Motion To Gesture To Sound: Mapping For Interactive Dance

Jan C. Schacher
Zurich University of the Arts
Institute for Computer Music and Sound Technology
Baslerstrasse 30
CH-8048 Zürich, Switzerland
+41 43 446 55 00
jan.schacher@zhdk.ch

ABSTRACT
Mapping in interactive dance performance poses a number of questions related to the perception and expression of gestures in contrast to pure motion-detection and analysis. A specific interactive dance project is discussed, in which two complementary sensing modes are integrated to obtain higher-level expressive gestures. These are applied to a modular non-linear composition, in which the exploratory dance performance assumes the role of instrumentalist and conductor. The development strategies and methods for each of the involved artists are discussed and the software tools and wearable devices that were developed for this project are presented.

Keywords
Mapping, motion sensing, computer vision, artistic strategies, wearable sensors, mapping tools, splines, delaunay tessellation.

1. INTRODUCTION
"To be a body, is to be tied to a certain world, [...] our body is not primarily in space: it is of it."
Maurice Merleau-Ponty, Phenomenology of Perception [1]

Mapping for interactive real-time audio systems finds its greatest challenge in the interactive dance context. Compared to an instrumentalists gesture repertoire or a typical visitor in an interactive installation, dancers have a much higher awareness and sensitivity to spatial and physical motion aspects. In a traditional choreography paradigm, these are used to generate an expression, either as an image on a stage or as tension arch in a dramaturgy. For the dancer the elements of body, motion, space and effort imply a much higher degree of complexity than for a musician. Interactive instant composing and realtime control of sound transforms the normal hierarchical relationship a dancer has been trained for. In this new situation the dancer embodies the musical interpreter and takes on an instrumentalists role, whereas the composing musician steps back and provides a meta-composition in the form of algorithms and the organization of the sonic materials. Gesture mapping builds the bridge between the domains of space, bodies and motion and the abstract structures of a real-time interactive audio composition. The mapping's main role is to incorporate elements and know-ledge about the perception of movement and gesture, which are subsequently applied to the extraction of features from the sensed data and translated to expressive control of the essential features responsible for the musical expression.

2. BACKGROUND
Dance has seen a development in the direction of open forms and indeterminate choreographies since the nineteen-fifties. In the last twenty years these methods have evolved to incorporate technology-based musical and visual compositions. The technical equipment used for sensing the dancers has closely followed the evolution of materials and available devices and has become quite easily accessible and ubiquitous. Abstraction of dance elements into motion graphs and dance notation, such as the Laban methods [2], provide a systematic and methodological basis for the exploration of the possibility spaces of motion and effort elements in combination with the technical interfaces. The fundamental issue of combining dance with interaction and music is that the experience, training and perception of the performers in the domains of dance and music do not cover the same emphasis of performing music or gesture. The relationship between musical compositional structures and dance languages ahs been a field of study for a considerable amount time. [3]
Marc Coniglio of Troika Ranch states: (2002) [4] "The problem […] is that, to really play [the music], and for the audience to see that the dancers are playing, you need to move like a musician. [T]he movement of the dancer needs to be in service of the sound. […] But this is not the gestalt that we perceive when we watch a dancer move. We really see energy – We're […] looking at the way that the dancer moves through space and the overall articulation of the movement."
The experience of the divergence in perceptual weight between individual motion elements and musically expressive entities is common to all work in this cross-disciplinary field. On a fundamental physiological level the expression and perception of sound and movement – sensed by the eye and the ear – differ so much that a straight parametric linking between the phenomena of each domain seems impossible. Strategies have to be developed and methods devised to transform expressive qualities from the idiom of dance to those effective in music. Research in the field of music interaction tries to show how to translate abstract motion and gesture information into more universally applicable affects end emotions such as Camurri's groundbreaking work with Eye's web [5] in the and in the context of Kansei (2002) [6]. Arfib et al. (2002) propose models of abstraction between a gesture and a sound synthesis algorithm [7]. The strategy is to use perceptual spaces such as related-to-gesture and related-to-sound perception parameter layers rather than pure gesture or sound synthesis parameter spaces. These can be completely decoupled from the specific

