
Statistical analysis of human body movement and group interactions in response to music.

Frank Desmet, Marc Leman, Micheline Lesaffre and Leen De Bruyn

IPEM, Department of Musicology, Ghent University, Belgium
fm.desmet@ugent.be

Summary. Quantification of time series that relate to physiological data is challenging for empirical music research. Up to now, most studies have focused on time-dependent responses of individual subjects in controlled environments. However, little is known about time-dependent responses of between-subject interactions in an ecological context. This paper provides new findings on the statistical analysis of group synchronicity in response to musical stimuli. Different statistical techniques were applied to time-dependent data obtained from an experiment on embodied listening in individual and group settings. Analysis of inter group synchronicity are described. Dynamic Time Warping (DTW) and Cross Correlation Function (CCF) were found to be valid methods to estimate group coherence of the resulting movements. It was found that synchronicity of movements between individuals (human-human interactions) increases significantly in the social context. Moreover, Analysis of Variance (ANOVA) revealed that the type of music is the predominant factor in both the individual and the social context.

Key words: Human body movement, Social Interaction, Statistical Analysis, Music Research, Embodiment.

1 Introduction

The analysis of human body movement is relevant for a number of research areas such as therapy and rehabilitation [1], sports [2], bioinformatics [3] and neurology [4]. Also in empirical music research, there is a growing interest in how the human body moves and responds to music [5, 6, 7]. However, the study of music-driven human body movement is complex because it has to deal with several factors that introduce variability on top of music-driven time varying data, such as the neural-muscular-skeletal variability among subjects, the variability in response patterns of single subjects due to learning and training, or the subjects' background (gender, culture) [8]. The present study relies on Leman's model of music communication, which is based on the notion of embodiment [9]. The human body is thereby considered as a natural

mediator between mind and physical environment. In this paper, we focus on Leman’s social factor [10] of the above music communication model by studying music-driven body movement of a group of people whose social interaction is taking place in ecological conditions. Music is thereby seen as a social phenomenon and the quantification of social interaction is considered to be a key factor for the development of future electronic mediation technologies and applications. So far, most studies on music-driven body movement have been carried out in controlled laboratory conditions, often with single subjects, limited to simple motor tasks [11, 12]. Although some studies have focused on group behavior, few studies have studied music-driven body movement in real life (ecological) environments [13]. In this paper we test the hypothesis that humans move more synchronous to the beat of the music and with each other in group than in an individual setting. To test the above hypothesis we rely on different methods yet all methods share a common approach, namely, the definition of similarity measures, which is then applied on the Multivariate Time Series (MTS) matrices. This paper is organized as follows: in section 2 the experimental design and data considerations are reviewed, section 3 deals with the analysis, discussion and conclusions are given in sections 4 and 5.

2 Experimental Design and Data Considerations

The experiment was carried out during the Accenta 2007 exhibition in Ghent, where groups of four subjects moved remote Wii sensors while listening to (recorded) music. Sixteen groups of four adolescents participated in the experiment (mean age sixteen). Audiences could watch the performances of these groups. Each group had to perform the task in two conditions, namely, an individual condition, where the participants were blindfolded, and a social condition, where the participants could see each other. Each group had to move in response to 6 pieces of music. Each piece lasted about 30 seconds. For a more detailed description of the experimental setup see [14, 15]. The acceleration data from the remote Wii sensors were registered wireless on-line on a laptop computer via the Bluetooth protocol in a PD patch and sampled at a 100Hz rate. Given the design of the experiment (16 groups, 4 participants per group, 6 musical excerpts, 2 conditions, 3 axes), this resulted in 2304 time series with length $N = 3000$ (30 seconds x 100 samples/second). In order to avoid the influence of hesitations and confusions at the beginning and the ending of the task, a 5-25 second interval of the time series was chosen for further analysis, instead of the recorded 0-30 seconds. The x, y, z dimensions of the accelerometer of the Wii sensor were further reduced to one single dimension, using the formula

