“3rd. Pole” – a Composition Performed via Gestural Cues

Miha Ciglar
Sound Artist / Student
University of Music and Dramatic
Arts Graz, Austria
++43 650 973 3947
miha.ciglar1@guest.arnes.si

ABSTRACT

“3rd. Pole” is a musical composition that is performed by a dancer, on a specially designed interface (instrument), based on motion tracking technology. This paper introduces the technical and artistic ideas behind the composition and outlines some of the main conceptual tendencies of my earlier work [8]. Several independent components like choreography, instrument design, sound design, formal concepts, etc. were in parallel development throughout the last three years, and came together in this piece. Some of those were already realized in individual projects and are joined now under a new concept of interdependence. A major component of this project however was the implementation of a real-time gesture follower / gesture recognition algorithm applied to full-body motion data. This should be considered as an autonomous module, not allied to this particular project exclusively. Therefore it can be treated and developed independently – as a multi-purpose data mapping strategy – for it has the potential to find further use in any kind of interactive (dance, music, theatre, etc.) performance scenario.

Keywords

Motion tracking, Gesture recognition, Haptic feedback

1. TECHNICAL DESCRIPTION

1.1 Infrastructure

The dancer is monitored by the *“Vicon 8”* motion capture system [15]. A brief description of the system and some common data mapping strategies for musical applications can be found in [3]. The *“Vicon 8”* system consists of 12 infra red cameras / sensors, placed around the dancer (performance area) and is able to track and extract the Cartesian x/y/z coordinates of light-reflecting markers on his body in 3 dimensional space, at a sampling-rate of up to 120 frames per second. In our case, the markers were arranged in groups, so that a characteristic constellation of 4 to 5 markers attached to the end of each limb (fig.1) would represent one central point from which we received our spatial coordinates. The trajectories of those coordinates were then used as an input for a gesture recognition algorithm, inspired by the concept of left to right Hidden Markov Model architecture [14], [11]. The algorithm was implemented in the real-time programming environment: PD (Pure data) [13] which is receiving the location data from the Vicon server through the OSC communication protocol [16].

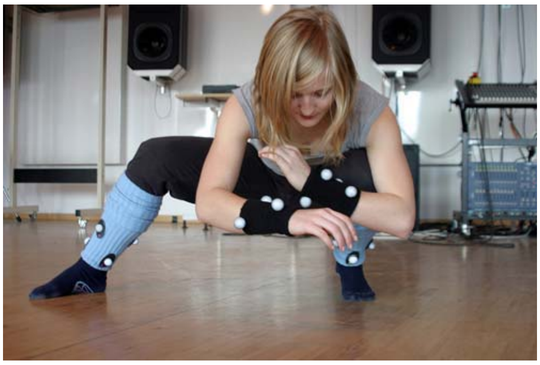

Fig.1: markers attached to the dancer’s limbs

The dancer receives feedback from the system in two ways. In form of music and in form of electricity, that is directly applied to his body, through a cable, he is holding in his mouth. Ideally the setup should be realized with a wireless unit so the dancer has more freedom to move and does not get tangled in the cable.

1.2 Gesture Recognition

1.2.1 Related work

There is a variety of different approaches dealing with gesture recognition in performing arts, deploying all kinds of sensing devices. Those are applied to movement patterns of dancers / performers directly [1], [4] or, if concerning musicians, to their conventional instruments like in [2] - showing an example of such an augmented instrument. Gesture recognition / classification techniques may also be of benefit in analytical applications like for example in music pedagogy [4], or in [7], where it was proposed to separate style and structure of full body gestures and to analyze stylistic differences between different gesture realizations. Further, there are also freely available tools for gesture recognition, like the MNM and FTM libraries [5] developed at IRCAM, for application within the MAX/MSP programming environment [12].

1.2.2 Initial directives

My goal was not to design a blind, – multi-purpose pattern recognition system, which could operate with any kind of multidimensional data, but to take in account the particular characteristics of human-body gestures with special consideration of the human perception capabilities and its tolerance against significant variations in different realizations. It is an attempt to simulate the human ability to read and recognize a gesture as an abstract sign, by drawing attention to the temporal progression of the relational statistics among selected body features. The key thesis that I was trying to work with is that a human body-gesture can be sufficiently described or abstracted by the trajectories of the inter-point (marker) distance variations.

