THE SOUNDING GESTURE: AN OVERVIEW

Amalia de Götzen

VIPS Group, Dept. of Computer Sciences
Univ. of Verona, Verona, Italy
degotzen@sci.univr.it

ABSTRACT

Sound control by gesture is a peculiar topic in Human-Computer Interaction: many different approaches to it are available, focusing each time on diversified perspectives. Our point of view is an interdisciplinary one: taking into account technical considerations about control theory and sound processing, we try to explore the expressiveness world which is closer to psychology theories. Starting from a state of the art which outlines two main approaches to the problem of ”making sound with gestures”, we will delve into psychological theories about expressiveness, describing in particular possible applications dealing with intermodality and mixed reality environments related to the Gestalt Theory. HCI design can indeed benefit from this kind of approach because of the quantitative methods that can be applied to measure expressiveness. Interfaces can be used in order to convey expressiveness, which is a plus of information that can help interacting with the machine; this kind of information can be coded as spatio-temporal schemes, as it is stated in Gestalt theory.

1. INTRODUCTION

Sound started to find its place in the broad field of Human-Computer Interaction in the last decade of twentieth century. In particular, there have been two important events that pushed sound research towards the area of non speech auditory output: the special issue of the HCI journal on non speech sound edited by William Buxton [1] in 1989 and the First International Conference on Auditory Display held in 1992 [2]. Thus, during the 1990’s the community has grown and has given birth to many research themes; many of them can still be considered new and open investigation fields: finding new auditory interfaces and displays in which gesture can take its place is one of them.

Gesture and sound seem naturally connected in a clear and obvious way: the image of instrument players learning to use their body in order to produce sound is indeed widespread and compelling enough. While each instrument needs specific gestures to be played in a correct and pleasant way, it is possible to find some invariant laws regulating gestures across all instruments. Computers can be thought as just another musical instrument, perhaps not new but still far from having a coded tradition related to musical gesture. Musical gesture can be simply thought as a gesture that produces sounds in an continuous feedback loop: this is a general definition that can be used in many interactive contexts besides the musical ones.

This kind of studies involves several research domains which are now getting to communicate and work together: human performance, auditory perception and signal processing are all involved in this investigation area.

2. GESTURES IN HUMAN-COMPUTER INTERACTION

The importance of gesture in the restricted community of Human-Computer Interaction which deals with sound is increasing. A significant sign of this increased attention can be found in carried out research in the field: in the last 5 years 2 research projects have been funded by the European Commission: MEGA (Multisensory Expressive Gesture Application) and the Cost287-ConGAS (Gesture CONtrolled Audio Systems) action. There are many reasons for that:

- this is an emerging subject which can benefit from advances in the other field (e.g. human performance, auditory perception, signal processing);
- it is a complex multi-disciplinary field which encompasses a strong scientific and technical side which does indeed include several kinds of communication, intelligent interaction, cognitive aspects, experimental psychology (related to gesture emotion, intention analysis and to a esthetic cognitive processes) augmented with humanistic research related to music. Communication of all these aspects happens through sound; however, the peculiarity of this communication is that messaging is esthetic rather than semantic (i.e. it cannot be coded in a unique and unambiguous way); this trans-disciplinary approach should be considered as a plus in a context like telecommunication where intelligent interaction and cognitive aspects are ever more relevant.

While both projects are dealing with gesture, they are quite different in scope, focusing in particular aspects of gesture interaction. The MEGA project (cf. http://www.megaproject.org), which ended in 2003, was centered on modelling and communication of expressive and emotional content in non-verbal interaction through the use of multi-sensory interfaces in shared interactive mixed reality environments: the attention was directed to the emotional content conveyed by gesture. In particular, the project focused on music performance and full-body movements as first class conveyors of expressive and emotional content. Main research issues were the analysis of expressive gestures (i.e. analysis of the expressive content conveyed through full body movement and musical gestures), the synthesis of expressive gestures (i.e. the communication of expressive content through computer generated expressive gestures, such as music performances, movement of virtual as well as real (robotic) characters, expressive use of visual media), the strategies for mapping the analysis data into multimodal output. MEGA showed a clear connection between gesture and expressive intention using approaches that integrate classical recognition techniques with novel analysis techniques by integrating consolidated scientific experiences along with theories from music performance and from choreography, the Laban's theory of effort. [3, 4].

