

Accessible Choral Ensembles for Visually Impaired Singers

Claire Galea
Faculty of ICT
University of Malta
claire.galea.13@um.edu.mt

Chris Porter
Faculty of ICT
University of Malta
chris.porter@um.edu.mt

Choral activities are generally led by a conductor who uses visual cues and non-verbal instructions to drive the performance. Song dynamics are generally well rehearsed however live performances may necessitate unrehearsed messages in order to correct errors or to introduce dynamics in response to external factors. These messages are communicated by the choir master just-in-time, to which however, visually impaired choristers have no access. This paper outlines an investigation into how technology can contribute to this end while presenting a solution which adopts optoelectronic devices for gesture recognition, real-time communication protocols and over-the-air haptic-feedback to enable participation while minimising adoption barriers via intuitive and low-friction interaction. Insights from both qualitative and quantitative techniques will be presented along with techniques used to understand, assess and evaluate the domain in an iterative series of interventions.

Human-computer interaction in choral activities, gesture recognition and interpretation, realtime communications

1. INTRODUCTION

Visually impaired singers are normally excluded from performing several activities due to their impairment. Choral participants generally require visual contact with a choir master who would be giving specific instructions to singers. The messages that the conductor communicates include starting and closing instructions, tempo and general dynamics such as volume and accents. In choral performances, singers are required to respond to the conductor's messages in a real-time manner.

In an interview with a professional conductor it was pointed out that "specific messages need to be communicated to the singers to correct errors that may occur during a performance" and therefore the argument that rehearsals are enough to include visually impaired singers does not hold. Not being able to respond to 'new' and unrehearsed instructions takes away the beauty of performing live based on unpredictable influences.

The motivation behind this project is to try to resolve the communication barrier that exists between a conductor and a visually impaired singer without adding excessive workload or adoption friction on any party involved. This work proposes a prototype which captures conductor gestures, in their natural form, which are then interpreted and transmitted over-the-air to singers as haptic

messages and in realtime. The idea is to allow visual and non-visual singers to respond to the same message at the same time - albeit over different channels. The chorister will be able to access visual messages, interpret the information and participate along with the other choristers.

To this end, this exercise was driven by the following overarching research question:

How can Human Computer Interaction (HCI) techniques and technologies aid in conducting a choral ensemble consisting of visually impaired singers?

A number of professional choral conductors were involved in the exercise to construct a better understanding of the domain, as well as to evaluate intermediate revisions of the prototype.

Literature on (1) concepts and techniques in choral singing, (2) challenges of accessibility in choral activities as well as (3) technology and human computer interaction had to be reviewed in order to establish the state-of-the-art as well as to inform the research effort. It is important to point out that this is early-stage and ongoing work and we're presenting this paper to encourage other researchers to think beyond the norm of accessibility research and to reach out to, and merge with, other domains that may benefit highly from research in interaction design

2. BACKGROUND

Van Weelden (2002) describes conducting as “a complex art that involves, effective nonverbal communication”. During rehearsals, conductors make use of both verbal and non-verbal communication to guide the ensemble. However, during a live performance, non-verbal communication is generally the only means by which a conductor is able to communicate with singers (McClung, 1996). Gestures or “human movements that are relevant to the production and comprehension of music” (Cox, 2010) are used by conductors to convey two types of messages (Durrant, 2012):

1. **Indicative** – gestures which are similar to non-musical signals and which are unambiguous to understand. Examples include the cut-off signal which indicates when singers or musicians should stop, as well as the entry cue which indicates the section or individual that is bound to start performing a part of the music piece.

2. **Connotative** – gestures that convey unplanned messages that relate to the sound quality and the character of the musical piece. These musical gestures can take different forms including hand gestures, facial expression and other body gestures.

Conducting gestures are bound to a specific space, namely the conducting plane, along which most of the information is communicated to the ensemble (DeVenney, 2010), and the conducting box, the area that the conductor works in and that approximately ranges from the bottom of the chin to waist level (see Figure 1).

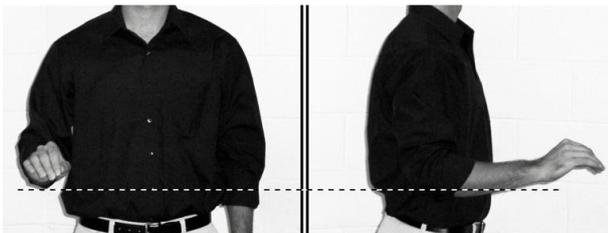


Figure 1 Front and side view of the conducting plane in the conducting box (Source: *Conducting Choirs, Volume 1* [9])

Table 1 outlines the gestures that have been considered for this investigation. These gestures were synthesised from interviews with domain experts as well as from choral conducting literature.