$$a_{t_i} = \sqrt{(a_x^2(t_i) + a_y^2(t_i) + a_z^2(t_i))}$$

where a_{t_i} is the global value at time i , and $a_{x(t_i)}$ the acceleration value for dimension x . Inspection of the resulting time series revealed differences in

the range of the accelerations (strong and weak responses). As the occurrence in time of acceleration changes is of interest in this analysis rather than the intensity the amplitude of the calculated accelerations was rescaled to an [0,1] interval. Due to the definition of the total acceleration the minimum value of the series is 0 hence rescaling was based on the division of the values by the maximum value in the corresponding time series.

3 Analysis

Before analysis, the time series were tested for stationarity, as it is well known that this condition has a great influence on the stability of correlation coefficients [16]. The Unit Root test was used to investigate possible deviations from stationarity. It was found that this assumption could not be accepted in the majority of the MTS. Possible explanations for this observation are drift of the Wii sensors due to the end of the lifetime of the batteries or failing Bluetooth connectivity. Therefore, trend removal was used to obtain stationary MTS. Dynamic Time Warping (DTW) was then used in order to deal with small anticipations and delays in human movement [17]. In this study, we apply a multivariate DTW and a similarity measure based on the cost function. Constraints were introduced in order to speed up DTW calculations. A Sakoe-Chuba band [18] with a width of 100 was selected which accounts for a 2.5 % range. A cumulative distance matrix was then used to find the optimal path, by applying dynamic programming. DTW was calculated for all 6 possible combinations in the MTS (4x2001) series within each group. Fig.1 shows a fraction of the time series of 1 group (four subjects) before (left panel) and after warping (right panel).

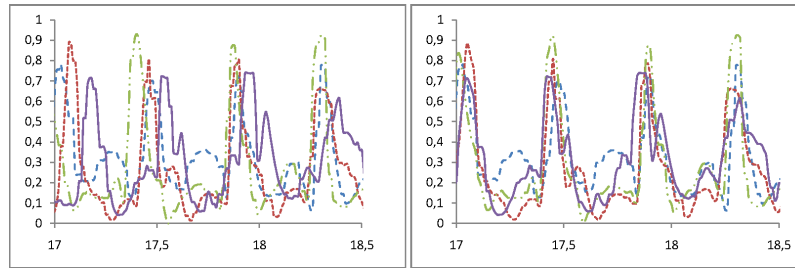


Fig. 1. Original (left) and corresponding warped (right) accelerations (Group1 Song1)

The warped MTS were then inspected for normality of the residuals of the differences between subjects within the group. It was found that, after warping, normality of the residuals could be accepted (Kolmogorov-Smirnov, $\alpha = 0.05$), which was not the case for the original series. The benefit of DTW was

investigated by comparing the cross correlations of the original and the corresponding warped series. In the example it can be seen that in the individual context Wii1, Wii3 and Wii4 move in a similar way whilst Wii2 shows a different pattern. Cross correlation on the original series show moderate to very low correlations with 2 non-significant values. The DTW data show higher and significant correlations in all inter subject combinations (Table 1). Each cell in this table represents the correlation (upper row) and the corresponding significance (bottom row) between the subjects. We define the similarity between subject movements as $S_{ij} = (1 - \text{Corr}_{ij})$ with S_{ij} between 0 and 1, low values indicating a high (closer related) inter subject synchronicity.

