Using Interactive Machine Learning to Sonify Visually Impaired Dancers’ Movement

| Dr Simon Katan  Goldsmiths University  London, United Kingdom  s.katan@gold.ac.uk |
| --- |

# ABSTRACT

This preliminary research investigates the application of Interactive Machine Learning (IML) to sonify the movements of visually impaired dancers. Using custom wearable devices with localized sound, our observations demonstrate how sonification enables the communication of time-based information about movements such as phrase length and periodicity, and nuanced information such as magnitudes and accelerations. The work raises a number challenges regarding the application of IML to this domain. In particular we identify a need for ensuring even rates of change in regression models when performing sonification and a need for consideration of how to convey machine learning approaches to end users.

## Author Keywords

Interactive Machine Learning; Accessible Interfaces; Dance; Sonification.

## ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces.

# INTRODUCTION

In working with visually impaired dancers, the absence of the visual modality raises a number of challenges for the choreographer. Discussions with Iris Karayan, a choreographer who works with a company of visually impaired dancers, highlighted difficulties with communication of movement to the dancers. Equally significant was the problem of gestural communication between the dancers themselves. Contemporary choreographic technique typically involves collaborative approaches to creativity where ideas are worked out through dancers’ physical responses. Restricted communication can seriously impede this sort of devising process.

We investigated how the synthesis of an additional aural modality might mitigate for the absence of visual communication. The research took place during the course of a single workshop with choreographer Iris Karayan and her company of visually impaired dancers. Through the use of custom built wearable computing which maps sensor input to real time synthesis, we were able to provide the dancers with a uniform aural reference for their movements. Our work yields a number of encouraging findings and raises a number of technical challenges for the application of IML to this domain.

# RELATED WORK

Prior work has successfully applied real-time sonification of motion and foot pressure to the teaching and training of dance students without visual impairments. In this case auditory display forms a “closed interaction loop” for use by the teacher and student to analyse particular components of predetermined dance moves which the student is attempting to learn. Sonification is identified as ideal for this application as a consequence of the high resolution of the human auditory system which “enables listeners to recognize small and even invisible changes in body movement” [6]. Further to this, Varni, Dubus *et al*. demonstrate the use IMU data from mobile phones to help “users to keep synchronization for longer” in “dyadic coordinated human rhythmic activity” [9]. In other work haptic feedback has been proposed as a way of enabling visually impaired audiences to experience the effects of kinaesthetic empathy which sighted audiences are able to access through vision [8]. Vibrotactile feedback sensed through the user’s hands is used to communicate qualitative aspects such as “softness of circular patterns.” Finally the ongoing EU H2020 project DANCE – Dancing in the Dark is currently carrying out multiple investigations into how “qualities of body movement can be expressed, represented, and analyzed by the auditory channel” [1].

First used in the context of Computer Vision [2], Interactive Machine Learning allows the design of interaction through the provision of examples gestures rather than the coding of responses. Examples of gestures (eg. a static pose recorded via several accelerometers) are recorded as training points and paired with the desired computer responses. A supervised learning algorithm then uses this data to build a model which can map human actions to computer responses in real time. The accuracy of the model is evaluated by the designer through use, and further training examples can be provided to improve it.

IML has been investigated by Fiebrink et al [3,4] in the context of expert musicians building new instruments, and by Katan et al [7] in the context designing specialised musical interfaces for users with disabilities. Whilst the former demonstrates that these systems can directly encode aspects of embodied practice and “feel right” to use, the latter highlights difficulties in designing and memorising custom gesture vocabularies, and demonstrates potential advantages of using IML to train rich continuous control spaces for free exploration by the user. Similarly this current work uses IML as a means generating rich continuous control spaces, but with the intention of providing a sonic reference for the dancer rather than a custom instrument.

# METHODS

The workshop lasted three hours and involved four dancers from Iris’ company and Iris. All dancers were registered blind, but visual impairments ranged from partial sight to no sight. Iris has no visual impairment. We used two identical custom wearable devices which independently sensed, interpreted and sonified users’ movement with localised sound. Each device comprised multiple components mounted on a four point adjustable harness. A unit at the back housed a Raspberry Pi model A+, inertial measurement unit (IMU), and 3 watt amplifier, whilst a 60 mm speaker was worn on the left shoulder. Two further IMUs were worn on each wrist using velcro straps and connected to the RPI via cables (Figure 1.).

The sonification was performed by a C++ application incorporating the Maximillian audio library [5], and FANN[[1]](#footnote-1) which implements artificial neural networks. Eight input features were used comprising pitch and roll for each of the IMUs and two features showing the difference in yaw between the IMU on the torso and the wrist worn ones. During the workshop we tried two contrasting synths. One used frequency modulation, and the other granular synthesis.