Laban and Lawrence use a 4D space to classify human movement based on effort in which the axes are exertion (light - strong), control (fluent - bound), effort (flexible - direct) and duration (sustained - quick); see Figure 1. Exertion is concerned with strength or weight (W); control with space (S), effort with flow (F) and duration with time (T). The eight basic (W,S,T) exertions are slashing, gliding, pressing, flicking, wringing, dabbing, punching and floating.

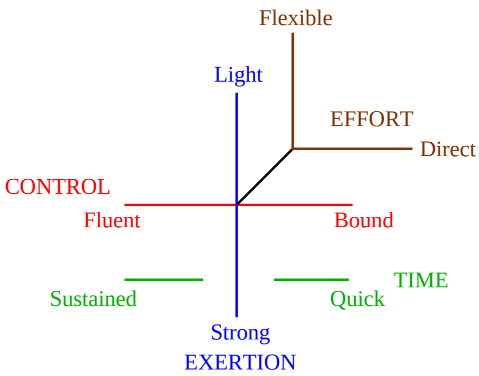

Figure 1: *Laban 4D space*

In this context, the meaning of gesture seems to be a very broad one: generally speaking, gesture is often referred to dancer movements and sometimes to specific body expressions, but gesture can be considered also a structure with definite semantics defined into an abstract space. Thus, a musical phrase can be considered a gesture which can express an emotion using only musical parameters, where the music is the abstract space. This connection between music and body movement is explicit in dance: a specific choreography can be used to better express a given musical work in a ballet and vice-versa; the emotional states carried by the music are the same ones expressed by body movement. Moreover, there are many psychological studies about music and emotion that maintain that music can represent the dynamic properties of emotions like speed, strength and intensity variations. In particular, Gurney asserts that music can express these emotions through the association of affinities between musical characteristics and body movements which can show them [5]. Furthermore, Imberty [6] underlines that there are some kinetic tension and release schemes which are typical for both body and emotion; this leads to think that movements and emotional states are a coherent set and gesture is a communication channel.

Starting from the previous definition of gesture many works have been carried out in the following different (though connected together) areas:

- Analysis and synthesis of expressive content in human movement and gesture;
- Analysis and synthesis of expressive content in musical gesture performance.

In both cases the analysis process starts from gesture-derived information (physical movements or audio signals), captured by sensors into a computing system. These physical signals may be in different formats: they may consist of time variant signals such as sampled audio signals, sampled signals from tactile, infra-red sensors, signals from haptic devices, or events such as MIDI-messages or low-level data frames in video.

Several low-level features can be extracted and processed statistically in order to carry out subsequent expression-related analysis.

Considering the audio domain, these low-level features are related to tempo (=number of beats per minute), sound level, spectral shape (which is related to the timbre characteristics of the sound), articulation (eg. legato, staccato), attack velocity (which is related to the onset characteristics which can be fast or slow), pitch, pitch density, degree of accent on structurally important notes, periodicity (related to repetition in the energy of the signal), dynamics (intensity), roughness (or sensory dissonance), tonal tension, and so on.

Expressive cues can be located and extracted measuring these low-level physical parameters (e.g., kinematics data such as position of body joints or tracking of points on the body silhouette); they include global measures (i.e., cues depending on full body movement, such as the contraction index and the quantity of motion), measures depending on the current position of body joints such as the stability index, cues inspired by Rudolf Laban's Effort Theory such as the directness index and the fluentness index, cues inspired by psychological studies such as the durations of pause and motion phases.

