Multimodal analysis of synchronization data from patients with dementia

Frank Desmet[*1], Micheline Lesaffre[*2], Joren Six[*] , Nathalie Ehrlé[#], Séverine Samson[+]

IPEM, Dept. of Arts, Music and Theatre Sciences, Ghent University Belgium
#Centre Hospitalier Universitaire de Reims, France
+ PSITEC, Université de Lille, UFR de Psychologie, Université de Lille-Nord de France, Villeneuve d'Ascq, France

[1]fm.desmet@ugent.be, [2]micheline.lesaffre@ugent.be

ABSTRACT

Little is known about the abilities of people with dementia to synchronize bodily movements to music. The lack of non-intrusive tools that do not hinder patients, and the absence of appropriate analysis methods may explain why such investigations remain challenging. This paper discusses the development of an analysis framework for processing sensorimotor synchronization data obtained from multiple measuring devices. The data was collected during an explorative study, carried out at the University Hospital of Reims (F), involving 16 individuals with dementia. The study aimed at testing new methods and measurement tools developed to investigate sensorimotor synchronization capacities in people with dementia. An analysis framework was established for the extraction of quantity of motion and synchronization parameters from the multimodal dataset composed of sensor, audio, and video data. A user-friendly monitoring tool and analysis framework has been established and tested that holds potential to respond to the needs of complex movement data handling. The study enabled improving of the hardware and software robustness. It provides a strong framework for future experiments involving people with dementia interacting with music.

I.   INTRODUCTION

A. Background

In the last decade, serious efforts have been made to determine the relationship between music and health and wellbeing. Lesaffre (2013) gives an overview of the challenges that arise from using new technologies and developing new methods when working in a domain that is unfamiliar. Especially interesting are the new possibilities that evolve from working with the embodied music interaction paradigm. This is particularly the case for specific target populations such as people with dementia. It has been argued that the use of reliable monitoring technology in a proper music interaction context may be beneficial for people with dementia (Lesaffre, 2017). Indeed, music is known to be useful in contexts where people have difficulties with verbal and emotional communication. Music stimulates the brain's reward centres while bodily movement activates its sensory and motor circuits. Strong links between music and motor functions suggest for example that sensorimotor synchronization to music could be an interesting aid for motor learning (Moussard, Bigand, Belleville, & Peretz, 2014). However, in view of the development of musical interventions in dementia, the rigorous methodological standards required are not always met (Samson, Clément, & Narme, 2015). This can partly be explained by the lack of custom tools that can support evidence-based research. Therefore, there is a need to develop monitoring and analysis tools that enable validating the efficacy of involving music interaction for the benefit of people with dementia.

B. Sensorimotor synchronization to music

Humans are known to have an advanced ability to synchronize movements (e.g. steps, hand and finger taps) to an external rhythm. In Repp and Su (2013) sensorimotor synchronisation (SMS) is defined as the coordination of rhythmic movement with an external rhythm. Repp and Su surveyed research in the field comprising conventional tapping studies, dance, ensemble performance, and the neuroscience of SMS. The ability to synchronise is considered as cognitively demanding (Bläsing, Calvo-Merino, & Cross, 2012; Dhami, Moreno, & DeSouza, 2014), and temporal regularities in music can entrain cognitive attentional resources (Jones & Boltz, 1989; Large & Jones, 1999). SMS is also considered to have a positive social and emotional significance (e.g. Wiltermuth & Heath, 2009).

Moreover, musical synchronization has proven effective in the rehabilitation of physical and social-emotional clinical disorders such as Parkinson's disease (Nombela, Hughes, & Owen, 2013). Fundamental studies on SMS found indications that physical strength and spatial references are important contributing factors (Leman, Moelants & Varewyck, 2013). However, due to the lack of appropriate tools for non-intrusive and objective measurement, there is hardly any evidence-based understanding of motoric, expressive and empathic responding to music of people with dementia.

C. Tools

In general, the measurement and analysis of human movements in a musical context is inherently multi-disciplinary and is a real challenge in science today. This multi-disciplinary research requires a methodology that combines both a bottom-up and a top-down approach. The bottom-up approach is concerned with the observation of body movement, which is based on sensing technologies. The top-down approach is concerned with the identification of music related and non-music related actions, based on the observer's interpretation in combination with survey data related to the participants.

Measuring movement in people with dementia is even less straightforward, because attaching sensors or other measurement devices to their bodies might elicit stress and

fear. Furthermore, capturing information about body movement requires a data acquisition system (a) that can measure, sample, and digitize physical properties; (b) that can send that information to a computer for further processing; and above all, (c) that is not invasive and usable in an ecological setting.