affordances of the gesture acquisition devices and the specific parameter spaces of the sound synthesis algorithms and help express essential expressive elements. In installation art the issues of sensing the audience interaction with sound and image through movement and gesture acquisition has led to quite sophisticated computer vision applications that have to address some of the same issues as the ones discussed here. Adaptive and evolving mappings form part of the strategies to enhance the audience engagement. [8]

3. MAPPING

The following overview builds on and extends upon much of the literature about mapping in the domains of electronic music and dance in the last decade. The main focus will be put on the special circumstances of integrating two or more differing sensing modalities into a Meta description, which expresses affect and emotion rather than physiology and physics.

3.1 Mapping as a Bridge

Mapping is a central subject in all forms of interaction. It forms the bridge between the actor and the materials being acted on. For an interactive music system it forms the essential connection that expresses the compositional ideas about form and structure. For an interactive dance concept the main challenge is to be able to grasp the expressive aspects of motion and translate them into meaningful expressive aspects of the sound work. When working with few channels of sensing data, this relationship can quite easily be established through experimentation and the principles applied to a variety of aspect in the sound processes. The larger the number of sensing parameters is and the more modes of sensing and interacting these cover, the more complex this task becomes.

3.2 Motion to Gesture Translation

One technique that has proven useful for unifying several sensing streams into semantically useful information (many-to-one-mapping) is to introduce the notion of the gesture as a Meta layer of motion. Gesture in interaction nowadays signify finger swipes on touch-enabled devices, such as cell-phones or table-top display, but on a more fundamental level a gesture is a sequence of movements that form a whole, a gestalt and can be recognized as semantic unit. A gesture in the context of dance can have a high degree of abstraction from everyday movements, but it can nevertheless convey information and emotion: "A particular emphasis is on affective, expressive, emotional information. In fact, it is the capability of interpreting expressive information that allows interaction of technology and art at the level of the language art employs to convey content and to provide the audience with an aesthetical experience." Camurri (2004) [9] On a language level the gesture could be compared to a sentence whereas individual movements would be equivalent to words or even phonemes. A gesture in the musical context is geared towards producing a certain result and is highly instrument specific. For a dancer the relationship leading from a gesture to a sound has no definition in the traditional paradigm of choreographies interpreting existing pieces of music. "These relationships may be established by viewing the body and space as musical instruments, free from the associations of acoustic instruments, but with similar limitations that can lend character to sound through idiomatic movements" Winkler (1995) [10]. The abstract gesture extracted and interpreted from the lower-level motion data becomes an entity with which to compose interaction and ultimately build musical pieces that convey a stronger relation to choreographic methods rather than to mere interaction through the links established by using a specific technology. (Figure 1.)

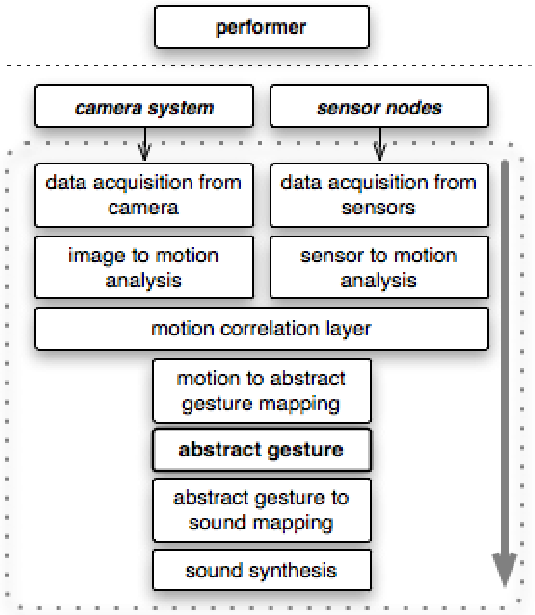

Figure 1. Schematic structure of a multi-tiered mapping.

