11
BP 133
91403 ORSAY, France
FirstName.FamilyName@limsi.fr

Serge de Laubier
Puce Muse
2 rue des Pyrénées, Silic 520
Wissous 91320 RUNGIS Cedex, France
sdl@puceuse.com

ABSTRACT

A new type of feedback is presented that involves both the auditory and visual modalities. It combines an audio resonant bandpass filter, a geometrically constructed mass-spring system and its graphical skin. The system shows a resonant behavior that is detailed in various parameter setups. Complex mass-spring topologies result in a coherent self-sustained audio-visual system that mimics gusts of wind blowing a veil and associated sound effects.

Keywords

Audio-visual composition, Transmodality, Feedback

1. TRANSMODAL FEEDBACK

The twentieth century has seen a very large body of work concerning the connection between the visual and acoustic modalities and, more specifically, between sound, music, light, and image. Most of these works can be classified as *transmodal*: either using images to generate sound, or analyzing sound and music to generate graphics that can in turn be used to modify sound and music [1].

Another line of artistic exploration concerns the connection of one modality with itself: the notion of *feedback*. First considered as a undesirable effect, audio feedback has been appropriated by pop musicians such as The Who and Jimmy Hendrix as an interesting ornamentation of their music in which their instrument (a guitar) was used as a control filter. Audio feedback can be considered as an *intra-modal* system that uses sound to generate sound.

Our purpose in this work is to explore the potentialities of the combination of trans- and intramodal communications in what we term *transmodal feedback*. How can a system for audio \leftrightarrow graphic feedback be designed, in which sonic output is used as input for graphical synthesis, that is in turn fed into the sound generator?

We first analyze some transmodal applications which offer interesting insights into the correspondences that can be established between the audio and graphic modalities. Then, a transmodal feedback system that combines physical modeling, graphical rendering, and a sound resonator

is presented. Last, several variations are proposed in order to illustrate different parametrizations and renderings.

2. TRANSMODAL CORRESPONDENCES

The correspondences between two modalities tend to be metaphorical when they are used for artistic and creative purposes, and tend to be more literal when they are used for control purposes.

In the metaphorical category, and connecting sound to graphics, is the work of Golan Levin. The sound (the voice) is transformed into illustrative graphical effects inspired from the cartoon world [9]. Similarly, rich graphical environments such as urban models can be easily associated with sonic interpretations [19]. Metaphorical representations introduce a distance between the source stimulus (image or sound) and its perceived effect. For this reason they are not appropriate for feedback effects which require better coherence between input and output.

Literal transmodal correspondences, that are better suited to feedback, are encountered in systems where a modality is used to control another one. One of the motivations behind these works is that human perceptual capabilities depend on the modality. For instance, vision is very good at distinguishing visual patterns in large sets of visual data, while audition is good at perceiving very brief sound variations.

Visual representation of music is a literal correspondence between graphics and audio that has its origin in the notation of music through scores. Digital media have offered new perspectives to interactive composition through the graphical representation of musical composition. It can be based on sophisticated musical theories such as Xenakis's theory for Iannis [5] or more abstract representations such as Sonos [16] or Metasynth [11]. Similarly virtual instruments are visual interfaces for music synthesis that focus on playability, direct manipulation, and real-time interaction [8].

Duality of sonic and visual representation is also well illustrated by visual representations of sound databases [15] that can help the user to build a mental map of the soundscape of the sample collection. The reverse combination of sound and graphics is abstract data sonification: the process of representing generic data by means of audio signals [2].

Since our purpose is to close the loop and allow reciprocal transmodal information exchange, we return to the notion of feedback in a resonating system before introducing our model of audio \leftrightarrow graphic feedback.

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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NIME 06, June 4-8, 2006, Paris, France
Copyright remains with the author(s).

