MOTION CAPTURE DATA AS MACHINE-READABLE NOTATION TO CAPTURE MUSICAL INTERPRETATION: EXPERIMENTING WITH MOVEMENT SONIFICATION AND SYNTHESIS

Julien Mercier
Media Engineering Institute (MEI),
School of Engineering
and Management Vaud, HES-SO
julien.mercier@hes-so.ch

Irini Kalaitzidi
Department of Computational Arts,
Goldsmiths University of London
eri.kalaitzidi@gmail.com

ABSTRACT

Musical notation can be described as an abstract language that composers use so that performers may interprets a score. Such notation comes before interpretation, is reproducible, and although it contains hints targeted at the performer on how to interpret, the actual performance is uniquely situated in time and space. The only way to record and re-experience a sound performance is to use microphones, which transform acoustic waves into an electrical signal and usually lose at least some spatial dimension in the process. This may be hindering in the field of experimental music, where physical limits of sound material may be put to the test. In this short paper, we discuss how motion capture could be an alternative to, or an expansion of the acoustic recording of a performance involving movement. By recording the performer's movements, some of the dimensions that make their interpretation singular (i.e., character, accentuation, phrasing, and nuance) are retained. A method capturing sound through movement may be interesting in the context of sound synthesis with deep learning and hold potential advantages over current methods using MIDI or acoustic, which either lack dimensions or are very sensitive to noisy data. We briefly discuss rational, practical, and theoretical foundations for the development of potentially innovative outputs.

1. INTRODUCTION

Traditionally, musical notation serves as a common, abstract language for musical composition that doesn't specify all aspects of musical performance. In a second time, various performers must seize this coded language and make various choices during the interpretation phase. It's a vector between different humans involved in the creation and interpretation of sound. The actual performance, which is situated in time and space, cannot be notated in such an abstract way; it can merely be recorded as an acoustic output: The acoustic waves are converted into an electrical signal by a transducer before it is sampled in order to be encoded digitally. During live performances, many of the spatial qualities of sound are usually lost in the process, because microphones typically capture sound stemming from various sources and merge them. In a studio setting, this is avoided by recording every instrument separately. Altogether, the acoustic recording process is currently the only format through which a performance and the interpretation it embodies are captured in order to be replicated afterwards.

2. MOVEMENT SONIFICATION WITH MOTION CAPTURE

Current advances in computer vision algorithms make it possible to capture and record the motion of a performer (whether hands, face, or whole body) in real time with great levels of accuracy, thus capturing different dimensions than those recorded by a microphone. By doing so, it is possible to capture the source impulse that yields the distinctiveness of a performance rather than its mere output, as is the case with acoustic recording. As early as 1938, Alexander Truslit's [1] used a rudimentary motion capture system to elaborate his theory on the gestural quality of musical interpretations. He researched the relationship between the motion of music and the perceptual processes of the audience. Wöllner [2] investigated Truslit's hypothesis by recording and comparing free movements and instructed movements and asking participants to determine which sound they had produced in a self-other recognition task. Motion recording has the potential to capture some of the subtle and sensory dimensions that make it an interpretation singular: character (the general hue given to the expression of a piece), accentuation (Agogik), phrasing, and nuance. It results in a type of numerical notation that can only be read by machines in order to re-enact a performance at a later stage. As such, capturing the performer's movement represents an interesting alternative method to acoustical recording for the technical repeatability of a performance. In fact, it may prove a particularly fitting method in some cases. It can be used for the automatic notation of improvised music without the need for post-hoc transcription, which usually doesn't do justice to the creativity of the exercise. If the instruments used by the performers are digital, then every unit of a similar model is identical, and the parameters used can be reproduced.

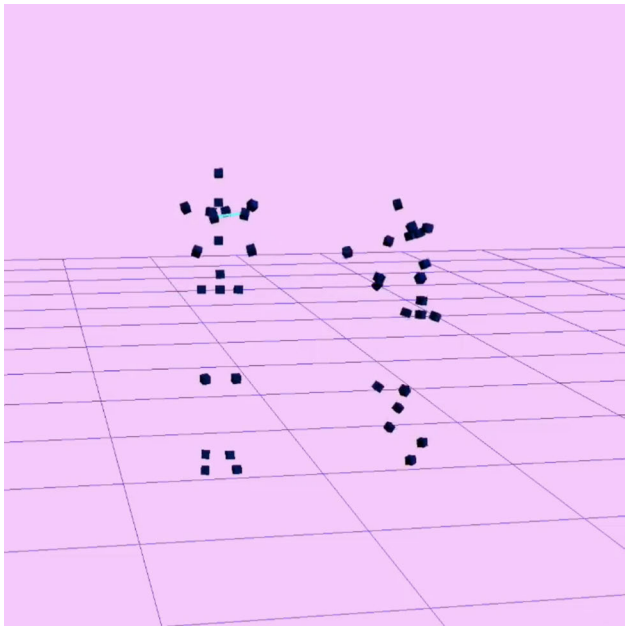

Figure 1. Once recorded by a MaxMSP program, the performers' movements can be played again and sonified with the same synthesizer. Provided a similar setup, there is no difference between the sounds produced during the original performance or its re-enactment.