2.1 State of the art

Louis Braille, known for his reading and writing system for the blind, also invented a notation for use with musical scores – allowing for blind musicians to read and interpret musical notation independently (Smaligo, 1999). The Braille Music system makes use of a combination of six-dot

embossed cells “to represent the pitch and the rhythm of each note” (Royal National Institute of Blind People, 2016). By referring to Braille Music notation, blind musicians are able to fully understand the notation of a music piece, however the learning curve for Braille Music is significant (Park, 2015). Menehan (2009) argues that during rehearsals blind choral singers prefer to resort to recordings (which give instructions on the tempo and dynamics) rather than Braille Music scores, as it is impossible for them to ‘read’ music notes together with lyrics simultaneously (Menehan, 2009). Moreover, the choir master might need to send unrehearsed instructions during a performance, therefore scoping the utility of Braille Music to the studying or rehearsal stages.

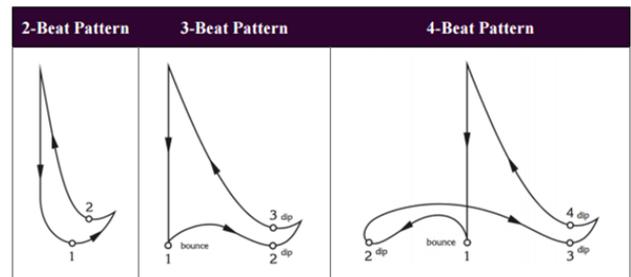


Figure 2 Tempo conducting patterns treated as gestures in this work (adapted from the *Conducting Course* available at <https://www.lds.org/music/conducting-music/conducting-course-book-and-audio-examples>, accessed April 2018)

Petar Matev, the conductor of the ‘Academician Petko Staynov’ choir, explains that his blind singers memorise musical notes along with their surroundings (the breathing and movement of the conductor) and the directions that he verbally gives during the rehearsals (Matev, 2010). Memorisation of musical pieces restricts the conductor’s role during a live performance where live interpretation is limited along with the flexibility to introduce new instructions in response to external factors. Depending on the dynamics of the performance, the conductor might need to send specific messages that were not rehearsed beforehand. We argue that memorisation during performances takes away all affordances that non-impaired singers enjoy, mainly related to guidance, emotions and expressions (facial or body postures) communicated by a conductor during an actual performance. Quoted in (Menehan, 2009), Alan Harler, a music director at the Mendelssohn Club of Philadelphia, explains that visually impaired singers “take their cues from the collective movement around them, especially the motion involved in breathing”. Also, choral conductor Osvaldo César Manzanelli (2010) states that gestures that are commonly used in choral conducting cannot be used to communicate with visually impaired people in a choir. While the use of hardly audible verbal communication made it possible for him to conduct his choir, Manzanelli argues that future tools may

Table 1: Conducting gestures considered in this work

Gesture	Description	Performing hand
<i>Preparatory beat</i>	<i>Hand movement that indicates that the singers should breathe in, in preparation to start singing</i>	<i>Right</i>
<i>Downbeat</i>	<i>A strong downwards movement that is performed at the first beat of every measure in the music piece</i>	<i>Right</i>
<i>Upbeat</i>	<i>The upward hand movement that is performed right after a downbeat</i>	<i>Right</i>
<i>Two-beat pattern</i>	<i>The pattern given for a 2/4 time signature music piece (see Figure 2)</i>	<i>Right</i>
<i>Three-beat pattern</i>	<i>The beat pattern given when conducting a music piece with a 3/4 time signature (see Figure 2)</i>	<i>Right</i>
<i>Four-beat pattern</i>	<i>Used with a 4/4 time signature (see Figure 2)</i>	<i>Right</i>
<i>Crescendo</i>	<i>Increase in loudness indicated by an upward movement with the palm facing up</i>	<i>Left</i>
<i>Diminuendo</i>	<i>Reducing loudness indicated by a downward movement with the palm facing down</i>	<i>Left</i>
<i>Closing/cut</i>	<i>Closing fist gesture which is sometimes accompanied by a circular wrist/hand movement</i>	<i>Left</i>

improve this communication (Manzanelli, 2010).

Audible cues may be disruptive and impractical, especially in larger settings. A professional conductor interviewed in the initial phases of this project also argued that “verbal communication is ideal for rehearsals, but it might disrupt performances that can be recorded or even the audience might hear the spoken instructions”.

Another practice used in this context is to have assistants next to visually impaired singers on stage, who transmit cues throughout the performance through physical contact (e.g. touching their arm or leg (Menehan, 2009)) following a predefined protocol. Instructions are therefore relayed by a human proxy, possibly introducing delays, errors, but most importantly – dependence on third parties.

3. RELATED WORK AND TECHNOLOGIES

Tools and techniques from HCI were adopted in the music domain to (1) improve digital music generation as well as to produce digital instruments, to (2) develop interactive music systems for education and entertainment purposes as well as to (3) support conductor training (Holland et al., 2013).

