Chapter 1
Analysis of the qualities of human movement in individual action

Paolo Alborno[1], Ksenia Kolykhalova[1], Emma Frid[2], Damiano Malafronte[1], and Lisanne Huis in 't Veld[3]

Abstract

The project was organized as a preliminary study for Use Case #1 of the Horizon 2020 Research Project "Dance in the Dark" (H2020 ICT Project n.645553 - http://dance.dibris.unige.it).

The main objective of the DANCE project is to study and develop novel techniques and algorithms for the automated measuring of non-verbal bodily expression and the emotional qualities conveyed by human movement, in order to enable the perception of nonverbal artistic whole-body experiences to visual impaired people. In the framework of the eNTERFACE '15 Workshop we investigated methods for analyzing human movements in terms of expressive qualities. When analyzing an individual action we were mainly concentrating on the quality of motion and on elements suggesting different emotions. We developed a system to automatically extract several movement features and transfer them to the auditory domain through interactive sonification.We performed an experiment with 26 participants and collected a dataset made of video and audio recordings plus accelerometer data. Finally, we performed a perception study through questionnaires, in order to evaluate and validate the system.

As real time application of our system we developed a game named "Move in the Dark", which has been presented in the Mundaneum Museum of Mons, Belgium and Festival della Scienza, Genova, Italy (27 November 2015).

University of Genova, Italy
e-mail: {paolo.alborno,ksenia.kolykhalova,damiano.malafronte}dibris.unige.it
KTH Royal Institute of Technology, Sweden
e-mail: {emmafrid}@kth.se
Maastricht University, Netherlands
e-mail: {e.huisintveld}@maastrichtuniversity.nl

Key words: Analysis of movements, Sonification, DANCE H2020, Movement qualities

1.1 Introduction

The proposed work is presented as part of the EU Project DANCE - Dance in the Dark (H2020 ICT Project n.645553 - http://dance.dibris.unige.it). The DANCE project aims to innovate state of art on recognition of emotions expressed by the movements of the human body, given human movement as social language. We propose to develop technologies of sensory substitution enabling the perception of dance performance, and in general of bodily movement, through interactive sonifications.

In previous studies it has been experimentally shown that movement qualities may communicate social relations and intentions, such as: emotional states [**?**], affiliation [**?**], cultural background [**?**], dominance [**?**], agreement [**?**], group cohesion[**?**], empathy[**?**].

Several approaches dedicated to automatically detecting emotional states from body movement were recently developed. In [**?**] the authors proposed to detect emotional states from low-level features of hand movements, such as maximum distance of the hand from the body, its average speed and acceleration. In [**?**], expressive qualities of the movement such as amount of motion, contraction and directness indexes as well as velocity and acceleration were used to classify four emotions (anger, sadness, joy, and pleasure) with dynamic time warping classifiers. Also Gross and colleagues in [**?**] analysed expressive quality (Effort-Shape model) and kinematics (range of motion and the velocity) of emotional hand gestures such as knocking: six different emotions were investigated.

Despite the increased interest in modelling emotional behaviour, analysis of expressive qualities of human full-body movement is rather under-investigated with respect, e.g., to facial expressions or prosody in speech.

Many previous studies, indeed, focus on emotion recognition based on audio and facial expression data. Therefore our project is in the direction of further investigation of expressive movement qualities of full body movements. The main focus was on single performers and the quality of the respective motion, rather than on the type and direction of the movements.

DANCE Use Case #1 served as a foundation for experiments' set up on sensory substitution by means of associating appropriate interactive sonifications to the extracted movement features. In our experiment, we used sonification to transform parameters of human movement qualities into sound in order to investigate the perception of users.

The remaining part of this paper is structured as follows. The next section describes the particular case of Use Case #1 of DANCE Project and section 3 presents movements features: their definition, computation and extraction.

After that, the following section presents the system architecture. Section 5 presents the perceptive studies for testing and validation of work of the system. Section 6 describes the interactive demonstration - game "Move in the Dark", implemented in the framework of the workshop.