1.2.3 Inter-point distance variation

A first version with 4 points (markers), which mark the ends of extremities (arms, legs) was completed, where the variation in distance between each pair of markers is taken as feature. By choosing this approach, we immediately get rid of the absolute coordinates in space and are not bound to a specific location or orientation of the performer inside the tracking area. Four markers would generate 6 distances between the markers, respectively a 6 dimensional vector space for modeling the state sequence that would classify a particular gesture. In this situation, where we work with 4 markers, we achieve a 50% dimensionality reduction: from **12** (4 * x,y,z coordinates) to **6** (inter–point distances), however this approach would not be so effective once we increase the amount of markers.

1.2.4 Spatiotemporal quantization

The gesture recognition system was designed not to distinguish between **intensity** as well as **temporal evolution** of different gesture realizations, which should allow for some degree of variation in the interpretation. The *inter-point distance variation* parameter set is a time dependant vector, defined by the sign information of the first derivatives (velocity) of the incoming signal. Its individual dimensions are thereby confined to 3 discrete states: (constant “0”, increasing “+1” and decreasing “-1” distance), making the system unsusceptible to gesture intensity. With 6 dimensions (distances) in 3 possible states respectively, we have a set of $3^6 = 729$ different state vectors to distinguish between. Whenever one dimension changes its state, the system would generate a new state vector, keeping the unchanged dimensions as they were, thereby allowing an arbitrary and even nonlinear temporal evolution of a gesture. After the algorithm has been trained, a gesture can be identified in real-time as it is being conducted and we get a continuous parameter describing the degree of completion of a particular gesture.

1.2.5 Gesture segmentation and state clustering

There is no perceptually relevant gesture segmentation taking place in this algorithm, because the state vectors described above are mostly being generated burst-wise. On the other hand, those bursts could be interpreted as indicators of transition points, of perceptually relevant segments. However, there is merely the succession order of those incoming state vectors that is of our interest here, but the exact timing (duration) information of the incoming state vectors is not captured by the algorithm and is not used for gesture classification, for we want to achieve freedom in the temporal interpretation of the gesture. The bursts or temporal clusters of different states occur due to large distance jumps through the vector space and would result from a change of movement direction of every single limb in relation to the others. If only one limb is activated (changing location) and suddenly alters its course, the usual consequence is a change of value in three dimensions (the relation to the other three static limbs). Ideally, in this case we would generate a single location change in the dimension space. In practice however, the individual dimension values would not change absolutely synchronous, resulting in a line of trace through the vector space, which is composed of instable (elusive) states and is pointing from one stable state to the other. This phenomenon would appear even more pronounced in the case of complex, full-body gestures. There is a slight variation in the sequence as well as the actual presence of those instable states in successive realizations of the same gesture, which is why we need to define a radius of tolerance (a cluster in the vector space) for each incoming state of the probe sequence. This radius was designed to exhibit a dynamic behavior, namely to allow for a specific degree of deviation from the currently compared state in the reference (exemplar) sequence, but simultaneously featuring indifference towards a specific “location” (the exact dimension) in which the deviation might occur.

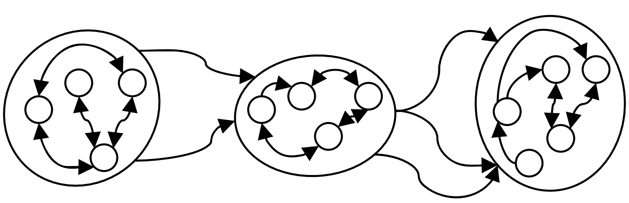

Fig.2: a sequence of three temporal clusters of state vectors (indicated by the small circles) with a dynamic tolerance radius, (indicated by the ellipsoids), representing the spatial clusters