Training was performed through multiple neural networks using many to one mappings - Each net mapped all the input parameters to one output parameter. The training itself was carried out prior to the workshop. Static poses were provided as training examples by the researcher and the model was iteratively tested and retrained until a desirable interface was achieved. The sonifcation used the interpolated output from the models to construct a continuous control space bounded by the original training examples.

We wanted to avoid prescriptive activities so that we could observe emergent behaviour from this additional modality.  We therefore adopted a workshop-based approach and collected qualitative data from our observations. We did not isolate the sonification from other modalities as we were interested in observing how it added to cross-modal communication.

The workshop was split into three open ended exercises. Firstly the dancers experimented with the devices on their own to build an understanding of the relationship between their movement and the sound. Secondly dancers worked in tandem taking it in turns to copy each other’s motion (Figure 2). The final exercise required Iris and the dancers to create memorable sequences of sound-movement pairs.

The dancers were interviewed about their experiences at the end of the workshop in a semi-structured interview. Three of the questions evaluated their response to the devices on a five point Likert scale (Table 1). A further two questions prompted for open-ended feedback. For further information please see the full feedback online.[[2]](#footnote-2)

## DESIGN DECISIONS

Sound was chosen over haptic feedback for a number of reasons. In contrast to haptic feedback, sound conveys location and agency whilst remaining communally perceivable. The interface design used localised sound so that the dancers could intuitively identify their sonification from others’ sounds. Synthesised sound offers considerably more perceptual possibilities than any of the currently available haptic feedback methods.

Informed by previous experiences, we decided to avoid the use of camera-based technologies for tracking. Our particular concerns with these methods was the requirement for the dancer to be mindful of the camera’s location and field of view which we considered particularly problematic when working with visually impaired dancers.

From discussions with Iris about the dancers’ abilities and needs before the project, it became clear that they were competent at orienting themselves in the space and the technology did not need to address this issue. For this reason we kept all sonification independent of location and orientation, focusing on sonifying the quality of dancers’ movements instead.



Figure 1. A participant using one of the devices

## RESULTS

The researchers’ observed the successful transfer of nuanced temporal information between participants through sonification which was additional to that communicated verbally or through touch. Findings for the transfer of information about movement at broader levels of detail were inconclusive. The researchers and participants encountered a number of difficulties with using IML to create regression models for sonifying movement, in particular the non-linearity of the models produced.

## DISCUSSION

During the first exercise participants oriented themselves through experimentation with the devices. We answered questions about how they worked. Many of these focussed on the nature of the mapping. For example, whether specific movements would trigger sounds and whether particular sensors were responsible for controlling individual parameters. Through this we observed that participants were confused by the combined mapping of the three sensors onto a single stream of audio. For example, they complained of a movement from one arm not causing the same sound as last time, when actually their other arm and torso were now in different orientations causing the sound to change.

Participants also complained about the non-linearity of the aural response. For example, they complained when large movements only resulted in small sonic changes and vice-versa, and also when encountering areas of the control space where the sound remain constant despite their movements.

During the copying exercise the researchers observed the transfer of additional nuanced information to that communicated verbally or through touch. This was most noticeable in the time domain. For example, we observed some dancers phrasing movements in synchronization (Figure 3.). We noticed imitation of subtle aspects of phrases such as the magnitude and phrasing, and also observed participants beginning to vary these through improvised call and response patterns. This is demonstrated in the video documentation between 0’00 and 0’30” . The participant on the right of the screen can be seen imitating the opposite participant who is tracing circles with her hand. Not only are the gestures copied but subtle inflexions such as deceleration towards the lowest point of the circle are also imitated in near synchronization. Successful transfer of a different gesture is demonstrated between 0’30” and 1’00.

We also observed participants using sonification to identify when their gestures had been copied incorrectly. Subsequently participants offered verbal corrections relating to timing and magnitudes of movement. This can be seen in the video documentation from 1’00 to 1’50[[3]](#footnote-3). The participant on the right attempts to copy the gesture of the left hand participant, but identifies that the gestures don’t match. Given the participants’ lack of sight, the sonification is the only possible modality by which this verification can have occurred.

Attempts to use the sonification to convey body related and geometric specifics of movement were less successful. In all but a few cases participants were unable to establish broad characteristics of the movement such as specific poses and turns from the sound alone. Instead participants resorted to verbal descriptions and touch for communication of these aspects. This can be clearly seen in the video from 1’50 to 2’30 where the dancer continues her attempts to copy the gesture resorting to verbal communication and ultimately requiring the assistance of the choreographer.

During the third exercise, the dancers were unsuccessful in identifying and organising discrete movement sound pairs. They were unable to conceptualise or repeat sequences. Our observations were inconclusive as to why they were unable to do this. Whilst the difficulties mentioned above might have contributed, it is equally possible that the task which was of a greater complexity than the other two, required more time than had been allotted in the workshop.