1.2 Experimental Scenario

A user who is temporarily deprived of vision is learning to recognize (her own as well as others') movement qualities only by using the auditory channel. Sensory deprivation is a mean to amplify user's capabilities and sensibility in recognizing individual movement qualities by means of the auditory modality. The structure of the described use case can be summarized in two phases:

Phase 1

The user (normal sighted and blindfolded) learns how the qualities of gesture are translated into sound, and comprehends how to exploit such an inner vision, induced by sound, of the movement features (e.g., energy, fluidity, rigidity, impulsivity and so on).

Phase 2

The user, who in Phase 1 has been familiarized with the mechanisms of interactive sonification of her own movements, is able to recognize movement qualities of another person just by listening the interactive sonification.

1.3 Movement Features

Starting from the literature from experimental psychology and HCI, e.g., [**?**], [**?**], [**?**], [**?**], humanistic theories and the arts, e.g. [**?**], [**?**] and from meetings with dance experts, a collection of expressive movement qualities was defined to be considered in the DANCE project, including smooth, light/heavy, fluid, impulsive, sudden/sustained, symmetric, contracted/expanded, energetic and synchronised.

In our experiment three main movement features were considered and extracted based on the data from accelerometers: Energy, Fluidity and Impulsivity.

To extract information about the user motion, we used four Nexus S commercial Android smartphones. Each Nexus S are equipped with InvenSense

MPU-6050 three-axis gyroscope and accelerometer. The accelerometer produces data in units of acceleration (distance over time2), and the gyroscope produces data in units of rotational velocity (rotation distance over time). The scale of sensibility of each of the MPU-6050 is set to +/- 2g. The smartphones were tied to the arms and legs of the user for a total of four accelerometers.

1.3.1 Energy

The quantity of energy spent by the user is computed from the total amount of displacement of the tracked joints. We define the Weighted Energy Index - WEI(f) as the weighted sum of the kinetic energy. Given the three dimensional velocities of the $i-th$ joint at time frame f, the total velocity of the joint at time frame f is defined by:

$$v_i(f) = \sqrt{(v_1^2(f) + v_2^2(f) + v_3^2(f)} \tag{1.1}$$

The energy produced by the single joint is computed by:

$$J_i(f) = 1/2 \cdot m \cdot v_i^2 \tag{1.2}$$

With $N = number$ of accelerometers, i.e. 4, and $m = 1$ for each joint.
The maximum value between the joints energies at each frame is selected:

$$M(f) = max[J_1, ..., J_N] \tag{1.3}$$

$$k = argmax[J_1, ..., J_N] \tag{1.4}$$

And the total amount of energy is:

$$WEI(f) = M(f) + \sum_{i \neq k} C_i(f), \tag{1.5}$$

Where C is the single contribute, $i = 1...N$, $i \neq k$ and T is a threshold value in [0,1] used to limit the WEI index.

$$C_i(f) = \frac{J_i(f)}{M(f)} \cdot \frac{(1-T)}{(N-1)} \tag{1.6}$$

1.3.2 Fluidity

We define a fluid movement as an uninterrupted movement characterized by low jerkiness and without abrupt variations in speed. Fluid movements are sometimes associated with slow and sluggish movements, in contrast with large and energetic body movement.

To measure fluidity we defined a Fluidity Index FI based on inverse of the jerkiness of the movement produced by the joints.

$$FI = \frac{1}{(\dot{a_1}^2 + \dot{a_2}^2 + \dot{a_3}^2)} \tag{1.7}$$

1.3.3 Impulsivity

Impulsivity can be defined as a temporal perturbation of a regime motion [**?**]. Impulsivity lacks of premeditation i.e. it is performed without a significant preparation phase, quickly and with a high energy change [**?**], [**?**].

We defined an impulsivity index II based on the inverse of the fluidity and abrupt changes of energy :

$$II = \frac{1}{FI} \cdot \frac{WEI(f)}{WEI(f-m)} \tag{1.8}$$

where m is fixed and arbitrary. The ratio between the values of the WEI index at frame f and frame f-m is used to evaluate whether the energy evolution contains quick changes or peaks, and it is proportional the impulsivity of the movement.