3. RESONANCE AND FEEDBACK

3.1 Audio and Video Feedback

Unimodal feedback is the process of capturing the signal produced by an emitter in a modality (typically a loudspeaker for sound) and reamplifying it. It is illustrated by Figure 1. It generally involves the contribution of an external trigger source that plays a more important role in video feedback than in audio feedback.

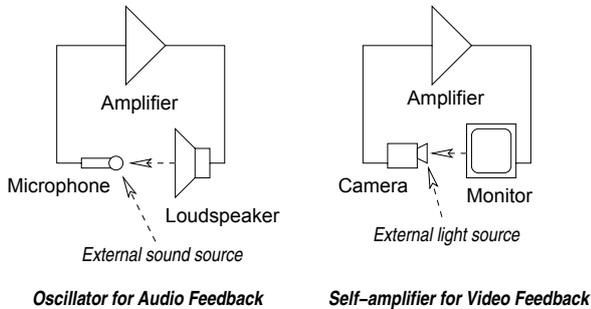


Figure 1: Unimodal Feedback.

The “classic” audio feedback (also known as Larson effect) occurs when an amplifier receives as input its own output. The loop results in an increasingly loud signal until the limits of the amplifier are reached. Audio feedback can be seen as an echo with very short delay defined by the characteristics of the system (distance between loudspeaker and microphone, amplifier, characteristics of I/O devices, room...), and transforms it into an oscillator. The selected frequencies correspond to the Barkhausen effect and are such that the input and output signals are in phase (with additive intensities) and the gain is slightly above 1. Since the amplified signal is mainly controlled by the characteristics of the system, the external sound source plays the role of a trigger and the output pitch is dominated by resonance frequencies.

What is known as video feedback is by nature very different from audio feedback since it relies only on gain and not on oscillation. For this reason, all colors are equally subject to amplification, contrary to audio feedback that amplifies a very narrow band of frequencies. Periodicity in video feedback occurs in space and not in time, and results in tiling or kaleidoscopic effects whose base graphical components are defined by the external signal. (Visual perception occurs in time and in space, but only spatial perception involves resonance and periodicity.)

3.2 Transmodal Feedback

In order to design the architecture of a transmodal feedback system, we must establish a reciprocal communication between a graphical and an audio application so that the signal emitted by one component is accepted by the other one. For this purpose we use networked applications and encapsulate transmitted data via network messages. The overall architecture is given in Figure 2 in which emitters have been preserved for human access to the system output, but sensors (microphone and camera) have been removed because they are not necessary any more (even though unimodal feedback could be combined with transmodal feedback).

The design of an audio↔graphic oscillator is not as straightforward as it is for a pure audio system. First there is a temporal inconsistency between the processing

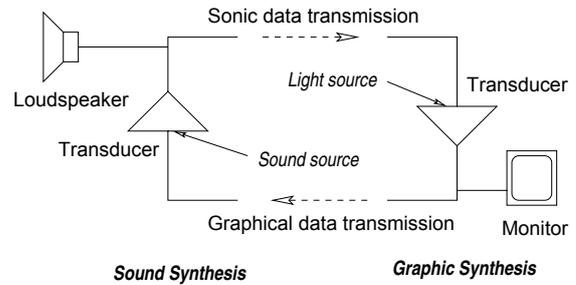


Figure 2: Transmodal Feedback.

delays of the audio, graphic, and communication systems. The processing delays in a graphic system are higher than or equal to the frame refresh rate (typically 40ms). They cumulate with the communication delays between the audio and graphic system (around 1ms). In an audio system, the delays are close to the period of the sound signal (a few μ s). The processing delays of an audio↔graphic system are controlled by the frame rate, and therefore greater than 40ms.

A second temporal inconsistency concerns the emitted signals. The phase of the visual signal is several orders of magnitude higher than the phase of the audio signal, which is in turn much higher than the delays involved in a looping audio↔graphic system. The system cannot work as an oscillator as discussed for an amplifier in pure audio feedback.