The synthesis process begins from the human silhouette extracted from the videocamera or microdance recordings (a simple black and white bitmap). Some of the spatio-temporal cues extracted from movement in the analysis phase are mapped into visual changes/morphing of the silhouette: In other words, movement cues such as "quantity of movement (momentum)" or "direct/flexible" are mapped into visual cues synthesized in the silhouette.

Most of the research done has then been focused on so-called semantic spaces or maps. A semantic map represents categories of semantic features related to emotion and expression on a predefined grid. Typically, a gesture is then a trajectory in this space, and each trajectory can be seen as a point in a trajectory-related (super)space.

Energy-velocity spaces have been successfully used to synthesize musical performance. The space is derived from perceptual experiments [7] and it has been used in synthesis of different and varying expressive intentions in a musical performance thus far. The energy-velocity space is correlated with legato-staccato properties versus tempo. In this space, positions are used to define MIDI parameters as well as audio signal parameters which control the timing and the dynamics of the notes to be played during a performance. The MIDI parameters typically control tempo and key velocity. The audio-parameters control tempo, legato, loudness, brightness, attack time, vibrato and envelope shape.

The main characteristic of this kind of abstract spaces is that they are very related to the concept that interpret the physical space which can be of different nature: it is a sort of multimodal space. The interesting think is that, for instance, I can analyze an image extracting some expressive characteristics that I can use to synthesize sound. We can do a well known close loop where the input is the gesture and the feedback is the audio; see Figure 2

All these elements show that gesture is a wide speculation topic in human-machine interaction research which can be studied in depth in systems that involve art and digital music [8] and, more generally, sound. Furthermore, audio systems (including Digital

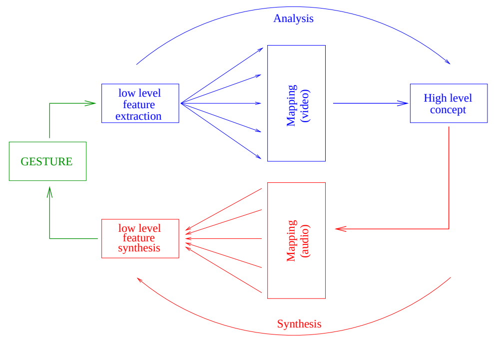

Figure 2: *Analysis/Re-synthesis throw high level mapping*

sound synthesis and processing) are a wide field in which the control aspect (which is the focus of the Cost287-ConGAS action — `http://www.cost287-congas.org`, started in 2003) still needs to be studied and to be linked with scientific investigation upon gesture. In fact, since real-time digital signal processing has become a reality for many digital sound effects (including sophisticated ones), an increased knowledge of gestural devices and their interaction with digital sound effects is now necessary: the control of digital audio systems by gesture goes in that direction.

On the other hand, in traditional musical situations gesture usually produces sound. The relationship between gesture and sound is unique, it is a cause to effect link. In computer music, the possibility of uncoupling gesture from sound is due to the fact that computers can carry out all the aspects of sound production from composition up to interpretation and performance. Real time computing technology and development of human gesture tracking systems may enable gesture to be introduced again into the practice of computer music, but with a completely renewed approach. There is no longer a need to create direct cause-to-effect relationships for sound production, and gesture may be seen as another musical parameter to play with in the context of interactive musical performances.

From another point of view, with the emergence of new types of interfaces and technologies, including wireless devices, touch-weight-pressure sensitive devices, virtual reality interfaces, force-feedback devices, etc., new types of human computer interfaces are being provided to enhance and supplement the keyboard/slider/-potentiometer controllers usually considered for digital sound production, getting closer to the wide variety of means and I/O modalities found in traditional music instruments.