1.4 System Architecture

EyesWeb.

The EyesWeb open platform (`http://www.infomus.org/eyesweb_ita.php`), designed at Casa Paganini - InfoMus research centre of University of Genoa, is a development and prototyping software environment for both research purposes and interactive applications. It supports multimodal analysis, real-time processing of non verbal expressive gestures, research on synchronization, coordination, and entrainment in dance and music performance. We used the EyesWeb platform as one of the core modules for developing the interactive demos and applications planned in the project. The system has been implemented in terms of a collection of software modules for the EyesWeb XMI platform. Extraction of features for analysis of affective behavior has been implemented as software modules part of the EyesWeb DANCE Library.

Max/MSP.

The software Max/MSP was used for interactive sonification. EyesWeb and Max/MSP were connected via the Open Sound Control (OSC) protocol [**?**].

Sonification was used to enable participants to understand specific qualities of their movement in real-time simply by listening to the synthesized sound generated by their own movement. We developed three synthesis models based on a granular synthesis: one for energy, one for fluidity and one for impulsivity. The granular synthesis technique was opted since this method allows for creation of a wide variety of complex sound textures. The granular synthesis method can be seen as a form of additive synthesis, in which the sound results from the additive combination of many smaller sonic grains [**?**]. A *grain* is a short sonic event with an amplitude envelope with the shape of a quasi-Gaussian bell curve, typically of the length 1-50 milliseconds [**?**]. We developed a flexible granular synthesis engine in which the sound buffer could be easily altered in order to produce many different types of sound textures.

The first two sound models (Energy, Fluidity) used the same sound buffer; a short recording of a section of string instruments. The third sound model (Impulsivity) used a buffer of a sound reminding of an explosion. Motion features were sent from the motion analysis in EyesWeb and mapped to the following parameters in the granular synthesis engine:

1. Energy
   - velocity mapped to maximum grain amplitude and overall amplitude of the outputted continuous stream of sound
   - grains of randomized length between 1 and 20 ms
   - inter-grain delay set to 12 ms; producing a continuous stream sound
2. Fluidity
   - velocity mapped to inter-grain time and amplitude as for Energy
   - fluidity mapped to grain envelope slope, inter-grain delay (randomized to 0-200 ms),
   - grains of randomized length between 10 and 190 ms.
3. Impulsivity:
   - mapped to triggering of single grains (randomized length between 1-50 ms) of sound buffer with sharp envelope.

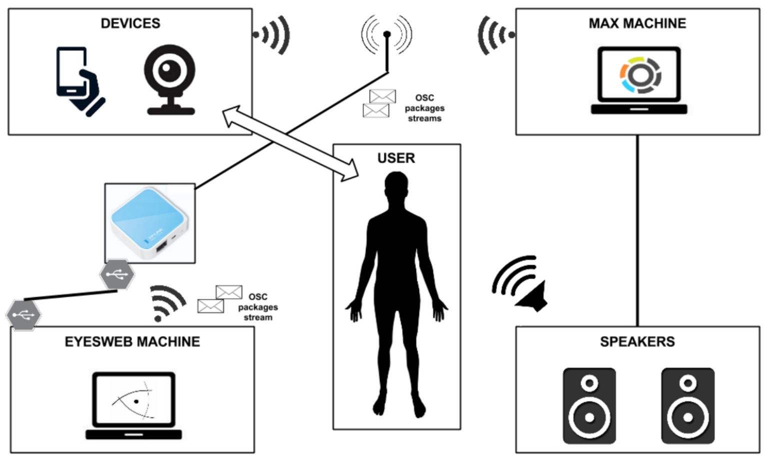

Fig. 1.1 System Architecture

1.5 Evaluation

In the first part of our experiment, participants explored how our system mapped their movements into sound and got acquainted with the sonifications. Then participants were asked to recognize the movement qualities only by listening to pre-recorded audio-files that were created using the same sonifications. In this way, participants were in "blind" conditions and we investigated the user's capabilities and sensibilities in recognizing individual movement qualities only by means of the auditory modality.