When comparing unimodal audio and video feedbacks, it appears that audio feedback offers a richer domain of experimentation because of its double nature: phase coincidence (the signal is tuned to the characteristics of the system) and self-reinforcement. It seems therefore desirable to build a system which will act as a resonator. Since we cannot work on the signal directly (because of the second temporal inconsistency), oscillations will concern higher level audio parameters such as envelope or pitch. Because of the first temporal inconsistency, the resonator frequency must be lower than 25hz, possibly much lower.

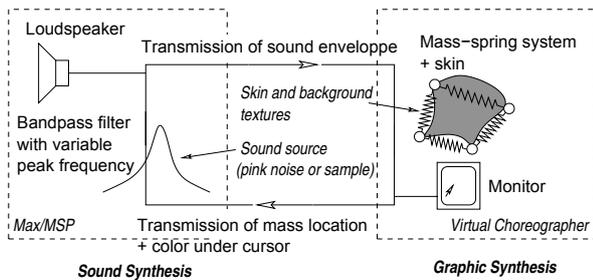
The architecture proposed in Figure 2 has no reason to be an oscillator if the transmitted data are not periodic. In order to equip the application with a generator of periodic signals, the graphical component is complemented with a mass-spring system (MSS) that directly controls the graphical output, and indirectly the sound generation. We now turn to the implementation of this architecture and its two major building blocks: a skinned MSS on the graphical part, and a resonator related to the MSS dynamics on the audio part. The application is named *GraphSon*.

4. AUDIO↔GRAPHICS FEEDBACK

The architecture of *GraphSon* is made of two networked applications: an audio patch under *Max/MSP* [10] that implements a resonant bandpass filter externally controlled by the speed and acceleration of the graphical elements, and a virtual 3D scene under *Virtual Choreographer (Vir-Chor)* [18] that is made of a skinned MSS parametrized by the sound envelope derived from the audio patch. Data exchange between these components is made through OSC. Figure 3 show the instantiation of Figure 2 in the case of *GraphSon*.

4.1 Graphical and Physical Components

Mapping is considered as an important issue in the de-


 Figure 3: *GraphSon* Architecture.

sign of virtual instruments and concerns the “intelligent” and sensitive association between a musician’s gestures and the control of his/her instrument. Mapping tends to be considered not just as an interface, but as an autonomous component in virtual instruments. Because of their intuitive and rich behavior, physical models can be used as mapping devices that produce complex and variable responses to stimuli: for instance, obstructions in particle flows and resulting collisions (*FlowField* [4]), or MSSs and their complex dynamics (*GENESIS* [3] or *PMPD* [12]).

Our interest for such systems in this work is not for the purpose of mapping human stimuli to musical synthesis, but for the introduction of a resonator in our audio↔graphic feedback loop parameter (Figure 3). The MSS associates input sound envelope values to a graphical output through an indirect mechanism.

In a MSS, the equation that controls the dynamics of a mass M_i that is linked to n_i masses $M_{i,j}$, is

$$m_i \cdot x'' = -d \cdot x' + m_i \cdot g_x + \sum_{j=1}^{n_i} k_{i,j} (d(M_i, M_{i,j}) - \text{lin}i_{i,j}) \quad (1)$$

in which m_i is the mass of M_i , d the viscous damping coefficient, g the gravity, $k_{i,j}$ the spring constants, and $\text{lin}i_{i,j}$ the lengths of the unstretched springs. Sound envelope e is used to modify dynamically two of the MSS characteristics: its damping factor and the spring elasticity

$$d = k_{damp} \cdot e \quad \text{and} \quad \forall i, j \quad k_{i,j} = k_{elast} \cdot e \quad (2)$$

The audio-visual effect is that high sounds result in a stiff and constrained MSS (mild and sustained wind in a non extensible veil), while low sounds result in a weak and free MSS (strong gusts of wind in a light and extensible veil). In the second case, the potential energy accumulated in the veil can be released suddenly and transformed in kinetic energy. Such a correlation produces perceptually plausible correspondences between audio and graphics [7].