	Wii1	Wii2	Wii3	Wii4	Wii1	Wii2	Wii3	Wii4	Wii1	Wii2	Wii3	Wii4
Wii1	1	0.012 0.589	0.565 0.000	0.368 0.000	1	0.228 0.000	0.711 0.000	0.558 0.000	1	0.752 0.000	0.782 0.000	0.644 0.000
Wii2		1	0.030 0.182	-0.144 0.000		1	0.258 0.000	0.224 0.000		1	0.767 0.000	0.670 0.000
Wii3			1	0.434 0.000			1	0.592 0.000			1	0.613 0.000
Wii4				1				1				1

Table 1. Example of correlation and significance values in individual and social conditions

To obtain a measure for group coherence, correlations were averaged for each group, song and condition. A geometric representation of S_{ij} can be used as a tool to classify groups. The plot is constructed by positioning the 4 participants of a group so that the distances (length of the lines) are proportional to the similarities. In the example (Fig. 2, left), the subjects handling Wii1, Wii3 and Wii4 move synchronous, while the subject that is handling Wii2 shows a different pattern in the individual condition. However, in the social condition (Fig. 2, right) the coherence improves (Wii2 moves along with the other participants).

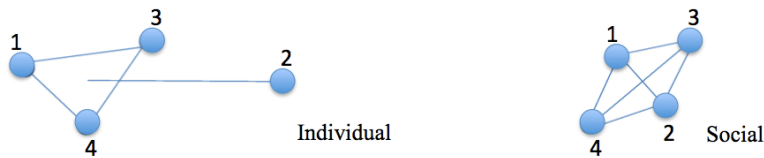


Fig. 2. Geometric representation of within group similarities

A plot of the obtained values (individual versus social) indicates a possible nonlinear trend (Fig. 3). The coherence of a group seems to be proportional

to the degree of difficulty of the song. For instance, song 4 was well known by the participants and had a clear beat. Even in the individual condition, subjects were able to synchronize very well with the music. Hence the improvement of the group coherence of the social condition was low in this case. On the other hand, song 3 was found difficult and unknown by the participants, resulting in a higher effect of the social condition on the group coherence. Univariate ANOVA analysis was used to investigate the effect of condition and song. Homogeneity of variance (Modified Levene, $\alpha = 0.05$) and normality (Kolmogorov-Smirnov, $\alpha = 0.05$) could be accepted.

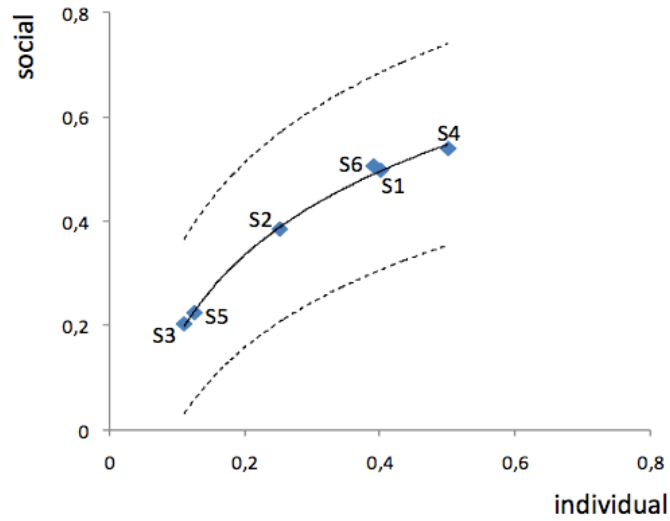


Fig. 3. Social versus Individual correlations (dashed lines represent 95 % interval)

The test indicates that the type of song was the dominant factor while the effect of condition is rather weak but significant. No interaction (condition x song) was observed. A Tukey analysis reveals that the songs can be grouped in 3 subsets (S3, S5), (S2) and (S1, S6, S4). Finally the DTW cost function was evaluated. Several cost functions have been proposed. For this study the total cost was based on the Euclidean distance of the corresponding warped (x_i, y_j) pairs. Univariate ANOVA was used to estimate the effect of song and condition on the DTW cost. It was found that the warp cost depends mainly on the song and that there is a small decrease in the social condition except for song 4, which has the lowest cost. In order to test the validity of the above method a separate experiment was set up. The subjects of this experiment were bachelor students of musicology (average age 22) who did the same experiment as the Accenta setup in the laboratory at IPEM. The group coherences were

calculated and compared with the results of the experiment. As only 3 groups were involved the results are only indicative but nevertheless they reveal some interesting information. First of all it can be seen that the social vs. individual coherences of the students are comparable with the results of the Accenta experiment (Fig. 4).