1.5.1 Participants

Two groups of participants were asked to provide a rating of the system. The first group of participants included 16 people (13 M, 3 F; different nationality; mean age 29 years, STD=5.67). The second group of participants consisted of 9 people (6 M, 3 F; mean age 26 years with STD=4.06).

1.5.2 Materials and method

Experiment consists of two phases. The first group participated in both phase 1 ("Learning phase") and 2 ("Validation of the system"). The second group participated only in phase 2. The perceptual evaluation was performed in order to compare the ability of trained and not trained users on understanding which kind of movement feature represented by a particular sonification and if the sound is intuitive enough to be perceived without prior knowledge. The two phases of the experiment can be described as follows:

Phase 1

“Learning phase”: participants were asked to move freely. Four Android smartphones were attached with armbands to their forearms and calfs. Participants had one minute for exploration of each movement feature (energy, fluidity/rigidity, impulsivity), i.e. they can freely move and listen to the sound produced by the system. After trial of the system, participants provided ratings, describing their personal opinion and experience they had intearcting with our system. The questionnaire using a six-points Likert scale is shown in Table 1.1.

The second group of evaluators were introduced as a control group; they did not participate in the first phase of experiment, i.e. the learning phase. Instead, they immediately proceeded to the questionnaire for evaluation of the short audio segments. We investigated the ability of not trained participants, to recognize which movement quality mapped into sound without prior knowledge.

Phase 2

“Validation of the system”: we constructed a questionnaire with short audio recordings that had been generated from actual movements, using our system. Both groups of participants (Trained and Control) provided answers to the questionnaire, trying to understand which movement feature sonification they hear, no video feedback was provided.

In the questionnaire with short audio segments we used two types of questions with respect to two different sonification models that we developed and used in experiment (fluidity and impulsivity). The sonification model of fluidity is based on the energy, therefore first type of questions asked our participants to chose two out of four conditions: ”low energy or high energy” and ”fluid or rigid” (where rigid is a way to express ”not fluid” in this experiment). Second type of questions concerned about impulsivity sonifications allowed to chose between ”impulsive or not impulsive”. A diagram of the experimental setup is shown in Figure 1.2. Apart from the interview data from both groups, we have collected video, audio and accelerometers data of the first group.

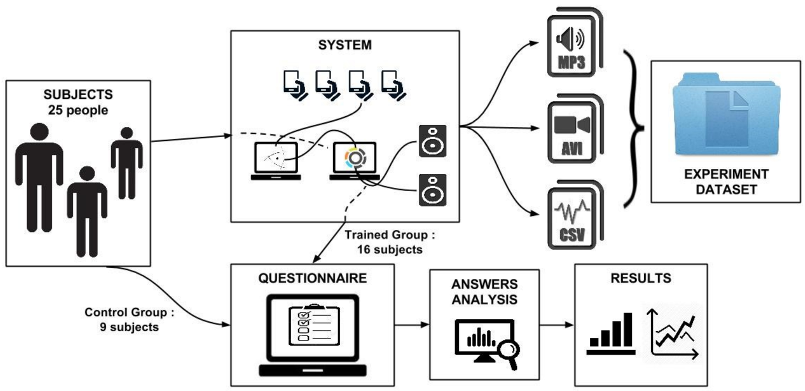

Fig. 1.2 Diagram of Experimental setup

1.5.3 Results

According to the collected feedback from the first group of participants, we were able to draw the following conclusions: Sound was responsive to the movement (mean 4 with STD 0.73) and easy to manipulate (mean 3.81 with STD 0.655). Learning of the system was interesting (mean 4.19 with STD 0.655) and intuitive (mean 4.06 with STD 0.929). Participants stated that it was easy to manipulate the sound corresponding to the following movement features: fluidity (mean 3.63 with STD 0.957), energy (mean 4.31 with STD 0.793) and that they could distinguish between rigidity and impulsivity sonifications (mean 3.63 with STD 1.147). According to respondents, they were at their own pace wearing the system equipment meaning that they were comfortable enough (mean 3.75 with STD 1).

In Table 1.1 detailed information of the feedback from participants is provided.