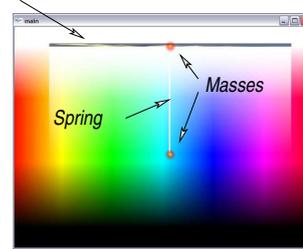
The graphical scene is implemented in *VirChor*. The `<graph>` element describes a MSS, and the `<patch>` element a Bezier patch. At each frame, a script is executed that reconnects the control points of the skin to the masses of the MSS. Two models are designed according to table 1 and illustrated by Figure 4. A quad is used as skin in the simplest model *GraphSon₂*.

The target application is *GraphSon_{4x4}* because it offers richer behaviors, and better graphical renderings and animations. It combines a 4×4 MSS with a grid topology and a bicubic Bezier patch defined by 16 control points (masses at nodes, springs for inter-connectivity). The simplest application is used for analyzing the parameter effects and resonating behaviors in section 5 under simpler experimental conditions and fewer parameters.

 Table 1: Parameters of *GraphSon* Instances.

Name	Masses (fixed)	Springs	Skinning
<i>GraphSon₂</i>	2 (1)	1	Quad
<i>GraphSon_{4x4}</i>	16 (2)	24	Patch 4×4

Handlebar for mouse control of fixed masses


 GraphSon₂

Handlebar for mouse control of fixed masses

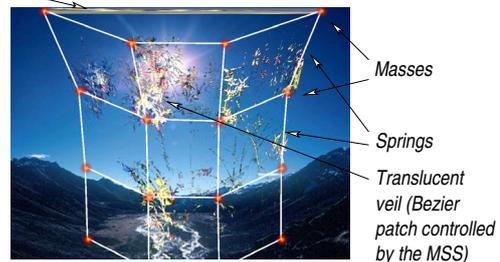

 GraphSon_{4x4}

 Figure 4: Two instances of *GraphSon*: Gestures are transmitted to the upper masses of a MSS that controls an animated translucent veil.

4.2 Audio Component

For audio-visual coherence purposes, the sound generated from the graphical output is intended to reproduce the noise of a veil in the wind. The effect is obtained by using a pink noise source (the wind) filtered by a digital bandpass filter that produces high pitch noise for strong gusts of wind.

The filter is controlled by its quality Q , its gain G , and its center frequency f_{res} . The higher the quality, the shorter the bandwidth, and the higher the output at the resonance frequency. The second order equation used for the filter is

$$y_n = G(x_n - r \cdot x_{n-2}) + c_1 \cdot y_{n-1} + c_2 \cdot y_{n-2} \quad (3)$$

r , c_1 , and c_2 are parameters calculated from f_{res} and Q .

In order to produce a satisfactory audio effect, the resonance frequency is controlled by the acceleration of masses in the bottom line. Strong accelerations of these masses correspond to high pitch output, giving the impression of a strong wind blowing the veil.

The resonance filter is implemented in *Max/MSP* with the `reson~` object that has 4 inputs: an audio signal and 3 digital values G , f_{res} , and Q . Equation (3) is taken from [10]. The frequency f_{res} is a linear function of the acceleration of one of the masses in the MSS. It is computed in the audio patch from the values of the mass location received from the graphical component. The output of `reson~` is the filtered audio input, pink noise produced by the object `pink~`. The envelope of the output audio signal, sent to the graphical component, controls damping and elasticity.

5. FEEDBACK CONTROL AND ANALYSIS

We now turn to the study of the resonating audio↔graphic feedback loop under various parameter values. The behavior of the feedback loop depends on several factors: the topology of the MSS, the parameters of the audio system including the nature of the base sound (noise or sample), the transmission delays through the network, and the motion of the controlled mass by the user. This section is intended to provide better insight of the basic echo resonance in the system in its simplest form: pink noise and a 2-mass 1-spring system. More detail is also provided on the parametrization of the system and its effect on the animation of the graphical scene and the audio output.