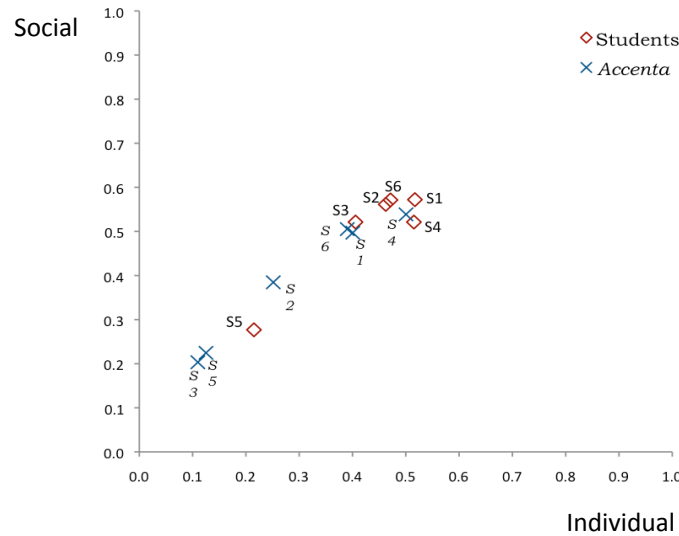


Fig. 4. Comparison of Accenta and student correlations

On the other hand, there are differences over the songs. Song 3 has much higher levels for subjects from musicology, than for the subjects of the Accenta experiment. This can be explained by taking into account that the students all have a musical education background and hence familiar with baroque music. Songs 1, 2, 4 and 6 all show a high coherence for the students indicating a possible effect of age. Song 5 shows an improvement but has low values when compared to the other songs. Song 5 had the most complex rhythm and was influenced by oriental elements. An effect of cultural background may be a possible explanation.

Alternatively, the human-music interaction was studied based on the amount of seconds the participants synchronized correctly with the nominal tempo of the music. This is calculated from the norm of the raw data for each block of two seconds by applying a fast fourier transform (FFT) over a 4-second moving window with a 2-second overlap. The dominant peak in the fourier transform is identified and compared with the nominal beats per minute (BPM) of the excerpt for deciding on the correctness of the synchronization. Also the half and double of the nominal BPM were considered as correct. For more detailed information about this method see [14, 15].

Based on the obtained scores, the impact of a social context on synchronization was studied using ANOVA analysis. Homogeneity of variances (Modified Levene test, $\alpha=0.05$) and normality (Kolmogorov-Smirnov, $\alpha=0.05$) could be accepted. Results show that synchronization results of the participants are significantly higher in the social condition compared to the individual condition (Anova, $\alpha=0.05$). The main effects are visualised in an interaction plot in Fig. 5.

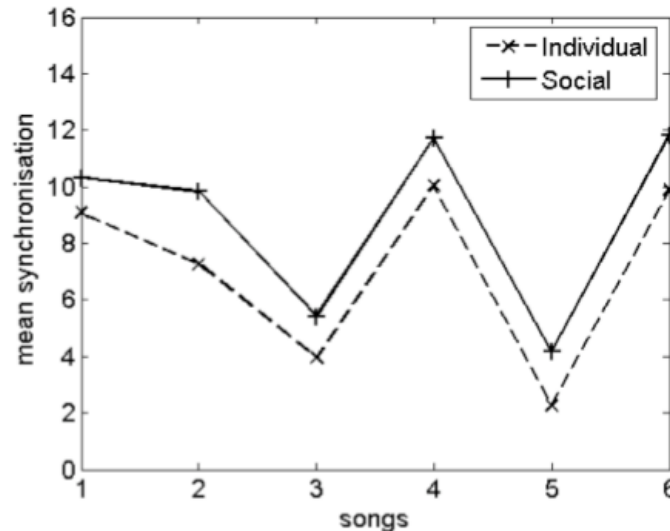


Fig. 5. Visualisation of the mean synchronization results per song in the individual and social condition

