P(l)aying Attention: Multi-modal, multi-temporal music control

Nicolas Gold
Dept. of Computer Science,
University College London
Gower Street, London
WC1E 6BT, UK
n.gold@ucl.ac.uk

Chongyang Wang
UCL Interaction Centre,
University College London
Gower Street, London
WC1E 6BT, UK
chongyang.wang.17@ucl.ac.uk

Temitayo Olugbade
UCL Interaction Centre,
University College London
Gower Street, London
WC1E 6BT, UK
temitayo.olugbade.13@ucl.ac.uk

Nadia Bianchi-Berthouze
UCL Interaction Centre,
University College London
Gower Street, London
WC1E 6BT, UK
n.berthouze@ucl.ac.uk

Amanda C. de C. Williams
Dept. of Clinical, Educational & Health Psychology
Gower Street, London
WC1E 6BT, UK
amanda.williams@ucl.ac.uk

ABSTRACT

The expressive control of sound and music through body movements is well-studied. For some people, body movement is demanding, and although they would prefer to express themselves freely using gestural control, they are unable to use such interfaces without difficulty. In this paper, we present the P(l)aying Attention framework for manipulating recorded music to support these people, and to help the therapists that work with them. The aim is to facilitate body awareness, exploration, and expressivity by allowing the manipulation of a pre-recorded 'ensemble' through an interpretation of body movement, provided by a machine-learning system trained on physiotherapist assessments and movement data from people with chronic pain. The system considers the nature of a person's movement (e.g. protective) and offers an interpretation in terms of the joint-groups that are playing a major role in the determination at that point in the movement, and to which attention should perhaps be given (or the opposite at the user's discretion). Using music to convey the interpretation offers informational (through movement sonification) and creative (through manipulating the ensemble by movement) possibilities. The approach offers the opportunity to explore movement and music at multiple timescales and under varying musical aesthetics.

Author Keywords

multi-temporality, music sonification, machine-learning, attention, interpretation, body movement

CCS Concepts

•**Applied computing** → **Sound and music computing;** •**Human-centered computing** → *Collaborative interaction;* **Auditory feedback;** •**Computing methodologies** → *Machine learning;* •**Social and professional topics** → People with disabilities;

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

1. INTRODUCTION

Attention is important. It can be beneficial (in terms of concentration), can be lost (e.g. through distraction caused by pain [23]), or even unrealised (attending to something without conscious thought). Interpreting movement and drawing attention to important areas of the body is part of a physiotherapist's role in helping their patient, supporting collaborative exploration and the management and improvement of the patient's physical capability [18]. Patient and therapist thus form a small "ensemble" (or in group therapy, a larger "ensemble" led or facilitated by the therapist). Andersson and Cappelen discuss this in terms of moving from a hierarchical power relationship between therapist and client towards a more active role for the client that in a music-making context, empowers the therapist also [3].

In ensemble music performance, the interpretative role is played to some extent by a conductor through their gestures. They "are often described as embodying the music" [12] and act (as Wigglesworth characterises it [26, p. 3]) as the connector between the composer, musicians, music and audience. Kumar and Morrison found that conductors' gestures affect the perception of a conducted performance [12] and Durrant and Varvarigou [8] discuss the need for conductors to have an appropriate gestural vocabulary that does not mislead performers or distract audiences. Conducting gestures are a means to cause an orchestra to behave in a certain way [16] and thus conducting is a potentially useful metaphor for linking movement, attention, interpretation, and music.

Although movement and music take place in time, time is challenging to those who find movement difficult. For example, fear may slow the start of a movement, interrupt it, or cause it to cease earlier than desired [9]. Wigglesworth comments that time "waits for no-one, cares for nothing" [26, p. ix] but also claims that we can control our perception of time through music.

On occasion the strictures of musical time may be a helpful challenge, but in many cases they may not be. Those with difficulty in moving may find great benefit from participation in musical activity yet be unable to coordinate their movements sufficiently accurately to participate in an ensemble. To use music-making as a support in this kind of situation we must therefore find dimensions of musical manipulation that can free us from the constraints of absolute time, without losing the benefit of feeling agency over the music. In this context we explore the case of chronic pain and attention-driven sonification of body exploration.

Chronic pain is a prevalent, disabling condition in which pain persists in the absence of tissue damage and is consistent with dysfunction in the nervous system [10, 21]. People with chronic pain find harmless everyday movement (e.g. in driving, laundry) challenging [21].