5.1 Basic System

If the simplest MSS (*GraphSon₂* presented in 4.1) is connected to the resonance filter fed with a pink noise a periodic behavior is observed, illustrated by Figure 5. In this figure two values are plotted that trace the dynamics of the audio and graphic systems:

- the height of the lower mass, the free mass since the other one is fixed to the handle (dotted line),
- the sound level which is used to control damping and spring coefficient (solid line).

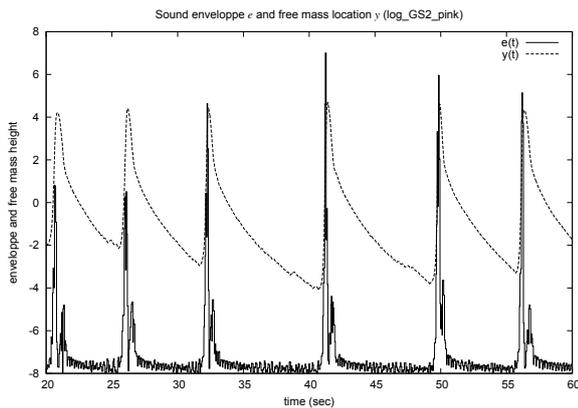


Figure 5: *GraphSon₂* Basic Resonating System (see upper part of Figure 4): 1 Fixed Mass, 1 Vertically Moving Mass, 1 Spring, and Pink Noise.

The basic behavior can be described as follows. When the mass reaches its lowest position (maximal extension of the spring), it slows down, decreases the pitch of the resonance frequency, and increases the Q of the filter. This results in a weaker sound that in turn decreases damping and spring coefficients. Because of low damping values, the MSS becomes more reactive to small movements of the lower mass and the spring then retracts very quickly.

The use of various sound samples does not modify significantly the behavior of the resonator, even though it has a strong impact on the audio output. Several tests were made with various kinds of music: piano romantic music, techno/world music, natural sound effects... but none had a strong impact on the system behavior. Such observations are coherent with resonating audio feedback, in which resonance is controlled by the system characteristics and the trigger sound plays a secondary role.

5.2 Color Parametrization

In order to provide the user with easy access to the parametrization of the audio system (and also indirectly on the graphical system), the red, green, blue components of the color under the mouse cursor are transmitted to the audio patch and associated with parameters of the audio resonator. The associations are made as follows:

- The green value controls a multiplicative factor of acceleration that defines f_{res} and also controls G ,
- the red value controls Q (the height and width of its bandwidth),
- the red and blue values bring an additional additive factor to f_{res} .

The color can be used in two ways. It can either be used as a control device for the user. If she/he moves the mouse cursor on the background image, various responses are obtained from the system. Color can also be used in a more passive way by placing the mouse cursor on the animated veil. Then the variation of colors under the mouse cursor results in dynamic modification of audio parameters that reciprocally modify the animation and rendering of the audio scene.

Various types of veil colorings are used to produce different color variations and thus different behaviors of the feedback loop. In Figure 4 above, two types of veils are shown. In the upper snapshot, a blended semi-transparent veil is used: from white opaque at the top to translucent at the bottom. The bottom snapshot shows a more complex rendering of the veil that is implemented through shaders: the veil color is the composite of several semi-transparent textures combined with masks. The transparency parameters of textures and masks are computed from dynamic geometrical characteristics of the veil and vary according to its dynamics.

The combined effects of color and veil motion are shown in Figure 6. The color under the mouse cursor is the blending of a red background color and the semi-transparent white color of the veil. Because of the high value of the red channel, the audio resonator is a sharp filter with a narrow bandwidth. Because of the veil motion, when the veil drops the color under the mouse cursor becomes whiter, which makes blue and green values higher, and thus tend to reactivate the audio system. The combination of these two effects gives the resonator of the feedback loop a smaller period than it had without the veil (compare Figure 5 for fixed pink color and mouse cursor outside the veil and Figure 6 with mouse cursor over the veil).

