FROM MOTION TO EMOTION: A WEARABLE SYSTEM FOR THE MULTIMEDIA ENRICHMENT OF A BUTOH DANCE PERFORMANCE

MICHAEL BARRY[4], JUERG GUTKNECHT[4], IRENA KULKA[3]
PAUL LUKOWICZ[1,4] AND THOMAS STRICKER[2]

[1]*Electronics Laboratory, Dept. of Electrical Engineering & IT*
ETH Zurich, Switzerland
lukowicz@ife.ee.ethz.ch

[2]*Laboratory for Computer Systems, Dept. of Computer Science*
ETH Zurich, Switzerland
gutknecht, tomstr@inf.ethz.ch

[3]*Hyperwerk - Interaction Design, Univ. of Applied Arts*
Basel Stadt und Land, Switzerland
irena.kulka@gmx.net

[4]*Institute for Computer Systems & Networks, UMIT*
Hall in Tirol, Austria
michael.barry, paul.lukowicz@umit.at

We present a mobile, multimedia system based on a network of body worn motion sensors, a wearable computer and a visualization engine that is used to produce a visual enhancement of Butoh dance performance. The core of the system is a novel motion classification scheme that allows us to capture the emotion expressed by the dancer during the performance and map it onto scripted visual effects. We describe the artistic concept behind the multimedia enhancement, the motion classification scheme and the system architecture. In an experimental evaluation we investigate the usefulness and the robustness of the wearable computer as well as the classification accuracy of the motion-sensing system. We also summarize the experiences with using the system for live performances on stage in several shows.

Keywords:

Communicated by: D Taniar

1   Introduction

Enhancing artistic expression has always been a fascinating application of multimedia technology. New systems are eagerly adapted by the art community. Beyond their artistic value such systems have often been on the forefront of technology exposing scientifically interesting problems and leading to innovative solutions in computing. This paper describes a system that demonstrates how the newly emerging wearable computing and sensing technology can be used to implement a dynamic, emotion driven multimedia enhancement of Butoh dance performance. Butoh is a mixture between free-form dance, performing arts and meditation,

as outlined Section 2.1. From a technical point of view, a Butoh dance could be characterized as a controlled, however amorphous motion of the whole body, through which the dancer expresses his/her moods and emotions. Emphasizing the artistic aspect of Butoh, it is often said that *"Butoh is what happens to dancing when the rational mind stays out of the way"* [1, 2]. Butoh performances are often accompanied by music and visual effects. However since improvisation is central to the Butoh concept, it is difficult to use statically predefined or pre-recorded effects. Instead it would be desirable to adapt the effects to the dancers performance. This is what our work aims to achieve using body-worn acceleration sensors and a wearable computer. In addition to our interest in arts, we look at the project as a first step in using some simple wearable sensors to correlate abstract characteristics of human motion with moods and emotions. (For a discussion of related work see Section 6).

At this time we have implemented a complete system that has been used for live performance on stage at several shows and carried out a systematic study of the recognition accuracy of the different motion characteristics during the dance. The paper starts with artists perspective of the problem, the idea behind the project, and the visualization concept (Sections 2.1 and 2). This artistic part leads to a 3-dimensional dance style classification scheme. The scheme is the interface to the recognition algorithm which is described and evaluated in Section . We then continue with the description of the wearable system used in our experiments and the visualization engine (Section 4). The paper closes with a summary of the impressions and experiences during several performances on stage.

2   Artistic Concepts

2.1   Butoh Dance

Butoh dance is a contemporary dance improvisation method originating in modern Japan. The method has already influenced contemporary western dance, performance and dance theater. Butoh's special combination of meditation with expressive articulation creates an altered mode of movement, working along a deeper mental and esthetic physiological automatism, without adhering to explicit forms. This abstraction makes it an interesting model system to study the structure of abstract esthetic feelings in general and to implement such structures in practical interactive applications. To the best of our knowledge no such studies have been done in the past on Butoh. A most recent progress in the research of dance expressivity came from recent related studies [3, 4] (See Section 2.4 and 6 for a detailed discussion). Although Butoh uses an unusual kind of movement for a dance style, it makes use of a particularly broad expressive spectrum and some of its expressive aspects can be compared to natural movement and possibly to expressivity in other dance styles [4].

2.2   Artistic Idea

Machine-enhanced artistic output is often considered to be interesting, as it is conceptually new and playful. Still it might be too discontinuous in terms of esthetics and content, to involve deeper levels of human perception and association. The idea of our work is to enhance the experience of Butoh performance by providing real-time biofeedback that is continuously related to an actual state of abstract and esthetic perception by humans, as expressed in the dance movement.

Rather than taking the simplistic approach of imitating human behavior, our goal remains

to invent artificial experimental structures that have an inspiring mental effect on the observer. Our system recognizes different esthetic patterns created by human motion, interprets them in terms of esthetic feeling, and translates this 'meaning' into machine-constructed visual effects. (See Section 5 for more detail.) The interpretation rules presently used, remain simple and are based on mirroring emotionally expressive states or contiguous sequences of such states. The feedback of a visual output during an improvisation provokes some fairly challenging changes in a performance process. Performers can adapt to the interactive instrument by means of their composition skills or by actually growing into a new motion technique [5], even including changes of their physical body scheme and proprioception. A growing community of performance-artists is actively exploring such possibilities [5]; however still little work is based on expressivity recognition and on an intuitive, 'natural' movement semantics [6, 3].

Within our framework of a variable composition structure based on a set of emotionally and esthetically tuned elements, performers can enter intuitively composed forms of improvisation, reflexive feedback processes or find new forms of interactive collaboration with visual artists. With our wearable sensing and visualization system, dance performers can include their own visual articulation in their improvisation work. The real-time visual output can finally serve as a complementary or interfering language for the spectators' interpretation of a performance, adding a further expressive and narrative level to the work.

2.3   Mapping motion to emotion

As depicted in Figure2, our multimedia enhancement system comprises the following physical and logical parts: A *body area network* made of acceleration sensors and a *wearable controller* processing the input into an abstract *motion space*, mapping it into an abstract *emotion space* and passing it on to a *visualization engine*

The major purpose of the abstract motion and emotion spaces is providing an properly abstracted framework for the mapping between "motion" and "emotion".

