Haptic and Visual feedback in 3D Audio Mixing Interfaces

Gelineck, Steven; Overholt, Daniel

Published in:
Proceedings of the 9th Audio Mostly Conference

DOI (link to publication from Publisher):
10.1145/2814895.2814918

Publication date:
2015

Document Version
Accepted author manuscript, peer reviewed version

Haptic and Visual feedback in 3D Audio Mixing Interfaces

Steven Gelineck
Aalborg University Copenhagen
Department of Architecture, Design and Media Technology
A. C. Meyers Vænge 15, 2450 Copenhagen, Denmark
stg@create.aau.dk

Dan Overholt
Aalborg University Copenhagen
Department of Architecture, Design and Media Technology
A. C. Meyers Vænge 15, 2450 Copenhagen, Denmark
dano@create.aau.dk

ABSTRACT
This paper describes the implementation and informal evaluation of a user interface that explores haptic feedback for 3D audio mixing. The implementation compares different approaches using either the LEAP Motion for mid-air hand gesture control, or the Novint Falcon for active haptic feedback in order to augment the perception of the 3D space. We compare different interaction paradigms implemented using these interfaces, aiming to increase speed and accuracy and reduce the need for constant visual feedback. While the LEAP Motion relies upon visual perception and proprioception, users can forego visual feedback with interfaces such as the Novint Falcon and rely primarily on haptic cues, allowing more focus on the spatial sound elements. Results of the evaluation support this claim, as users preferred the interaction paradigm using the Falcon with no visual feedback. Furthermore, users disliked active haptic feedback for augmented perception of 3D space or for snapping to objects.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: Audio input/output.; H.5.2 [**User Interfaces**]: Input devices and strategies.

General Terms
Human Factors, Design, Experimentation

Keywords
User Interface, 3D Audio, Mixing, Haptic Feedback, Mid-air gestures, Leap Motion, Visualisation, Haptic User Interface, Novint Falcon

1. INTRODUCTION
Spatial 3D audio is becoming more and more common in cinema, games and other artistic contexts [12, 10]. Technology has matured and commercial standards for 3D audio are

AM '15 October 07 - 09 2015, Thessaloniki, Greece
ACM 978-1-4503-3896-7/15/10 ...$15.00.
http://dx.doi.org/10.1145/2814895.2814918

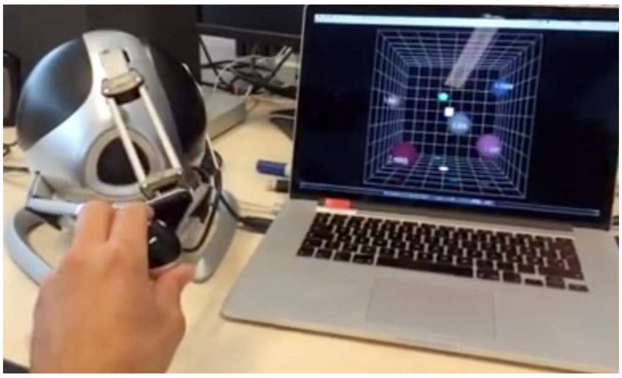

Figure 1: The Novint Falcon used to manipulate the position of sound sources in the 3D virtual space.

fairly well established[1 2]. While research on user interfaces for mixing of 3D audio is still in its infancy, some interesting approaches are emerging. Commercially, user interfaces currently implement traditional faders and rotary knobs as well as 2D joysticks for positioning and manipulating audio sources in three dimensions. However, more progressive commercial interfaces include the EVO with 3DAW from Fairlight[3], which adds support for the LEAP motion controller enabling users to position audio sources using free hand movements.

A recent study [5] investigated different approaches comparing the LEAP motion controller to a more traditional mixing console interface. This study suggested that while there was great potential in using an integral controller with 3 degrees of freedom, where movements are directly coupled to movements of audio sources in 3D space, there were challenges to do with especially precision, ergonomics and fine-tuning because of the lack of tangible or haptic feedback of the interface. Additionally, the results of the study suggested that engaging in interfaces that rely heavily on visual feedback decreases the listening experience (a phenomenon also observed by Lech & Kostek [8]).