3.3 Multi-Tiered Mapping

Working with multiple different sensor input presents an extended playing field, where the abstracted gesture can be made to express a richer meaning. Spatial scopes of movements are an excellent example of the superposition of two perspectives in the gesture data acquisition. On the one hand the global space of the stage offers a topographical approach to motion mapping via camera based motion tracking. On the other hand sensors placed on the body that react to physiological inputs in a body centric space provide detailed information in a limited scope but expressive scale. This type of complementary layering of sensed information can be readily extended, but the complexity increases and the readability of a mapping from an outside point of view diminishes. Special attention has to be paid to contradictory information delivered by differing sensing mechanisms; an indication of the influence and bias exerted by the affordances the sensor channel on the acquired data. A robust data reduction, correlation and weighting algorithm is needed to help assuage this problem, but aesthetic and perceptual decisions must still be made by the composer to get the aesthetically meaningful results.

4. INTERACTIVE DANCE PROJECT “MOVE2”

In some ways the Constellation of the Project “Move2” reflects the typical setting for an interactive dance piece. A composer, a dancer and digital artist collaborate to build the piece. The premise starts from a modular interaction-driven composition that receives its definitive shape only during a performance. The dancer has a background in improvised performance as well as traditional dance training. Both the composer and digital artist share a history of exploratory music performance and open form music composition. To reach stage-worthiness a joint development process is undertaken and is in fact still ongoing. After establishing the technical basis of the interactive system, the work-process comprised many sessions of experimentation and trials before a form and the means for the interactive dance piece was found.

4.1 Composer Strategies

The composer has to be aware of the non-linearity of a performer driven musical form. The sonic material is organized in ways, which make recombination in almost all permutations possible. Expressive low-level musical synthesis parameters such as gain, filter frequency or lfo-rates are exposed to be driven by the dancers gestures. Discrete triggering events shape the overall form of the piece; the decisions about selection and removal of sonic materials lie mostly in the performers "hands".

4.2 Performer Strategies

For the performer the one big challenge is the non-repeatability of certain musical combinations. The high degree of mnemonic training acquired for learning and repeating long choreographies is of limited use here. The learning process is one of memorizing guiding principles of the interaction and has to lead to a familiarity with the musical materials and the combination afforded by the interactive system. The tendency to develop a gesture language geared towards the musical mapping is one of the pitfalls to be avoided. It is more important to gain an understanding of the rules of play and to enter into the exchange without needing to assert total control over the entire interaction process.

4.3 Mapping Developer Strategies

For the digital artist the role is one of a mediator or translator. The mapping and bridge building is primarily an act of communication between the idioms of dance and music. Being rooted in one of the two helps but part of the process is to learn to communicate the essential elements in terms valid in both domains. Technically the mapping has to be adaptive and flexible and the amount of extracted data just enough to provide a rich and diverse set of gestures and interactive situations without flooding the musical processes with less than meaningful data. The development of these techniques happens in a cyclical process, where new strategies can be introduced periodically but has to become persistent enough for both the composer and the performer to be able to integrate the new possibilities into the existing fabric of the interactive piece. To find a method that permits the extraction of salient features in a wide variety of ways, and apply it in an interesting variety of was s probably to be the hardest part of the development of interactivity and the mapping. The best strategy has proven to build layers of abstraction, each concerned with one domain of complexity. Raw motion data for example can provide data for interesting mappings when connected to sound primitives such as amplitude, but are much more difficult to apply to more differentiated parametric controls to sound aspects such as timbre and micro-time chunk processes like the ones used in granular synthesis processes.