The feedback from the participants was inconclusive as to whether the devices helped them understand and communicate movement (Table 1). One participant commented that the sonification helped her understand her own movement through moving her wrists, but was not helpful for communicating with others. Another commented that “repetition helps me notice the difference in the sound.” Another of the participants commented that when you listen to the sound for the first time “you don’t understand it. You need a lexicon, a vocabulary.” All participants agreed that they enjoyed being able to control sound through movement and that the sound encouraged them to move in new ways.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statements** | **User A** | **User B** | **User C** | **User D** |
| “I enjoyed controlling sound through movement.” | 1 | 1 | 1 | 1 |
| “Sound helped understand mine and other's movements better” | 3 | 5 | 3 | 5 |
| “Sound encouraged me to move in new ways.” | 1 | 1 | 1 | 1 |

Table 1. Participant responses – a score of 1 indicates strongly agree.

Figure 2. Call and response copying of gestures

## Macintosh HD:Users:kimonsatan:Desktop:bd2.png

Figure 3. Performing a gesture in synchronisation

## CONCLUSIONS

Our observations demonstrate that sonification supports the communication of movement between visually impaired dancers by allowing the transfer of nuanced qualitative information in the time domain. The sonic interaction allowed participants to identify subtle differences between each others’ movements and make micro-adjustments to synchronise them in real time. This facilitates more fluid interaction between the dancers, ultimately increasing their independence from the choreographer.

Sonification was less useful with regards to communicating and conceptualising other aspects of the movement such as direction or what limbs were being moved. This might be improved by addressing a number of challenges regarding the application of IML in this context.

Firstly the non-linear regressions produced by the neural network increased in sensitivity as the input parameters approached training points. Whilst ideal for pose recognition, this proved disadvantageous for our application where interpolation between training points is more important for the end application. As a result it was difficult to train the regression to evenly cover all of the control space. Areas where the sound remains constant despite changes in control parameters proved difficult to eliminate through editing training points. This was borne out by the complaints of the participants. Whilst employing linear regression models would improve the outcome, the situation calls for exploring training approaches which focus on the shape of the regression as a whole rather than individual points.



Secondly the questions and comments of the participants demonstrated difficulty in understanding how the input from multiple sensors mapped into a single stream of audio output. Driven by the assumption of a parametric model, participants tended to work with isolated movements focussed around individual limbs. Multiple approaches might be taken to mitigate for this effect. One would be to employ separate sound streams for each sensor. However, this risks a sonic output which is too complex to be easily interpreted by the dancers. Alternatively a physical interface could be designed to not unduly draw users’ attention to individual sensors. Such an interface might employ conductive materials to detect poses without drawing focus towards specific areas of the body. Irrespective of these solutions, this problem highlights the need to develop more accessible vocabulary and metaphors of machine learning techniques so that users can intuitively understand how to control them.

## REFERENCES

1. Dance.dibris.unige.it. (2016). DANCE EU ICT H2020 Project. [online] Available at: http://dance.dibris.unige.it/ [Accessed 12 May 2016].
2. Fails, J. and D. R. Olsen. 2003. Interactive machine learning. Proc. IUI, 39–45.
3. Fiebrink,R.,P.R.Cook Trueman.D, 2011.Human model evaluation in interactive supervised learning. Proc. CHI, 147–56.
4. Fiebrink,R.,D.Trueman,C.Brittetal. 2010.Toward understanding human-computer interactions in composing the instrument. Int’l Comuter Music Conf.
5. Grierson, Mike, Kiefer, Chris. 2011. Maximillian: An easy to use, cross platform C++ Toolkit for interactive audio and synthesis applications in Proceedings of International Computer Music Conference 2011.
6. Grosshauser T., Blasing B., Spieth C., Hermann T., 2011, Wearable sensor based real-time signification of motion and foot pressure in dance teaching and training, J. Audio Eng. Soc., Vol. 1, No. 1, 2011 October
7. Katan, S, Grierson, M, Fiebrink R, 2015 Using Interactive Machine Learning to Support Interface Development Through Workshops with Disabled People Proc. CHI
8. Lycouris, S, Timmons, W, Ravenscroft, J & Wright, M, Choreohaptic Experiments, 2012, Artefact, University of Edinburgh, Edinburgh, UK; Somatics and Technology Conference 2012,
9. Varni Giovanna, Dubus Gael *et al*. ,Interactive sonification of synchronization of motoric behavior in social active listening to music with mobile devices, J. Multimodal User Interfaces, Vol 5. No. 3, 2012 May

1. http://leenissen.dk/fann/wp/ [↑](#footnote-ref-1)
2. http://igor.gold.ac.uk/~skata001/hearingdance/results.html [↑](#footnote-ref-2)
3. http://igor.gold.ac.uk/~skata001/hearingdance/video.mp4 [↑](#footnote-ref-3)