To meet these requirements a force plate system was developed aiming at providing a balance between functionality for the patient, sensitivity for measuring smaller movements, and data reliability. The system consists of a square wooden force plate (90 x 90 cm), mounted on a frame containing four weight sensors, one in each corner. The four calibrated sensors provide weight values and are read out individually by Arduino Due, an open- source prototyping platform, which can be used to calculate movement direction and quantity of movement. A software program was developed that can read multiple force plates simultaneously. The system was tested for the first time in a study that investigates spontaneous movement response to music in people with dementia. A detailed description is provided in Lesaffre, Moens, and Desmet (2017).

There is hardly any research that investigates sensorimotor synchronization to music in people with dementia. Experimental research on synchronization abilities typically collects information from different types of instruments, measurement techniques, and experimental setups. The increasing availability of several acquisition tools generates complex datasets, which are not easy to handle and therefore require new analysis frameworks.

This paper describes the analysis framework developed for processing sensorimotor synchronization data obtained from multiple measuring devices, such as a pressure sensor data, audio and video recording. The data was collected during an explorative study, carried out at the University Hospital of Reims (F), involving individuals with dementia.

II. EXPERIMENT

A. Participants

16 participants (13 female and 3 male; range 79 - 94 years; mean MMSE = 14,21) were recruited for this study. The study was carried out in accordance with the approved guidelines of the University Hospital of Reims. Each subject provided informed consent prior to participation.

B. Experiment design

Synchronisation was tested in the following conditions: (1) in the presence of a musical beat vs. a familiar song of the same tempo, (2) under auditory, visual and audio-visual conditions, and (3) in live vs. recording conditions (see Table 1). The 9 conditions were presented in randomised order. Throughout the experiment patients were encouraged to tap along with the music or performer.

**Table 1. Conditions used in the experiment.**

Auditory & visual	Auditory	Visual
Tapping & Pulse (video)	Pulse (audio only)	Tapping (video)
Tapping & Pulse (live)	--	Tapping (live)
Song (video & audio)	Song (instrumental)	--
Song (live & audio)	Song (audio recording of live performance)	--

Patients were encouraged to tap along with the music or performer. Each session took about one hour, including picking up and taking back the patient to his or her room.

The participants were exposed to a cheerful familiar song (Ah! Le petit vin blanc), a musette waltz with a tempo of 84 bpm. This tempo is in agreement with the spontaneous motor tempo of normal persons of the similar age, as described by McAuley, Jones and Holub (2006).

The experimental setup consisted of two force plates developed at Ghent University (IPEM), each with a chair mounted on it, one for the patient and one for the performer (see Figure 1). The two boards were placed in front of each other. The force plates have each four sensors (one in each corner) to measure the movement of the person sitting in the chair. A small table is mounted on each chair providing a comfortable position for tapping along with the music. Below this table was a microphone so that the tapping of a participant was measured as an audio signal. Two webcams were placed so that from both balance boards a video was recorded during the experiment. Furthermore, a projection screen with a projector behind it was placed at the backside of the experimenter board in order to enable pre-recorded video projections needed for the video conditions of the experiment.
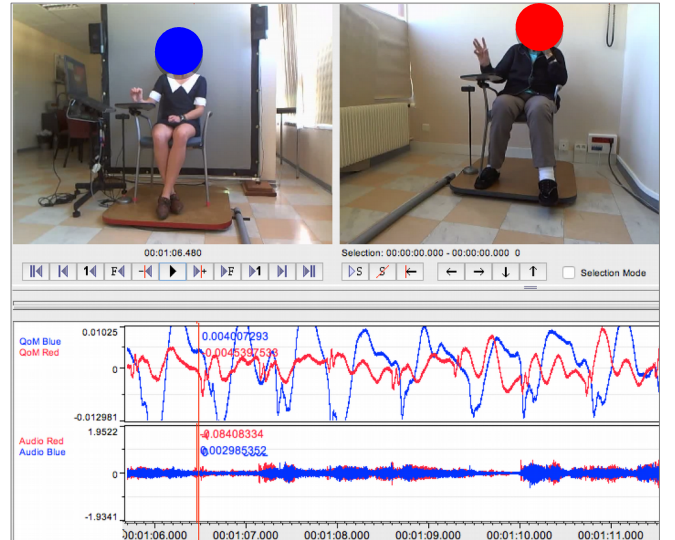
Figure 1. Example of the analysis framework in ELAN, representing video of performer (top left) and subject (top right), normalized intensity of quantity of movement time- series of subject (red) and performer (blue); and audio recording of the tapping.