Taking an artistic perspective, we addressed the problem of relating motion to emotion in three steps: (1) definition of esthetically and expressively relevant motion criteria covering a versatile motion system, (2) construction of an organized system of emotions, (3) finding an intuitive connection between a specific emotion state and a specific motion state.
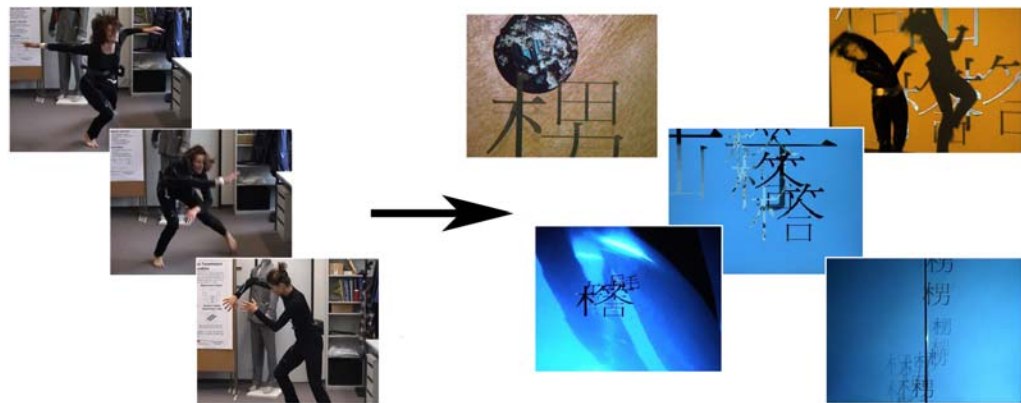

Fig. 1. Figure: Essential steps of our artistic concept. Sensing the motion, recognition of the patterns, and mapping them into scripted visual effects

Naturally, there are many ambiguities in this mapping: A restricted number of motion states represents a unrestricted continuum of feelings. Furthermore there is often more than one way to express one kind of feeling. A particular difficulty arises from the fact that Butoh differentiates highly similar esthetic, motoric and proprioceptive states. A slight change in associative shading of motion or a slight change of tension of the hand can lead to a radical change in the expressed feeling. This means that any conventional, rigid motion classification system remains far too coarse to capture the structure of Butoh. Nevertheless we believe that it makes sense to construct and to use an approximate structure in the sense of playing a composition that reflects some limited set of dimensions of emotion.

Motion and emotion are separate languages with generally hidden grammars and there is neither a formalism nor a simple logic nor any simple theoretical rationale in experimental psychology behind our choice of expressively relevant dimensions and the motion-emotion mapping. Experimental psychology is currently just at the beginning to observe such relations in respect to a commonly understood small set of basic emotions. As far as we know, no system or mapping has yet been established for the much larger amount of esthetic feelings related to motion. For these reasons and because our study was originally designed for a subjective artistic purpose rather than for a scientific study in psychology, we decided to start from a very subjective description of movements, esthetic emotions and their corresponding mappings as experienced by a Butoh dancer. In future studies, it would be possible to extend some aspects of this semantic complex with all the proper methods used in psychology research. A comparison of our descriptions with other systems and with available experimental results, suggests that our motion space is comparable to aspects of the Laban theory of effort (see Section 6 for more detail), to which many authors loosely relate as it is also derived from dance experience. Based on our experimental experience with performances we can say that many spectators perceived the visual effects spontaneously as correlated and enhancing - without any knowledge of the concepts behind our system.

2.4   The motion space

Butoh's expressivity is structured according to spatial references of a higher order. As a consequence, computing trajectories out of the acceleration data would not be a viable solution to capture the essence of Butoh.

We established a systematic protocol of different improvised dance sequences of reproducible expressive states or *categories*. Our simplified model has been derived from a systematic subjective description of the dancer's mental model of motion (imagery). Initially this lead to an extensive hierarchical system of 3 basic aspects, branching into finally 50 finely differentiated hierarchical 'dimensions'. The three most important ones are (a) space aspects, (b) trajectory aspects and (c) motoric aspects. In a refinement we distinguish between (1) space (direction and orientation of an imagined forcefield), (2) kinesphere (volume of action and center of imagined forces) (3) main symmetries (of forcefield and of action), (4) typical trajectories' form (relative directionality, fragmentation, shape) (5) trajectories' 'texture', fine articulation, : (6) typical motor condition (typical muscle tension, velocity, dynamics, rhythmic aspects), (7) posture. For practical reasons we then restricted the system to the basic aspects and the most common of the 'main dimensions'.

A condensation and rearrangement of these initial descriptive ideas resulted in a further simplified system of three dimensions, characterized by the three expressive criteria: *intensity*,
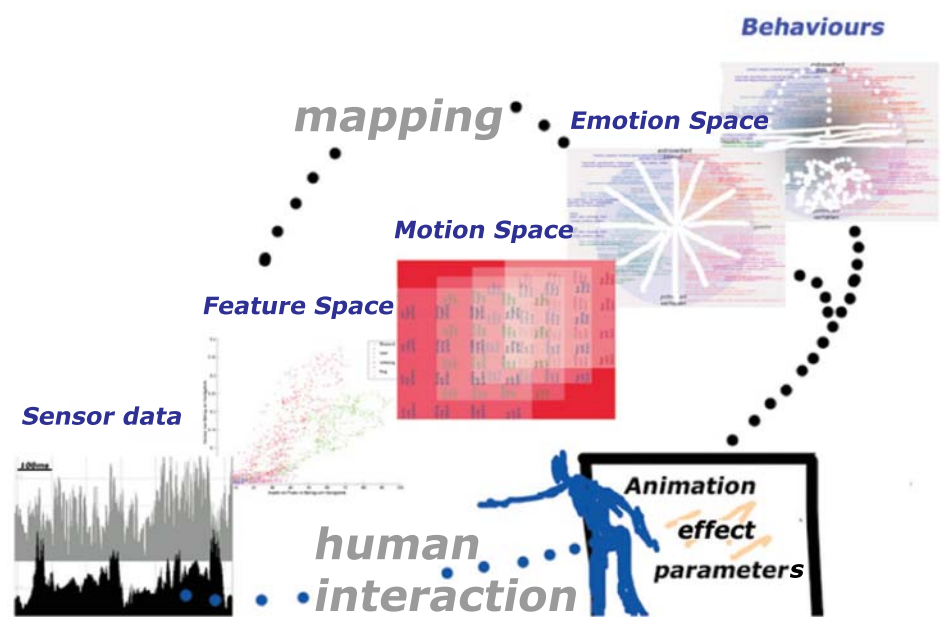
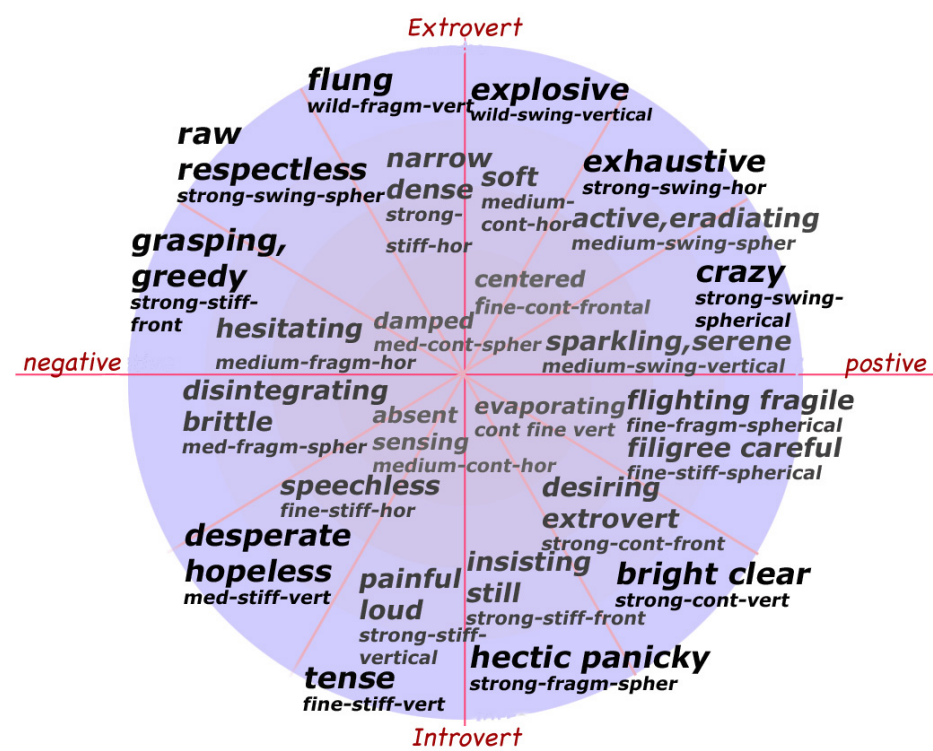Fig. 2. Mapping schema between sensor data, the motion patterns and the emotion space and visual effects.