5. IMPLEMENTATIONS

The technical system for "Move2" is fairly straightforward in terms of the components being used. Two laptop computers share the load of the work. On the first computer all sensor data acquisition from wearable sensors and computer vision algorithms takes place, the data is conditioned and the mapping is processed. The resulting discrete control events and continuous control streams are transmitted via MIDI to the second machine, where the data is applied to a commercially available track- and mixing-desk based audio software, which synthesizes or combines existing materials in a highly modular fashion.

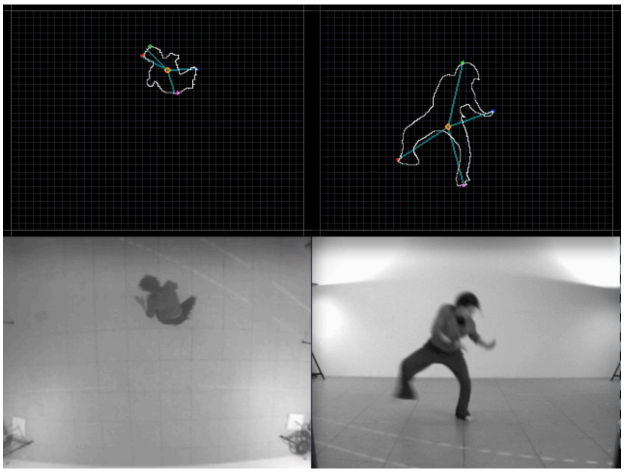

Figure 2. Dancer on screen with two vertical and frontal views each displaying body contour and cardinal point analysis.

5.1 Computer Vision

For the image capture part, an industrial firewire camera is used, which is capable of delivering 60 frames per second at VGA resolutions. It is positioned along the central axis of the stage and oriented horizontally into the performance space. The image shows the body of the dancer with the perspective of the audience, precisely the way bodies are perceived in a natural environment. A vertical camera position is added to obtain a more accurate view of the absolute position on stage. The zenith perspective lacks the depiction of the body shape necessary for gesture extraction but gives a better topographical view of the stage, which is essential information when working with spatialized audio. A computer vision algorithm based on OpenCV [11] is implemented in Max/MSP Jitter [12] using the cv.jit library [13] and extended with custom code written in C. This custom process computes the convex hull of the body's contour through a Voronoi/Delaunay triangulation, and extracts the five cardinal points of the body and the centroid of the hull. From this reduced set of points many salient features can be extracted: position and orientation in space, including correcting for the foreshortening generated through the camera perspective, speed and acceleration of the centroid and the points at the extremities, the body span – contraction and expansion in both the lateral and vertical dimensions. The reduction of information with this algorithm is quite effective, while maintaining a notion and knowledge about body shape. Further work is underway to refine the contour analysis to continuously track the points at the tips of the limbs and the head through convexity defects analysis of the contour and body posture estimation [14]. (Figure 2.)

5.2 Wearable sensors

A set of wireless sensor bracelets is used to detect body-space information in a very accurate fashion and with a fairly high sampling rate. The bracelets are built from readily obtainable electronic components and consist of a three-axis accelerometer, a three-axis gyroscope, a small-scale micro-controller board, a low-voltage XBee wireless transmitter and a small rechargeable Lithium polymer battery. These are mounted onto a flexible band and can be attached to the limbs using Velcro straps. The small size and compactness of the wireless sensor nodes makes it ideal for the dancer to wear several of these on the extremities of the limbs. With an update rate of more than 100 Hz and a sensitivity in the range of 1.5 g for the accelerometer and the three-dimensional gyroscope integrated in on the board, these sensors are ideal for detecting the attitude in space at rest and the acceleration and rotation energy expended on the limb during motion. (Figure 3).