The overall goal of such systems is to enhance the naturalness of human-computer interaction through more cognitive and intuitive interfaces in a field like digital sound production that requires an amount of communication precision and detail which is not usual in other Human-Computer Interaction (HCI) systems. This poses several challenges to the design and control of interaction between systems and humans where sound production is involved. Summing up, the control of sounds using gestural devices goes in two directions: on one hand it recovers a several centuries-long tradition tied to instrumental playing which proved to be extremely rich in nuance and detail while on the other it improves a new way to use machines to interact with sounds: it is more direct and natural, and it opens new and unexplored possibilities.

A pioneering work on gestural input devices was carried out during the analog electronic music period. The invention of the Theremin [9] anticipated the development of new expressive instruments. A noticeable aspect is that this first electronic musical instrument, a simple oscillator performed without physical contact, could produce very subtle and varied sounds because the generated sound reflected more the expressive quality of human movement than its own timbre quality.

Further recent works, using new technologies, have been done in this field [10, 11, 12, 13, 14] and a lot of new instruments with new control metaphors have been developed.

While the control aspect of the interaction is deemed interesting, the research interests are not limited to digital musical instrument control which can be considered just the first step of a deeper investigation. A few aspects have to be further investigated to find a way to use evaluation methods well known in HCI (e.g. Fitts' law[15]) in a multimodal context, dealing with motion, sound and expressivity (e.g. the Gestalt approach).

3. EMOTION FROM MOTION: EXPRESSIVE GESTURES

The previous paragraph faced the problem of defining what gesture is and how it can convey expressiveness. The 'MEGA approach' is quite different from the 'Gestalt one' that we are about to consider now.

The gesture movement expressivity in fact has been deeply investigated by Gestalt theorists [16] and is concerned with the study of functional relations such as the perception of causality. Gestalt psychology stated that the phenomenical world is made by a lot of elements that interact with each other by means of a series of functional relationships. Many of these relationships form our common experience: for example, if we hit an object maybe something else will move, some shapes are difficult to grasp, pouring the wine into a glass, etc.. Situations like those mentioned above do not involve just spatial and kinetics aspects but also functional relationships: in the last example for instance we don't see just the wine that changes position, but we actually see the wine *coming out* from the bottle neck and *going down* into the glass! Thus, functional relationships are the weaving of the phenomenical world, which is the world which can be described completely and naturally by means of direct experience. Through the Gestalt approach functional relationships give sense to everything which is around us: we know what things *are* by observing what they *do*. Psychology is then the science of actions and behavior and it has become essential to study perception like a phase of the action. In this context it may occur that an object influences a behavior just because of its meaning which is an element of perception itself. Such perception depends on the relationships which the object have with other objects: temporal and spatial relationships are experienced as functional relationships, be it causality or finalization.

3.1. Causality perception

Interesting studies have been carried out by Michotte [17] on the perception of causal relationships from which observations and researches about movement expressiveness are derived. These observations involves tertiary qualities of the objects which are often defined as *physiognomic*, that is "informing on the nature" of the object or event. We can in fact distinguish between:

- *primary qualities: dimensions, shapes, weight etc.*;
- *secondary qualities: color, taste, affective valence*;

- *tertiary qualities : good/bad, gloomy/happy, threatening/attractive*.

Michotte found out that precise timing is needed to achieve perceived causality. The best-known experiment and result is about the launching effect: Michotte studied the perception of causal relationships between two light spots that move always along the same line with a variety of velocity patterns. See Figure 3.