III. ANALYSIS FRAMEWORK

The aim of the analysis framework was to develop a method for the analysis of the tapping time series in different conditions; to evaluate the hardware and software robustness, and experimental setup; and to formulate recommendations for future experiments.

A. Data considerations

Using a hardware setup based on the two force plates developed at IPEM (see supra), in combination with webcam

video capture and audio recording of tapping, audio time series in different conditions were measured. After a data control and repeated viewing of the video and audio data four participants had to be excluded from the analysis due to incomplete data, for example by lack of response and missing conditions.

Data were recorded at 8000 Hz. The obtained data structure consisted of two csv files containing the force plate sensor data and the audio of the tapping, two audio signals (wav) of the tapping and two mp4 video files from the webcam.

The SyncSink application (Six & Leman, 2015) was used to synchronize the data. After synchronization an initial ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) structure for each participant/condition combination was setup in order to enable data inspection and determination of the start and stop time for analysis. Every ELAN structure contains the video of the performer (with audio), the video of the participant (without audio) and the audio recordings of the tapping (see Figure 1).

Finally, the csv files were imported in Matlab (Mathworks, 2014) using a toolbox developed at IPEM. In each participant/condition data structure a Matlab folder was added containing all results from Matlab calculations. By doing so data integrity was assured.

B. Analysis method

The aim of the analysis was to determine the amount of movement, the regularity of the tapping, and the synchronization of the tapping with the external tapping source of the condition.

To begin with, the sensor data were trimmed to new start and stop time positions obtained by inspecting the ELAN structures. Trimming was necessary due to additional unwanted noise at the beginning and the end of the experiment and to select the parts of the time series where actual tapping occurred.

Calculation of Quantity of Motion (QoM)

The four sensor outputs of the force plate were used to determine QoM. The method used was as follows: First, Cartesian coordinates [(1),(2)] were calculated so that the center of the balance board equals [0,0] and the positions of the sensors are at [-0.5,0.5], [0.5,0.5], [0.5,-0.5] and [-0.5,-0.5] (see Figure 2).

Figure 2. (a) Force plate layout and definition of coordinates, (b) polar plot example

$$x_i = \left(\frac{S1_i + S4_i}{\sum\limits_{s=1}^{4} S_{si}} \right) + 0.5 = \left(\frac{S2_i + S3_i}{\sum\limits_{s=1}^{4} S_{si}} \right) - 0.5 \quad (1)$$

$$y_i = \left(\frac{S1_i + S2_i}{\sum\limits_{s=1}^{4} S_{si}} \right) + 0.5 = \left(\frac{S3_i + S4_i}{\sum\limits_{s=1}^{4} S_{si}} \right) - 0.5 \quad (2)$$

(a)   (b)
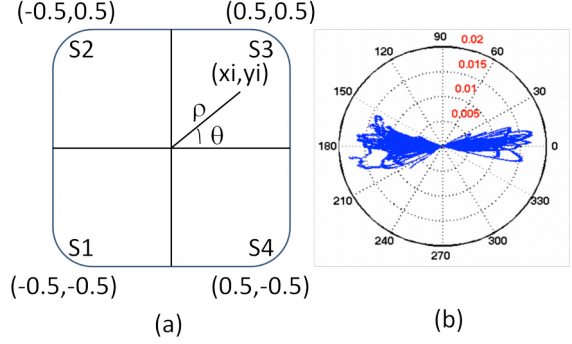
(Where s is the sensor and i is the result at point i in the time series.)

The resulting Cartesian coordinates are then transformed to polar coordinates [(3),(4)]:

$$\rho = \sqrt{x_i^2 + y_i^2} \quad (3)$$

$$\theta = atan2\left(x_i, y_i\right) \quad (4)$$

Second, the polar coordinates are translated so that the point of gravity is [0,0]. The latter is done to enable comparisons between subjects by making the recorded movement independent from the point of gravity of the participants.