As can be seen in Fig. 5, the songs themselves have a great impact on the synchronization results. A multiple comparison Tukey analysis shows that participants score significantly lower for songs 3 and 5 than for songs 1, 4 and 6, while the results of song 2 are somewhere in between. This can be explained by the rhythmical complexity of the songs: songs 1, 4 and 6 are pop songs with a very clear beat, songs 3 and 5 can be interpreted either binary or ternary, whereas song 2 can only be interpreted binary but has an unclear beat.

4 Discussion

For the MTS data in the experiment presented here, in which subjects were asked to synchronize with the beat of the music, identical results are obtained with the analysis of human-human synchronisation based on DTW and the

analysis of human-music synchronisation based on FFT (Fast Fourier Transform) [14, 15]. This indicates the validity of the DTW-based method to analyze movements to music of multiple subjects, which can now be applied to the study of human-human interaction in more generic movements to music. Although the data were collected in an ecological setting with several unknown sources of variance, it was shown that the effect of song and condition can be quantified. The impact of the characteristics of the music is the predominant factor and this is in agreement with the model of musical communication on which this study is based. Humans can indeed decode the intentionality of the music and translate the energy input into movements as a function of musical content. Both in the individual and social context this can be quantified. In the social condition there is a benefit as a consequence of imitation effects during the social interaction. The results clearly show that group coherence improves when people move together. Whether this is only the result of direct human-human interaction or that other factors such as presence of public play a role, is not yet clear. The results of the group coherence measure are in good accordance with the analysis of the other variables derived from the collected data. Up to now it is not possible to separate the human-human interaction from the human-environment one in an ecological setting. In order to improve this type of experiment a new experimental design was proposed. In this design the subjects are not blindfolded in the individual condition, but separated using screens, and 10 songs with carefully selected properties are chosen. A drawback from the Accenta data was the lack of consistent information of the participants' background. The pre-survey could not be used for analysis due to unbalanced results and information about the experience of the subjects during the experiment (post-survey) was not available. In future experiments the use of properly designed surveys need to be included in the experiments. Large scale user studies for use of the analysis of human-music relationships has been proven to be of great importance [19]. Finally additional measurements such as video analysis and more sensors per subject are recommended in order to refine the analysis. On the analytical level improvement of the applied techniques and the use of new methods will be investigated. At this moment Correlation optimized DTW based on PCA analysis [20] and quantification of complexity and determinism of the movement data are tested [21]. Another important issue is the reduction of calculation time of DTW. Several methods will be tested in the nearby future [22].

5 Conclusion

In this study different statistical techniques were tested for the analysis of human movements to music. Using DTW in combination with CCF and ANOVA it was found that the type of music is the dominant factor of the inter group movements as a response to music stimuli. The effect of condition is low but significant, even in the ecological setting of the experiment. Using DTW and

CCF it is possible to quantify interactions and to classify groups by coherence. The outcome of this study enables also to define a statistical path as a tool for researchers and as a guideline for appropriate experimental designs for future research.

6 Acknowledgements

This research has been conducted in the framework of the MEFEMCO (Methodological foundations of embodied music cognition) project (2008-2011) with support of the Fund for Scientific Research of Flanders (FWO), and the Emcomettecca (Embodied music cognition and mediation technology for creative and cultural applications) project, Methusalem -BOF Ghent University.