Fig. 3. Our mapping between the motion categories and the emotion space using polar coordinates

flow and *direction* with four levels each. We defined each dimension in terms of a set of features that can be derived from the tracked data (see Section 3). We collected sensor data patterns of wrist, upper arm and upper leg) movement and matched them with a given combination of the abstract expressive criteria. The net result is a three-dimensional abstract space, partitioned into 4*4*4 = 64 different categories with the following factors and levels:

Intensity.   As the name indicates, this dimension captures the intensity (reflected in speed and also frequency) of the motions which can be: (1) Fine (extremely weak), (2) Medium (average, normal, relaxed, weak), (3) Strong (forceful), (4) Wild (violent, extremely strong).

Motion Direction   The motion direction captures the principal axis of the imaginary force-stream used in Butoh imagery, and corresponds basically to the axis or plane towards which the hand motions are statistically most frequently oriented in a given expressive state. It can be

1. Frontal. This denotes a motion where the arms move around a forward directed horizontal axis, passing mostly frontally with respect to the body, with exceptions when the body is turning 'in an imaginary tunnel'. The dance also contains a lot of arm motions going towards and away from the body.

2. Horizontal. This signifies a motion where the dancer imagines a horizontally irradiating force field. The arms mostly move rather stretched out laterally from the body in a horizontal plane, often seemingly turning around a vertical axis passing through the body, (Since with this imagery the dancer tends to turn like a structure floating around horizontally).

3. Vertical. Here the arms perform a lot of up and down motions towards an imaginary vertical axis through the body.

4. Spherical. Means an imagery of spatially unrestricted and spherically irradiating forces. A good balance between vertical and horizontal motions.

Motion Flow   The motion flow dimension attempts to describe our intuitive notion about motions being smooth, fragmented (hectic) or swinging as reflected by the temporal profile of velocity or acceleration. Possible values for this dimension are:

1. Rigid, (hard, resistant, containing elongated pauses or slow decelerations and accelerations sometimes with edgy direction changes)

2. Continuous (smooth, gliding, fluent, relatively calm)

3. Swinging (dynamic, flexible)

4. Fragmented (staccato, discontinuous, breaking directions) Contains a lot of sudden stops and accelerations.

2.5   The emotion space as our narrative structure

Studies investigating artificial emotions and expressivity face the problem of how to define or describe a certain nonverbal expressive state. The *circumplex model of emotions* is widely used in psychology to build a topological model of a continuous *emotion space* [7]. More complex models underlying the semantics of emotion are an issue of current psychological research [4, 8].

We basically adapted the circumplex model to our study with Butoh and arranged the emotions in segments of a circular plane spanned by two orthogonal dimensions, representing *pleasantness* (horizontal, x axis from negative, painful dark to positive, bright,) and *activation* (vertical, y axis from introvert to extrovert) as depicted in Figure 2 (right). Thus, the increasing radius from the center represents the *intensity* of a rather similar emotional quality. Finally, a third dimension was added in order to be able to capture altered narrative levels and different styles of scenery. Roughly spoken, an animation sequence is a path in the emotion space that may or may not jump to different levels, depending on the expressive states traveled through. It should again be noted that we treat the emotion space as a subjectively organized narrative structure and not as an objectively derived circumplex model.

The same is true for our mappings from motion to emotion and to visual effect (animation). So far, the mappings result from individual subjective annotations that can be subject to an artistic process of rearrangement modification of meaning and even replacement, depending on a specific experience or on a new thematic focus.

2.6   Artistic visual design

A crucial task is the design of animation sceneries corresponding to the different segments in the emotion space. Originally, we aimed at using static photographic background with moving abstract symbols, filled with photographic or graphic contents, inspired by a meditative imagery. In our prototype, we modified this idea slightly and are now using Chinese characters instead of abstract symbols. We treat the characters symbolically and as abstract 'actors'. They are chosen and interpreted on a purely esthetic basis and treated as merely moving areas and spots that may vary dynamically in size, contents and brightness, so to create abstract patterns interacting with the photographic fore- and backgrounds.

3   Recognition Methodology

The recognition problem outlined in Section 2.4 is quite different from the typical activity recognition tasks for which acceleration sensors are often used [9, 10].

Thus for example an arm moving into a horizontal position followed by a vertical oscillating motion and by the arm being lowered again might signify that the user has greeted someone through a handshake [11]. Unlike typical activity recognition that can be reduced to a simple trajectory classification task, our butoh dance style classification task can be characterized as follows:

1. The aim is to recognize certain abstract characteristics of the motion rather then certain motion sequences. Thus depending on the way they are conducted two motions with identical trajectories could represent different classes.

2. In general, for every classification dimension, the dancer is likely to maintain a certain

class of dancing for anything between tens of seconds to a few minutes. However this does not mean that every single motion conducted during this period will actually belong to this class. Instead a certain period needs to be classified according to the dominant motion type. This means that the system needs to be able to identify transitions between periods with different dominating motion classes.

3. At times the classification can be ambiguous, in particular where neighboring classes are concerned. As an example while the distinction between e.g. wild and fine is always clear, it is less clear when a motions stops being wild and starts being strong. This means that the exact point of transition between two different periods is not alway exactly defined.

From the above considerations it can be seen that the problem at hand involves two different time scales. The first one concerns individual movements, which are on the order of 1 sec. The second one concerns segments belonging to a single class which last tens of seconds up to a few minutes. This has lead us to the following two step recognition methodology:

Individual Movement Characterization: In a sliding window chosen to fit the first time scale (approx 1 sec) appropriate features are computed from the sensors signals. The features are a physical representation of the three classification dimensions.