In chronic pain treatment through physical therapy, there are various temporal perception needs: those of the patients themselves, a therapist, others in a group setting and so forth. These audiences, although aligned in their overall purpose, need to experience a patient's movements in different ways and at different temporal scales. Patient and physio may both have analytical needs, but the patient's experience of movement (as an embodied process) will be vastly different to the physio's (who may wish to slow the movement in order to better understand the detail of its progress, or speed it up to enhance perceptions of capability and self-efficacy). In discussion, both may wish to modify the movement speed retrospectively to explore specific aspects or overall trends.

Music can be helpful in reducing procedural pain (e.g. dental surgery) and anxiety associated with procedures or medical conditions (e.g. [5]), and while there are anecdotal accounts of the usefulness of music in chronic pain, as an aid to distraction and relaxation, its possible use in supporting everyday activity has been largely overlooked (but see [13]).

To use music in the scenario we address here, what is required is a sonic representation of movement that (1) minimises or eliminates the need for music-synchronous action on the part of the patient (but permits it if desired), (2) reveals an interpretation of that movement (provided here by a machine-learning system), (3) permits its exploration by the patient or others, and (4) maintains musical coherence throughout. The sonification needs to be simultaneously informational and experiential: informational [19] so that aspects of movement are revealed, and experiential to ensure it increases self-efficacy or a sense of being capable, and possibly induces changes in behaviour [17, 20]. For musically expressive applications, the retention of music in the sonification is clearly essential.

The remainder of this paper presents the design of a framework (and implementation thereof) of a sonification designed to support all these aspects and allow the multi-temporal, multi-modal exploration of interpreted body movement in a musical setting with potential for both analytical and creative uses. The patient is, in essence, a conductor: guiding the ensemble toward and away from features of their (the patient's) movement (and by extension their body and perception of their body) through their gestures as interpreted by an observer, in this case a machine learning system.

2. FRAMEWORK DESIGN

To support the desired range of information and experiential needs of patients and therapists, we have developed a simple model-based musical sonification framework (entitled "P(l)aying Attention") in which data streams are mapped to part volumes in an ensemble of thirteen instruments (thirteen is an arbitary choice resulting from our current data: more or fewer streams could be used).

There are three key (and linked) design considerations: how to manipulate the sonification/music, when to manipulate the sonification/music, and how to obtain the interpretation of movement that determines what changes should be made. The relative weight of these considerations may vary depending on the relative emphasis placed on musical outcomes vs interpretative outcomes.

Manipulating music without disturbing musical time requires an orthogonal channel of change, in this case the relative volume of the instrumental part for each channel of data. This approach sits between *parameter mapping sonification* and *model-based sonification* in Walker and Nees' taxonomy [24]. Taken without the machine-learning aspect, it is a parameter-mapped sonification connecting attention-score to part-gain. However, with the machine-learning component included, it is more a model-based sonification where a patient's movements cause responses in the machine-learning system that are themselves sonified.

The general approach could be adapted for use with larger groups e.g. dyads or triads. The degree of interpretation and its representation could be altered, e.g. for a dyad, one might be interested in comparing the difference in absolute position between therapist and patient (or between a pair of dancers) and thus use that difference as the driver of part gain, but also in the difference in the interpretation of body part importance (that might also be visually represented).

In addition to considering the nature of sonification control, one must consider *when* such control should be applied temporally. We identify a number of temporal scales and contexts that need to be supported:

- *Movement-synchronous:* movement is analysed in real-time as it happens.
- *Movement-asynchronous:* movement is analysed after it has happened.
  - *Movement-replay:* movement is analysed by replaying forwards as recorded.
  - *Movement-reversed:* movement is analysed by replaying it backwards.
  - *Movement-scaled:* movement is analysed on a scaled timing, e.g. half-speed.
- *Music-synchronous:* movement corresponds to musical features (e.g. beats).
- *Music-asynchronous:* movement does not correspond to musical features.
- *Discursive-free:* movement is analysed in free time and in any direction for the purpose of discussion.

It is likely that some of these modalities will be combined e.g. a patient may undertake music-synchronous movement with a therapist undertaking movement-synchronous analysis. On the 'opposite' end of the scale, a patient may undertake music-asynchronous movement and discursive-free analysis.

