INTERACTIVE SONIFICATION OF MOVEMENT QUALITIES – A CASE STUDY ON FLUIDITY

Paolo Alborno, Andrea Cera, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Corrado Canepa, Gualtiero Volpe, Antonio Camurri

University of Genova, Casa Paganini – InfoMus, DIBRIS
paoloalborno@dibris.unige.it, andreawax@yahoo.it, steto84@infomus.org, Maurizio.mancini@unige.it, radoslaw.niewiadomski@dibris.unige.it, corrado@infomus.org, gualtiero.volpe@unige.it, antonio.camurri@unige.it

ABSTRACT

The EU H2020 ICT Project DANCE investigates how affective and social qualities of human full-body movements can be expressed, represented, and analysed by sound and music performance. In this paper we focus on one of the candidate movement qualities: Fluidity. An algorithm to detect Fluidity in full-body movement, and a model of interactive sonification to convey Fluidity through the auditory channel are presented.
We developed a set of different sonifications: some follows the proposed sonification model, and others are based on different, in some cases opposite, rules. Our hypothesis is that our proposed sonification model is the most effective in communicating Fluidity. To confirm the hypothesis, we developed a serious game and performed an experiment with 22 participants at MOCO 2016 conference. Results suggest that the sonifications following our proposed model are the most effective in conveying Fluidity.

1. INTRODUCTION

The EU H2020 ICT Project DANCE investigates how affective and social qualities of human full-body movements can be expressed, represented, and analysed by sound and music performance. DANCE addresses research challenges such as: is it possible to perceive movement expressive qualities in dance through the auditory channel? Can we imagine concrete ways to "listen to a choreography", "feel a ballet"? If we can capture the inner and intimate expressive qualities conveyed by movement to an external observer, these qualities might be made manifest through other sensory modalities such as, for example, the auditory one. In such a way, by closing her eyes and by listening to the auditory representation of movement qualities, a user can be made aware of some information, which is hidden in the movement and may be difficult to be perceived otherwise.

Interactive sonification is receiving a growing relevance in the scientific and artistic communities: it is used in rehabilitation, sensory substitution, perception enhancement, and human-computer interfaces (Dubus & Bresin, 2013).
The importance of sonification in communicating movement qualities is observable since the silent movies era, where sonifications were improvised by a pianist playing while observing what was happening on the screen (Hermann, Hunt, & Neuhoff, 2011). With the development of more sophisticated techniques the role of sonification in movies became more and more important: sonifications are now used in a broad range of scenarios (e.g., to convey off-screen events, to cover cuts and scene transitions, signal flashbacks, and direct the watcher's attention) (Hermann, Hunt, & Neuhoff, 2011).

This paper focuses on designing and experimenting different interactive sonifications of a specific quality of movement: *Fluidity* (i.e., how to perceive movement Fluidity by the auditory channel). We start from our recent proposal of a computational model (and the corresponding EyesWeb software module) of Fluidity. In this work we adopt a conceptual framework recently proposed by Camurri and colleagues for the analysis of expressive movement qualities (Camurri, et al., 2016).

2. AUTOMATED ANALYSIS OF MOVEMENT QUALITIES: FLUIDITY

Fluidity is often considered as a synonym of "good" movement (e.g., in certain dance styles), and is different from "smoothness", which is referred to the movement of a single joint. Furthermore, Fluidity is one of the properties that seem to contribute significantly to perception of emotions (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004).
Caridakis and colleagues (Caridakis, et al., 2007) investigated fluidity of hands trajectories, and computed it as the sum of the variance of the norms of the hands' motion vectors. Piana et al. (Piana, Stagliano', Camurri, & Odone, 2015) studied human motion trajectories and defined a Fluidity index based on the minimum jerk law.

Starting from literature on biomechanics and psychology, and by conducting interviews and movement recordings with experts in human movement such as choreographers and dancers, we propose the following definition of Fluid movement [6] (performed by a part of the body, by the whole body, or by a group of dancers behaving as a single organism):

- the movement of each joint is smooth, following the standard definitions in the literature of biomechanics (Viviani & Flash, 1995) (Morasso, 1981);
- the energy is free to propagate along the kinematic chains (e.g., from head to trunk, from shoulders to arms; in a group from a dancer to another) according to a coordinated wave-like propagation. That is, there is an efficient propagation of movement along the kinematic chains, corresponding to a minimization of the dissipation of energy.

