.

D1.7 Models and Algorithms - Phase 2

Project No	GA824160
Project Acronym	EnTimeMent
Project full title	ENtrainment & synchronization at multiple TIME scales in the MENTal foundations of expressive gesture
Instrument	FET Proactive
Type of action	RIA
Start Date of project	1 January 2019
Duration	48 months
WP responsible	UNIGE
Due date	Month 30

TABLE OF CONTENTS

1. INTRODUCTION

This document presents an update to the first results on models and algorithms presented in D1.6 (released in December 2020), on multi-temporal predictive models elaborated in the context of dyadic and group synchronisation in human ensembles, integrating fast-time signatures of participants (e.g., eigen frequency, amplitude), multisensory coupling functions (of visual, auditory, haptic and tactile types), and their consequences on entrainment and synchronization (phase and period), as well as low-time psychological and social modulators (e.g., mood and attitudes, likeability, rapport, social competences and emotion), and expressive qualities of gesture.

As detailed in D1.6, specific predictions are tested in WP2, based on modelling and experimental effort to uncover the emergence of dyadic and group synchronisation (e.g., Alderisio et al., 2017; Zhong et al., 2018) at different levels in the EnTimeMent multiscale approach.

The models and algorithms that are being developed by the consortium can be divided into two families: approaches based on machine learning, and approaches based on computational models and algorithms based on different techniques, such as the integration of graph theory and game theory to measure the joint origin of human movement (Kolykhalova et al 2020), and computational models of the individual motor signature (Slowinski et al 2016).

In section 2 we present an initiative of the EnTimeMent ML Technical Group on periodic internal meetings on the role of ML in EnTimeMent. The following sections present updates with respect to the specific models and algorithms presented in D1.6, Section 2.

2. MACHINE LEARNING FOR MOVEMENT ANALYSIS: INTERNAL PERIODIC MEETINGS

The Consortium organized a technical group of researchers on Machine Learning, to continue to analyze and exploit the themes described in the previous deliverable D1.6. In this section we report the new initiative involving the EnTimeMent Consortium, started at the beginning of the third year of the project. It consists of periodic meetings of the internal technical group, focusing on brainstorming sessions on the role of Machine Learning in the modeling of multiple temporal scales in the prediction and analysis of human movement. Specific research scenarios are addressed and shared among partners, and useful feedback are shared between the whole EnTimeMent community. In particular, several topics have been discussed so far with a special interest on discussion on shallow/deep learning models and on how these models can be applied in different contexts faced in the EnTimeMent project.

To sum up, a first (online) meeting was held in February to define the internal working group and the organization of the periodic meeting: then, three meetings took place the last Friday of April, May and June.

In the first meeting/brainstorming session, KTH showed their state-of-the-art and shared with the community how deep learning models can be used to assess different group interactions.

Next, on May, UNIGE showed the pipeline they used in two cases of study (TELMI dataset – "*are we able to automatically distinguish the skill-level of each violin player?*" or on the ellipsis dataset – "*can we automatically distinguish who draw an ellipse?*"). The idea is the same already reported in the previous deliverables where a first benchmark is achieved using shallow models and secondly traditional deep learning models are used to observe improvements with respect the first benchmark. Finally, deep learning models able to handle multiple temporal scales are exploited observing the goodness of this approach respect the other two which already belong to the state-of-the-art.

In the last meeting (end of June 2021), the main argument of interest was about developing interfaces between humans and various machine learning models that have emerged from previous presentations. Moreover, the discussions have been documented in order to further help to plan the execution front of what will emerge in next meetings.

3. RELATION BETWEEN COMPUTATIONAL MODELS OF MOVEMENT ANALYSIS AND BRAIN FUNCTION

UM published two papers and one internal report that are directly relevant for understanding the relation between computational models of movement analysis and brain function, and to test the biological plausibility of computer models.

In Poyanas, Vaessen, de Gelder (2020) we computed postural and kinematic features from affective whole-body movement videos and related them to brain processes. Using representational similarity and multivoxel pattern analyses, we showed systematic relations between computation-based body features and brain activity. Our results revealed that postural rather than kinematic features reflect the affective category of the body movements. The feature limb contraction showed a central contribution in fearful body expression perception, differentially represented in action observation, motor preparation, and affect coding regions, including the amygdala. The posterior superior temporal sulcus differentiated fearful from other affective categories using limb contraction rather than kinematics. The extrastriate body area and fusiform body area also showed greater tuning to postural features. The discovery of midlevel body feature encoding in the brain moves affective neuroscience beyond research on high-level emotion representations and provides insights in the perceptual features that possibly drive automatic emotion perception.