The final component in the sonification design is obtaining the interpretation of movement. People with chronic pain, aiming to protect themselves from increased pain or feared injury, may move cautiously in ways that are not efficient, and can contribute to longer term disability. Their fear and anxiety toward the pain and injury lead to the adoption of different strategies during functional activities. Specifically, their body parts are engaged in inefficient and biomechanically unnecessary ways, which could be typically observed for the use of a specific body part at different activity stages.

The machine learning model [25] used in this study was originally proposed to detect the protective movement behaviour of people with chronic pain, by paying attention to the salient body configurational and temporal evidences. The input to the model comprises 13 joint angles computed from the position of 26 full-body joints. During training, the attention mechanism learns to give more weight to the body parts (represented by joint angles) and temporal stages most informative for discriminating protective from non-protective movement behaviour. The training of the model is based on motion capture data of people, including those with chronic pain, engaged in everyday functional activities (e.g. sit to stand, stretch forward), and physiotherapists' annotation of occurrences of protective and non-protective behaviour in

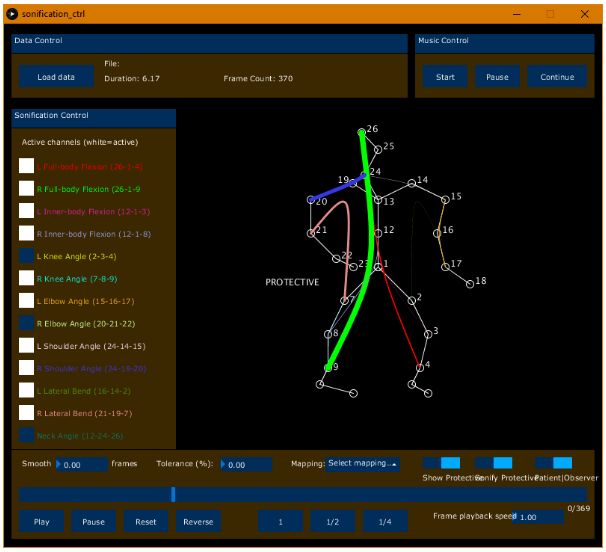

Figure 1: Proof-of-concept implementation with data loaded, three channels disabled, and the timeline partway through.

this data. The model is realised as weights along each of the joint and time dimensions of the movement data. These are normalised in the range zero to one and output as numeric values, thus can be simply used as a gain value for musical parts. For details on the model architecture and detection evaluations see [25] and for a description of the dataset used for the study please refer to [4].

3. PROOF OF CONCEPT

A proof-of-concept implementation of the P(l)aying Attention approach has been realised in Processing [1]. Stems for each of the thirteen parts were recorded using MIDI instruments in Ableton Live [2], frozen and exported, and then mixed to mono (to enable panning with Processing's sound library). Looping behaviour is provided within the Processing library itself. Two sound sets have been produced: one in the style of Afro-Cuban percussion, drawing on patterns described by Uribe [22] and augmented by one of the authors (Gold) to bring the part-count to thirteen; the other an arrangement of parts of the Pachelbel Canon (based on selections from the original score [14] and again, augmented by Gold). The purpose of this implementation was to investigate whether changes in movement interpretation are observable in the music under an absolute mapping of attention score to part gain. This requires that each part has sufficient continuity in its musical material that it could be associated to a particular joint group by the user or therapist even if sounding alone (or prominently, in comparison to other sounding parts). Although thirteen parts is not unusual in music, generating that many that are individually identifiable, playing constantly, and sufficiently congruent to work as a whole was somewhat challenging. Attention to body location is driven by the machine learning model.

Figure 1 shows the user interface for the proof-of-concept implementation. The column of buttons on the left-hand side represent the active data channels, each mapped to a particular audio loop and panned either hard left or hard right to create channel separation and support easier association of musical material with a joint group. Data is loaded into the system using the buttons at the top left, and music playback controlled at the top-right. Data is explored using the panel at the bottom of the screen, supporting playback at real-time speed or at a scale factor thereof, or the ability to 'scrub' through the timeline freely in either direction. Data frames are sampled at 60Hz and are here sonified without smoothing or aggregation. The representative human figure shows the connections between the 26 joints that are assessed by the machine-learning system.

The coloured lines of different weights are Bezier curves plotted between the joints in each group, with line weights varying in proportion to the attention scores. This graphical representation provides a second modality by which movement can be understood in conjunction with the sound, helping those using the system to a better understanding of movement, a stronger association between instrumental part and joint group, and an easier route to exploring the creative and expressive possibilities of the approach.