2.1. A Simplified Computational Model of Fluidity

In this work, we measure movement Fluidity from IMUs (Inertial Measurement Units) data. Data can be extracted from sensors located on two different body joints. We consider the two IMUs H^R and H^L, placed, respectively, on the users' right and left wrists [1].

[1] To extract the movement data we used the X-OSC sensors (X-IO Technology) that provide 9-axis inertial measurements of, respectively, the participant's right and left hand.

Starting from the raw inertial measures we compute the linear acceleration, i.e., the acceleration detected by the device minus the component corresponding to the gravity.

First, we compute two low-level movement features:

Jerkiness: at frame f, hands linear accelerations $H^N_{Lin_{x,y,z}}$ with $N\epsilon\{R,L\}$ are read; we calculate the squared jerk of the hands by deriving the linear acceleration components and summing them:

$$J^N = \left(\dot{H}^N_{Lin_x}\right)^2 + \left(\dot{H}^N_{Lin_y}\right)^2 + \left(\dot{H}^N_{Lin_z}\right)^2$$

Then, we normalize J^N over a buffer of 20 values:

$$J^N_{tot} = \frac{\sum_{i=1}^{20} J^N_i}{Max(J^N)}$$

Kinetic Energy: it is computed as the global kinetic energy of the wearer's hands, whose mass is approximated to 1 for sake of simplicity. Each velocity component is obtained by integrating the corresponding linear acceleration component:

$$E^N = \frac{1}{2}\left[\left(\int H^N_{Lin_x}\right)^2 + \left(\int H^N_{Lin_y}\right)^2 + \left(\int H^N_{Lin_z}\right)^2\right]$$

By integrating the linear acceleration components, we obtain the velocity components, necessary to compute kinetic energy. As before, we normalize the resulting value:

$$E^N_{tot} = E^N / Max(E^N)$$

Finally, we evaluate the user's fluidity of movement as:

$$F^N_{tot} = 1/[J^N_{tot}/E^N_{tot}]$$

Fluidity Index FI is the mean of the fluidity computed on the two hands:

$$FI = (F^R_{tot} + F^L_{tot})/2$$

Both the above-mentioned features (*Jerkiness* and *Kinetic Energy*) belong to Layer 2 of the conceptual framework described in (Camurri, et al., 2016).

3. SONIFICATION STRATEGY

Is it possible to sonify a movement in a way that the listener can perceive fluidity even without seeing it?

Our objective is to investigate strategies for sonifying Fluidity. To this aim, we propose a sound synthesis model and a set of sonifications. The sonifications are generated by controlling some of the variables that characterize the model. We defined sonifications consistent with the fluidity sonification model presented in Section 3.2 and. We also developed different sonifications, different or in contrast with the sonification model we propose.

3.1. Sonification Background

To identify the best approach we focused on two different sources of inspiration.
On the one hand, we analyzed the state of the art in the expression of extra-musical qualities in sound design and electroacoustic music. Works of Wishart (Wishart, 1986), Tagg (Steedman, 1981), Middleton (Middleton, 1993), Kahn (Kahn, 1999) on cross-modality, studies by Carron (Carron, Rotureau, Dubois, Misdariis, & Susini, 2015) on the sound designers' production techniques to convey specific extra-musical meaning provide a very useful and rich background of methodological guidelines for sonically rendering the Fluidity of a movement.

On the other hand, we took inspiration from cinematographic works. Sound design in cinematography can indeed provide a popular vocabulary, representing a largely shared way to associate sound and physical qualities (Hermann, Hunt, & Neuhoff, 2011).
We selected few sequences from very well know blockbusters such as The Matrix (Warner-Bros), The Fantastic Four (Marvel-Studios), Alien (Twentieth-Century-Fox-Film) in which movement Fluidity is clearly shown and sonically underlined: an example is given by the sequence in The Matrix movie (Warner-Bros) where the main character connects to the Matrix for the first time and his body becomes liquid.

We empirically analyzed the sequences to find which sound cues and features are mainly used to induce the sensation of fluidity in the spectator.

On the basis of the described background, we then identified a set of common elements to "fluid" sounds: smooth attack and release curves, smooth dynamic profiles (i.e., no audible jumps in dynamics and no cuts), and smooth timbral evolution.
In particular, the timbral content is close to the sound produced by flowing water, sounds audible underwater or sounds of bubbles. Often these sounds are more pitched than noisy, even if with non-harmonic relation between partials.