Transition Detection: To map the first time scale events into the second time scale classification of dance periods, a Hidden Markov Model (HMM) is defined for every classification dimension. Each model takes features defined for the corresponding dimension in a single sliding window as observables. A state or a group of states is then taken to correspond to a certain class and the Viterbi algorithm is used to determine the current state. Appropriate choice of the model parameters makes sure that a transition between states (or state groups) takes place only when the dominant style has changed and is not unduly influenced by variations of individual motions.

3.1   Individual Movement Classification

The individual movement classification involves three issues: the decision on the placement of sensors, the choice of features to be computed from those sensors and a choice of an appropriate window size.

Sensor Placement Close analysis of a number of dance sequences and interviews with the performer have revealed that the key information about dance style can be found in the arm's motion. Although the performer has argued that leg motions are irrelevant, signals from the upper legs have been also found to be useful for the separation of some classes. This is due to the fact that leg movements are used mostly to compensate for arm-motions allowing the dancer to keep her balance. Thus they are correlated with the style.

As a consequence of the above we have decided to use sensors placed on the wrist, the upper arm and the upper leg. Since, with respect to the dance style, there is no different between right and left arm/leg, sensors were placed only on the right upper arm, wrist and leg.

Feature Selection The features are the physical representation of the motion characteristics used for the dance style classification. As such they could in principle be derived from physical considerations. As an example motion power is obviously related to the energy contained in

the signal. Similarly information about the motion direction should be contained in the ratio of the vertical and horizontal acceleration components. However due to the subjective nature of some classifications and the fact that the acceleration signal without any supporting position-information is only indirectly related to many of the characteristics finding features that provide perfect separation purely from physical considerations is not feasible. For this reason a two step process has been employed for feature selection. First a number of features have been selected based on physical considerations. These are

- Standard statistical evaluation parameters specifically standard deviation (STD), mean, median, variance, maximum and RMS (root mean square) of the acceleration signal.

- A time domain analysis of the number and size of peaks contained in the acceleration signal. The peaks were derived using a standard hill climbing algorithm

- Frequency-domain-analysis based on an exponential fit of the logarithm of the amplitude of the Fourier transform of the signal. Assuming

$$amplitude(frequency) = A \cdot e^{b \cdot freqeuncy} + c$$

the three parameters A, b and c were fitted and used as features.

The above are defined on each acceleration axis: a_x, a_y, a_z, and the norm ($\sqrt{a_x^2 + a_y^2 + a_z^2}$ of each sensor.

In a second step, those features were tested on a number of dance sequences to interactively determine which combinations provide best separation for which classification dimensions. Example of such evaluation for a good set of features on a sequence containing movements of one style only is shown in Figure 5b. As a result the analysis following feature combinations have been found:

Intensity The intensity-classification does not depend on the motion direction which means that most information is contained in the norm. In terms of features mostly intensity and energy related features such as RMS mean and peaks are relevant. The features used are:

- *Wrist:* std, RMS, number of peaks, and mean peak size on the norm, mean peak size on the X- and Y- direction, RMS on the Z-direction

- *Upper Arm:* mean and mean peak size on the norm; mean peak size on the Y- and Z-axis

- *Leg:* mean peak size on the X- and Y-direction

Direction: The direction is mostly determined by the ratio of X-, Y- and Z-components of the individual sensors. The leg sensor is not relevant. The features used are:

- *Wrist:* median, STD on the norm, median on the X-and Y-direction, maximum on the Z-direction

- *Upper Arm:* mean peak size and RMS on the Y direction

Flow: Flow is mostly derived from dynamics related features in particular frequency domain parameters and peaks analysis.

- *Wrist:* number of peaks, mean peak size and Fourier fit parameters for the norm; number of peaks and Fourier fit parameters for the Z-direction

- *Upper Arm:* number of peaks, mean peak size and Fourier fit for the norm

- *Leg:* number of peaks, mean peak size and Fourier fit for the X-direction

Window Size The size of the jumping window needs to correspond to the first time scale which means a length of about 1 sec. Experiments have shown a length of 2 sec corresponding to 20 samples to be most efficient. In addition to reduce the computational complexity a pure sliding window has been replaced by a jumping window which was moved by 50 samples on each time step.

3.2   HMM Based Transition Detection

HMMs are often used for the modeling of time dependent signals and have been quite successful in such applications as speech or motion recognition. In most motion recognition applications individual states are used to model different phases of the motion with the observation probabilities accounting for the possible trajectory variations. Thus mostly sequential models with a separate multi-state model for every class are used. Since in our case there are no distinguishable phases an ergodic model with a single state for every possible class was first defined for each classification dimension. For the performance evaluation the mixed model alternatives were also considered. The difference is apparent in Figure 4.
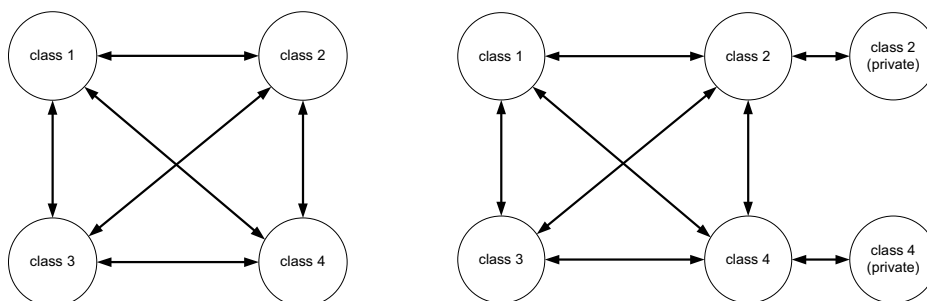

Fig. 4. The HMM models used for classification on all three axes. Left picture shows the purely ergodic model. On the right a mixed models with some private states is shown.

The observables are the features described above with the observation probabilities being trained from labeled dance segments (see section 3.3). Best results were obtained with Gaussian mixture distributions. The ratio of transition probabilities to the probabilities of remaining in a given state reflect the empirical knowledge on the average duration of a dance segment of a certain class. For the classification the features were fed into the model and the Viterbi algorithm was used to determine the most probable state sequence. The state sequence was then translated into classification by labeling every time segment during which the model remained in a given state with the corresponding class.

Model Improvements Later experiments have shown that for certain classes two state representations are more appropriate. Thus an additional 'private' state was assigned to some class representations as shown in Figure 4. The training of the resulting two state model included both the observation probabilities in both states per class and the transitions

between the different states. The transition probabilities to the states corresponding to other classes were the same as in the single state case. The classes that were assigned a two state model are: swinging, rigid and fragmented for the flow, horizontal for the direction and strong and fine for the intensity. In section 3.3 the effect of this improvement are shown.