3.1 Systems

Bertini and Carosi’s *Light Baton* (Bertini and Carosi, 1993) as well as Marrin and Paradiso’s *Digital Baton* (Marrin and Paradiso, 1997) are both aimed at guiding computers in real-time music-generation via hand-gestures, while Borchers’ *WorldBeat* (Borchers, 1997) as well as Borchers’ et al.’s *Personal Orchestra* (Borchers et al., 2004) provide a baton based interface for interactive music exhibits. In the field of conductor training, Lee et al. presented *You’re the Conductor* (Lee et al., 2004), an interactive music systems for children while

Fonteles et al. (Fonteles, Silva and Rodrigues, 2015) designed and implemented an interactive system aimed at aiding users understand the musical compositions for an orchestral ensemble. In this system, the gestures of the conductor are detected using the *Leap Motion* controller and represented in a 3D space in order to assist “conductors in calibrating their movements” (Fonteles, Silva and Rodrigues, 2015), however this implementation is limited to visualising single gestures, one at a time. *One-Man Orchestra* by Tsui et al. developed a similar system that is intended to act as a training platform for aspiring conductors (Tsui, Law and Fu, 2014). The system uses two *Leap Motion* devices to recognise the conductor’s gestures. Each *Leap Motion* is connected to a computer and the data captured from each controller is combined before gestures are interpreted. This information is then used to control a number of smart phone devices, each representing a specific orchestral instrument or section.

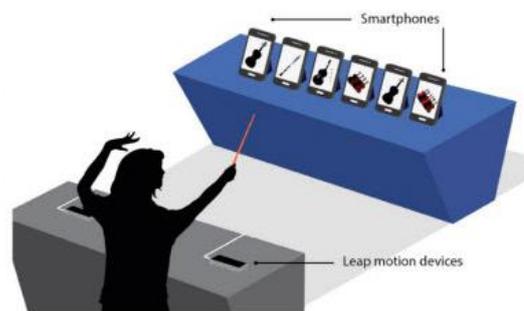


Figure 3 The One-Man Orchestra System (Source: *One-Man Orchestra: Conducting Smartphone Orchestra* [33])

The systems outlined above have made use of a number of sensors, including infrared, pressure, electromyography (EMG), motion and others, while information is transmitted across different channels, such as Bluetooth and TCP/IP. The choice of

sensors and channel needs to be informed by the context of use as well as the users making use of the systems. In choral conducting the basic needs are high-precision unobtrusive gesture recognition and real-time data transmission and user feedback.

For this purpose a number of technologies were evaluated, and a review is presented in the following section.



Figure 4 The WorldBeat interactive conducting system (Borchers, 1997)

3.2 Technologies

Gesture recognition is the process whereby gestures or poses are captured for consumption by a receiver (Sarkar, Sanyal and Majumder, 2013; Mitra and Acharya, 2007). The main goal behind gesture recognition is to recognise gestures performed by a human and use this information as instructions to other processes or devices (Badi and Hussein, 2014). Several gesture controllers exist which are capable of acquiring data related to hand and body movement, including the *Myo* armband (including EMG and omni-directional motion sensors), *Nod* finger ring (able to capture hand and finger movement) and *Leap Motion* (using infrared light and optical sensors to detect objects in a 3D space). The latter device was deemed to be the least obtrusive while offering the highest level of resolution since it captures both arms and hands, together with data about palm and finger position. On the other hand *Myo* is limited to one arm with limited hand/finger data resolution, while *Nod* is also limited to a single finger with limited resolution as to the conductor's overall posture. Having said that, although both *Myo* and *Nod* are both wireless, therefore affording more conductor flexibility, *Leap Motion* provides an acceptable range of detection akin to what conductors are already used to, without the need to introduce any wearable devices – affording a suitable and natural working space for the task at hand (see Figure 5).

The *Leap Motion* gesture controller captures limb properties and movement information on a frame-by-frame basis up to a height of two feet from the device itself. Similar to the *Leap Motion* Controller, PMD Technologies have developed *CamBoard Pico*. This device comes with a 3D gesture camera that is capable of sensing fast and subtle finger and

hand gestures that are within 20cm to 100cm from the device (Lai, 2014).

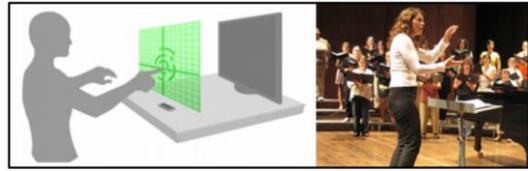


Figure 5 Range of Detection provided by Leap Motion Controller (Images adapted from: <https://www.codeproject.com/KB/game/1007186> and <https://music.washington.edu/areas/choral-conducting>)

Even though *CamBoard Pico*'s features are promising, it is significantly more expensive than a standard *Leap Motion* controller (by a factor of seven). Nonetheless, *Leap Motion*'s 60cm as well as *CamBoard Pico*'s 100cm height limit might still not be enough for certain conductors who over-emphasise gestures during a performance. Furthermore, optoelectronic gesture recognition devices making use of infrared technology may be affected by external sources of light that may hinder sensor precision.