In de Gelder, Solanas (2021) we propose a novel approach for studying the brain's ability to gather survival-relevant information from seeing conspecific body features. Specifically, we propose that behaviorally relevant information from body expressions is coded at the levels of midlevel features in the brain. These levels are relatively independent from higher-order cognitive and conscious perception of bodies and emotions. Instead, our approach is embedded in an ethological framework and mobilizes computational models for feature discovery.

In Zhan, Goebels, de Gelder (2021) we investigate the subjective and the objective side of action and emotion perception (https://doi.org/10.1101/2021.04.15.439961 doi: bioRxiv) represented in whole body images using both ultra high field fMRI and computational analysis of subjective verbal reports. We examined the representational geometry of bodily action- and emotion-understanding by mapping individual subjective reports with word embeddings, besides using conventional univariate/multivariate analyses with predefined categories. Dimensionality reduction revealed that the representations for perceived action and emotion were high dimensional, each correlated to but were not reducible to the predefined action and emotion categories. With searchlight representational similarity analysis, we found the left middle superior temporal sulcus and left dorsal premotor cortex corresponded to the subjective action and emotion representations. Furthermore using task-residual functional connectivity and hierarchical clustering, we found that areas in the action observation network and the semantic/default-mode network were functionally connected to these two seed regions and showed similar representations. Our study provides direct evidence that both networks were concurrently involved in subjective action and emotion understanding.

4. COMPUTATIONAL MODELS AND ALGORITHMS BASED ON THE INTEGRATION OF GRAPH AND GAME THEORIES

This section presents new results based on the extensions of the models and algorithms developed by UNIGE based on the integration of graph and game theory, described in D1.6 and published in Kolykhalova et al. (2020).

Besides the frameworks discussed in D1.6, the investigation of the Origin of Movement (OoM) is useful also in other contexts. As an example, in rehabilitation the detection of the origin of movement can help in enabling a patient to learn how to perform a movement (e.g., how to stand up from a chair) correctly, in order to avoid injuries. In other words, the patient could learn (or re-learn) how to perform a movement by avoiding a specific hurting joint to be the origin of movement. As a second example, in the case of athletes or musicians, the comparison among teachers and students of the origin of movement and of its propagation could help the latter to imitate the former's behavior, in order to improve their performance. Similarly, by comparing movements performed by the same person in different repetitions of the same gesture, one could make that performance highly reproducible, which is needed again in the case of athletes and musicians (e.g., a pianist may want to perform a technical gesture in the same way both in the study room and during a concert).

The model and the methodology described in D1.6 is under and extension in several directions:

- Exploiting a more complex skeletal structure (for which each cluster of joints is associated to a specific joint in the simpler 20-joint skeletal structure), in such a way to allow the analysis of movement at a finer interacting spatio-temporal scale, adopting a multiple-scale approach.

- Using movement-related features different from speed (or belonging to a higher dimensional feature vector) to compute the Shapley value, on order to get a comparison with the results obtained in Kolykhalova et al. (2020) using speed as a feature.

- Incorporating multiple temporal scales. For example, one can look at a fast temporal scale as a first step of the analysis of the origin of movement, then at a slower temporal scale where one can analyse the origin of movement at a higher level.

- Using the time series of the OoM feature derived from the vector of Shapley values at each time instant, in order to train a binary classifier to discriminate different gestures, or to distinguish movements performed either by teachers or students.

- Applying the developed methodology to analyze the emergence of the OoM when two individuals or small groups are involved in the movement itself. In this case the graph nodes are not anymore the joint of an individual person, but each person of the group is a node of a more complex "organism" formed by the social group. This extension of the theoretical framework is also in the direction of observing movement at different spatio-temporal scales.

Preliminary results on some of these extension are contained in (Matthiopoulou et al., 2020) and were presented in a joint UNIGE-EuroMov paper at the ACM ICMI 2020 EnTimeMent Workshop. The additional features, beside speed, investigated therein are tangential acceleration and angular momentum. Moreover, the above-mentioned work investigates the loss in information associated with the reduction of the original 62-joint skeletal structure to a 20-joint one, after a manual clustering of joints performed on the original structure. An additional feature, called “mass distribution”, is defined theiren, too. Loosely speaking, such feature quantifies how much each cluster of joints behaves as a rigid body. A small coefficient of variation of that feature is then associated with a small loss of information when moving from the more complex skeletal structure to the less complex one. Further developments of this research will be presented in September 2021 at ODS - International Conference on Optimization and Decision Science (Gnecco et al., 2021), allowing a fruitful cross-fertilization with people working in the area of optimization.