3.3   Systematic Accuracy Evaluation

To evaluate the classification performance of our method a dance scene several minutes long has been recorded for each possible combination of classes. All together to cover the three dimension, each consisting of four classes, 64 scenes were recorded. All scenes were recorded on video and reviewed by the dancer to verify the labeling. Some scenes, for which the dancer was found not to have consistently held the required dimension constant were re-recorded.

For training and testing each recorded scene was partitioned into 20sec long segments. Of those segments 70% were randomly used for training while the remaining 30% were withheld for testing.

Training For each class of every classification dimension the observation probabilities for the corresponding state need to be trained from the recorded data. In addition to this, the transition probabilities of the model with additional 'private' states need to be determined. To this end the 70 % training segments from all scenes, in which a particular classification dimension was fixed to a given value were used in a standard Baum-Welsh iterative training procedure.

Testing Procedure To test recognition accuracy the 30% segments reserved for this purpose were concatenated in a random sequence. This sequences represented a possible dance performance in which the style would vary in a certain way. It was fed into the three HMMs (one for each classification dimension) and evaluated using the Viterbi algorithm. The resulting state sequence, with each state corresponding to a class, was then compared to the ground truth. The latter was given by the scenes from which the individual segments were taken. Since the results had considerable variations depending on the random choice of sequences they were averaged over 25 sequences. An graphical example of the evaluation is given in figure 5.
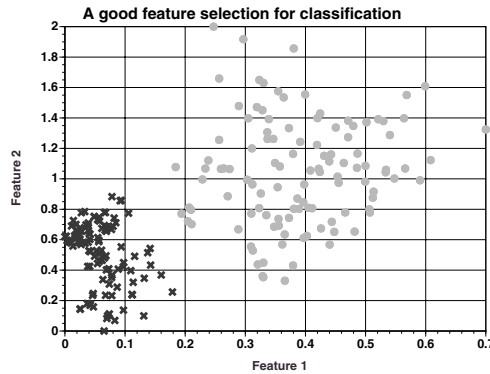

Fig. 5. Example of features separation on an using STD vs. number of peaks for a sequence containing two intensity levels.
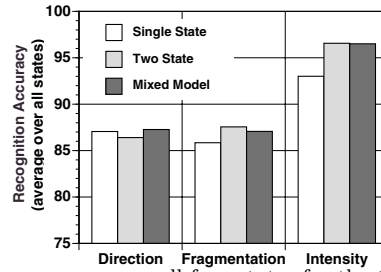
Fig. 6. Average recognition accuracy over all four states for the three models direction, flow and intensity.

3.4   Recognition Results

The overall recognition accuracy is shown in Figure 5. The figure also shows a comparison between the three models used. It is obvious, that the average recognition of the intensity is very model-dependent, whilst direction and flow recognition are quite stable.

Flow Obviously rigid and swinging are often mistaken. As rigid is seen by the dancer to have lots of sub-states, that were neglected, by assembling them to a single state in our simplified state-model, this is probably the main explanation for their confusion. There is also some confusion between swinging and fragmented, which we hope to be able to eliminate by smoothing the data.

Directions Mainly Frontal and Horizontal are confused, which is comprehensible, as both movements take place in a Horizontal plane. The vertical and spherical confusion can be explained by the fact, that vertical contains both upward and downward motion, so that the change between these two directions is interpreted as spherical movement.

Intensity Recognition of intensity is extremely good. There is some confusion between wild and strong however, which could be explained by the fact, that strong movements are at the limit of the dancers capabilities, so that its quite difficult to increase energy to reach a wild state.

Conclusion In summary we see that with few exceptions the recognition rates are over 90%. Taking into account the fact that the classification is at time ambiguous and might differ from artist to artist this can be considered satisfactory. This is even more so since most errors occur between similar categories as discussed above.

The detailed results can be seen in the confusion matrices in Table 7.

flow

state	swing	rigid	frag.	cont.
swing.	93.9987	25.7113	10.4210	1.3207
rigid	3.4914	72.5205	5.4978	0.6310
frag.	0.6556	0.7224	83.8877	0.1665
cont.	1.8543	1.0458	0.1934	97.8817

intensity

state	wild	strong	medium	fine
wild	98.0040	10.9807	0.1872	0
strong	1.9938	89.0118	0.6387	0
medium	0.0021	0.0075	99.0833	0.0647
fine	0	0	0.0908	99.9353

directions

state	front	vert	horiz	spher
front	96.3836	0.6637	17.2152	3.3347
vert	0.4375	94.2264	0.4744	11.0621
horiz	1.7827	0.2700	75.2993	2.4161
spher	1.3962	4.8399	7.0111	83.1871

Fig. 7. The confusion matrices for the three classification axes.

4   System Implementation

From a hardware perspective, our system comprises: a number of sensor nodes plus a central controller, a wearable controller and a stationary recognition and visualization engine. Roughly speaking, these components form a pipeline that generates, coordinates, analyzes, interprets and visualizes a stream of motion data, sampled at a sufficiently high rate (100 Hz in most experiments). We shall explain these components in some detail in the next sections.

4.1   The Sensor System

The sensor nodes used are provided by the Pad'Net (Physical Activity Detection Network) wearable sensor network developed at the ETH and described in detail in [12]. It consists of multiple sensor nodes interconnected in a hierarchical network. The purpose of a sensor node is to provide a physical interface for different sensor types, to read out the corresponding sensor signal, to provide certain computation power for signal preprocessing and to enable communication between the other sensor nodes in the network. Figure 9 shows such a sensor node with its corresponding block diagram. For the experiments three 3D-accelerometers (ADXL202E from Analog Devices) were used. The analog signals from the sensor were low-pass filtered (fcutoff=50Hz), AD-converted with 12Bit resolution using a sampling rate of 100Hz.