1.2.6 Adaptive filtering

Before the state vectors are determined, the incoming signal is low-pass filtered. The exemplar sequences that serve as the reference for later recognition are recorded with fixed filter parameters, and should be conducted as clear and evenly as possible. Later, in the recognition phase, the length of the integration window of the low-pass filter is being adapted in real-time, according to the overall acceleration value of the incoming signal. This approach is related to the idea of gesture segmentation described in [6], where the parsing of body motion into different gestural segments is based on the interpretation of acceleration values of the incoming signal. Although, the segment lengths and their durations are not relevant in our algorithm, the information about the location of transition points was found to be a useful parameter for the adaptive filtering of incoming data. Here, the sum of absolute values of the second derivatives of single dimensions of the incoming signal is the criterion for the choice of the number of integrands in the filter. The amount of states generated by the system depends on the size of the integration window processing the incoming signal. To assure a satisfactory inter-gestural discrimination, and a sufficient intra-gestural (variation) tolerance, we need to “code” the incoming data with redundant information where less is being generated by the nature of the signal (slow movements, few coarse changes). Whereas on the other hand, we need to reduce the amount of data being generated at high signal acceleration values (transition points) in order not to loose track of the gesture progression due to an excess of data. Experimental results have proven this strategy to outmatch a system with a static filter design.

1.2.7 Time warping

Although the adaptive filtering component should foster the disaggregation of temporal clusters and an equal state density distribution along gesture progression, there are still situations where the proportional variations of individual gesture segments exceed the threshold of correct recognition. If the current state of the probe signal, for example, does not match the currently compared state of the recorded reference sequence, neither its values would fit inside the probe-state clusters tolerance radius, the incoming state vector is being passed on to a time warping function, which compares it against a certain neighborhood of states. If this function finds a match in the values of the neighboring states of the reference sequence, it time warps the probe signal to it and updates the index of the state that is to be compared next.

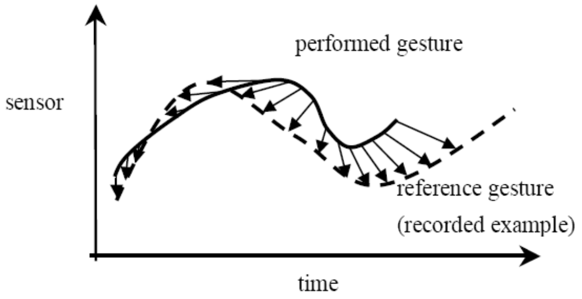

Fig. 3: Time warping the probe- to the reference gesture – image taken from [4]

1.2.8 Identification process

One of the intentions of this project was to blur the causal relationship of movement and sound, as it is usually the case when we apply direct mapping between sensor data and musical parameters. However, the approach of generating musical parameters via gestural cues should not restrain the control data to discrete values emerging at the end of a successful completion of a predefined gesture. The goal was rather to stick to the possibility of generating continuous output data, but to restrain it to accompany only specific choreographic material. Thus we are expecting to work with a continuous output parameter describing the degree of completion of a particular gesture in real-time. The algorithm does not need to output probability values or to show a degree of deviation from the temporarily observed state, since the acceptable deviation limits are already integrated in the clustering radius, the time warping function, etc. described above. We are not interested in how strong a deviation really is as long as it is inside a carefully chosen tolerance radius considering an adequate inter- and intra-gestural discrimination / tolerance. Each gesture in our prerecorded gallery has its own module, continuously monitoring the input feed. If the initial state of a gesture is being detected, the attention is put to the next and so on, for as long as the break condition is not exceeded. If this is the case, the algorithm stops tracking the gesture and returns in the initial state to continue looking for the beginning of the gesture again. As soon as it turns out that a gesture is not the one we are looking for, the algorithm needs to be ready to accept a new "candidate" sequence. Not all the incoming data needs to be assigned to a particular prerecorded gesture, and therefore we are not selecting the highest likelihood among our reference sequences to match the probe sequence. Thus the dancer is able to provoke an expected sonic result by selecting its choreographic material in real-time and to avoid sonification of his actions (different from the recorded gestures) if not desired.