Theoretical issues associated with the classification of large data sets such as those arising in the present project were investigated in (Kůrková and Sanguineti, 2021). A probabilistic model of the relevance of classification tasks was proposed therein. Correlations of classifiers with input–output functions implemented by connectionistic models used in machine learning were estimated, with particular attention to the effects of increasing sizes of sets of data to be classified.

References

Gnecco, G., Sanguineti, M., Camurri, A., Bardy, B., and Mottet, D. (2021). Comparing Features for Detecting the Origin of Movement based on a Graph-Theoretical Cooperative Game Model. To be presented at the International Conference Optimization and Decision Science (ODS 2021)

Kůrková, V., and Sanguineti, M. (2021). Correlations of Random Classifiers on Large Data Sets. Soft Computing, to appear. DOI: 10.1007/s00500-021-05938-4

Matthiopoulou, O., Bardy, B., Gnecco, G., Mottet, D., Sanguineti, M., and Camurri, A. (2020). A Computational Method to Automatically Detect the Perceived Origin of Full-Body Human Movement and its Propagation. In Companion Publication of the 2020 ACM International Conference on Multimodal Interaction (ICMI 2020), pp.449-453.

Matthiopoulou, O., Gnecco, G., Sanguineti, M., Camurri, A., Bardy, B., and Mottet, D. (2020). Detecting the Perceived Origin of Full-Body Human Movement via Shapley Values Games on Graphs. International Conference Optimization and Decision Science (ODS 2020)

Kolykhalova, K., Gnecco, G., Sanguineti, M., Volpe, G., and Camurri, A. (2020) Automated analysis of the origin of movement: An approach based on cooperative games on graphs, IEEE Transactions on Human-Machine Systems, vol. 50, pp. 550 – 560. DOI: 10.1109/THMS.2020.3016085

5. MECS - THE MULTI-EVENT-CLASS SYNCHRONIZATION ALGORITHM (UNIGE)

Synchronization is a fundamental component of computational models of human behavior, at both intrapersonal and inter-personal level. Event synchronization analysis was originally conceived with the aim of providing a simple and robust method to measure synchronization between two time series. UNIGE developed a novel method extending the state-of-the-art of event synchronization techniques: Multi-Event-Class Synchronization (MECS), already described in D1.6. The model has been refined and completed, described in a paper submitted to a top-ranked IEEE Transactions (currently under revision). MECS measures synchronization between relevant events – belonging to different event classes – that are detected in multiple time series. Using MECS, synchronization can be computed between events belonging to the same class (intra-class synchronization) or between events belonging to different classes (inter-class synchronization). In our paper (Volpe et al, submitted paper) , we also show how our technique can deal with macro-events (i.e., agglomerations of events satisfying specific temporal constraints) and macro-classes (i.e., agglomerations of classes). Finally, our submitted paper presents a case study in which we exploit MECS to compute synchronization between multimodal channels “on-the-fly”. In particular, the proposed example shows how synchronization between respiration and full-body movement of a person explains different movement qualities such as fluidity and impulsivity. Next steps in EnTimeMent will include the extension of our MECS algorithm to compute event synchronization at multiple time-scales. An approach to this problem was proposed recently (Eero et al 2017).We believe MECS can support research on synchronization processes both at an intra-personal and at an inter-personal level. Implementing MECS as a module in the EyesWeb platform makes it freely available to the scientific community, contributing to shed light on phenomena that range from neuronal activity, to human behavior analysis, up to social interaction and cooperation in teams.

References

Gualtiero Volpe, Paolo Alborno, Maurizio Mancini, Radoslaw Niewiadomski, Stefano Piana, Antonio Camurri (2019) MECS – The multi-event-class synchronization algorithm. JAM8 – Intl Joint Actions Meeting, Genoa.

Gualtiero Volpe et al (submitted) MECS – The multi-event-class synchronization algorithm.