Sensor Placement Close analysis of a number of dance sequences and interviews with the performer have revealed that the key information about dance style can be found in the arm's motion. Although the performer has argued that leg motions are irrelevant, signals from the upper legs have been also found to be useful for the separation of some classes. This is due to the fact that leg movements are used mostly to compensate for arm-motions allowing the dancer to keep her balance. Thus they are correlated with the style.
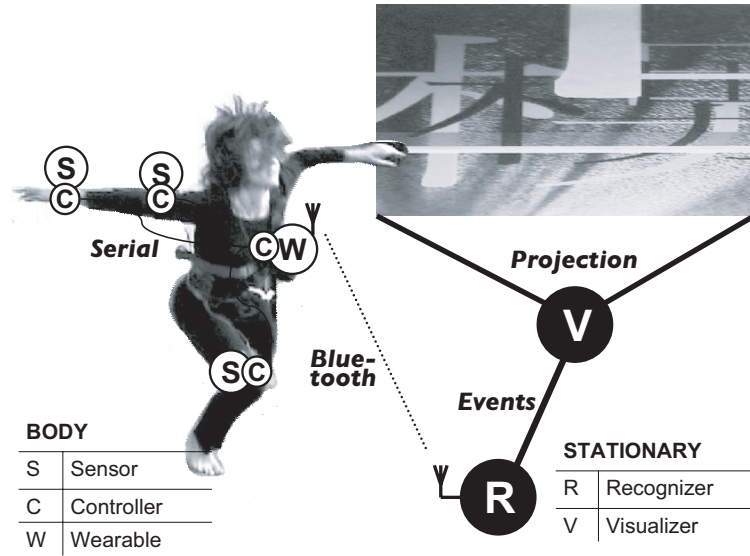
Fig. 8. Overall system architecture with the sensors, the wearable and the visualization machine.
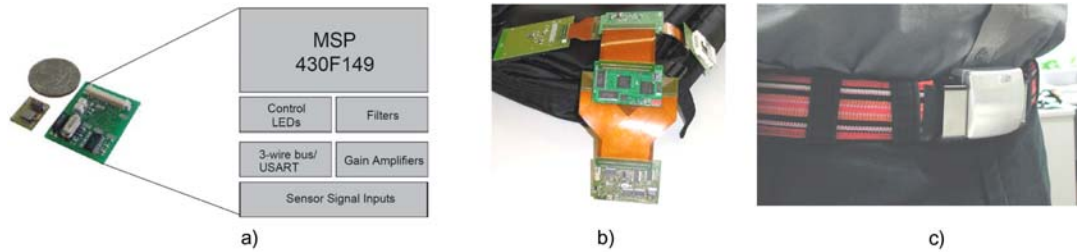

Fig. 9. The components of the wearable subsystem: left a node of the PadNET sensor network, center the boards of the WearARM wearable computer and right the QBIC system packaged in a belt buckle.

As a consequence of the above we have decided to use sensors placed on the wrist, the upper arm and the upper leg. Since, with respect to the dance style, there is no different between right and left arm/leg, sensors were placed only on the right upper arm, wrist and leg.

4.2   The Wearable Controller

The choice of a wearable controller depends on the amount of processing that it needs to do. In the simplest case it just needs to collect the raw signals from the sensors and send them the visualization system. For this case the top level node of the PadNet hierarchy was connected to a Bluetooth module.

In general however, it is desirable to perform parts or all of the pattern classification task on the wearable system. This has two reasons. The first one is technical: transmitting raw data from all sensors to a stationary machine over a wireless network requires more energy then doing the classification locally and transmitting selected events. The second one is

conceptual. Different performers might want to have their personal mapping of motions to classes and visualization events. Thus it makes sense for the recognition to be done by a personal device which is fully controlled by the user.

For the above reasons experiments were conducted using different mobile and wearable devices. In addition to the IPAQ PDA the WearARM (see figure 9) system developed by our group [13] and the MASC wearable used in the 2Wear EU project [14] were tested. In the future our next generation wearable, the QBIC [15] will be used. Alternatively to its function as a real-time data streaming device, the wearable was used by the dancer off-line to record sample motion patterns. Obviously, the resources provided by a device small enough to be worn comfortably by a dancer are scarce compared to typical portable or stationary hardware. For this reason and considering our plans of future research in power awareness, we refrained from using down-scaled standard software and developed a custom runtime kernel instead, with an emphasis on ultimate resource efficiency. The resulting system is a fully managed and modular runtime, programmed uniformly in a high-level language called *Active Oberon*, a descendant of Pascal and Modula-2. On top of the kernel, we implemented a memory file management and up streaming functionality based on the L2CAP layer of the Bluetooth protocol stack.

The Stationary System The two functional components of the stationary animation system are the *recognition/analyzer module* and the *visualization engine*, with an event oriented interface. The visualization system has been designed and implemented for this project from the ground up, and it relies on the same runtime kernel as the wearable controller.

4.3   The Visualization Engine

Corresponding to the typical, silent nature of Butoh dance, we decided in favor of a *visual* approach to feedback, in contrast to the more common audio oriented systems. After extensive discussions, we agreed on a two-level, event based animation policy. Events of both levels are generated by the recognition/analyzing subsystem and sent to the visualization engine via TCP/IP. Basically, first level events are used to select the animation scenery corresponding to the current expressive category, while second level events reflect motion features such as root mean square or number of peaks and control animation parameters within the current scenery, for example the size and color of actor objects, their trajectory and speed. The benefit of the two-level visual system is the option of creating interesting overlays of direct kinematic feedback and less direct symbolic effects. Obviously, the handling of events arriving at a high pace in real time is a demanding task that relies on super-efficient processing. For this reason and for the sake of flexibility, we refrained from using off-the-shelf software and developed a custom animation system instead [16], running on top of a custom operating kernel called AOS [17].

The animation system is characterized by the following highlights:

- Each scenery is a stack of (arbitrarily nested) *views*, each view with its own *contents*, *properties*, *event* specifications and *sub-views*.

- Animation sceneries are specified statically as *scripts* in the form of XML-documents, internalized at loading time and interpreted dynamically at run time.

- Typical contents of sceneries comprise pictures and vector graphics, for example Chinese character glyphs.

It is worth noting that the animation system supports a rich variety of scenarios by providing full flexibility regarding the use of filling patterns (for example, it allows feeding life video into a vector graphic) and the nesting of contents. The system exports a set of built-in event types, among them the scenery-selector *Emotion-Event* (see first XML sample document below) but it also allows users to extend this set by arbitrary event types (see second XML sample document).

```
<Animation scriptversion="1.0">

<SizeX>1024</SizeX><SizeY>768</SizeY>
<Left>0</Left><Top>0</Top>

<Title>3D Emotion Space</Title>
<Description>Scenery selector</Description>
<Author>Martin Gernss</Author>

<IViewCollection>
 <Views>
  <IViewRectangle>
   <Properties>
    <SizeX>1024</SizeX><SizeY>768</SizeY>
    <Color><B>255</B></Color>
   </Properties>
  </IViewRectangle>
  <IView3DSpace>
   <Properties>
    <Speed>0</Speed>
   </Properties>
   <Events>
    <EmotionEvent>
     <Coords>
      <X><EventData1/></X>
      <Y><EventData2/></Y>
      <Z><EventData3/></Z>
     </Coords>
    </EmotionEvent>
   </Events>
   <Views>
    <ExternView fname="World.XML" X=1 Y=1 Z=0/>
    <ExternView fname="Trans.XML" X=2 Y=1 Z=0/>
    ...
    <ExternView fname="Dark.XML" X=0 Y=0 Z=1/>
   </Views>
  </IView3DSpace>
 </Views>
</IViewCollection>
</Animation>                     (a)
```

```
<IViewCollection>
 <Properties>
  <TranslateX>3</TranslateX><TranslateY>0</TranslateY>
  <Wrapping/>
 </Properties>
 <Views>
  <IViewChineseGlyphEx>
   <Properties>
    <PosY>250</PosY><SizeX>300</SizeX><SizeY>150</SizeY>
    <Glyph><No>15322</No></Glyph>
   </Properties>
   <Events>
    <ContentEvent subtype="0">
     <Glyph><No><EventData1/></No></Glyph>
    </ContentEvent>
    <ContentEvent subtype="1">
     <SizeX><EventData1/></SizeX><SizeY><EventData2/></SizeY>
    </ContentEvent>
   </Events>
   <Views>
    <IViewCollection>
     <Properties>
      <TranslateX>3</TranslateX><TranslateY>2</TranslateY>
      <Wrapping/>
     </Properties>
     <Views>
      <IViewStillImage>
       <Properties>
        <SizeX>800</SizeX><SizeY>600</SizeY>
        <Image>Pommes.png</Image>
       </Properties>
      </IViewStillImage>
     </Views>
    </IViewCollection>
   </Views>
  </IViewChineseGlyphEx>
  ...
  </IViewChineseGlyphEx>
 </Views>
</IViewCollection>              (b)
```

Fig. 10. Listing of the XML Script controlling the visualization. The code part (a) selects a scenery and the code part (b) describes a scenery and specifies the real time effects.