Our hypothesis is that Fluidity can be sonified using continuous sounds with a high value of spectral smoothness, evolving timbrically and dynamically with continuity, without audible steps. Sounds with low/medium spectral centroid are the first choice.
On the contrary, sounds with a very high spectral centroid, even if continuous and smooth, may remind of non-fluid phenomena (i.e., friction, noise), and convey the impression of something moving against a resistance.

3.2. Fluidity sonification model

The model we propose is based on granular synthesis: it easily allows to change the basic sound materials (buffers), to obtain a wide timbral variety.

The model is parametrized as follows:

- Fluidity index FI to detect fluid movements, on a large temporal scale (1-3 seconds).
- Kinetic Energy EI (the mean between E^R and E^L) to detect little, short and fast changes in the movements which otherwise would not have been considered by FI, potentially causing a lack of synchronization between the visual flow and the sonification.

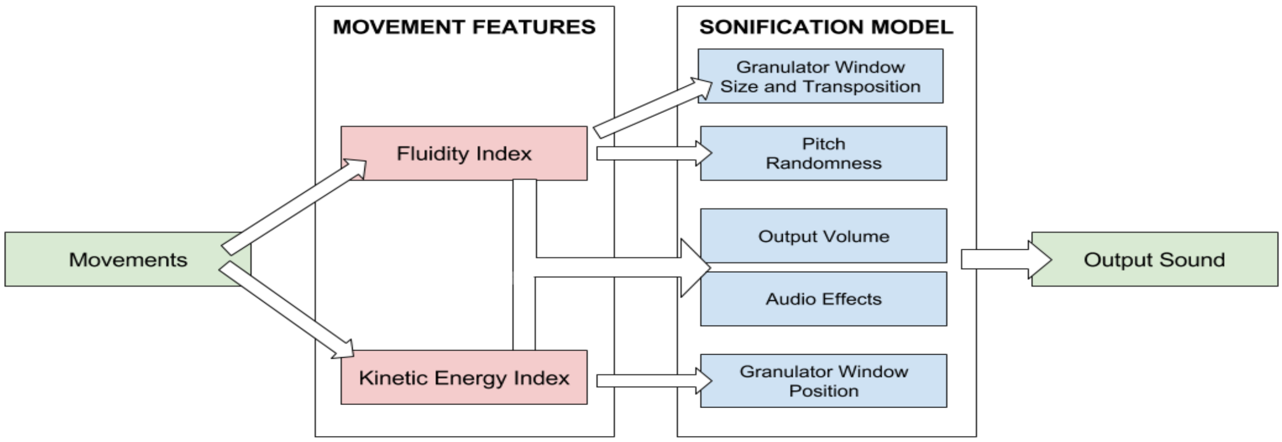

Figure 1. Schematic representation of the sonification system: Kinetic Energy index and Fluidity index control the generation of sound.

Starting from the hypotheses described so far, we designed seven different sonifications and we tested them by mapping them to a short (33s) recorded dance sequence in which the dancer was instructed to characterize her movements by high and continuous fluidity.

Sonifications have been rendered in order to assess several degrees of correspondence between the qualities of the dance and the sound. The granular synthesis model is implemented as MAX/MSP patches and uses a 20 seconds long buffer.
The granulator window position is modulated by the value of the Kinetic Energy index. Fluidity index is then used to make small displacements and to tune the window size.
Moreover, a combination of the two indexes controls the output volume and some audio effects (comb filters, delays) that are applied to the sound before the final audio synthesis (see Fig. 1).

The mapping between the movement indexes and the granulator's parameters is designed to provide a smooth and fluid control of the model patch (low-pass filtering and temporal interpolations with ramps of at least 100 milliseconds duration).

The “ideal” buffers for conveying Fluidity are based on either pitched and harmonic or inharmonic sounds. The frequency content of the buffers varies over time: in their initial portion (explored by the granulator's window when the energy of the movement is low) the buffers are fitted with a low centroid, and in the final portion (explored by the granulator's window when the energy of the movement is high) the centroid is higher.

Fluid movements characterized by low energy (slow, calm) are sonified with low centroid and dark timbre. Fluid movements characterized by high energy (fast, circular) are sonified with higher centroid and brighter timbre. High energy is translated to a richer sound in high frequencies, brighter, “energetic”, and vice-versa. The final portion of the buffer is exploited for movements that are very energetic and presumably moving towards less-fluid qualities.

3.3. Sonifications

We developed seven different sonifications: they differ in the type of buffer (“ideal” or “wrong/contrasting”) and type of mapping (“good” or “bad”); “ideal” buffers have been designed to comply with the description of fluid sounds given in the previous section. On the contrary, “wrong” buffers have been designed following the opposite criteria. “Good” mappings are characterized by smooth transitions and continuity while “bad” ones use steps and discontinuities.