4. DESIGN METHODOLOGY

Interaction design revolves around “creating user experiences that enhance and extend the way people work, communicate and interact.” (Preece, Rogers and Sharp, 2002). A meeting with a national voluntary organisation set up to support visually impaired people was held to gather information about challenges their members encounter in general, as well as to discuss the research problem. Furthermore, a discussion was held revolving around musical activities that their members participate in.

Analysis of the problem domain was conducted with experts in the field to gather knowledge on choral conducting and to better understand the research problem. This analysis included the use of semi-structured interviews with conductors, which were then transcribed and systematically analysed to set the basis for the requirements development process. Braun and Clarke's Thematic Analysis (Braun and Clarke, 2006) was used as a Qualitative Data Analysis (QDA) method to extract qualitative ideas from within the dataset. Thematic Analysis kicks off with a familiarisation stage on the data collected from the semi-structured interviews. This data is then coded, and themes are extracted from these codes. Themes are then reviewed, named and defined, and finally a write-up of the analysis is produced (Guest, 2012). The ideas obtained from the semi-structured interviews were coded, the frequency of the codes were elicited and the extracted data was then used to drive design decisions taken during the development of the

potential solution. Conductors and singers were also consulted during the various prototype evaluation iterations. These evaluations were composed of direct observation with conductors and singers using the technology in a controlled environment. This also involved the collection of audio and video data, as well as structured feedback obtained in debriefing sessions. This feedback was in turn analysed thematically and outcomes were used to inform subsequent interventions. A System Usability Scale (SUS) questionnaire (Sauro and Lewis, 2012) was also used across iterations. This simple usability measure was presented to participants after each intervention to provide a baseline usability score over time.

5. ACE – ACCESSIBLE CHORAL ENSEMBLES

Since different conductors have their own conducting styles, posture and mannerisms, it was necessary to first abstract gestures to their primitive features. The gesture interpreter would determine the type of gesture based on these features, without relying on calibration for each and every conductor – given that conventional conducting gestures are used. The interpreter adapts itself by (a) understanding the starting position of the conductor's hands within the conducting plane (adjusting for height or posture) and (b) predicting conductor intentions based on the most abstract features for different gestures. For instance, by reading the position, movement and velocity of the conductor's right-hand as well as for each individual finger every 50ms, the gesture interpreter could determine the conductor's intentions before the entire gesture is performed (see Table 1). This would otherwise cause a lag in communication considering the time span between the conductor's intention materialising in the completion of a gesture, its recognition and finally transmission to the singers' device. Having the interpreter react at the intention stage and the start of the gesture (with a level of confidence about the gesture) and communicate this command before the full gesture is performed and recognised minimises the risk of late singer reactions. This element of prediction was also necessary to respect the real-time requirement arising from the data. One expert conductor stated that "a gesture needs to be communicated to the singers in a real-time manner, and any delays will affect the outcome of the performance". Prediction was also necessary to mitigate risks of lags arising from communication protocol limitations and receiving devices' processing power.

5.1 ACE architecture

Figure 6 outlines the main architectural components necessary for an end-to-end

prototype. ACE makes use of *Leap Motion* as a gesture controller, capturing six basic gestures performed by the conductor, namely: 'Breathe', 'Start singing', 'Up/Downbeat', 'Crescendo', 'Diminuendo' and 'Closing/cut'. Tempo as well as beat patterns are composed of upbeat and downbeat gestures. For instance, the 3-beat pattern used to signal the 3-beat waltz rhythm is composed of an upbeat movement and two downbeat movements, all of which cross a virtual horizontal line on the conductor plane.

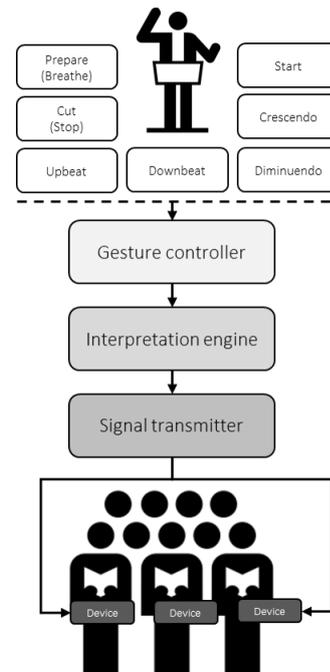


Figure 6 High-level end-to-end architecture

The data obtained from this device is processed and the corresponding gesture is identified. Conductors can also see a visualisation of the gestures being captured on a tablet/laptop placed on the podium, to which the gesture controller is connected. This gesture is translated into its equivalent command, which is then passed to the chorister via a WebSocket connection. The visually impaired choristers' devices are connected to this WebSocket connection which in turn translate the stream of commands into the corresponding signals. This architecture consists of two client-side applications, and an orchestrator which receives interpreted signals and broadcasts such signals to any connected device over a WebSocket connection. Several communication channels and protocols were evaluated for this work, including HTTP-based real-time technologies such as traditional polling and long-polling. Both techniques have their own benefits and shortcomings, however shortcomings such as lost updates and connection overheads respectively drove the investigation towards different alternatives. On the other hand, the WebSocket (WS) protocol is a duplex communication channel that allows for efficient two-