Finally, the total QoM (5) is calculated as follows:

$$QoM = \frac{\sum\limits_{i=1}^{N} \rho_i}{T} \quad (5)$$

(Where T is the duration of the time series)

The division by T makes the results independent of the duration of the time series, so that the values of QoM can be compared between subjects/conditions.
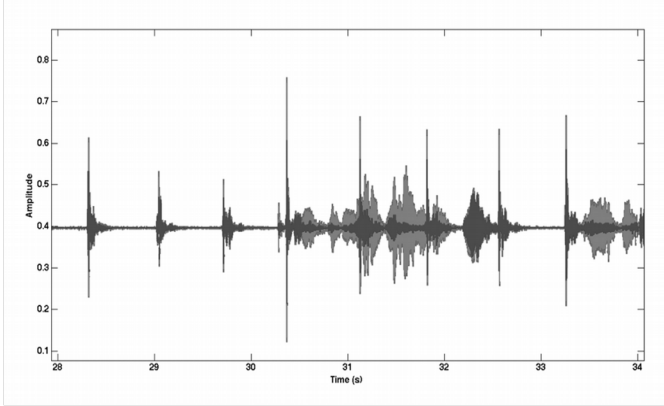
Peak detection of taps

The tapping was recorded by means of a microphone mounted under the table. The microphone recorded not only the taps but also any other sound (e.g. talking, music, metronome and surrounding noise) that was present during the experiment.

Therefore, it was challenging to detect peaks in the complex mixed signals. Furthermore the sampling frequency was high (8000 Hz), so that the original signal was blurred with shoulders, spikes etc. Figure 3 shows an example of a typical signal in this dataset with indication of the parameters used in the peak detection algorithm (O'Haver, 1997).
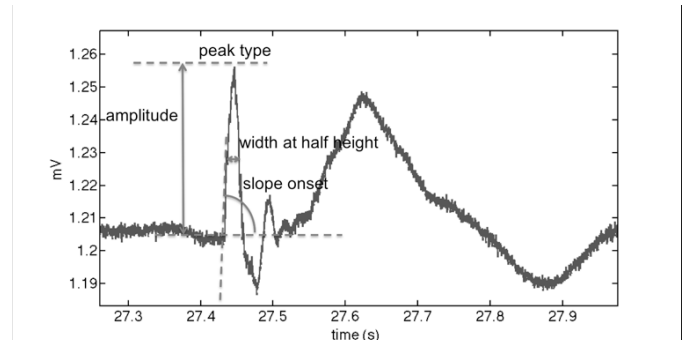
Figure 3. Example of peaks in the obtained signals. Detection is based on amplitude, slope onset, width at half high, and peak type.

Before the peak detection algorithm could be applied it was necessary to clean up the complex signals. First, the noise had to be removed. Tapping peaks occurred as wavelets in the signal so that wavelet decomposition based on the Hilbert transform was used to filter the data (Hahn, 1996). Second the data were filtered and smoothed using a Savitzky-Golay filter (Savitzky & Golay, 1964). An example of this process is shown in Figure 4.

Figure 4. Example of wavelet decomposition and Savitzky-Golay filtering. The gray signal is the original series, the black signal is the resulting filtered and smoothed series.

Peak detection was based on the Matlab PeakFinder toolbox (O'Haver, 2009). Here peak detection depends on amplitude, slope onset, width at half high, and peak type (eg. Gaussian). An initial set of these threshold parameters was used as seed. The 4 parameters were then manually adapted to obtain correct peak detection.

Inter Tap Interval (ITI) variability

The first variable calculated from the detected peak intervals is the variability of the intra-subject tapping sequences. The standard deviation of inter-response intervals for a trial was calculated by measuring the mean inter-response interval and by then expressing the variability of intervals around that mean in terms of the standard deviation (SD_{ITI}).

$$ITI = \sum_{i=1}^{N-1} \frac{\lfloor t_{i+1} - t_i \rfloor}{N-1} \tag{5}$$

$$SD_{ITI} = \sqrt{\frac{1}{N}\sum_{i=1}^{N-1}\left(t_i - ITI\right)^2} \tag{6}$$

SD_{ITI} provides information on how regular a participant tapped. Histogram analysis was used to detect outliers. The presence of outliers due to high ITI values is the result of a participant who interrupted tapping during the task.

Sensorimotor Synchronization (SMS)

In this experiment the external rhythm sources depended on the conditions. This means that for each condition the modeling had to be adapted (Elliott, Chua, & Wing, 2016). The event asynchronies are used to calculate the synchronization.