The conditions under which the performer's movements are turned into sound (including the spatial setup) can be faithfully reproduced across different places and times. In the case of experimental music, the spatial qualities of the sound are critical to ensuring proper perception, and acoustic recording systems can sometimes reach their physical limits. In specific cases where digital musical instruments can be parameterized in absolutely identical ways, recording a performance through motion capture may be equivalent to "higher-dimensional MIDI data" and enable a more naturalistic reproduction of the sound produced by the performance. Each instrument has its own specificity, and an approach based on motion recording may prove interesting in various situations.

3. USE CASE: DANCE IMPROVISATION SONIFICATION

The use of motion capture in combination with dance and movement sonification is the object of multi-disciplinary research and encounters a variety of applications in data sonification and gesture interfaces, and it is observed through frameworks such as affective computing [3]. Movement sonification fosters new perspectives for practice-based artistic research [4], and it is an effective setup to achieve emotion-to-sound translations and is investigated by composers and sound designers to communicate affective information [5]. Movement sonification is often used to expand motion natural acoustics, and there are many ways to transform movement data into sound [6]. Real-time movement sonification has also been used in the field of gestural rehabilitation [7].

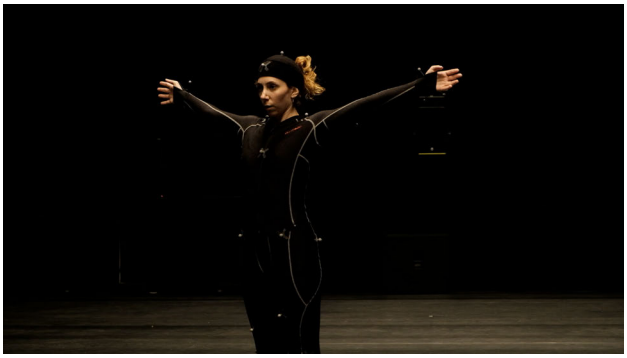

Figure 2. Screen capture of the performer wearing the motion capture suit and improvising with real-time auditory feedback. A video of the setup is visible at the following link: vimeo.com/783283941

We experimented with motion capture and movement sonification during an artistic research residency at the Institute for Computer Music and Sound Technology in 2020. We created a series of simple synthesizers with MaxMSP controlled by body motion through the use of a large motion capture system. The distance between the performer's hands and feet controlled simple sine-wave oscillators, while their X and Y position in the room controlled additional granular synthesizers. We recorded sessions during which a dancer performed various improvised choreographies intuitively, guided only by real-time sonification feedback. This type of auditory feedback has a positive influence on perception and motor movement, such as weight distribution, joint angles, and jumps are recognized through sound. By practicing, someone can determine through sound whether a complex movement was correctly executed. [8]. Our synthesizers were designed to turn specific types of movement into simple sounds. As such, re-enacting the performance by re-playing the motion data sequence produces rigorously similar outputs as the actual performance did, if the track is played on similar hardware, somehow inverting the notational process: from the performer to the composer. The motion capture data kept idiosyncrasies, which in this context are an operationalization of the concept of interpretation. Under different circumstances, the method could possibly be used to capture the hand movements of a pianist [9] and encapsulate some usually hidden dimensions of interpretation in this particular type of machine notation. However, the gestures we experimented with have a particular type of expressiveness, and the method might be only suited to the context of emerging virtually-controlled music performance [10].