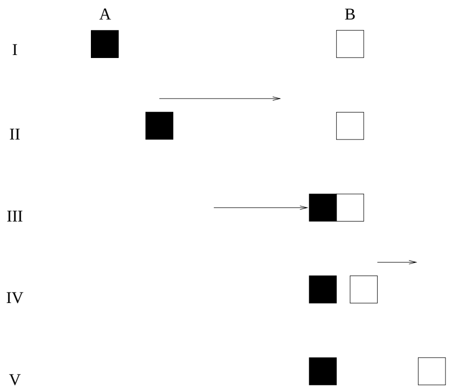

Figure 3: *Michotte's launching effect (from Vicario 2004)*

Under particular experimental conditions Michotte found out that the A movement appear as active, while the B movement appear as passive. Both movements are simple, identical translations into space; however, under specific timing relationships between them they acquire an expressive quality which leads to perceived expressive individualities (or tertiary qualities) of both elements. This kind of characteristics are qualities of the perceptive structure and, because of that structure, objects and events are expressive. A very interesting propriety of that structures is their *intermodality*: a structure can maintain the same qualities in every sensorial modality: we perceive the same expressive quality with different senses (this is why, for example, we say that a voice (hearing) is sour (taste)). It is the principle for metaphor creation. An appropriate example of this property is shown in Figure 4.

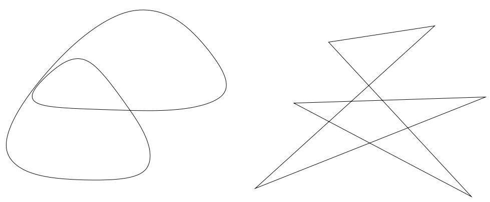

Figure 4: *The intermodality of expressive qualities: Who is Takete and who is Maluma? (from Köhler 1984 [18])*

In this example the absolute majority of people say that Maluma is the round figure while Takete is the other one. This is a fine example since both images and words make no sense to anyone, but because of their similar structure they obtain a coherent result. The angular figure has rapid changes like the sound of Takete, while the round figure has a continuous line without abrupt discontinuities, like the sound of Maluma. This kind of studies is very interesting for Human-Computer Interaction which deals with multimodality very often. It is a stimulus to find structures similar to the ones discovered by psychologists about, for instance, the perception of causality in sound: it is clear that we express emotions through our body with gestures, voice, etc., because these means use the same structure of the emotion itself. We can understand emotions by observing body expressions; furthermore, it clarifies whether objects and movements have such similar expressive qualities.

Work made in expressive movement perception can be very useful to delve into the research on expressive control in the auditory domain: this work looks all the more promising if it is focused on the profile of variations of sounds rather than on their spectral composition, because then it could be possible to translate many of the results obtained in that field to sound.

Michotte studies were carried out in the '50s, and a lot of further investigation have been done since then by Levelt, Minguzzi, Kanisza and Vicario and some other researchers [19, 20, 16]. The question which is still open is to find the main factors of the phenomenical rendering of causality. Michotte found out that *movement amplification* was the common structure of all causality phenomena while Minguzzi, for instance, found out that the common element was the conservation of the kinematic status.

3.2. Applications

Based on the Gestalt approach some application has been developed in information visualization and in animation [21, 22, 23, 24]. In fact motion has revealed itself as a perceptually rich and efficient display dimension: it may prove useful in visualizing complex information because of its preattentive and interpretative perceptual properties, reducing some of the over-use of the representation techniques. As Michotte discovered, the key factor to causal perception is movement amplification, where the movement of the motor object (A) extends into that of the projectile (B). To verify this Michotte considered the perceived effect of A's movement on a qualitative change (appearance, disappearance, change in form or color) of object B. When there was no opportunity for movement amplification, no causality was remarked. Thus, merely causing objects to appear and disappear in temporal and spatial contiguity is insufficient for the impression of causality (as in flashing them on and off the screen in close coincidence); some form of *kinematic integration* must occur. It is quite natural to apply these results to new visual interface designs, which perhaps seem to be the counterpart of the audio-ecological approach. Using again Michotte results (such as the importance of movement speed in perception) Amaya et al. used signal processing techniques to analyze emotion in motion [25]. They captured movements from subjects performing two types of human activity, drinking from a cup and knocking on a door, in three emotional contexts (angry, sad and neutral). They identified two attributes which varied considerably over the different emotional movements: speed (frequency) and spatial amplitude (the range, or size of the motion). They divided the movements into basic periods (e.g hand to cup, cup to mouth, cup down, hand back); determined the speed of the end effector along its trajectory in both the angry and neutral movements., and calculated speed transforms for both the neutral and angry movement by integrating the longest period along the trajectory and dividing it into frame templates. Such transforms can then be applied to a new, neutral movement by time-warping the frame distribution, interpolating between frames where needed. The spatial amplitude intensity for each joint is calculated for each movement period and nonlinear signal amplification is used to apply the amplitude transform to generate new joint positions from the new, neutral movement. They tested the approach by deriving angry and sad transforms from the cup-drinking data. Then they applied the transforms to the neutral knocking data, and found a close match between the generated and the real (motion-captured) angry knocking motion data.