Most of the multi-temporal requirements are met by the proof-of-concept implementation: data can be explored freely or under time constraint, and in any direction and at different speeds. Data for whatever the current frame constitutes at any moment (whether transient during playback, or static during 'scrubbing') is reflected in the audio and visual renderings. Real-time rendering is not yet implemented thus of the scales and contexts identified earlier, only movement-synchronous is not currently supported.

Durrant and Varvarigou [8] identify movement as a tool for a kinesthetic approach to musical expression. Since our implementation enables control over a pre-recorded ensemble through interpretation of body movement, rather than focusing on analytical concerns it is possible for a user to direct the ensemble through their movement and draw out aspects of the music accordingly. They would be "co-conducting" in partnership with the machine-learning system that is interpreting their gestures, and developing a vocabulary of conducting gestures that lead to particular interpretations in the machine learner and thus changes in ensemble balance.

4. RELATED WORK

Gesture, movement, time, sound, interpretation, and attention have long been studied in a variety of contexts. Dubus and Bresin present a systematic review of mapping strategies for sonifying physical quantities [7], and Schaffert et al. review the use of sound in the context of sports and rehabilitation [15]. Previous work has considered sound and music for chronic pain support [13, 17, 18, 19]. There is a large body of work relating to artistic uses of gestural sound and music interfaces. Of relevance to the techniques used here, examples include Caramiaux et al. who explore gesture to sound mapping using machine learning [6], Siegel who reports experiences of using motion-tracking technology to control multi-speaker sound diffusion [16] and Lee and Yeo who report work using dancers' respiration to control the volume of separate tracks within a multitrack MIDI File [11].

5. FUTURE WORK AND CONCLUSIONS

This paper has presented a framework and proof-of-concept implementation of a sonification of interpreted body movement, designed to support analytical understanding and with the potential for creative applications in music.

Future work will include extending the implementation to allow variations in the relationship between data and sonification (e.g. smoothing, aggregation, relative mapping), and user interface enhancements (fully assignable colour, pan position, audio, and more advanced animation), and new modalities (dyadic representation, and real-time data). Empirical studies will be undertaken to determine the applicability of the approach in a range of scenarios, generative music directly derived from body movement may be used to 'personalise' the resulting multi-track audio, and we intend to explore longer musical forms and timbral control.

6. ETHICAL STANDARDS

This work is funded by the EU Horizon 2020 FET Proactive project "EnTimeMent", grant number: GA824160. The authors declare that they have no conflicts of interest (financial or otherwise) that would affect the work reported here. Data used in this work was gathered with informed consent under various approvals relating to the EmoPain project: UCL - 5625/001, UCLIC/1516/012/Staff Berthouze/Singh; NRES Committee - 11/0514 and used in accordance with UCL Research Ethics Committee approval number: 5095/001 (EnTimeMent project).

7. REFERENCES

[1] Processing, n.d. https://processing.org, last accessed 29 Jan 2020.

[2] Ableton AG. Live, n.d. https://www.ableton.com/en/live/, last accessed 29 Jan 2020.

[3] A.-P. Andersson and B. Cappelen. Musical interaction for health improvement. In K. Collins, B. Kapralos, and H. Tessler, editors, *The Oxford Handbook of Interactive Audio*, pages 247–262. Oxford University Press, July 2014.

[4] M. S. H. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, and et al. The Automatic Detection of Chronic Pain-Related Expression: Requirements, Challenges and the Multimodal EmoPain Dataset. *IEEE Trans. Affect. Comput.*, 7(4):435–451, Oct. 2016.

[5] J. Bradt, C. Dileo, and N. Potvin. Music for stress and anxiety reduction in coronary heart disease patients. *Cochrane Database of Systematic Reviews*, (12), 2013.

[6] B. Caramiaux, J. Françoise, N. Schnell, and F. Bevilacqua. Mapping through listening. *Computer Music Journal*, 38(3):34–48, 2014.

[7] G. Dubus and R. Bresin. A systematic review of mapping strategies for the sonification of physical quantities. *PLOS ONE*, 8(12):1–28, 12 2013.

[8] C. Durrant and M. Varvarigou. Perspectives on Choral Conducting: Theory and Practice. In G. F. Welch, D. M. Howard, and J. Nix, editors, *The Oxford Handbook of Singing*. Oxford University Press, April 2019.