6. CAPTURING HUMAN MOVEMENT AND SHAPE INFORMATION FROM SMALL GROUPS TO EXTRACT EXPRESSIVE AND SOCIAL FEATURES – USING MARKER-LESS TECHNIQUES

As an update to our previous study as described in section 1.6 we are now performing the same studies with an updated version of the pose estimation algorithm, AlphaPose (v0.4.0). To quickly recap, our method involves the conversion of time domain data to the frequency domain. Our previous methodology involves the FFT. To test alternative methods, where there shall be no constraints pertaining to the assumption of the linearity as is the case of FFT, we are also exploring the possibility of using HHT. The two methods being fundamentally different, their results will be interesting to explore.

We have also looked at expanding our previous data set by including a wider range of annotated videos, which will allow us to better understand the range of our hypothesis. Additionally, we also plan to perform an analysis of the audio of our datasets for the sections under observation using the MIR toolbox in Matlab. This audio analysis will be helpful in interpreting the PLV results and will be a novel aspect of this study.

These efforts will soon be condensed into a paper that we plan to publish soon.

References

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. Human brain mapping, 8(4), 194-208

Presacco, A., Forrester, L. W., & Contreras-Vidal, J. L. (2012). Decoding intra-limb and inter-limb kinematics during treadmill walking from scalp electroencephalographic (EEG) signals. IEEE Transactions on neural systems and rehabilitation engineering, 20(2), 212-219

Skogstad, S. A. V. D., Nymoen, K., Høvin, M. E., Holm, S., & Jensenius, A. R. (2013). Filtering motion capture data for real-time applications

Varni, G., Mancini, M., Fadiga, L., Camurri, A., & Volpe, G. (2019). The change matters! Measuring the effect of changing the leader in joint music performances. IEEE Transactions on Affective Computing

L. Bishop, "Collaborative musical creativity: How ensembles coordinate spontaneity," Frontiers in psychology, vol. 9, p. 1285, 2018

L. Bishop and W. Goebl, "Communication for coordination: gesture kinematics and conventionality affect synchronization success in piano duos," Psychological Research, vol. 82, no. 6, pp. 1177–1194, 2018

7. HIERARCHICAL HUMAN ACTIVITY RECOGNITION AND PROTECTIVE BEHAVIOUR DETECTION (PUBLISHED IN [WANG ET AL. 2021])

Joint modelling of functional movement and cognitive/affective movement is a practical approach to machine perception as it enables comprehensive appreciation of scenes. However, a more fundamental value is based on the fact that cognitive/affective experiences are usually not isolated, but rather typically occur within everyday functioning. For certain applications, the cognitive/affective experience is even particular to the functional movement being performed. This is certainly the case for protective behaviours (i.e. behaviour intended to protect from harm or exacerbation of pain, typically bodily expressions [Sullivan et al. 2006]) which are embedded in the performance of the movement that the person finds challenging. In such cases, knowledge about the aim of the functional movement (e.g. the activity goal) can enable interpretation, or identification in the first instance, of cognitive/affective behaviour. Cognitive/affective expression could occur at the different levels of abstraction in human movement abstraction from discrete [Kleinsmith et al. 2011] (pose) to increasingly continuous (gesture, action, interaction, activity) [Edwards, Deng, and Xie 2016][Karg et al. 2013]. Thus, whereas on one hand, there is value in integrating activity recognition in affective behaviour detection, on the other hand, it may be beneficial to account for the different temporal scales of these two tasks.

This is one of the main motivations behind our Hierarchical Human Activity Recognition and Protective Behaviour Detection (Hierarchical HAR-PBD) architecture. Figure 1 gives an overview of the architecture. One of the primary characteristics of the architecture is the separate spatial and temporal encodings for human activity recognition and protective behaviour detection while the prediction for human activity recognition system further feeds into the protective behaviour detection task. The same underlying architecture (a Graphical Convolution and Long Short-Term Memory neural network, GC-LSTMNN) is used for both tasks, and the human activity recognition module is first pre-trained before being integrated in the protective behaviour detection module. The weights of the human activity recognition module are frozen in the training of the latter so that the human activity recognition can inform protective behaviour detection without the module for the latter enforcing its temporal scales on that for the former. The underlying architecture, the GC-LSTMNN, is described fully in deliverable D3.7 while results of experiments on the Hierarchical HAR-PBD are reported in deliverable D2.2.