Next, we checked whether the prior knowledge of the system inuenced the recognition of the audio sonifications.

First, we calculated the proportion of the correct answers, when participants could recognize the movement quality from short audio segments. Table 1.2 displays the percentage of correct recognitions by the two groups (Trained and Control) of participants for each movement feature.

An ANOVA was conducted independently for each analysed movement feature. The results of the analysis between groups are the following:

Low Energy (F=0.566, p=0.459); High Energy (F=0.566, p=0.459), which showed no significant difference between trained and control groups.

For analysis of the movement feature Fluidity/Rigidity, between the groups, a four one-way ANOVAS has been conducted, comparing all four conditions: rigid low energy (F= 10.015, p=0.004), rigid high energy (F=0.007, p=0.933), fluid low energy (F=1.659, p=0.210), fluid high energy (F=0.131,

Table 1.1 Feedback about the system. (6-points Likert scale from 0 - not at all to 5 - very much)

Question	**Min**	**Max**	**Mean**	**STD**
How responsive was the sound to the actions you performed?	3	5	4.00	0.730
How well could you manipulate the sounds?	3	5	3.81	0.655
How compelling was your sense of moving around in the system?	3	5	3.81	0.655
Did you at your own pace wearing the system equipment?	2	5	3.75	1.000
I thought learning how to use the system was interesting	3	5	4.19	0.655
I thought learning how to use the system was intuitive	2	5	4.06	0.929
I thought learning how to use the system was confusing	1	4	2.00	1.155
I thought the sounds were pleasing	1	5	3.13	1.258
It was easy for me to manipulate the sounds related to fluidity	2	5	3.63	0.957
I had to put too much strength into creating an impulse	1	4	2.44	1.153
It was easy for me to differentiate rigid and impulsive movements	2	5	3.63	1.147
It was easy for me to manipulate the sounds related to energy	3	5	4.31	0.793
I could notice the difference between using one, two or more limbs during the energy experiment	1	5	2.88	1.310

p=0.720). According to results we found that trained group of participants are significantly better in recognizing rigid low energy movements with respect to Control group (that had no prior knowledge about system and did not hear any sonifications prior interview).

Analysis of the ability to recognize the movement quality Impulsivity did not show a significant difference between the groups: No Impulse (F=1.341, p=0.259) and Impulse (F=0.001, p=0.978).

Therefore, we were able to conclude that the chosen sound models were intuitive for both groups, with and without prior knowledge.

1.6 Demo Game "Move in the Dark"

As application of our experimental system, we developed interactive multiuser and multimodal game named "Move in the Dark".

Table 1.2 Proportion of correct answers in recognition of the movement qualities from auditory stimuli (first group of participants - "Trained", second group of participant - "Control")

Group	**Feature**	**Varirity**	**Mean**	**STD. Err**	**95% Confidence Interval**	
					Lower Bound	Upper Bound
Trained	Rigid	Low	0.547	0.051	0.441	0.652
		High	0.797	0.061	0.670	0.923
	Fluid	Low	0.828	0.049	0.726	0.930
		High	0.375	0.069	0.232	0.518
	Impulse	No	0.937	0.0360	0.860	1.0144
		Yes	0.890	0.0393	0.806	0.974
	Energy	Low	0.750	0.00	0.750	0.750
		High	0.9375	0.036	0.860	1.014
Control	Rigid	Low	0.278	0.068	0.137	0.419
		High	0.806	0.081	0.637	0.974
	Fluid	Low	0.722	0.066	0.586	0.858
		High	0.333	0.092	0.143	0.524
	Energy	Low	0.722	0.050	0.606	0.837
		High	0.9167	0.041	0.8206	1.0128
	Impulse	No	0.861	0.0605	0.721	1.0007
		Yes	0.888	0.0291	0.829	0.950

Two users compete with each other, blindfolded. Listening to the sound around them, they have to start moving accordingly to what they hear. The better the user is able to translate her body gestures, the more points the user is gaining in order to be the winner. The type of sound is randomly chosen between sonifications of fluid and impulsive movement. Each sound is synthetically generated by the system and the volume is controlled by a random value representing the energy. The diagram in Figure 1.3 describes the game step by step.