The seven sonifications belong to four different groups:

A. Two sonifications generated using an “ideal” buffer and “good” mapping, identified by numbers 2 and 4
B. A single sonification generated using a “ideal” buffer and “wrong” mapping, identified by number 5
C. Two sonifications generated using an “ideal” buffer, but with “bad” mapping, numbers 1 and 3
D. Two control sonifications, identified by numbers 6 and 7

Sonifications belonging to Group A contain the patches designed to sonify Fluidity in the best possible way.
In this group the buffers are based on a sound material characterized by absence of audible steps, in timbral and dynamic evolution and by high values of spectral smoothness and low centroid, according to the observation that fluid movements show no jerks, sudden stops or sudden changes of direction.

Group B contains a single sonification. It uses the same mapping of Group A, but the buffer is designed in the opposite way: the average spectral smoothness value is low, showing a chaotic behavior and the centroid of the buffer is higher in the initial portion then it decreases.

Group C sonifications are based on the same buffer of Group A but make use of a non-fluid mapping characterized by a ten-step discretization of the indexes values controlling the granulator's parameters, furthermore interpolation ramps length is decreased from 10 to 5 milliseconds.

Group D contains two control sonifications, generated with a different synthesis technique (not based on granular synthesis):

- *Microsounds* is designed to provide a sonic behavior deliberately contrasting with the idea of Fluidity. It is based on the superposition of four short loops made of percussive sounds (each loop is between 400 and 700 milliseconds). Kinetic Energy controls the speed of each loop, the playback rate and amplitude of the samples and the cutoff frequency of a low-pass filter.

The result is a contrasted and irregular sonic material, which segments the sound continuity into a myriad of micro-events.

- *Pink Noise* conveys a fluid, smooth timbral profile, realized with a simple technique, less evocative than granular synthesis. The sound is rendered by a pink noise generator and a low-pass filter. The cutoff frequency and the output volume are controlled by the energy while and the filter slope is piloted by the fluidity index.

Sonification Name	ID	Group
Ideal_buffer_inharmonic_bad_mapping	1	C
Ideal_buffer_inharmonic_good_mapping	2	A
Ideal_buffer_pitched_bad_mapping	3	C
Ideal_buffer_pitched_good_mapping	4	A
Wrong_buffer_good_mapping	5	B
Microsounds	6	D
Pink Noise	7	D

Table 1: Sonification table

4. EXPERIMENTAL SETUP

We developed a software platform to create and flexibly configure several serious games to evaluate how users "hear" a movement or a dance. A detailed description of the platform architecture can be found in (Kolykhalova, Alborno, Camurri, & Volpe, 2016).
For this experiment, we generated an instance of this platform to validate the seven sonifications described in Section 3.

4.1. Experimental Scenario

We carried out the experiment on a group of 22 adult participants.
To engage the users in the experiment and facilitate them in focusing on the task, we designed the experiment as a competitive game between two players/participants.
An entire game session consisted of listening to seven sonifcations produced from a pre-recorded short dance performance (not visible to the participants).
Sonifications are presented in a random order at each new session.
While listening, each player was asked to move freely following what they listened to. Players were not aware of the origin of the sonic material they listened to. To avoid mutual influence between the players and to increase their sensitivity on the auditory perception they were blindfolded before starting the experiment.

The original 9 axis IMU motion data recorded during the dance performance, were used to generate the sonification, and, at the same time, each participant/player motion data were received from two IMU sensors on her wrists.
Both the recorded dancer and the player's motion information were used to compute the value of the movement qualities described in Section 2.1, but only the data from the dancer were sonified.

At every time instant, the Performance index ($PerformIndex_{pi}$) of each player is computed as:

$$Perform\ Index_{pi} = \frac{F_{pi}}{F_d} \qquad i \in [1,2]$$

where F_{pi} and F_d are the player's and the dancer's fluidity index respectively. This index is used to compute each players' game score.

The score reflects how much each player is able to "understand" the dancer movement's qualities (in this case Fluidity) through the sound. Our hypothesis is that sonifications following our model will communicate more efficiently to the players the original quality of movement, resulting in higher scores.