Table 2: Gestures, prediction engine rules and Leap Motion API data points used

Gesture(s)	Movement prediction	API hand-data used
<i>Breathe and start singing</i>	<i>Conductor's right hand in pinching position and accelerating in an upward direction. The maximum height reached determines the point from which the start singing instruction will be observed, activated after any downward movement at any speed and distance (normalising for subtle and dramatic movements)</i>	<i>pinchStrength (providing a value in the range [0.. 1] for the strength of a pinching pose), palmVelocity (providing the rate of change of the palm position in millimetres per second) and palmPosition (which gives the distance from the Leap Motion origin to the centre of the palm in millimetres).</i>
<i>Tempo</i>	<i>The engine determines the conductor's initial vertical hand position (conducting plane) and creates a virtual horizontal barrier which is used to determine upbeat and downbeat gestures at any velocity and relative height – respecting individual conductor's own style</i>	<i>palmVelocity and palmPosition (initial [x,y,z] coordinates as observed from the first gesture)</i>
<i>Crescendo</i>	<i>The engine determines and upward left hand movement relative to the current hand position with an open palm facing upwards.</i>	<i>palmNormal (which identifies the pointing direction of the palm, and thus is used to determine whether the palm is facing upwards or downwards), an array containing finger position data and palmVelocity</i>
<i>Diminuendo</i>	<i>Engine follows a mirrored version of the prediction rules used in the crescendo gesture.</i>	<i>palmNormal, an array containing finger position data and palmVelocity</i>
<i>Closing/cut</i>	<i>The engine predicts this gesture when the right hand is closed beyond a certain threshold, irrespective of the hand's x, y or z positions and velocity.</i>	<i>grabStrength (which gives the strength with which a hand is forming a grab pose, with values ranging from 0 (open hand) to 1).</i>

way communication between a client and a server in a stateful manner. This protocol allows for a persistent TCP connection to be established over which messages may be broadcast with minimal overheads associated with traditional HTTP requests, resulting in very low latency. WS messages are sent as data frames over an established duplex channel (Fette and Melnikov, 2011).

Communication channels such as Bluetooth, WiFi and 3/4G were investigated, however WiFi was deemed to be an ideal channel based on (1) availability, (2) ease of setup and use, (3) transmission rates, (4) non-restrictive physical layout, (5) support for many concurrent connections, (6) latency, (7) channel robustness and (8) running costs.

5.1.1. Conductor application

The conductor application combines the gesture controller together with the interpretation engine within a web-based Javascript client. This client receives frames generated by the gesture controller which frames are then sent to the interpretation engine before transmitting to connected clients. *Leap Motion's* JavaScript API was used, producing a Frame object containing lists of entities, such as hands, fingers, and tools, as well as basic gestures and motions. The interpretation engine (see Figure 6) reads these frames and attempts to interpret palm and hand data according to the pre-selected set of gestures. The interpretation engine starts by determining which hands are included in the frame.

Based on this condition, the respective gesture functions related to the detected hand(s) are traversed to predict whether any known gesture is being performed by the conductor. In the eventuality where both hands are detected, the functions related to both the right hand and the left hand are executed in parallel. It is also assumed that conductors will use each hand for a specific set of gestures, of course adjusting for left and right-handedness. The interpretation engine has been scoped down to a few basic gestures as shown in Table 2. Once gestures are predicted (see Table 2) these are then broadcast to any connected device via the signal transmitter.

5.1.2. Signal Transmitter

The signal transmitter adopts a real-time architecture based on the WebSocket protocol. This is implemented as a SignalR server, on top of the ASP.Net framework. A WebSocket server can continuously push content to connected clients over established duplex channels. SignalR makes use of WebSocket transport when possible, and falls-back to other transport techniques (such as long-polling) when the WebSocket protocol is not supported by the client's browser. When singers load the web application on their mobile devices (by visiting a bookmarked webpage), they are automatically registered with the SignalR server, which maintains a list of connected devices. The conductor application also connects to the SignalR server as a client, however it is also able to communicate upstream – sending interpreted gestures to be broadcast to all connected chorister

devices. For this project, the SignalR server was hosted on a cloud infrastructure (Azure). Client devices are registered to a broadcast-list based on a unique identifier. An IP range can be used to group clients together based on proximity, however grouping could also be specified using unique choir names to which devices are associated. This registration initiates a WebSocket connection handshake with the SignalR server which in turn adds clients to its broadcast list. The signal transmitter is configured to broadcast any message it receives from the interpretation engine to the respective broadcast list.