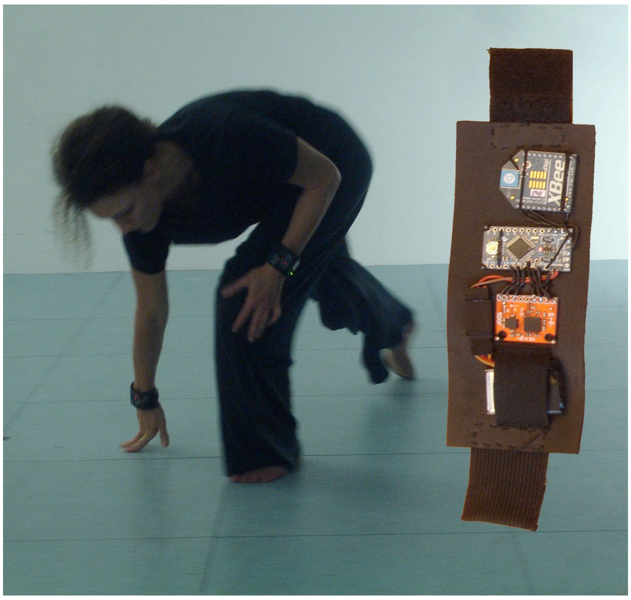

Figure 3. Dancer wearing wireless sensor bracelets; inset the details of a bracelet with the LiPo battery, 5DOF sensor, Arduino Mini Pro, Xbee Module (bottom to top).

The Nime community has seen quite number of wireless sensor applications in the last few years [15] many of them using custom-built sensor electronics and transmitters. More advanced multi-node systems also exist that are capable of reconstructing the full body model from inertial and accelerations measurements. [16] The bracelets presented here inherit from that tradition but present an evolution in the direction of a more simple and easy to integrate project, something that is achievable by artists without engineering degrees (such as the author).

5.3 Mapping Tools

To complement these data acquisition devices a set of extensions for the Max/MSP programming environment has been developed that facilitate some of the basic processes used in typical mapping layers. The first is a lookup-table based transfer function called icst.map that accepts arbitrarily complex sets of points and maps a given input range to a desired output range. A companion tool for this is implemented in an object called icst.spline and offers interpolation algorithms in the form of splines of different complexities from arbitrary sets of control points. These splines also provide a generic set of interpolation routines for an arbitrarily high dimensionality. An earlier version of a tool for graphically defining bezier and bspline curves was used for generating sound trajectories in spatialized music. [17] The spline algorithms implemented here range from simple to higher order polynomials such as (in order of increasing complexity) the classic bspline, the catmull-rom, cardinal, hermite, and advanced kochanek-bartels splines but also offer basic cubic and linear interpolation. Using these processes in the transfer functions of one-to-one, one-to-many, two-to-many and many-to-many mapping connections has proven particularly interesting for adaptive functions where weightings need to be changed in real-time for entire regions of the data-streams; moving one control point on a spline is a simple yet very effective way to transform an entire range of the transfer function. (Figure 4.)

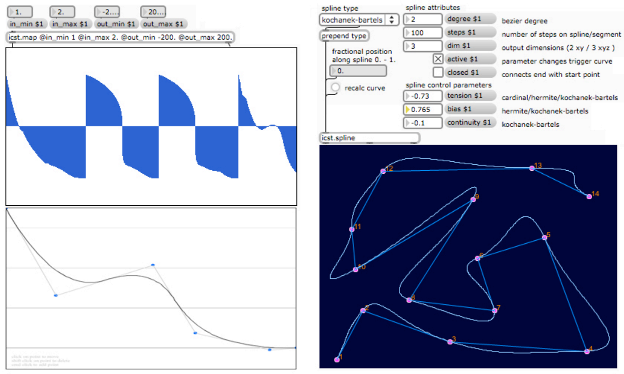

Figure 4. MaxMSP implementation of scaling function with adaptive behavior using bspline-curves changing through several cycles (left) and interpolation using a Kochanek-Bartels spline with bias, continuity and tension control simulating inertia and lag (right).