Here we extend the research by investigating the use of a haptic feedback device for assisting the user not only in augmenting the feel for the 3D space, but also potentially en-

[1] `http://www.auro-3d.com`
[2] `http://www.dolby.com/us/en/brands/dolby-atmos.html`
[3] `http://www.fairlight.com.au/product/3daw/`

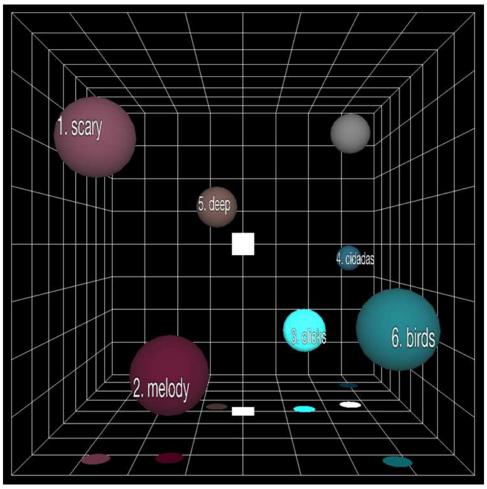

Figure 2: The GUI displays a 3D cube representing the available positioning space. Each virtual sound source is represented by a coloured sphere and the cursor is represented by a grey sphere. Shadows help the user perceive depth positions of each sphere.

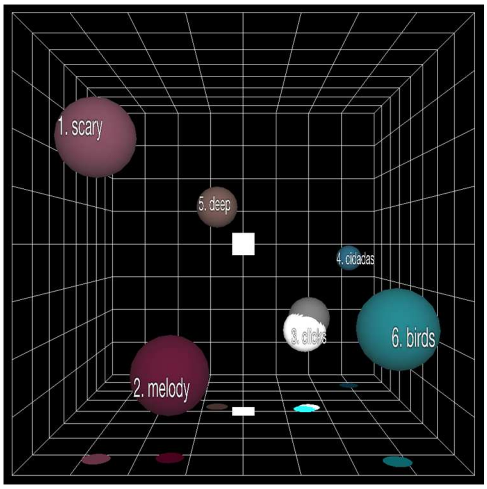

Figure 3: When the cursor is close enough to a sound source it can be selected and it turns white.

abling faster and more accurate selection and positioning of different audio sources. Finally, the presented work explores how to make the user less dependent on constant visual feedback (as seen when sound engineers close their eyes while adjusting parameters on a traditional mixing console). In this work sound is represented binaurally (with headphones using a generalised Head Related Transfer Function (HRTF)), but the mixing techniques implemented can easily be applied to other 3D object-based sound rendering systems, such as the multi-speaker variants Dolby® ATMOS, DTS:X, and the upcoming MPEG-H standard [7].

The paper begins by presenting related works leading to the considerations we have had in terms of designing different solutions implementing haptic feedback for 3D mixing. Next, we describe the system implementation focusing on how the force feedback schemes have been implemented, followed by an informal evaluation carried out as an initial pilot study of the potential of those different interaction schemes. Finally, we conclude on the findings of the evaluation an provide suggestions for further work within the area.

2. RELATED WORKS

Spatial positioning of audio sources in 1D (stereo) and 2D (surround sound) is carried out using dedicated controllers that fit well to the job at hand—rotary knobs for simple panning between two speakers, or 2D joystick/interactive surface for surround sound panning. Here the degrees of freedom of the controller fits well with the degrees of freedom of the audio space. The challenges appear, when moving to a third dimension, which requires 3 degrees of freedom. In an article by Mathew et. al [10] they conclude that there is a *"lack of well-suited controllers"* for object based spatial audio production systems. After conducting an extensive survey on how artists produce for spatial audio, Peters et. [15] found that a key importance when producing spatial audio is controllability via external controllers.

Several approaches to mid-air musical control, including the control of spatial audio, exist [14, 16, 9, 4]. After having explored mid-air gesture controls [5] we were interested in incorporating haptic feedback. The technology has matured and several devices and software libraries are available, including the Open Source Haptics library [1] used in in this study — see more in the *Implementation* section. Sinclair provides a good overview of related works on using haptic feedback devices for musical purposes [17]. Further advances in this area include [13, 2, 3]. Most relevant for the research presented here is the work by Melchior et. al [11], which compares different forms of haptic feedback with mouse and slider input for controlling different movements of audio sources in 3D space. We build upon their research by exploring methods for augmenting the perceptual feel of the 3D space (by haptically providing information about the current 3D position), for fast and precise selection and positioning of audio sources and for decreasing the dependency on visual feedback when mixing for 3D audio.

3. SYSTEM DESIGN

The haptic 3D audio mixer is built around a virtual 3D space, where one is able to move and position virtual sound sources in relation to a centred listening position (see [5] for more details[4]). For rendering of the 3D audio, we use a binaural audio setup built in Max/MSP based on the binaural spatial panning tool, which is part of the Spat software suite[5]. Here it is possible to control the distance, azimuth angle and elevation angle for an arbitrary amount of audio channels. For the exploratory evaluation presented later we used 6 different sound sources each representing differences in timbral, rhythmical and frequency content.