The sample script in Figure 10a acts as a selector of sceneries. Essentially, it consists of a 3D space view corresponding to the "emotion space" of the Butoh dancer. Its *Events* section specifies how each emotion event selects a view in the emotion space corresponding to the three coordinates *X*, *Y* and *Z* as delivered by the arguments *EventData1*, *EventData2* and *EventData3* respectively.

The script in Figure 10b depicts an example of an XML animation script, a collection consisting of two *Chinese glyph* views, moving along a horizontal line at constant speed is specified, the first of which, in turn, consists of another collection of just one view, this time

a still image used for filling the glyph. The event section of the Chinese glyph advises this view to translate the user defined *ContentEvent* subtypes 0 and 1 into a change of the glyph (code) and its size respectively.

We should add that, thanks to the clearly defined event interface, the custom animation engine can easily be replaced by any other. In particular, we recently experimented with Max/MSP as an alternative visualization back-end.

5   Presentation Experience

Performances were shown in autumn 2003 in the performance center for multi-medial arts "plug.in" in Basel, Switzerland, in the "Disappearing Computer Jamboree" 2003 at the institute of media design in Ivrea, Italy and during the 150 years celebrations of the ETH in Zurich. The performances had a demonstration character. The dancer felt that the interaction with the system added dramaturgic tension to the show and the simultaneous interaction with the image and public was challenging. The reception by the public showed that people were perceiving some correlation of movement and image. They were looking for the nature of correspondences between dance and the image and were expecting possible interpretations of the visual story . Our experience stresses the point that as an addition to the dance, the visual language should be relatively simple esthetically well-readable and intuitively interpretable so that the public can grasp some essence of the interaction in an immediate way. Some parallels of visual movement and the movement of the dance is interesting. Our future visual representations will explore several choreographic possibilities. During the performances and in our tests, the wearable computer proved to be a non-disturbing and easily portable device that the dancer could relatively easily learn to control.

In visual and performance art, meaning, expression and re-interpretation are a priori ambiguous and in a constant process of subtle change. While our system could detect and feed back either 'emotionally consequent' or 'irrational' human interpretations at the level of an esthetic approximation, it will, as any machine will, always deviate from its precise nuance. In contrast to verbal language, this actually poses a conceptual problem in a refined abstract esthetic perception, where each nuance per se is absolutely significant. This is an issue that has to be addressed and accounted for in the artistic strategy, either by consciously taking advantage of the discrepancies, or by choosing a stable esthetic and thematic focus or by perpetually catching deviating mental and visual episodes.

Following a very different approach, the direct streaming of a range of features from the originally expressive criteria to the animation system might lead to results that are pattern-wise related without being obvious and that can inspire without giving an interpretation. This position fits well the concept of Butoh. An abstract pattern or texture-like visual language might well support the constructive character of this approach.

With this experience we can consciously make use of structural invention, of restricted esthetic tendencies and of randomness in the future.

6   Related Work

The interdisciplinary nature of our work means that it is related to a number of different research areas and that the space constraints of a conference paper will prevent us from presenting even a nearly complete survey. As particularly relevant we consider the general

Fig. 11. Visually enhanced Butoh dance performance in the multi-media performance space "plug-in" in Basel, Switzerland. A similar performance and demonstration of the technology was given at the Institute of Media Design in Ivrea, Italy.

Fig. 12. Performance given at the 150 year celebrations of ETH Zurich

motion analysis with wearable sensors, emotion analysis using wearable sensors, wearable arts related applications, and other dance, in particular Butoh analysis and visualization attempts.

Many authors in the field refer loosely to the *Laban Theory of Movement* [18]. Laban's theory of effort represents the quality (effort) of a motion in an abstract 4-dimensional space.

Our three dimensional system loosely matches Labans concept, yet the whole concept and definition of space is different. What Laban corresponds to strength we correspond to intensity. Laban's aspect of time and flow and trajectory-space is what we have merged into a concept of flow. Our concept of direction is similar to Laban's dynamosphere. Laban's expressive dance concept and the modern dance are built upon explicit body forms and directions, whereas Butoh uses referential forms and force-fields hat constrain free movement. Accordingly to this, the body centered movement area or 'kinesphere' in Butoh is not oriented relative to the body, but relative to outer space.

So far motion analysis using acceleration sensors has been mostly applied to two areas: activity recognition (e.g. [9, 11]) and medically motivated biomechanical analysis (e.g. [19, 20]) To our knowledge so far acceleration sensors have not been used for emotion analysis. Instead, the author of [21] mostly uses physiological parameters such as galvanic skin response pulse. Here the majority of work has been done in the context of so called affective computing [22]. In the multimedia area the work on emotion analysis aims at video classification [23] or at the extraction of emotions from sound and gesture [24, 25]. An interesting dance visualization system emphasizing the localization in space is the body brush which is based on infrared illumination ([26]). A more general framework for the vision based recognition of gestures and enhancement of artistic expression is described in [27]. Using acceleration sensor technology, the TGarden Project [28] explores behavior in artificially constraining costumes and audio-visual spaces, following a less semantic approach, by direct translation from low level gestural parameters into low level parameters of the visual language (video effects). In the artistically sensitive and innovative artwork of Levin and Lieberman generative graphics are controlled by voice sound patterns, optimally merging constructive rules with intuitive 'synaesthetic' perception. Of particular interest to our work is an effort to extract emotion based information from dance and movement sequences using video signal described in [29, 3, 30, 31, 4]. These studies refer to a limited amount of psychological emotion categories related to a 'naive' body language (happiness, anger, fear etc) while in our approach we aim at differentiating and recognizing a larger amount of more abstract esthetic feeling states and connotations typically created by dancers and perceivable by public. In a recent and ongoing study investigating the objective psychological nature of such movement perceptions, the generality of some Butoh expressive categories was shown in terms of recognition and sorting by other subjects[4].

7   Conclusion and Future Work

We have shown that visual animations based on the emotion classification can be used in a artistic performance in such a way that is perceived as an enrichment by the artist and the audience.