The Demo game has been designed with a use of the EyesWeb Platform, Max/MSP Software, Graphical Editor, EyesWeb Designer and accelerometers of 4 Android Smartphones.

A Move in the Dark session consists of three parts:

Prepare.

At the beginning of the game two users can specify their names and how long they want to play. Before starting the game, the user should be blindfolded.

Hear and move.

The system will now start to produce sonifications related to movement qualities as fluid, rigid and impulsive. The user interprets the sounds and translate

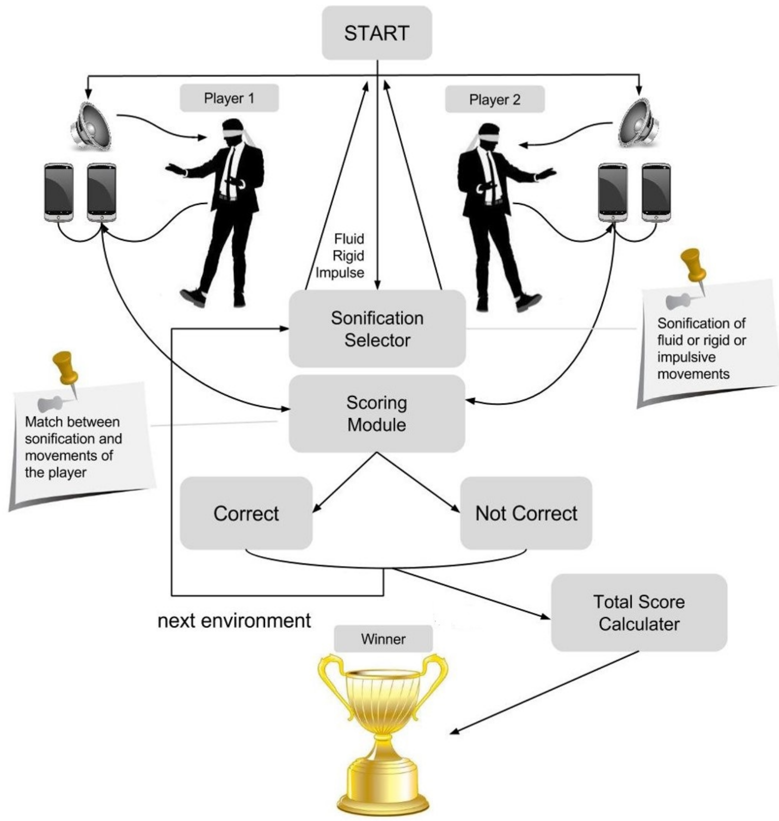

Fig. 1.3 Game "Move in the Dark". The diagram of the game structure

them into movements. Meanwhile the players are moving, graphs related to each considered quality are displayed into the game interface for the audience.

Check your scores.

Depending on how much each player has been able to correctly translate the sonification into the related feature, a winner will be determined by the system. At the end of the game, the acquired scores are displayed.

The "Move in the Dark" game has been presented in the Mundaneum Museum of Mons, Belgium and Festival della Scienza, Genova, Italy (27 November 2015).

1.7 Conclusions and future work

The project at the eNTERFACE Workshop 2015 was carried out in collaboration between three institutions: University of Genova (Italy), KTH Royal Institute of Technology (Sweden) and Maastricht University (Netherlands) at the premises of the Numediart Lab in the University of Mons, Belgium. It is considered as a preliminary study for Use Case #1 of H2020 DANCE Project.

We acquired a deeper understanding of the computational models and theories meaningful to achieve the goals of the DANCE Project, as well as the paths for investigation in the following work related to extraction of high level movement qualities from the sensors data.

With the stated objectives in mind, we developed a working system able to describe a number of movements qualities and translate them into the auditory domain, that was successfully perceived by the participants who took part in the experimental study.

In the future we are planning to define new algorithms related to additional features detection and adapt the implemented techniques to be effectually used with video, Kinect and MoCap data.

1.8 Acknowledgements

This work was partially supported by the EU H2020 ICT DANCE Project no.645553.