5.3 Complex System Behavior

We now turn to *GraphSon_{4x4}*, a MSS made of a 4×4 grid of masses that controls a bicubic Bezier patch. As for simpler systems, we plot on the same graph the location of the lowest left mass and the audio level.

Because of the more complex internal dynamics of its MSS, the loop resonance is not as clear as it is in the case of a 2-mass 1-spring system. The veil has its own internal short term dynamics that combine with the longer term loop dynamics. The loop resonance is easier to detect for low color values associated with a soft filter (Figure 7).

If the audio system receives a bright color associated with high parameter values for gain and quality, the output level is higher and higher values are sent for damping and spring coefficients. The veil has short amplitude movements with very short periods. Periodicity is much more

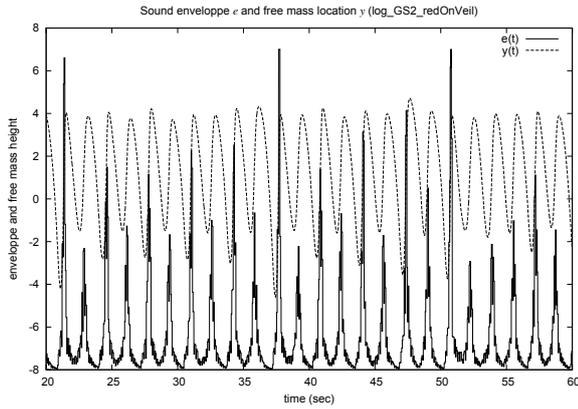


Figure 6: MSS Topology of Figure 5 and Pink Noise, with an Additional Control through the Color under the Mouse Cursor. The Cursor is Located over a Semi-transparent White Veil and a Red Background Color.

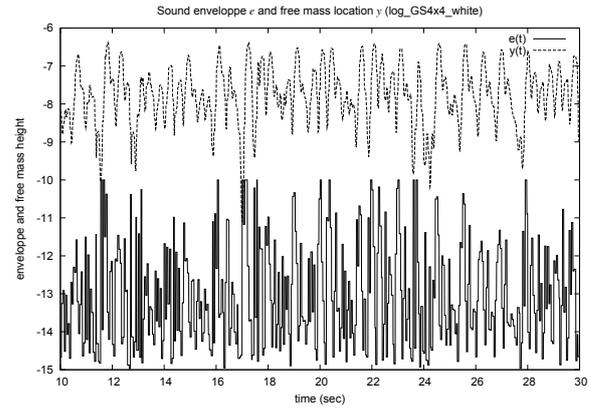


Figure 8: MSS Topology of Figure 7 and Pink Noise. Mouse Cursor on White Background Color (1,1,1).

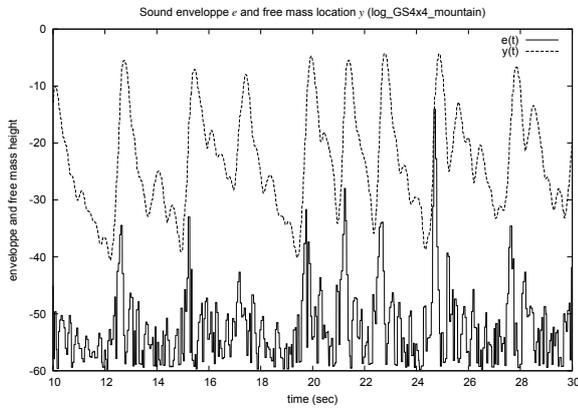


Figure 7: *GraphSon*_{4x4} A MSS made of a Grid of 16 Masses (see lower part of Figure 4): 2 Fixed Masses, 14 Vertically Moving Masses, 24 Springs, and Pink Noise. Mouse Cursor on Dark Background Color (0,0.02,0.07).

a decreasing slope followed by a short peak.