Similar works have been done with gesture and sound [26] from analogies between music performance and body motion. The aim was to identify relevant gestures in sound control in order to develop sound models that respond to physical gestures: the authors found out that subjects correctly classified different types of motion produced by the model. Substituting then sampled sounds with physic-modeled sounds a model of expressive walking and running has been built [27]. This kind of study can be very useful, for instance, in VR applications and, in this specific case, in order to characterize footsteps of avatars controlled in accordance to gestures dynamics to produce natural and expressive sound variations and to investigate how sound feedback can affect vision.

The cross modal effect has been deeply investigated: auditory and visual spatial information, originating in the same event, usually results in one unified percept and the interaction represent a prerequisite for phenomenal causality [28]. Several cases of visual influence on auditory perception are reported, such as the McGurk effect [29], while fewer ones report the opposite influence [30, 31] (e.g. the disambiguation provided by audio stimuli when a visual ambiguity is introduced). Recent work has been carried out by Gusky and Troje on audiovisual phenomenal causality [32]: following the Michotte method, the authors found out that the causality perception increases when additional auditory or visual information marks the onset of target motion. Similar studies have been carried out using physical modelling with continuous auditory feedback [33, 34]: continuous sound feedback can emphasize causality.

The main purpose of these works is to introduce interactive cartoon models of everyday sound scenarios with a new control approach through audio-visual-tangible systems. The idea is that human interaction in the world is essentially continuous, while the majority of sounds that right now are used in computer environments are totally unnatural (e.g. triggered samples). A few practical examples of that interaction have been elaborated by the SOb project (`http://www.soundobject.org`), developing rolling sound models based on physical models of impact [35]:

- *the invisiball*: a thimble acts as the sender and the receiving antenna is placed under a 3D elastic surface. Finger position in the 3D space is detected in real-time and it is used by the algorithm controlling the rolling movement of a ball;
- *the ballancer*: the user has to balance a ball tilting a rectilinear track, and the modeled sound of the ball rolling over the surface of the track along with a tactile-visual response provides the feedback.

These examples of interaction have demonstrated that everyday sounds can be very useful because of their familiar control metaphor: no explanation nor learning are necessary [36]. Moreover, it is clear that continuous audio feedback affects the quality of the interaction and that the user makes continuous use of information driven by sounds to adopt a more precise behavior.

4. CONCLUSIONS

After novel expressivity paradigms and gesture analysis methods have been developed in an effort to map human gestures into quantitative scales (e.g. Fitts' law, steering law, etc.), it has become crucial to deal with new, often specific sets of control parameters.

Furthermore, the role of multi-modality and multi-sensory communication will be central in the design of the next generation interfaces. As a consequence, non-speech communication will play an important role inside the information stream established between machines and users [37, 38].

Acoustic events play an important role in our general perception of the environment and can have a strong impact on our affects. This is especially true when acoustic cues are used to enhance or to complement the visual modality:

- audio feedback can be very informative when visual feedback is missing;
- sound conveys information about the environment (e.g. the material of a stroked surface, the brand of a car, the genre of a person walking or the voice of a person).