4.2. System implementation

The game platform and this specific instance were implemented in the EyesWeb XMI platform. EyesWeb XMI (CasaPaganini-InfoMus) allows for real-time recording, synchronization, and real-time processing and analysis of multimodal data.
It includes a collection of software modules, and a visual development language enabling users to build applications and graphical interfaces. EyesWeb modules support analysis of nonverbal motor behavior, including motion trackers, real-time extraction and analysis of motion qualities, trajectory analysis, time-series analysis, machine learning, and analysis of affective and social interaction.

5. DATA ANALYSIS

Sonification scores are computed as the sum of all the Performance indexes of all players, for each one of the seven sonifications. To take into account the player subjective expressive style, sonification scores were normalized to the maximum score obtained by each player.

The evaluation of the sonifications is based on the differences among the sonification scores given by the game to all players.

The game's total score is calculated as the sum of each player's scores during a whole gaming session, but it remains relevant only to entertainment purposes i.e., to find out which of the two players has won the competition and it does not represent an interesting factor with regard to the validation of the sonifications.

5.1. Statistical Analysis

A statistical analysis on the sonification scores was performed, to confirm the hypothesis that sonifications following our model are the best candidate to convey fluidity through the auditory channel.

To test our hypothesis a one-way repeated-measures ANOVA was conducted with one within-subject measure: Condition (1-7) as sum of the Performance Index for the different sonfiications as dependent value.
Since Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated ($p < 0.005$), a Greenhouse-Geisser correction was used ($\varepsilon = .507$).

Within-subject analysis showed a significant effect of Condition F(3.041 ; 63.852) = 14.081 ; p<0.001 after application of Greenhouse-Geisser correction of Sphericity).

Sonification Group	***Sonification ID***	***Sum***	***Mean***	***Variance***
A	2	19,85777	0,902626	0,010792
A	4	19,94846	0,906748	0,010937
B	5	14,07463	0,639756	0,08897
C	1	17,78638	0,808472	0,03178
C	3	18,52351	0,841978	0,011421
D	6	11,45464	0,520666	0,091935
D	7	18,01019	0,818645	0,023744

Table 2: Sum, average, and variance of the scores obtained by the seven different sonifications.

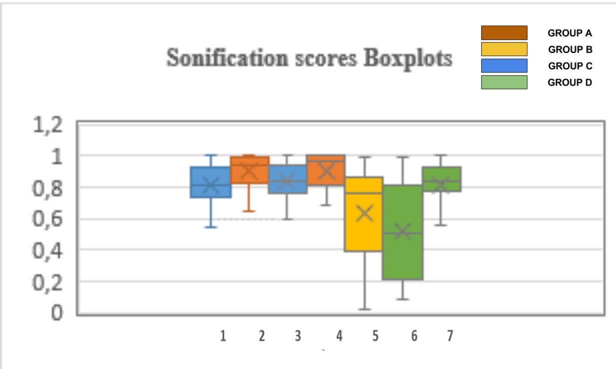

Figure 2: score boxplots

Next, the effect of Condition was analyzed using a post hoc tests with Bonferroni[2] correction.
Post hoc comparisons indicated that the Performance Index sum in Condition 6 was significantly lower than in Condition 2 (p<0.001), Condition 3 (p<0.005) and Condition 4 (p<0,001).
Moreover, comparisons indicated that the Performance Index sum in Condition 5 was significantly lower than in Condition 2 (p<0.005) and Condition 4 (p<0.005).
In addition, sonification 6 showed a significant difference from sonification 7 (p<0.001).

Results suggest that sonifications in groups A and C better convey movement fluidity than sonifications in group B. In group D *Microsounds*, as expected, did not perform as good as Group A and C sonification. We did not observe the same behavior for the *Pink Noise* sonification.

[2] Bonferroni correction was applied by multiplying the p-value of the least significant differences (LSD) by the number of tests, i.e., 21)