These results show that strong gestures can control the system while they are executed and for a short time afterward, but the system quickly returns to its periodic behavior when the excited state is over.

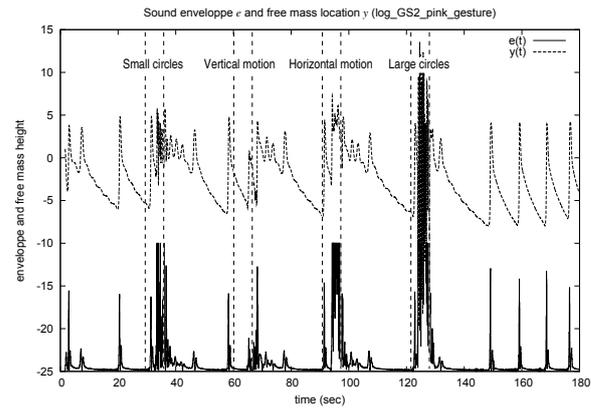


Figure 9: Gesture Mapping with *GraphSon*, Conditions of Figure 5.

difficult to detect in the resulting motion and audio signal (Figure 8).

As for the previous simpler MSSs, the audio signal does not play an important role in the dynamics of the system. Other parametrizations of veil and sound should be considered if the purpose is to influence more strongly the loop resonance by the audio signal.

5.4 Combination with Gesture

The simplest system (*GraphSon*₂ and the mouse cursor on a static pink color) has an autonomous resonance that is shown Figure 5. If this system is manipulated by an operator who controls the location of the fixed mass (the upper mass), the system behaves as follows (see Figure 9):

1. during gesture control, the output follows the constrained motion of the upper mass (the values between the two vertical dotted lines),
2. when the manipulation is completed, the system has a transient chaotic behavior (5 to 10 seconds),
3. finally the periodic resonance restarts and begins by

6. SYNTHESIS AND PERSPECTIVES

In this study, we have presented a model and an application that build an audio↔graphic feedback loop made of a MSS and its visual skinning, and a resonant bandpass filter. Audio level is used to control the physical system dynamics, while mass acceleration controls the filter characteristics. In addition, color under mouse cursor directly parametrizes the filter and indirectly modifies the MSS reactivity. The loop actually behaves like a resonant system with a period between 2 and 5 seconds. Periodicity is better observed on a simple MSS or in quiet situations (soft filter and dark color).

Further studies could be carried out:

- The system dynamics can be studied formally in the simple case by taking into account the internal characteristics of the audio and graphic systems and the information propagation delays between the two components. The output of the formal study should

be then compared with the dynamics observation in the computer model.

- The artistic or industrial applications of such an audio-visual environment for the realistic or non-realistic rendering of natural phenomena such as wind can be further investigated. Current works tend to study separately graphical and sonic modeling [13], but we are convinced that deeper investigations of the perceptual correlations between sound and image in the modeling of such natural phenomena are promising directions of research [7, 6]. It is therefore necessary to design new generations of audio-visual environments such as the one presented in this study to offer a framework for such studies on multi-modal modeling and perception.
- For sound creation purposes, richer parameter sets and richer topologies could be taken into consideration: other MSS topologies such as the ones explored by PMPD for audio-visual composition [12], other audio patches with physical modeling of wind phenomena such as the ones used for musical instruments [17], other color parameters such as hue, saturation, and value, and more complex visual renderings through physical cloth modeling or shaders and BTF textures.
- If the purpose is to design a virtual instrument that uses the feedback resonance for graphical and audio synthesis, gesture-based control should be investigated more deeply, possibly with haptic feedback [14]. High speed in graphical rendering through bitmap animation or decoupling of mass-spring animation and associated skinning would yield higher resonance frequencies in the audio↔graphic loop and produce interesting audio-visual patterns.