In order to correlate several layers of sensor data, the simple parallel transformations of streams becomes soon impractical. It is necessary to create a space of sensor-streams that can be traversed simultaneously. With this goal in mind the transfer-mapping tools are extended to higher dimensional maps, using algorithms such as the Voronoi and Delaunay tessellations [18] to calculate relationships of points to certain regions of a point graph. One immediate application is to use these tessellation algorithms to generate the convex-hull from a set of contour-points in a video-image but the main usage is to extract neighborhood relationships from two-dimensional fields of points (Figure 5). This approach continues in the direct lineage of prior work presented by Van Nort in his Library of Maps tools [19], Bevilacqua with the MnM toolbox [20] and Bencina with the Metasurface [21], but I believe it provides a different, more generic approach in terms of the details of working with vectors or matrices of points and is above all directly applicable to body contours presented in the rasterized pixel-matrices of video-images.

5.4 Example Mapping Sequence

The mapping of a movement by the dancer, for example the lifting of the right arm with a certain emphasis and speed, could happen as follows. The information traverses the entire mapping system to produce a sound fragment designated by the composer to express the quality of rapid lifting. The chunking of sensor-data happens in several different time frames simultaneously and is applied on different compositional levels to affect the musical result. After extracting the raw sensor data from a motion chunk in the image domain, the speed and upward direction of the left cardinal point is recognized as expressing a possible fast lifting gesture. The potential gesture is stored and compared to the gesture result from the accelerometer data of the bracelet. If they match, the results of the cardinal points and the accelerometer analysis are then combined and weighted in order to determine which gesture has taken place. Once a correspondence of a sufficiently high degree is established a discrete event and a number of qualifying descriptors are transmitted to the composition interpreter, which determines how to produce and articulate the sound gestalt most fitting with the gesture found.

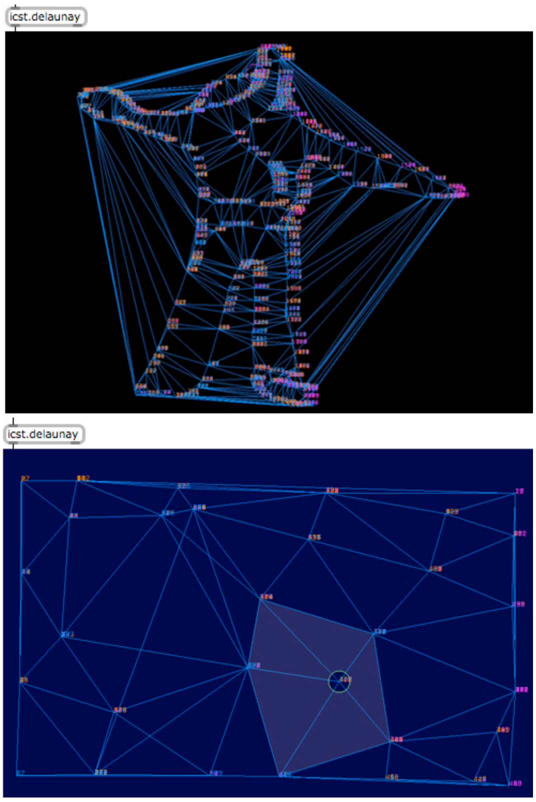

Figure 5. Body contour and convex-hull extracted from a video-image (above) and spatial neighbor relationship calculated on a set of points using the Delaunay tessellator in MaxMSP Jitter (below).

6. CONCLUSION

Multi-tiered mapping poses a number of interesting challenges in several of domains, not the least being the perceptual impact of interaction. The use of an abstracted semantic gesture model can facilitate the integration of complementary, sometimes-contradictory sensor data. Defining perceptual rather than technical criteria for qualifying a mapping both in the motion/gesture domain and in the musical domain help to avoid the typical pitfalls of overly simplistic or exceedingly complex mapping relationships. Efforts are to be undertaken to establish a systematic approach to these abstractions, and the tools that are available from other domains have to be brought into play for the benefit of a more intelligent, emotionally engaging and aesthetically satisfying motion to gesture to sound translation.