5.1.3. Client application

Since WebSockets are supported by most modern web browsers, it was decided to write client applications as web applications. This removes the need for choristers to install native applications on their devices. Client apps are loaded via a URL after which service workers (running in the background) take over and execute the listening and notification algorithms.

Humans use their five senses subconsciously in daily routines, but visually impaired people have to “rely on sensory information from audition and touch” due to the loss of their visual sense (Wan et al., 2010). Studies (Badi and Hussein, 2014; Sathian and Stilla, 2010) have shown that the non-visual (auditory and tactile) abilities of visually impaired people are enhanced when compared to those of sighted people. In their study, Wan et al. (2010) showed that visually impaired people performed better in tasks involving auditory perception compared to people with visual abilities. For these reasons, and given the capabilities of modern web-browsers, it was decided that the choristers’ client application should afford audio and haptic feedback to transmit instructions using the browsers’ built-in media capabilities¹ and the Vibration API² respectively (see Figure 7). Haptic feedback was used for instructions such as tempo changes, close/cut, crescendo and diminuendo, whereas in-ear audio feedback was used to signal cuing instructions, mainly breathe (i.e. prepare) and start singing (i.e. first downbeat).

6. USER EVALUATION

This interactive system was taken through several stages of evaluation, each of which informed the design for subsequent interventions. Initially the interpretation engine was presented to and evaluated by three professional conductors. The

interpretation engine was improved to better capture subtle gesture nuances.



Figure 7 Initially, the distribution of instruction signals were split across two senses

Eventually, an on-screen instruction visualiser was developed and used to test the interpretation engine’s output on individual singers (with no visual impairment). This was necessary to understand the time taken between when a gesture is performed and when this is interpreted, displayed on screen and reacted to by singers. Finally the end-to-end prototype was evaluated with a small ad-hoc choir which was called up a number of times within a span of one month. Direct observation, video recordings and exit questionnaires were used during each session.

6.1 Professional conductors

Three professional conductors were individually invited to review the first stage of the overall system – the gesture interpretation engine. These sessions were recorded (audio only) and data was transcribed for further analysis. The following were the main points arising from these sessions:

The 600mm range of motion imposed by *Leap Motion* might be problematic for some conductors.

- Downbeat signal should be emphasised over the upbeat signal as singers take the downbeat as a reference point for tempo.
- Although most gestures are standard, a level of familiarisation was still necessary.
- Choir section cuing was necessary when conducting large groups with different vocal sections (sopranos, altos, tenors and basses).
- Certain conductors might want a delay between when the gesture is performed to when the signal is transmitted.
- Conductor profiles and calibration may be necessary for more complex gestures.

Usability issues were fixed while recommended features were also added (e.g. downbeat signal) before moving on to the next evaluation stage.

¹ <https://developer.mozilla.org/en-US/docs/Web/API/HTMLMediaElement> (accessed January 2017)

² https://developer.mozilla.org/en-US/docs/Web/API/Vibration_API (accessed January 2017)

6.2 Singers and Musicians

Another three participants were individually invited to evaluate the speed at which gestures are interpreted and communicated. At this stage a visual representation of interpreted gestures was provided on-screen to which participants had to react while singing a simple tune. A level of frustration arising from cognitive workload was observed when on-screen instructions were introduced during a performance.

6.3 Choir Iterations

Finally two further iterations were planned with a small choir. A song with which everyone was familiar was selected (the national anthem) so as to control for additional learning curves. All choristers were blindfolded and given a mobile device (and an ear-piece) which was pre-registered with the signal transmitter. Sessions consisted of a familiarisation period, and two tasks. During the first task the song had to be performed without changes in dynamics (i.e. consistent tempo) whereas in the second task new dynamics were introduced during the performance without any prior knowledge of when or what these might be (e.g. changes in tempo, stopping and starting mid-way as well as crescendos and diminuendos).

Audio and video data was captured throughout each iteration, along with exit questionnaires to record usability perceptions. After each session, participants also engaged in a discussion about their experience while using the system. Data from each iteration was used to improve the overall design, which changes would then be tested in subsequent iterations. The following are the main issues encountered during each iteration.

6.3.1. Iteration 1

In this session, the three participants were blindfolded and equipped with the initial prototype. All participants commented that, in both tasks, there was an overwhelming amount of information being received while performing, split between audio signals and device vibrations. They all agreed that some signals could be eliminated, transmitting changes to dynamics rather than all dynamics (e.g. transmit changes in tempo rather than sending all upbeat and downbeat signals). Participants also remarked that audio signals were generally received with a random delay, making it difficult to follow gestures such as 'breathe' and 'start'. They argued that these are very important gestures and should be transmitted in a synchronised and real-time manner. It was also noted that contrary to the audio signals, haptic feedback was generally received in a timely manner. Participants agreed that the 'breathe' and 'start' instructions can also be translated into haptic signals, and that they would prefer to hold a second

device for these instructions rather than receive signals over a separate channel employing a different sense.