7. ACKNOWLEDGMENTS

This study has benefited from a research collaboration on virtual instruments between the authors and Hugues Genevois (LAM), Brian Katz (LIMSI), and Norbert Schnell (IRCAM). Many thanks to Brian Katz, Jean-Baptiste Thiebaud (Queen Mary Univ.), and the three anonymous reviewers for their comments on a preliminary version of the paper.

8. REFERENCES

- [1] Audiosculpt. <http://forumnet.ircam.fr/349.html>.
- [2] S. Barrass and G. Kramer. Using sonification. *Multimedia Systems*, 7(1):23–31, 1999.
- [3] C. Cadoz, A. Luciani, J.-L. Florens, and N. Castagné. ACROE-ICA: Artistic creation and computer interactive multisensory simulation force feedback gesture transducers. In *Proceedings of New Interfaces for Musical Expression (NIME'03)*, pages 235–246, 2003.
- [4] T. Chen, S. Fels, and T. Schiphorst. Flowfield: Investigating the semantics of caress. In *ACM Special Interest Group on Graphics and Interaction (SIGGRAPH'02)*, page 185, 2002.
- [5] T. Coduys and G. Ferry. IanniX: Aesthetical/symbolic visualisations for hypermedia composition. In *Proceedings International Conference Sound and Music Computing (SMC '04)*, 2004.
- [6] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. West, G. Pingali, P. Min, and A. Ngan. A beam tracing method for interactive architectural acoustics. *The Journal of the Acoustical Society of America (JASA)*, 115(2):739–756, 2004.
- [7] J. K. Hahn, J. Geigel, J. W. Lee, L. Gritz, T. Takala, and S. Mishra. An integrated approach to motion and sound. *The Journal of Visualization and Computer Animation*, 6(2):109–124, 1995.
- [8] S. Jordà. Sonigraphical instruments: From FMOL to the ReacTable. In *Proceedings of New Interfaces for Musical Expression (NIME'03)*, pages 70–76, 2003.
- [9] G. Levin and Z. Lieberman. In-situ speech visualization in real-time interactive installation and performance. In *Proceedings of NPAR'04*, pages 7–14, 2004.
- [10] MaxMSP 4.5 reference manual. <http://www.synthesisters.com/download>.
- [11] MetaSynth. <http://www.uisoftware.com/MetaSynth/>.
- [12] A. Momeni and C. Henry. Dynamic independent mapping layers for concurrent control of audio and video synthesis. *Computer Music Journal*, 30(1):49–66, 2006.
- [13] S. Ota, T. Fujimoto, M. Tamura, K. Muraoka, K. Fujita, and N. Chiba. "1/f β noise-based real-time animation of trees swaying in wind fields. In *Proceedings of Computer Graphics International (CGI'03)*, 2003.
- [14] X. Rodet, J.-P. Lambert, R. Cahen, T. Gaudy, F. Gosselin, and F. Guédy. Sound and music control using haptic and visual feedback in the PHASE installation. In *Proceedings of New Interfaces for Musical Expression (NIME'05)*, pages 109–114, 2005.
- [15] D. Schwarz. Recent Advances in Musical Concatenative Sound Synthesis at Ircam. In *Workshop Audio Mosaicing: Feature-Driven Audio Editing/Synthesis*. (ICMC'05), Barcelona, 2005.
- [16] A. Sedes, B. Courribet, and J.-B. Thiebaud. Visualization of sound as a control interface. In *Proceedings of the 7th International Conference on Digital Audio Effects (DAFX'04)*, 2004.
- [17] J. Smith. Virtual acoustic musical instruments: Review and update. *Journal of New Music Research*, 33(3):283–304, – 2004.
- [18] Virtual Choreographer 1.2 reference manual. <http://virchor.sourceforge.net/html/>.
- [19] P. Waters and A. Rowe. Alt-space: Audio-visual interactive software for developing narrative environments. In *Proceedings CADE2004*, 2004.