The study of the reaction of sounds to gestures will lead to more sophisticated ways to produce sounds. The typical structure of musical/instrumental gesture (division between control and sound production, that is between audio feedback, haptic feedback and visual cues, and sound production) has to be carefully studied and extended to lead to better (i.e. more natural) uses of sound effects.

Recent studies have shown that signal processing techniques can be fruitfully used to control the expressive content of a sounding object through gesture [39]. A challenge task of the research effort can be the desire to place the emotion expression in the foreground, providing a model with a method enabling the interaction between users and machines through sound and musical cues.

5. REFERENCES

[1] W. Buxton, "Introduction to this special issue on nonspeech audio," *Human-Computer Interaction*, vol. 4, no. 1, pp. 1–9, 1989.

[2] G. Kramer, *Auditory Display*, Addison-Wesley, 1994.

[3] A. Camurri, S. Hashimoto, M. Ricchetti, R. Trocca, K. Suzuki, and G. Volpe, "Eyesweb - toward gesture and affect recognition in interactive dance and music systems," *Computer Music Journal*, vol. 24, no. 1, pp. 57–69, Spring 2000.

[4] R. Laban, *Effort*, Macdonald & Evans Ltd., London, 2^{nd} edition, 1947.

[5] M. Budd, *Music and the emotions: the philosophical theories*, Routledge ed., London, 1992.

[6] M. Imberty, *Entendre la musique. Semantique psychologique de la musique*, Dunod, Paris, 1979.

[7] A. Camurri, G. De Poli, M. Leman, and G. Volpe, "A multileyered conceptual framework for expressive gesture applications," *Proc. of MOSART Workshop*, 2001.

[8] M. Wanderley and M. Battier, *Trends in Gestural Control of Music*, Ircam, Paris, France, 2000.

[9] T. Winkler, "Making motion musical: gestural mapping strategies for interactive computer music," in *Proc. of Int. Computer Music Conf.*, 1995, pp. 261–264.

[10] P. Cook, "Remutualizing the instrument: Co-design of synthesis algorithms and controllers," in *Proc. of Stockholm Music Acoustics Conf. 2003 (SMAC 03)*, Stockholm, Sweden, 2003.

[11] D. Arfib, J. M. Couturier, and L. Kessous, "Gestural strategies for specific filtering processes," in *Proc. of Int. Conf. on Digital Audio Effects (DAFX-02)*, Hamburg, Germany, 2002.

[12] S. Jorda, "Interactive music system for everyone: Exploring visual feedback as a way for creating more intuitive, productive and learnable instruments," in *Proc. of Stockholm Music Acoustics Conf. 2003 (SMAC 03)*, Stockholm, Sweden, 2003.

[13] R. Bresin, S. Dahl, M. Marshall, M. Rath, and B. Moynihan, "Controlling the virtual bodhran - the vodhran," in *Proc. of Stockholm Music Acoustics Conf. 2003 (SMAC 03)*, Stockholm, Sweden, 2003.

[14] D. Howard and S. Rimmel, "Gesture-tactile controlled physical modelling music synthesis," in *Proc. of Stockholm Music Acoustics Conf. 2003 (SMAC 03)*, Stockholm, Sweden, 2003.

[15] S. Gibet, J. Kamp, and F. Poirier, "Gesture analysis: Invariant laws in movement," in *Gesture-Based Communication in Human-Computer Interaction*, A. Camurri and G. Volpe, Eds., Berlin, 2003, Springer.

[16] G. Vicario, *Psicologia generale - I fondamenti*, 2004.

[17] A. Michotte, *La perception de la causalité*, Louvain: Publications Universitaires, 1946.

[18] W. Köler, *La psicologia della Gestalt*, Feltrinelli, Milano, 1984.

[19] W. J. M. Levelt, *Motion braking and the perception of causality*, pp. 244–258, Louvain: Publications Universitaires (Studia Psychologica).