7. ACKNOWLEDGMENTS

The author would like to thank the composer Marie-Cécile Reber and the dancer Angela Stöcklin for their continued dedication to this project.

8. REFERENCES

[1] Merleau-Ponty, M., *Phenomenology of Perception*, Gallimard, Paris, 1945, Routledge & Kegan Paul, London, 1962, p.171

[2] Laban, R. v. (1963). *Modern Educational Dance,* Macdonald & Evans Ltd., London

[3] Hodgins, P. (1992) *Relationships between score and choreography in twentieth-century Dance: Music, Movement, and Metaphor* Paul Lewiston, N.Y. E. Mellen Press

[4] DeLahunta, S. "An *Interview with Mark Coniglio"* http://www.sdela.dds.nl/sfd/isadora.html

[5] Camurri, A., Hashimoto, S., Ricchetti, M., Ricci, A., Suzuki, K., Trocca, R., Volpe, G., *"EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems*", Computer Music Journal, v.24 n.1, p.57-69, April 2000

[6] Camurri, A., Trocca, R., Volpe, G., *"Interactive Systems Design: A KANSEI-based Approach"* in: Proceedings of the 2002 Conference on New Interfaces for Musical Expression (NIME02), Dublin, Eire

[7] D. Arfib , J. M. Couturier , L. Kessous and V. Verfaille, "*Strategies of Mapping between Gesture Data and Synthesis Model Parameters Using Perceptual Spaces*", *Organised Sound* 7(2): 127–144, 2002 Cambridge University Press. UK

[8] Schacher, J.C, "*Action and Perception in Interactive Sound Installations: An Ecological Approach*", in: Proceedings, New Interfaces for Musical Expression, NIME 09 Conference, Pittsburgh 2009

[9] Camurri, "*Multimodal Interfaces for Expressive Sound Control*" in the Proceedings of the 7th Int. Conference on Digital Audio Effects (DAFX-04), Naples, Italy, 2004

[10] T. Winkler, "*Making Motion Musical: Gesture Mapping Strategies for Interactive Computer Music*" in the Proceedings of the 1995 International Computer Music Conference.

[11] http://sourceforge.net/projects/opencvlibrary/

[12] www.cycling74.com

[13] http://www.iamas.ac.jp/~jovan02/cv/

[14] Mittal, L. Zhao, L. S. Davis "*Human Body Pose Estimation Using Silhouette Shape Analysis*" Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003

[15] Lee, E., Enke, U., Borchers, J. de Jong, L. "*Towards Rhythmic Analysis of Human Motion using Acceleration-Onset Times*", in: Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07), New York, NY, USA

[16] Majoe, D., Kulka, I., Schacher, J.C, "*Towards An Interactive Multimedia Experience For Club Music And Dance*" in: Proceedings, International Conference on Mobile Computing and Multimedia MOMM, Kuala Lumpur, Malaysia, 2009

[17] Schacher, J.C "*Gesture Control of Sounds in 3D Space*" in: Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07), New York, NY, USA

[18] M. Gold and M. Dakowicz. "*Kinetic Voronoi/ Delaunay drawing tools.*" In Proceedings 3rd International Symposium on Voronoi Diagrams in Science and Engineering (ISVD 2006), pages 76-84, Banff, Canada, 2006.

[19] VanNort, D. Wanderley, M., "*The LoM Mapping Toolbox for Max/MSP/Jitter*" in: in: Proceedings of the 2006 International Computer Music Conference ICMC 06, New Orleans, USA

[20] Bevilacqua, F., Müller, R. Schnell, N. "*MnM: a Max/MSP mapping toolbox*" Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05), Vancouver, BC, Canada

[21] Bencina, R. (2005). "*The Metasurface: Applying Natural Neighbor Interpolation to Two-to-Many Mappings*" in: Proceedings of the 2005 Conference on New Interfaces for Musical Expression (NIME05), Vancouver, BC, Canada