At the conceptual level our artistic project provides technology applicable to a wider scope of applications. We are currently studying modifications to our classification and recognition scheme that deal with casual everyday motions rather instead of dedicated expressive motions in a dance.

There are limitations of our system at several levels of abstraction. The proposed motion to emotion mapping is ambiguous to some extent by its design. The reduction of motions and emotions to our set of categories, their mapping as a linear representation and the criteria of separation could be questioned asking for more specific psychological research to justify our

assumptions. Still our categories seem to be pronounced enough to perform a quality test of the recognition and most of the classes are consciously created and reproducibly recognized. We are currently investigating an enhanced real-time analysis in more quantitative terms, in respect to the tradeoff between recognition quality and the delay of the real-time recognition. To ensure the effect of the visual enhancement, the recognitions delays need to be either minimized or otherwise conceptually integrated.

Finally we achieved to develop a physically robust wearable system in a novel application. The intensity of some of the motion patterns put a hard test to the prototype equipment. Several phases of debugging and re-engineering were required until the hardware, the system software and the wireless communication equipment were robust enough for a live performance. We experienced that the new QBIC device embedded in a belt buckle and some newer wireless sensors have further improved performer comfort and allow more complex data processing right at the source of data.

Acknowledgments

References

1. M. Holborn and E. Hoffman, *Butoh: Dance of the Dark Soul* Aperture Publishers, New York (1987).
2. K. Nanako and T. Hijikata, *The Words of Butoh*, The Drama Review **44**, pp12-28 (2000).
3. A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, *Multimodal Analysis of Expessive Gesture in Music and Dance Performances*, Gesture-Based Communication in HCI, LNCS **2915**, (2004).
4. L. MacFarlane, I. Kulka, and F.E. Pollick, *The Representation of Affect Revealed by Butoh Dance*, Psychologia **47**, (2004).
5. D. Söke and L. Martina (Eds), *Dance and Technology*, Alexander Verlag, Berlin (2002).
6. R. Wechsler and R. Weiss, title = *EyeCon Palindrome*, http://eyecon.palindrome.de/, (2001).
7. R. Plutchik and H. Conte, *Circumplex Models of Personality and Emotions*, APA Books, Washington (1996).
8. PHYSTA, *Review of Existing Techniques for Human Emotion Understanding and Applications in Human-Computer Interaction*, PHYSTA Project Report Summary, http://www.image.ece.ntua.gr/physta/, (1998-2001).
9. C. Randell and H. Muller, *Context awareness by analysing accelerometer data*, 4th Intl. Symp. on Wearable Computers, Digest of Papers, pp175-176 (2000).
10. J. Mantyjarvi, J. Himberg and T. Seppanen, *Recognizing human motion with multiple acceleration sensors*, 2001 IEEE International Conference on Systems, Man and Cybernetics **3494**, pp747–752 (2001).
11. N. Kern, B. Schiele, H. Junker, P. Lukowicz, and G. Tröster, *Wearable sensing to annotate meeting recordings*, Personal Ubiquitous Comput. **7**, pp263–274 (2003).
12. H. Junker, P. Lukowicz, and G. Tröster, *PadNET: Wearable Physical Activity Detection Network*, Proceedings of the 7th International Symposium on Wearable Computers, New York, USA, pp244-245 (2003).

13. P. Lukowicz, U. Anliker, G. Tröster, S. Schwartz, and R. W. DeVaul, *The weararm modular, low-power computing core*, IEEE Micro, **21** (2001).

14. J. Gutknecht et al., *2wear - project final report*, The Disappearing Computer DC, http://2wear.ics.forth.gr (2004).

15. P. Lukowicz and S. Ossevoort and M. Lauffer and F. Macaluso, and G. Tröster, *Overview of the QBIC Wearable Computing Platform*, Digest of Papers. International Forum on Applied Wearable Computing (2004).

16. M. Gernss, *Animations - An Animation Software System for Bluebottle AOS*, Term Project, CS Departement, ETH Swiss Institute of Technology, Zuerich, http://bluebottle.ethz.ch (2003).

17. P. Muller, *A Multiprocessor Kernel for Active Object-Based Systems*, Proceedings of JMLC 2000, pp263–277 (2000).

18. *R. Laban and F.C. Lawrence, Effort*, publisher= Macdonald and Evans, London (1947).

19. T. Tamura, Y. Abe, M. Sekine, T. Fujimoto, Y. Higashi, and M. Sekimoto, *Evaluation of gait parameters by the knee accelerations*, Proc. of the 1st Joint BMES/EMBS Conf., p828 (1999).

20. M. Sekine, T. Tamura, T. Fujimoto, and Y. Fukui, *Classification of walking pattern using acceleration waveform in elderly people*, Engineering in Medicine and Biology Society, **2** pp1356–1359 (2000).

21. R. W. Picard, E. Vyzas, and J. Healey, *Toward Machine Emotional Intelligence: Analysis of Affective Physiological State*, IEEE Trans. Pattern Anal. and Mach. Intell., **23**, pp 1175–1191 (2001).

22. R. W. Picard, *Affective computing: challenges*, Int. J. Hum.-Comput. Stud., **59**, pp55-64 (2003).

23. H.-B. Kang, *Affective content detection using HMMs*, Proceedings of the eleventh ACM international conference on Multimedia, pp259–262 (2003).

24. S. Mocrieff, C. Dorai, and S. Venkatesh, *Affective Computing in film Through Sound Energy Dynamics*, Proc. ACM MM'01, pp525–527 (2001).

25. A. Camurri, R. Trocca, and G. Volpe, *Interactive Systems Design: A KANSEI-based Approach*, Proc. of NIME 2003, (2002).

26. H. H. S. Ip, H. Young, and A. C. C. Tang, *Body-Brush: A Body-driven Interface for Visual Aesthetics*, Multimedia'02, pp664–665 (2002).

27. F. Sparacino, G. Davenport, and A. Pentland, *Media in Performance: Interactive spaces for dance, theater, circus and museum exhibits*, IBM Systems Journal **39** p479 (2000).

28. X.W. Sha, Y. Serita, J. Fantauzza, S. Dow, G. Iachello, V. Fiano, J. Berzowska, Y. Caravia, D. Nain, W. Reitberger, and J. Fistre, *Demonstrations of expressive softwear and ambient media*, Proc.of UbiComp03 .

29. R. Natatsu, M. Tadenuma, and T. Maekawa, *Computer technologies that support Kansei expression using the body*, Proc. of the 9th ACM Intl. Conf. on Multimedia, pp358–364 (2001).

30. MIDAS, *MIC Interactive Dance System*, MIC.ATR Media Integration & Communications Research Laboratories, http://www.mic.atr.co.jp/organization/dept3/pdf/midas.pdf

31. F. E. Pollick, *The Features People Use to Recognize Human Movement Style*, Proceedings of the 5th Workshop on Guesture and Sign Language base HCI, (2003).