[20] G. Kaniza, G. Finzi, G. Miguzzi, E. Marigonda, G. Tampieri, S. Wehrenfenning, G. Vicario, and M. Zanolla, *Ricerche sperimentali sulla percezione*, Università degli studi di Trieste, 1968.

[21] P. A. While and A. Milne, "Impressions of enforced disintegration and bursting in the visual perception of collision events," *Journal of Experimental Psychology: General*, vol. 128, pp. 499–516, 1999.

[22] C. Ware, E. Neufeld, and L. Bartram, "Visualizing causal relations," in *Proc. of IEEE Symp. on Information Visualization*, pp. 39–42.

[23] L. Bartram and C. Ware, "Filtering and brushing with motion," *Information Visualization*, vol. 1, no. 1, pp. 66–79, 2002.

[24] L. Bartram, C. Ware, and T. Calvert, "Moving icons: detection and distraction," in *Proc. of Interact Conf.*, Tokio, Japan, July 2001.

[25] K. Amaja, A. Bruderlin, and T. Calvert, "Emotion from motion," in *Proc. Conf. on Graphics Interface '96*, Toronto, Canada, 1996, pp. 222–229, Canadian Information Processing Society.

[26] R. Bresin and S. Dahl, "Experiments on gestures: walking, running, and hitting," in *The Sounding Object*, D. Rocchesso and F. Fontana, Eds., Firenze, 2003, Mondo Estremo.

[27] F. Fontana and R. Bresin, "Physics-based sound synthesis and control:crushing, walking and running by crumpling sounds," in *Proc. of Colloquium on Musical Informatics*, Florence, Italy, May 2003.

[28] J. Lewald, W. H. Ehrenstein, and R. Guski, "Spatio-temporal constraints for auditory-visual integration," *Behavioural Brain Research*, vol. 121, pp. 69–79, 2001.

[29] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, pp. 746–748, 1976.

[30] L. Shams, Y. Kamitani, and S. Shimojo, "Visual illusion induced by sound," *Cognitive Brain Research*, vol. 14, pp. 147–152, 2002.

[31] K. Watanabe and S. Shunsuke, "When sound affects vision: effects of auditory grouping on visual motion perception," *Psychological Science*, vol. 12, no. 2, pp. 109–116, March 2001.

[32] R. Guski and N. Troje, "Audiovisual phenomenal causality," *Perception and Psychophysics*, vol. 65, pp. 789–800, 2003.

[33] M. Rath and D. Rocchesso, "Informative sonic feedback for continuous human controlling a sound model of a rolling ball," in *Proc. of Int. Workshop on Interactive Sonification*, Bielefeld, 2004.

[34] D. Rocchesso, F. Avanzini, M. Rath, R. Bresin, and S. Serafin, "Contact sounds for continuous feedback," in *Proc. of Int. Workshop on Interactive Sonification*, Bielefeld, 2004.

[35] M. Rath and F. Fontana, "High-level models: bouncing, breaking, rolling, crumpling, pouring," in *The Sounding Object*, D. Rocchesso and F. Fontana, Eds., Firenze, 2003, Mondo Estremo.

[36] S. Brewster, *Non-speech auditory output*, Lawrence Erlbaum, 2002.

[37] F. Avanzini and D. Rocchesso, "Controlling material properties in physical models of sounding objects," in *Proc. of Int. Computer Music Conf.*, La Habana, Cuba, September 2001, pp. 91–94.

[38] F. Fontana, D. Rocchesso, and E. Apollonio, "Acoustic cues from shapes between spheres and cubes," in *Proc. of Int. Computer Music Conf.*, La Habana, Cuba, September 2001, pp. 278–281.

[39] S. Canazza, G. De Poli, C. Drioli, A. Rodà, and A. Vidolin, "Audio morphing different expressive intentions for multimedia systems," *IEEE Multimedia*, pp. 79–83, July-September 2000.