Figure 8 Evaluation with three singers using both audio and haptic signals

All three participants rated the system usability as 'Moderate' through the SUS questionnaire, indicating that further refinements could be performed on the system being developed.

6.3.2. Iteration 2

Following the first iteration, it was observed that audio feedback was causing more confusion. This was also the result of an unacceptable delay between signal transmission and audio feedback via the browser. Although the signal was being received in a timely manner, the browser was taking an additional 500ms to 1 second before executing the 'play audio' command, even with small audio files. For this reason audio feedback was abandoned in favour of a second device providing a second set of haptic feedback. Further optimisations on gesture recognition, particularly for tempo related gestures, were also carried out. Issues encountered in the second iteration:

- Certain haptic signals were not always distinguishable enough, especially when gesture were performed in short successions or in pulses (e.g. quick crescendo).
- Although most dynamics were interpreted, transmitted and reacted to correctly, including crescendos, diminuendos and start/stop commands some singers experienced missed tempo signals. This caused a significant amount of confusion especially when singers started going out of sync. This issue was primarily noticed on low-spec devices, as such behaviour was not experienced in higher-end phones.

Following this session, further improvements were made on the overall system design, including optimisation on the interpretation engine to improve handling of concurrent gestures as well as the development of more distinguishable haptic

feedback patterns to better distinguish between short-lived instructions (e.g. crescendos and diminuendos in quick succession). This was achieved through the use of different vibrating-patterns (e.g. pulsating vibrations for diminuendo and continuous vibrations indicating crescendos). The ratings obtained from the SUS questionnaire presented at the end of the session indicate an improvement in the overall usability of the system. In a final round of evaluation, an increased level of acceptance was reported from all participants. Overall, the participants commented positively on the use of haptic feedback on two separate devices across which all instructions are communicated.

7. CONCLUSIONS AND FUTURE WORK

This project looked into the possibility of developing an inclusive space for visually impaired singers to be full participants in choral activities. This study provided us with insights and experiences to confidently conclude that with minor improvements, ACE could be adopted in the field.

The architecture presented in this paper offers visually impaired singers the opportunity to participate in choral ensembles without: (a) dependence on facilitators, (b) the need for audible verbal cues and (c) without limiting live performances to a set of pre-rehearsed dynamics. Furthermore ACE affords conductors the ability to conduct an ensemble which includes visually impaired singers without the need to alter conducting gestures. In terms of technology, ACE adopts readily available infrastructures (i.e. WiFi) and technologies (i.e. web enabled devices and WebSockets) to deliver a real-time environment necessary to communicate just in time instructions. An iterative development approach was taken for the implementation of a prototype, in which feedback from experts and choir members was collected using both qualitative and quantitative methods. This feedback was crucial in the development of ACE. Conductors who participated in this project expressed interest in ACE for their choral activities, thereby encouraging inclusion of visually impaired singers. Future work is of course necessary, and we envisage the following aspects, in no order of preference:

- Inclusion of more complex gestures, including advanced time signatures (e.g. 6/8 or 5/4) and choir section cuing (e.g. preparing baritones).
- Better range of motion, while investigating the possibility of using two gesture controller devices. Two devices were at one point being used in this project, however it was decided to scope out this aspect in favour of other core challenges.

- Conductor calibration could allow for better gesture interpretation across conductors, who may use different postures or motions for similar commands.
- Working with a group of visually impaired singers to elicit in-depth experience feedback.
- Investigate the use of wearable devices to minimise dependency on smart-phones (e.g. smart watches and custom-built wrist-bands).

Participants in this study were not part of the target population and the condition had to be simulated through blindfolding. Although we consider this to be a limiting factor when it comes to user evaluation, this study is considered to have wider reaching implications, taking several components of the problem at hand into account, including (a) an understanding of the state of the art, (b) domain-specific gesture recognition (to reduce adoption friction for choral conductors), (c) real-time and ubiquitous technology (to reduce adoption friction for participants and ensembles) as well as (d) human factors and ergonomics.

We understand that sensory skills and sensitivity to haptic feedback may be more accentuated within the target population (see section 5.1.3) also due to more experience in making use of haptic information in other contexts. We therefore expect this system to afford higher communication resolutions (i.e. gesture types and frequency) when adopted by visually impaired choristers. Given the size of the country in which this study was carried out (Malta), finding volunteer visually impaired choral singers was very difficult, even though national organisations were involved from the outset. We do believe however that the overall set of insights generated through this work lends itself directly for adoption, and further research, with visually impaired singers wishing to be part of an accessible choral ensemble.

8. REFERENCES

- Badi, H. S., & Hussein, S. (2014). Hand posture and gesture recognition technology. *Neural Computing and Applications*, 25(3), 871-878. doi:10.1007/s00521-014-1574-4"
- Bertini, G., & Carosi, P. (1993). Light baton system: A system for conducting computer music performance. *Interface*, 22(3), 243-257. doi:10.1080/09298219308570635
- Borchers, J. O. (1997). WorldBeat: Designing a baton-based interface for an interactive music exhibit. Paper presented at the Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA. 131-138.