Figure 5 shows the different intervals used to calculate the synchronization.
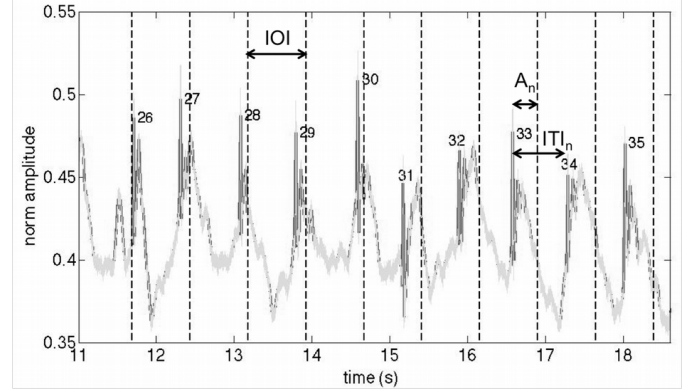
Figure 5. Example of tapping time series and definition of intervals. IOI is the inter onset interval of the external rhythm source, ITI$_n$ is the nth Inter Tap Interval and A$_n$ is the nth event asynchrony.

$$SD_{asy} = \sqrt{\frac{1}{N}\sum_{i=1}^{N-1}\overline{\left(A_i - \acute{A}\right)}} \tag{7}$$

IV RESULTS

QoM was calculated for all selected participants (N = 12) and for all conditions. The results show that most tapping was performed in the condition with live singing performance. The least tapping occurred in the pulse conditions without live performance.

In order to evaluate the hardware, software and the analytical method 4 participants in 3 conditions were selected for a full analysis. The conditions were live tapping of the performer to a pulse (condition 2), pre-recorded video of the performer hand tapping and singing (condition 3), live performance involving hand tapping and singing (condition 4). Table 2 shows the results of the analysis.

Table2. Summary of results

ID	Condition	QoM (mV)	ITI	SD$_{asy}$ (mV)
22	2	20.10	1.35	234
22	3	16.64	1.21	498
22	4	31.74	0.98	369
23	2	24.74	0.92	269
23	3	32.95	1.42	375
23	4	45.02	1.12	296
24	2	5.62	1.23	245
24	3	5.15	1.14	258
24	4	4.57	1.13	269

32	2	19.35	1.21	452
32	3	27.77	1.09	398
32	4	39.40	1.05	475

V DISCUSSION AND CONCLUSION

The aim of this exploratory study was testing new monitoring tools and developing an analysis framework for multimodal synchronization data.

Experimental design

Due to several issues related to the experimental design, such as involving patients with varying types of dementia and degrees of cognitive impairment (MMSE range 3-23), a complete analysis of all participant/condition combinations could not be done. Some participants did not move at all or tapped only during short times. It was observed the ability to bodily respond is related to the degree of cognitive impairment, suggesting that in future experiments on synchronization patients with severe cognitive impairment should not be included.

Hardware and software

In general the newly developed hardware was stable during the whole experiment. Only in 2 cases there was a drift of a sensor and in 1 case a sensor failed. The main drawback of the hardware is the recording of the taps by means of microphones. The microphones register a lot of noise from different sources during the experiment. This results in time series, which are difficult to synchronize and to analyze. A further refinement is to calibrate the sensor output in order to obtain meaningful values for QoM instead of millivolt (mV).

Meanwhile, an adapted system has been developed for follow-up experiments. The tapping is now registered using a sensor mounted in the table, and additional sensors are added in a footplate on top of the force plate to register the leg and feet movements. Furthermore a new device is added to the system generating 4 beeps at the start of the experiment. The beeps are simultaneous recorded by the webcams and by the interface, which records the sensor data. This adaptation makes it possible to generate synchronized files automatically.

The software performed well. However, a workaround had to be established because there was no direct connection with a server available at the location of the experiment so that the data were not accessible at other (distant) locations. An ftp server was used that caused very long download times.

The original mp4 video files from the webcams were not compatible with ELAN software. Conversion to an mp4 format suitable for ELAN was possible using ffmpeg but in the resulting mp4 files the audio was out of sync. A solution was found by conversing mp4 to mov (using ffmpeg).

The SyncSink application worked mostly fine. Some difficulties to synchronize occurred when the tapping was very weak and a lot of audio noise was present in the data.

Methodology

Up till now, the analysis framework enables to calculate QoM, ITI and SD_{asy}. The bottleneck in this exploratory study was the peak detection. Detecting peaks in complex mixed signals is not straightforward. In some cases the peak detection failed due to very soft tapping in combination with a lot of background noise. In other cases the audio from the tapping could not be synchronized with the audio from the videos. Selecting the appropriate thresholds for wavelet decomposition and Savitzky-Golay filtering is still a trial and error procedure.