[9] M. Geisser, A. Haig, A. Wallbom, and E. Wiggert. Pain-related fear, lumbar flexion, and dynamic emg among persons with chronic musculoskeletal low back pain. *The Clinical Journal of Pain*, 20(2):61–69, Mar-Apr 2004.

[10] IASP Task Force on Taxonomy, editor. *Classification of Chronic Pain*. IASP Press, Seattle, 2nd edition, 1994.

[11] J.-s. Lee and W.S. Yeo. Real-time modification of music with dancer's respiration pattern. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, Ann Arbor, Michigan, 2012. University of Michigan.

[12] A. B. Kumar and S. J. Morrison. The conductor as visual guide: Gesture and perception of musical content. *Frontiers in Psychology*, 7:1049, 2016.

[13] J. W. Newbold, N. Bianchi-Berthouze, N. E. Gold, A. Tajadura-Jiménez, and A. C. de C. Williams. Musically informed sonification for chronic pain rehabilitation: Facilitating progress & avoiding over-doing. In *Proc. CHI Conf. on Human Factors in Computing Systems*, CHI '16, page 5698–5703, New York, NY, USA, 2016. ACM.

[14] J. Pachelbel. Canon and Gigue; v (3), cemb; D-Dur; P 37, 1838. Owning institution: Staatsbibliothek zu Berlin - PK, http://resolver.staatsbibliothek-berlin.de/SBB0000A73000000000.

[15] N. Schaffert, T. B. Janzen, K. Mattes, and M. H. Thaut. A review on the relationship between sound and movement in sports and rehabilitation. *Frontiers in Psychology*, 10:244, 2019.

[16] W. Siegel. Conducting sound in space. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, pages 376–380, Copenhagen, Denmark, 2017. Aalborg University Copenhagen.

[17] A. Singh, N. Bianchi-Berthouze, and A. C. de C. Williams. Supporting everyday function in chronic pain using wearable technology. In *Proc. CHI Conf. on Human Factors in Computing Systems*, CHI '17, page 3903–3915, New York, NY, USA, 2017. ACM.

[18] A. Singh, A. Klapper, J. Jia, A. Fidalgo, A. Tajadura-Jiménez, N. Kanakam, N. Bianchi-Berthouze, and A. C. de C. Williams. Motivating people with chronic pain to do physical activity: Opportunities for technology design. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, CHI '14, page 2803–2812, New York, NY, USA, 2014. ACM.

[19] A. Singh, S. Piana, D. Pollarolo, G. Volpe, G. Varni, A. Tajadura-Jiménez, A. C. de C. Williams, A. Camurri, and N. Bianchi-Berthouze. Go-with-the-flow: Tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. *Human–Computer Interaction*, 31(3-4):335–383, 2016.

[20] A. Tajadura-Jiménez, M. Basia, O. Deroy, M. Fairhurst, N. Marquardt, and N. Bianchi-Berthouze. As light as your footsteps: Altering walking sounds to change perceived body weight, emotional state and gait. In *Proc. 33rd Ann. ACM Conf. on Human Factors in Computing Systems*, CHI '15, page 2943–2952, New York, NY, USA, 2015. ACM.

[21] I. Tracey and M. C. Bushnell. How neuroimaging studies have challenged us to rethink: is chronic pain a disease? *The Journal of Pain*, 10(11):1113–1120, 11 2009.

[22] E. Uribe. *The Essence of Afro-Cuban Percussion and Drum Set*. Warner Bros. Pub., Miami, FL, 1996.

[23] J. Vlaeyen, S. Morley, and G. Crombez. The experimental analysis of the interruptive, interfering, and identity-distorting effects of chronic pain. *Behaviour Research and Therapy*, 86:23–34, 11 2016.

[24] B. N. Walker and M. A. Nees. Theory of Sonification. In T. Hermann, A. Hunt, and J. G. Neuhoff, editors, *The Sonification Handbook*. Logos Verlag, Berlin, Germany, 2011.

[25] C. Wang, M. Peng, T. A. Olugbade, N. D. Lane, A. C. de C. Williams, and N. Bianchi-Berthouze. Learning temporal and bodily attention in protective movement behavior detection. In *Proc. 8th Int. Conf. on Affective Computing and Intelligent Interaction: Workshops and Demos (ACIIW'19)*, pages 324–330, Cambridge, UK, 2019. IEEE.

[26] M. Wigglesworth. *The Silent Musician: Why Conducting Matters*. Faber & Faber Ltd., London, UK, 2018.