1.2.9 Results and observations

It is to say that the algorithm is still in development at this time and all the constellations of different parameters were not extensively tested yet. The tests that we made up to now showed following results: Through careful tuning of the algorithm parameters, it was possible to achieve around 80% correct identifications – (4 out of 5 identical gestures (including variation factors) were recognized to 100%). At the same time, the inter- gesture discrimination was kept under 70%, i.e. no more than 30% of a "false" (arbitrary) gesture were identified as one of the reference gestures.

It is obvious that an approximation of a gesture through four points on a human body is not very accurate and satisfactory. Further, the concept of inter-point distance variation usually does not discriminate between mirror-inverted gestures, etc. We also discovered that it is possible to work with gestures of varying complexity levels (from robotic to more fluent and natural choreographies), but it is very important to maintain an equal degree of complexity in all gestures that we want to identify, since the algorithm tuning parameters depend strongly on gesture complexity. The selected choreographic vocabulary has to exhibit as much diversity between its single elements (gestures) as possible and the algorithm parameters need to be tuned according to it. However, if we take in account the specific conditions and limitations of such an approach, we can still develop a well distinctive choreographic language / vocabulary that might even set of a new and unique – system-conditioned aesthetic of movement.

1.2.10 Future work

For now there is still a lot of testing and tuning work to be done with this particular approach. In the further development of the gesture recognition system, I would still like to stick the basic principles of spatiotemporal quantization described in this paper, but to put more focus on the state-bursts (the temporal clusters described above) in the recognition process. Perhaps more reliable information could be gained, by disregarding the exact temporal progression of the state vectors, and by analyzing the temporal progression of state clusters instead. Then the statistics of state occurrences in such a cluster would be compared to each other in different gesture realizations. Since it was found out that the clusters mark the transition points of gesture segments, they consequentially include all the directional information of the preceding as well as the following segment.

1.3 Feedback

By moving through space, the dancer conducts actions in three spatial dimensions plus one temporal dimension. A fundamental part of the musical composition is the function that translates those actions to a two dimensional space (a time varying amplitude (the audio signal)), and will undergo a detailed discussion later in the text. The dimension of amplitude refers to the (fast changing) electronic signal waveform corresponding to the sound being generated and projected. In addition to the sonification of the electronic waveform, which produces an auditory feedback, the dancer is also exposed to an alternative instance of the same signal. This instance is the (amplified) signal itself, in its primary (the electronic) domain. The connection with the dancer is established by a cable which he is holding in his mouth. This concept of direct electronic signal-feedback was already applied and discussed in my earlier compositions and interface designs [9], [10]. It enables the dancer / performer to experience an alternative impression of the induced sound. Since it is electricity we are dealing with here, the dancer would feel a pain with waveform (sound amplitude) dependant intensity. Therefore, we need to be very careful with the amplification of the signal in order not to seriously harm the dancer.

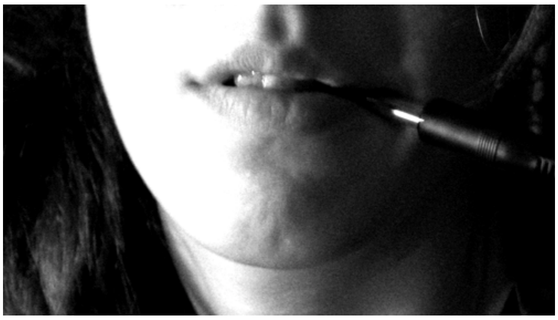

Fig. 4: the dancer with the audio-output cable in her mouth

2. ARTISTIC CONCEPTION

In a dance performance, there are usually 2 elements (visual and audible) that need to be arranged and put into a contrasting or harmonizing etc. context. The title *"3rd. Pole"* should indicate the inclusion of a third, a haptic component contributed by the electronic current running through the dancer's body. He is exposed to a situation where he is in absolute decision power and needs to consider and outbalance all three elements (poles). Like already mentioned, we have the induced sound respectively its electronic abstraction, which is in direct contact with the performer's body. This enables a different corporal perception and interpretation of the caused sound, since now the performer does not only have the audible but also a haptic reference - i.e. pain, caused by the electric current - for the choice of his following actions. Therefore, also the process of composition or better to say, the final arrangement of pre-composed material is only possible in real time, since we are interested in an alternative arrangement of the choreographic and musical progression, which is inspired by all three "poles" together. A pre-composed form or sequence of events would not make any sense, apart from satisfying possible sadistic tendencies of the composer.