4. DISCUSSION

We ask ourselves whether data captured while playing an instrument based on body movements may open interesting perspectives in the field of sound synthesis with deep learning. To our knowledge, movement sonification and movement capture have not been used to train deep-learning models for sound synthesis. Current methods ei-

ther rely on the use of musical notation data (i.e., sequences of notes) or on the use of sampled audio data formats. As mentioned earlier, both have limitations:

- In the case of synthesis of musical notation data (i.e. MIDI), new sequences are generated by using recurrent neural networks (i.e., LSTM), which are good at memorizing past states of sequential data. Outputs take the form of new sheet music. The advantages is that it produces "clean" data, while the main drawback is that the data contains few dimensions and doesn't encapsulate any of an interpretation's dimensions. Synthesis based on musical notation has famously been used by smartphone company Huawei to "complete" Schubert's Symphony No. 8, although the result was deemed unsatisfactory [11].
- In the case of audio file data synthesis, it is usually transformed into graphical data first (i.e., Mel spectrograms) and processed with computer vision algorithms which synthesize new spectrograms. These are then translated back into sound samples, and the result is usually extremely sensitive to the quality of the recording. The advantage of this approach is that it may encapsulate human interpretation dimensions, but the drawback is that it produces very noisy outputs. Unless training on uncompressed data of sounds recorded in an anechoic chamber with close-to-perfect conditions, results are still far from satisfactory.

We formulate the hypothesis–without being able to answer it as of now–that using motion capture data whose recorded movements produced consistent sounds could be an interesting approach to training a model capable of generating new sequences, using the same techniques as musical notation synthesis (LSTM or Transformer architectures), while encapsulating dimensions that are related to the interpretation as represented in the training data. It would benefits from the advantages of both mentioned methods: clean, notation-based outputs and the encapsulation of interpretation dimensions. It would also avoid their pitfalls: unlike most notation systems, motion capture data would retain multiple dimensions related to the singularity of an interpretation, while remaining immune to noise in the training data the way sound synthesis through spectrogram is. It may also be that some other unsuspected sensitive dimensions would be lost in translation. This paper makes no claim as to the actual results yielded by this approach. Rather, it proposes to lay down and discuss rational, practical and theoretical foundations for future development of a system aimed at innovative outputs in the field of sound synthesis with deep learning.

Acknowledgments

The authors would like to thank the ZHdK's Institute for Computer Music and Sound Technology for hosting our artistic research: Eric Larrieux, Tobias Gerber, Martin Fröhlich, and Kristina Jungic.

5. REFERENCES

[1] H. Brandner and M. Haverkamp, *Alexander Truslit: Gestaltung und Bewegung in der Musik.* Reprint of the original edition 1938. Augsburg: Wißner, 2015.

[2] J. Hohagen and C. Wöllner, "Movement Sonification of Musical Gestures: Investigating Perceptual Processes Underlying Musical Performance Movements," in *Proc. of the 13th Sound & Music Computing Conference (SMC)*, Hamburg, 2016.

[3] S. Landry, J. Ryan, and M. Jeon, "Design Issues and Considerations for Dance-Based Sonification," in *Proc. of the 20th International Conference on Auditory Display (ICAD2014)*, New York, 2014.

[4] A. Giomi, "Somatic Sonification in Dance Performances. From the Artistic to the Perceptual and Back," in *Proc. of the 7th International Conference on Movement and Computing (MOCO '20)*, New York, 2020, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/10.1145/3401956.3404226

[5] S. Landry and M. Jeon, "Interactive Sonification Strategies for the Motion and Emotion of Dance Performances," *Journal on Multimodal User Interfaces*, vol. 14, pp. 167–186, 2019.

[6] A. Effenberg, "Movement Sonification: Effects on Perception and Action," *IEEE MultiMedia*, vol. 12, no. 2, pp. 53–59, 2005.

[7] T. H. Nown, P. Upadhyay, A. Kerr, I. Andonovic, C. Tachtatzis, and M. A. Grealy, "A Mapping Review of Real-Time Movement Sonification Systems for Movement Rehabilitation," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 672–686, 2023.

[8] T. Grosshauser, B. Bläsing, C. Spieth, and T. Hermann, "Wearable Sensor-Based Real-Time Sonification of Motion and Foot Pressure in Dance Teaching and Training," *Journal of the Audio Engineering Society*, vol. 60, no. 7/8, pp. 580–589, 2012. [Online]. Available: https://www.aes.org/e-lib/online/browse.cfm?elib=16369

[9] W. Goebl and C. Palmer, "Temporal Control and Hand Movement Efficiency in Skilled Music Performance," *PLOS ONE*, vol. 8, no. 1, 2013. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0050901

[10] A. Bouënard, M. M. Wanderley, S. Gibet, and F. Marandola, "Virtual Gesture Control and Synthesis of Music Performances: Qualitative Evaluation of Synthesized Timpani Exercises," *Computer Music Journal*, vol. 35, no. 3, pp. 57–72, 2011. [Online]. Available: https://doi.org/10.1162/COMJ_a_00069

[11] D. Szelogowski, "Generative Deep Learning for Virtuosic Classical Music: Generative Adversarial Networks as Renowned Composers," 2021. [Online]. Available: https://arxiv.org/abs/2101.00169