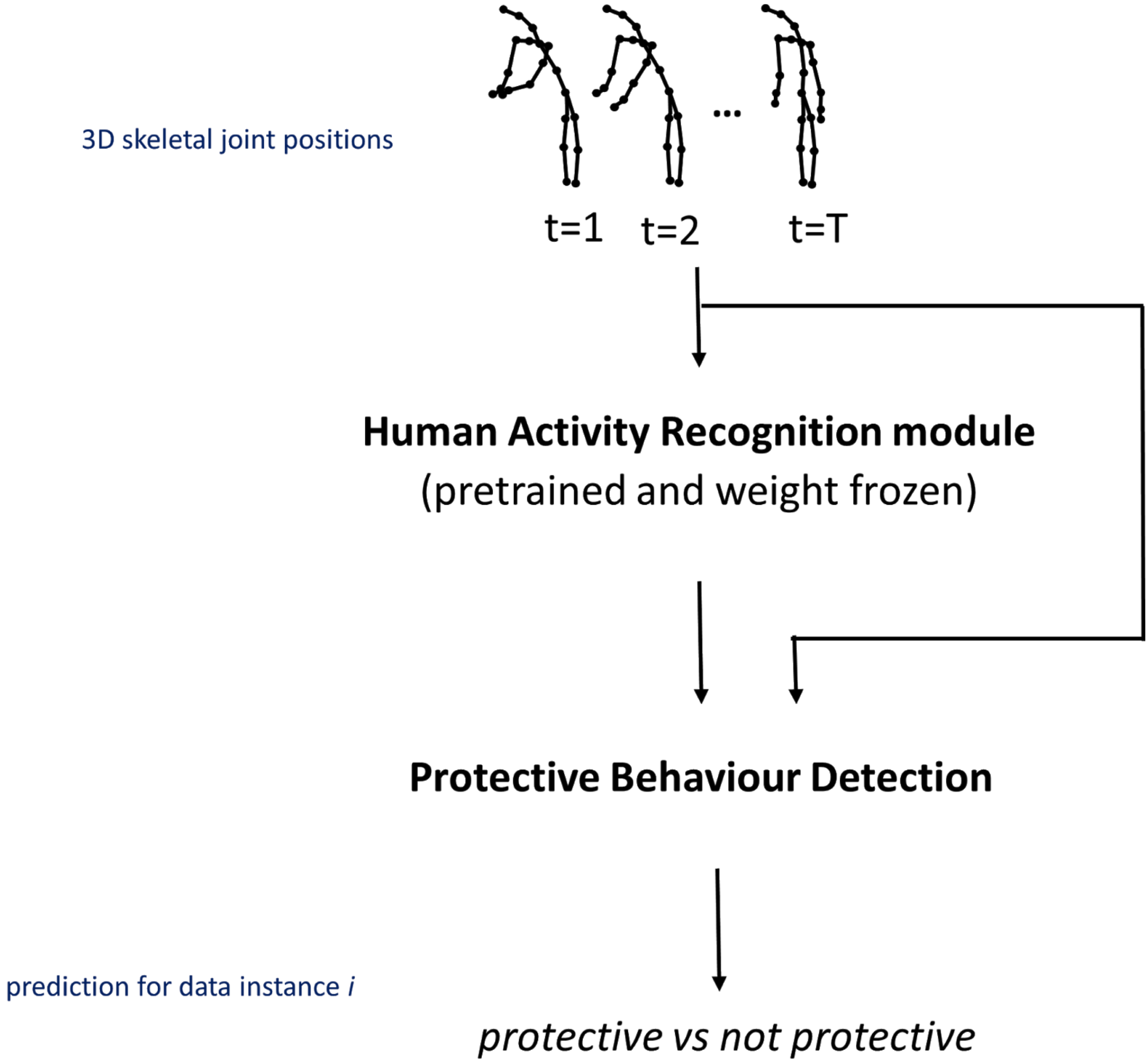

Figure 1. Hierarchical Human Activity Recognition and Protective Behaviour Detection

References

Chongyang Wang, Amanda C De C Williams, Nicholas D Lane, and Nadia Bianchi-berthouze. 2021. “Leveraging Activity Recognition to Enable Protective Behavior Detection in Continuous Data.” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5 (2): 1–24.

Michael J L Sullivan, Pascal Thibault, Andre Savard, Richard Catchlove, John Kozey, and William D. Stanish. 2006. “The Influence of Communication Goals and Physical Demands on Different Dimensions of Pain Behavior.” Pain 125 (3): 270–77.

Andrea Kleinsmith, Nadia Bianchi-Berthouze, and Anthony Steed. 2011. “Automatic Recognition of Non-Acted Affective Postures.” IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 41 (4): 1027–38.