3. CONCLUSION AND FUTURE WORK

The focus of this paper was rather on the interfacing concept and the interactivity of the system. A second major component of this project besides the gesture recognition system was sound design, which was not discussed here at all. Those components however are not bound to each other, so the project presented here is not meant to be considered as a sealed (finished) entity. It can be developed further independently in both, the artistic (musical, choreographic) and/or technological domains. *"3rd. Pole"* is only a first manifestation of an artwork and stands for one of many possible results that can be achieved in the future.

4. ACKNOWLEDGMENTS

This project has been supported by STEIM – www.steim.org – in Amsterdam (NL), by offering me a residency, during which I completed most of the sound design for this composition. The most important part however contributed the Institute for Electronic Music (IEM) – www.iem.at – in Graz (A), by providing the facilities and technical infrastructure. Special Thanks goes to Dr. Gerhard Eckel and David Pirro for providing theoretical opinions and IOhannes Zmölnig for technical assistance.

5. REFERENCES

[1] Aylward R. and Paradiso J., "Sensemble: Awireless, compact, multi-user sensor system for interactive dance" Proc. of the International Conference on New Interfaces for Musical Expression (NIME 06), Paris, France, 2006.

[2] Bevilacqua, F., Fléty, E., Lemouton, S., Rasamimanana, N., Baschet, F. "The augmented violin project: research, composition and performance report" Proc. of the International Conference on New Interfaces for Musical Expression (NIME 06), Paris, France, 2006.

[3] Bevilacqua, F., Dobrian, C. "Gestural Control of Music Using the Vicon 8 Motion Capture System", Proc. of the International Conference on New Interfaces for Musical Expression (NIME 03), Montreal, Canada, 2003

[4] Bevilacqua, F., Fléty, E., Guédy, F., Leroy, N., Schnell, N. "Wireless sensor interface and gesture-follower for music pedagogy", Proc. of the International Conference on New Interfaces for Musical Expression (NIME 07), New York, NY, USA, 2007

[5] Bevilacqua, F., Muller, R., Schnell, N. "MnM: a Max/MSP mapping toolbox ", Proc. of the International Conference on New Interfaces for Musical Expression (NIME 05), Vancouver, Canada, 2005.

[6] Bevilacqua, F., Cuccia, D., Ridenour, J. "3D motion capture data: motion analysis and mapping to music", Proceedings of the Workshop/Symposium on Sensing and Input for Media-centric Systems, Santa Barbara CA, 2002

[7] Brand, M. "Style machines", In Proceedings of SIGGRAPH, New Orleans, Louisiana, USA, 2000

[8] Ciglar, M. homepage: http://www.ciglar.mur.at

[9] Ciglar, M. "I.B.R. Variation III." Proceedings of the EMS – Electroacoustic music Studies Network Conference – Beijing, China - October 2006

[10] Ciglar, M. "Tastes Like…" Proceedings of the ACM Multimedia Conference. Singapore, November 2005

[11] Jie Yang, Yangsheng Xu, Chen, C.S. "Human action learning via hidden Markov model" IEEE Transactions on Systems, Man and Cybernetics, Part A, Jan, 1997.

[12] Max/MSP programming environment http://www.cycling74.com/products/maxmsp.html

[13] Puckette, M. "Pure Data" Proceedings of the ICMC, 1996

[14] Rabiner, L. R. and Juang, B. H., "An introduction to hidden Markov models," IEEE Acoust. Speech Sign. Process. Mag. 3 (1986) 4-16.

[15] Vicon 8 motion capture system: http://www.vicon.com/entertainment/technology/v8

[16] Wright, M. "Open Sound Control: an enabling technology for musical networking" Organised Sound, 2005/12/01, Volume 10, Issue 3, p.193-200, (2005).