- Borchers, J., Lee, E., Samminger, W., & Muhlhauser, M. (2004). Personal orchestra: A real-time audio/video system for interactive conducting. *Multimedia Systems*, 9(5), 458-465.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. doi:10.1191/1478088706qp063oa
- Cox, A. (2010). Musical gestures: Sound, movement and meaning. rolf inge godoy and marc leman (eds.). *Musicultures*, 37(0)
- DeVenney, D.P. (2010). The Right Hand. In *Conducting Choirs, Volume 1: The Promising Conductor*. Roger Dean Publishing Company, a Lorenz Company.
- Durrant, C. (2018). *Choral conducting*. New York: Routledge.
- Fette, I., & Melnikov, A. (2011). The WebSocket protocol. (). IETF: Internet Engineering Task Force; Internet Engineering Task Force (IETF). Retrieved from <https://tools.ietf.org/html/rfc6455>
- Fonteles, J. H., Silva, S., & Rodrigues, M. A. F. (2015). Gesture-driven interaction using the leap motion to conduct a 3D particle system: Evaluation and analysis of an orchestral performance. *SBC Journal on Interactive Systems*, 6(2), 11-21.
- Guest, G. (2012). *Applied thematic analysis. Introduction to applied thematic analysis* (pp. 3-20). Thousand Oaks, California: SAGE Publications, Inc.
- Lai, R. (2014). Exclusive: PMD's CamBoard pico XS is the tiniest gesture camera we've ever seen <https://www.engadget.com/2014/01/10/pmd-camboard-pico-xs-gesture-camera/> (accessed April 2018).
- Lee, E., Nakra, T. M., & Borchers, J. (2004). (2004). You're the conductor: A realistic interactive conducting system for children. Paper presented at the In Proceedings of the NIME 2004 Conference on New Interfaces for Musical Expression, 68-73.
- Manzanelli, O. C. (2010). Conducting a blind choir: An experience from argentina. *Europa Cantat Magazine: Music and Handicap*, 3
- Marrin, T., & Paradiso, J. (1997). (1997). The digital baton: A versatile performance instrument. Paper presented at the Proceedings of the International Computer Music Conference, 313-316.
- Matev, P. (2010). The academician petko staynov choir: A mixed professional choir of the blind in sofia, bulgaria. *Europa Cantat Magazine: Music and Handicap*, 3
- McClung, A. C. (1996). The relationship between nonverbal communication and conducting: An interview with rodney eichenberger. *Choral Journal*, 36(10), 17.
- Menehan, K. (2009). With access and accommodation for all. *The Voice of Chorus America*, 32(2)
- Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3), 311-324. doi:10.1109/TSMCC.2007.893280
- Park, H. (2015). How useful is braille music?: A critical review. *International Journal of Disability, Development and Education*, 62(3), 303-318. doi:10.1080/1034912X.2015.1020921
- Royal National Institute of Blind People. (2016). Braille music, <http://www.rnib.org.uk/braille-and-moon-%E2%80%93tactile-codes/braille-music>.
- Preece, J., Rogers, Y., & Sharp, H. (2002). What is interaction design? *Interaction design* (1st ed., pp. 1-31). New York, NY, USA: John Wiley & Sons, Inc.
- Sarkar, A. R., Sanyal, G., & Majumder, S. (2013). Article: Hand gesture recognition systems: A survey. *International Journal of Computer Applications*, 71(15), 25-37.
- Sathian, K., & Stilla, R. (2010). Cross-modal plasticity of tactile perception in blindness. *Restor Neurol Neurosci*, 28(2), 271-281. doi:10.3233/RNN-2010-0534
- Sauro, J., & Lewis, J. R. (2012). Quantifying the user experience: Practical statistics for user research Elsevier/Morgan Kaufmann. Retrieved from <https://books.google.com/mt/books?id=QncH02-bvjoC>
- Smaligo, M. A. (1999). Resources for helping blind music students. *Future Reflections*, 18(1)
- Tsui, C. K., Law, C. H., & Fu, H. (2014). One-man orchestra: Conducting smartphone orchestra. Paper presented at the SIGGRAPH Asia 2014 Emerging Technologies, Shenzhen, China. 11:1-11:2. doi:10.1145/2669047.2669049 Retrieved from <http://doi.acm.org/10.1145/2669047.2669049>
- Wan, C. Y., Wood, A. G., Reutens, D. C., & Wilson, S. J. (2010). Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia*, 48(1), 344. doi:<http://dx.doi.org/10.1016/j.neuropsychologia.2009.08.016>
- Van Weelden, K. (2002). Relationships between perceptions of conducting effectiveness and ensemble performance. *Journal of Research in Music Education*, 50(2), 165-176. doi:10.2307/3345820