Michael Edwards, Jingjing Deng, and Xianghua Xie. 2016. “From Pose to Activity: Surveying Datasets and Introducing CONVERSE.” Computer Vision and Image Understanding 144: 73–105. https://doi.org/10.1016/j.cviu.2015.10.010.

Karg, Michelle, Ali Akbar Samadani, Rob Gorbet, Kolja Kuhnlenz, Jesse Hoey, and Dana Kulic. 2013. “Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation.” IEEE Transactions on Affective Computing 4 (4): 341–59.

8. SONIFICATION IN SCENARIO 2 (UCL)

In work led by a UCL student collaborator under the supervision of Nicolas Gold, we (UCL) have developed a new prototype sonification that maps movement data directly to musical properties (Neubauer et al, 2021). The resulting sonification has four levels designed to address a range of factors in pain-related movement: anxiety reduction (through ambient, slow attack/release sounds), fostering feelings of achievement and progress (chime sounds for movement threshold achievement), and objective self-monitoring (filtered pink noise related directly to angular velocity of joint movement, and parameters of a granular synthesiser modified to convey energy through sound coherence and decoherence). The combined soundscape creates a complex and manipulatable whole. At present, pre-recorded data is replayed as if live, and we are working to link the live sensors directly to the sonification engine. Participant-based evaluation is planned to follow.

References

Neubauer, L., Gold, N.E, Olugbade, T., Williams, A. CdeC., Berthouze, N. (2021) “Functional Musical Sonification for Chronic Pain Support”, Ubiquitous Music 2021, Porto (and online), Sept 6-8, to appear.

9. INTERACTIVE SONIFICATION IN SCENARIO 3 (UNIGE)

We developed an interactive sonification model of movement expressive qualities, and a corresponding software prototype for the *DanzArTe – Emotional Wellbeing Technology* project, as a part of Scenario 3 of EnTimeMent (https://www.lavanderiaavapore.eu/2021/03/23/danzarte-welfare-territoriale/).

The model addresses the “fear of falling” in mobility exercises for frail subjects. Three objectives have been proposed in the design of interactive sonification: (i) to create an interactive low-stress experience of sound; (ii) to reassure participants; (iii) to enhance the perception and performance of fluid movements.

The first objective is addressed by observing guidelines for low-intrusiveness sound design (A. Cera, N. Misdariis, 2021), with particular attention to the relation between foreground and background, dynamic smoothness and avoidance of pitch repetition patterns.

The second objective is addressed by eliciting a low-arousal / positive valence state, using low-dissonance harmonies, low roughness in timbres, smoothness in dynamics, low complexity in sonic structure, and other strategies following criteria widely studied in scientific literature on emotion and music (e.g., Eerola and Vuoskoski 2013).

The third objective is addressed by exploiting and further extending state of the art criteria for fluidity sonification (Alborno et al 2016; Niewiadomski et al 2019), with particular attention to spectral flux, dynamic control of spectral centroid, smoothness of dynamic profiles, and other sonic variables.
In the proposed model, fast and medium temporal scales concern expressive features including kinetic energy, movement fluidity, body symmetry, which create modulations of the sonifications' surface and spatial movement.
Features at a slower temporal scale concern user engagement, and model the intrusiveness of the sonification, in order to arouse or relax the participant.
Details on research results and evaluations of the proposed model are described in a paper in preparation.

References

A.Cera, N.Misdariis (2021) INTRUSIVENESS, ANNOYANCE AND SOUND DESIGN
LINKs - series 5-6, pp.186-191.

T.Eerola, J.K.Vuoskoski (2013) A REVIEW OF MUSIC AND EMOTION STUDIES: APPROACHES, EMOTION MODELS, AND STIMULI, Music Perception.

P.Alborno, A.Cera, S.Piana, M.Mancini, R.Niewiadomski, C.Canepa, G.Volpe, A.Camurri (2016) INTERACTIVE SONIFICATION OF MOVEMENT QUALITIES – A CASE STUDY ON FLUIDITY. Proc. ISon 2016 - 5th Interactive Sonification Workshop, CITEC, Bielefeld University.

R.Niewiadomski, M.Mancini, A.Cera, S.Piana, C.Canepa, A.Camurri (2019) Does embodied training improve the recognition of mid-level expressive movement qualities sonification?, Journal on Multimodal User Interfaces, v. 13, n. 3, pp. 191--203, ISBN/ISSN: 1783-8738.