References

1. Nayak S. and Wheeler B. and Shiflett S. and Agostinelli S. Effect of Music Therapy on Mood and Social Interaction Among Individuals With Acute Traumatic Brain Injury and Stroke. *Rehabilitation Psychology*, 45(3): p. 274-283. Educational Publishing Foundation, 2000.
2. Martin AJ Motivation and Engagement in Music and Sport: Testing a Multidimensional Framework in Diverse Performance Settings. *Journal of Personality*, 76(1): p. 135-172, 2008.
3. Buldyrev, SV and Goldberger, AL and Havlin, S. and Mantegna, RN and Malsa, ME and Peng, CK and Simons, M. and Stanley, HE Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. *J. Physical Review E*, 51(5): p. 5084-5091, 1995.
4. Machulda, MM and Ward, HA and Borowski, B. and Gunter, JL and Cha, RH and O'Brien, PC and Petersen, RC and Boeve, BF and Knopman, D. and Tang-Wai, DF and others Comparison of memory fMRI response among normal, MCI, and Alzheimer's patients. *Neurology*, 61(4): p. 500-506, 2003.
5. Castellano, G. and Bresin, R. and Camurri, A. and Volpe, G. User-centered control of audio and visual expressive feedback by full-body movements. *Proc. of ACHI*, 2008.
6. Bernhardt, D. and Robinson, P. Interactive control of music using emotional body expressions. , 2008.
7. Thaut, MH, McIntosh, GC, Rice, RR, Miller, RA, Rathbun, J. and Brault, JM Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Movement Disorders.*, 11(2): p. 193-200, 1996.
8. Stergiou, N. Considerations of Movement Variability in Biomechanics Research. *in Innovative Analyses of Human Movement.*, p. 29-62, Human Kinetics, 2004.
9. Leman, M. Embodied Music Cognition and Mediation Technology , Mit Press, 2007.
10. Leman, M. and Desmet, F. and Styns, F. and Van Noorden, L. and Moelants, D. Embodied listening performances reveal relationships between movements of player and listeners. *Enactive/074th International Conference on Enactive Interfaces*, p. 19-24, 2007.

11. Toiviainen, P. and Snyder, JS Tapping to Bach: Resonance-Based Modeling of Pulse. *Music Perception*, 21(1): p. 43-80, UnivCalifornia Press, 2003.
12. Boone, RT and Cunningham, JG Children's expression of emotional meaning in music through expressive body movement. *Journal of nonverbal behavior*, 25(1): p. 21-42, 2001.
13. Clayton, M. and Sager R. and Will U. In time with the music: The concept of entrainment and its significance for ethnomusicology. *ESEM Counterpoint*, 1: p. 1-82, 2004.
14. De Bruyn, L. and Leman, M. and Moelants, D. and Demey, M. and Desmet, F. Measuring and Quantifying the Impact of Social Interaction on Listeners Movement to Music. *Proceedings of the 5th International Symposium on Computer Music Modeling and Retrieval*, p. 298-305, 2008.
15. Demey, M. and Leman, M. and Bossuyt, F. and Vanfleteren, J.. The Musical Synchrotron: using wireless motion sensors to study how social interaction affects synchronization with musical tempo. *Proceedings of the 8th International Conference on New Interfaces for Musical Expression.*, 2008.
16. Yang, K. and Shahabi C. On the Stationarity of Multivariate Time Series for Correlation-Based Data Analysis. *Proceedings of the Fifth IEEE International Conference on Data Mining*, p. 805-808, 2005.
17. Parsons, TW Voice and Speech Processing. *McGraw-Hill College*, 1987
18. Sakoe, H. and Chiba S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Signal Processing*, 26(1): p. 43-49, 1978
19. Lesaffre, M. and De Voogdt, L. and Leman, M. and Baets, BD and Meyer, HD and Martens, J.P. How potential users of music search and retrieval systems describe the semantic quality of music. *Journal of the American Society for Information Science and Technology*, 59(5): p. 695-707, 2008
20. Tomasi, G. and Berg F. and Andersson C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18: p. 231-241, 2004
21. Sarkar, A. and Barat P. Multiscale Entropy Analysis: A New Method to Detect Determinism in a Time Series. *Arxiv preprint physics*, 0604040, 2006.
22. Dixon, S. An on-line time warping algorithm for tracking musical performances. *Proceedings of the International Joint Conference on Artificial Intelligence*, 2005.