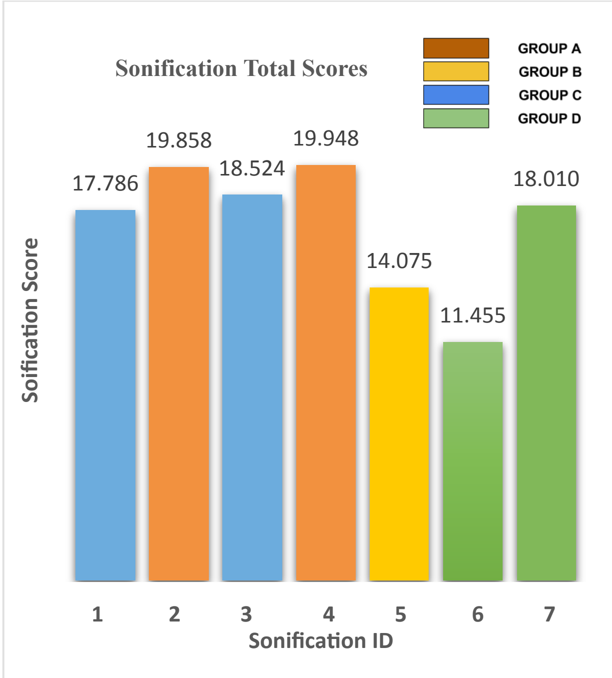

Figure 3: Sonification total scores on 22 players. Each column represents the sum of the Performance Indexes achieved by all the participants among the seven sonifications.

6. CONCLUSIONS

The results from the experiment confirmed the validity of our sonification model as a promising starting point to convey Fluidity: sonifications in Group A obtained significantly better scores and seem to be more effective. As expected, the control sonification *microsounds* resulted less effective, followed by Group B (that is designed deliberately contrasting with the model). No statistically significant difference between Groups A, C and the control sonification *Pink Noise* was found. These results are similar to those presented in (Frid, Bresin, Alborno, & Elblaus, 2016).

Future work will include further refinement of the model repeating the experiment by forcing some parameters of the group B sonifications to be more different from those in group A.
An ongoing work in the EU DANCE project includes a collaboration with the choreographer Virgilio Sieni to define and compute a set of mid-level movement qualities and to translate them into the auditory domain. Results from this work will be presented in public events in April 2016.

ACKNOWLEDGEMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n. 645553 (DANCE).

7. REFERENCES

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In *Gesture-based communication in human-computer interaction* (pp. 20-39). Springer.

Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., et al. (2016). The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. *Proceedings of the 3rd International Symposium on Movement and Computing* (pp. 6:1--6:7). New York, NY, USA: ACM.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., et al. (2007). Virtual agent multimodal mimicry of humans. *Language Resources and Evaluation , 41*, 367-388.

Carron, M., Rotureau, T., Dubois, F., Misdariis, N., & Susini, P. (2015). Portraying sounds using a morphological vocabulary. *EURONOISE 2015.*

CasaPaganini-InfoMus. (n.d.). *EyesWeb XMI*. Retrieved from http://www.infomus.org/eyesweb_eng.php

Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PloS one , 8*, e82491.

Frid, E., Bresin, R., Alborno, P., & Elblaus, L. (2016). Interactive Sonification of Spontaneous Movement of Children—Cross-Modal Mapping and the Perception of Body Movement Qualities through Sound. *Frontiers in Neuroscience* , 10.

Hermann, T., Hunt, A., & Neuhoff, J. G. (2011). *The sonification handbook.* Logos Verlag Berlin.

Kahn, D. (1999). *Noise, water, meat: A history of sound in the arts.* MIT press.

Kolykhalova, K., Alborno, P., Camurri, A., & Volpe, G. (2016). *A serious games platform for validating sonification of human full-body movement qualities.* MOCO 2016: 39:1-39:5.

Marvel-Studios. (n.d.). *Fantastic Four*. Retrieved from http://www.fantasticfourmovie.com/trailer

Middleton, R. (1993). Popular music analysis and musicology: bridging the gap. *Popular Music , 12*, 177-190.

Morasso, P. (1981). Spatial control of arm movements. *Experimental brain research , 42*, 223-227.

Piana, S., Stagliano', A., Camurri, A., & Odone, F. (2015). Adaptive Body Gesture Representation for Automatic Emotion Recognition. In *Transactions on Interactive Intelligent System* (p. in printing). ACM press.

Steedman, M. (1981). Kojak, 50 Seconds of Television Music: Towards the Analysis of Affect in Popular Music. By Philip Tagg. Göteborg: Musikvetenskapliga Institutionen, 1979.(Studies from the Dept of Musicology, Göteborg, 2) 301 pp. *Popular Music , 1*, 185-187.

Twentieth-Century-Fox-Film. (n.d.). *Alien*. Retrieved from http://www.foxmovies.com/movies/alien

Viviani, P., & Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance , 21*, 32.

Warner-Bros. (n.d.). *The Matrix*. Retrieved from http://www.warnerbros.com/matrix

Wishart, T. (1986). Sound symbols and landscapes. In *The language of electroacoustic music* (pp. 41-60). Springer.

X-IO Technology. (n.d.). Retrieved from http://www.x-io.co.uk/products/x-osc