Adding two new variables could enhance the methodology. First, the number of actual taps of a participant compared to the maximum number of taps in the reference source could be added. The percentage of actual taps compared to the maximum number of taps is then a measure of tapping activity. Second, following Stergiou nonlinear dynamics (chaotic structures) could be added to the analysis (Stergiou & Decker, 2011). Apart from this, Approximate Entropy (ApEn) and/or Sampling Entropy (SampEn) of the tapping time series could be an added value to the classical variance analysis.

Taking into account these findings, a follow-up experiment has been designed that focuses on synchronization abilities in people with dementia (see Ghilain, Schiatura, Lesaffre, Desmet et al, 2017).

REFERENCES

Bläsing, B., Calvo-Merino, B., Cross, E. S., Jola, C., Honisch, J., & Stevens, C. J. (2012). Neurocognitive control in dance perception and performance. *Acta psychologica, 139*(2), 300-308.

Dhami, P., Moreno, S., & DeSouza, J. F. (2014). New framework for rehabilitation–fusion of cognitive and physical rehabilitation: the hope for dancing. *Frontiers in psychology*, 5, 1478.

Elliott, M. T., Chua, W. L., & Wing, A. M. (2016). Modelling single-person and multi-person event-based synchronisation. *Current Opinion in Behavioral Sciences, 8*, 167-174.

Hahn, S. L. (1996). *Hilbert transforms in signal processing*. Boston: Artech House inc.

Ghilain, M., Schiaratura, L., Lesaffre, M., Desmet, F. & Samson, S. (2017). Music and movement synchronization in people with dementia. *Proceedings of the 25th anniversary Conference of the European Society of the Cognitive Sciences of Music.*

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological review, 96*(3), 459.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological review, 106*(1), 119.

Leman, M., Moelants, D., Varewyck, M., Styns, F., van Noorden, L., & Martens J.-P. (2013). Activating and Relaxing Music Entrains the Speed of Beat Synchronized Walking. *PLoS One, 8*(7).

Lesaffre, M. (2013) The power of music and movement to reinforce well-being. in: Lesaffre, M. & Leman,M (Eds.) *The power of music. Researching musical experiences.* Leuven: ACCO Academic.

Lesaffre, M. (2017). Investigating embodied music cognition for health and well-being. From theory to therapeutic approaches. In R. Bader. (Ed.). *Springer Handbook of Systematic Musicology.* Berlin, Germany: Springer.

Lesaffre, M., Moens, B., & Desmet, F. (2017). Monitoring music and movement interaction in people with dementia. In M. Lesaffre, P.-J. Maes & M. Leman (eds.), *The Routledge Companion to Embodied Music Interaction.* New York: Routledge.

Nombela, C., Hughes, L. E., Owen, A. M., & Grahn, J. A. (2013). Into the groove: can rhythm influence Parkinson's disease? *Neuroscience & Biobehavioral Reviews*, *37*(10), 2564-2570.

Mathworks, I. (2014). MATLAB: R2014a. *Mathworks Inc, Natick*.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*(3), 348.

Moussard, A., Bigand, E., Belleville, S., and Peretz, I. (2014). Music as a mnemonic to learn gesture sequences in normal aging and Alzheimer's disease. *Frontiers in Human Neuroscience*. 8:294

O'Haver, T. (1997). A Pragmatic Introduction to Signal Processing. Retrieved from https://terpconnect.umd.edu/~toh/spectrum/index.html

O'Haver, T. (2009). Peak finding and measurement. https://terpconnect.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic Bulletin & Review, 20*(3), 403-452

Samson, S., Clément, S., Narme, P., Schiaratura, L., & Ehrlé, N. (2015). Efficacy of musical interventions in dementia: methodological requirements of nonpharmacological trials. *Annals of the New York Academy of Sciences, 1337*(1), 249-255.

Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry, 36*(8), 1627-1639.

Six, J. & Leman, M. (2015) Synchronizing multimodal recordings using audio-to-audio alignment. *Journal of Multimodal User Interfaces* 9:223. doi:10.1007/s12193-015-0196-1

Stergiou, N., & Decker, L. M. (2011). Human movement variability, nonlinear dynamics, and pathology: is there a connection? *Human movement science, 30*(5), 869-888 %@ 0167-9457.

Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal, 2006*(146), 10.

Wiltermuth, S. S. & Heath C., (2009). Synchrony and cooperation. *Psychological Science*. 20, 1–5.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research.* In Proceedings of LREC 2006 (pp. 1556– 1559). Genova, Italy.