The virtual space is visualised as a 3D cube in which audio sources can be positioned by the user — see Figure 2. Each sound source is visualised as a coloured sphere with a text in front of it communicating the number and name of the sound source (for instance *"3. clicks"*). An additional grey

[4]videos: `http://media.aau.dk/~stg/3dAudioMixing.html`
[5]`http://forumnet.ircam.fr/product/spat/`

sphere represents the position of the controller (the *cursor*), when not interacting directly on any of the sound sources. Once the user moves this cursor close to a sound source, the coloured sphere of that sound source turns white indicating that the sound source can be selected and thereafter manipulated — see Figure 3. Finally, a white cube represents the location of the listener in the center of the 3D space.

3.1 Control Schemes

The focus of this paper is on assessing *different* approaches to using haptic feedback while manipulating sound sources in 3D. Therefore we have implemented several different control schemes, which all provide different forms of haptic feedback. Each input scheme is implemented using the Novint Falcon[6], which is commonly used as a game controller but also heavily used as a low cost solution in academic contexts. Additionally, we were interested in contrasting this to mid-air non-haptic interaction, which is why we have implemented an input scheme based on the LEAP motion controller[7]. Thus, the following six control schemes were developed:

- **LEAP Motion controlled** - users control the position of the audio source by hovering their hand above the LEAP motion. They select/deselect sound sources by pressing/releasing the space bar on a computer keyboard.
- **Falcon with no feedback** - users control the position of the audio source with the position of the Falcon controller. No force feedback is induced, so the user only feels the passive mechanical haptic feedback of the robotic arm itself. The user presses/releases a button on the device using the same hand as is used to manipulate the translation of the controller in order to select/deselect the desired audio source.
- **Falcon with pull towards centre** - same control scheme as above, but with a slight gravitational pull towards the centre of the virtual space. This was implemented in order to provide haptic feedback as to where in the 3D space one is located. The feedback emulates that of having an elastic band pulling towards the center (the further away from the center, the more force is applied).
- **Falcon with pull towards centre and snapping to sound sources** - same control scheme as above, but with a slight snapping to sound source locations as one moves close enough. This emulates the sort of haptic feedback one feels in the centre position of most stereo panning knobs. Note that the snapping only occurs when no sound source is selected as it is only meant to help the user select a sound source (when actually moving a sound source around one wants to move freely without sudden movements each time one passes another source.).
- **Falcon with pull towards centre, snapping to sound sources and/or key press to target sound sources** - same control scheme as above. Additionally we provide the user with an option to press a key on the computer keyboard (1-6), which induces a force that pushes the haptic device to the position of the sound source associated with the number that was pressed. The idea here is to enable the users to take their eyes off the screen and concentrate solely on positioning and listening.
- **Falcon with pull towards centre, snapping to sound sources and/or key press to target sound sources and no visual feedback** - same control scheme as above. However, here we remove any form of visual feedback, forcing the subjects to rely solely on the haptic and auditory feedback.

The way in which sound sources are selected here differs from [11], in which users "re-center" the virtual cursor to the position of the controller with a button press (this way the user does not have to move to the sound source to select it). There are drawbacks with both approaches, however we chose the approach described above to avoid the problem of sometimes "re-centering" the cursor at a physical boundary of the actual controller.

Several other haptic feedback schemes could be imagined including restricting the user to certain planes on demand, or to auto-aligning to other sound source's x/y/z positions (as seen in graphics software such as Photoshop and Illustrator). An interesting approach is also to first define a trajectory path and secondly trace along the path paying more attention to timing (as implemented in [11]).

Finally, we must note that when adding more forces to the system, one also increases the risks of instability leading to oscillation because the haptic feedback acts as a force feedback loop and because different forces are added together. With careful tweaking of the parameters, the goal for us has been to make the haptic feedback subtle enough to avoid oscillation while still being strong enough to be useful for the interaction. Future studies will include more structured testing of the optimal parameters in order to experimentally understand this balance.

4. IMPLEMENTATION

In this section we briefly provide an overview of the overall system implementation, after which we describe how force feedback for center pull, snapping and sound source targeting was implemented.

The graphical interface and LEAP motion input was implemented in Objective-C/Cocoa using OpenGL for displaying the graphical interface and the LEAP SDK for handling input. As mentioned earlier, the audio engine was implemented in Max/MSP using the binaural spatial audio objects that are part of the Spat software package from Ircam (we use the standard KEMAR HRTF). Finally, the position and button press data received from the Falcon and the force data sent to the Falcon is controlled in the same Max/MSP patch using Max objects distributed as part of the Open Source Haptics library[8]. All the haptic functionality is thus programmed in Max/MSP, which acts as both

[6] `http://www.novint.com/index.php/products/novintfalcon`
[7] `https://www.leapmotion.com`
[8] `http://openhaptics.org/`

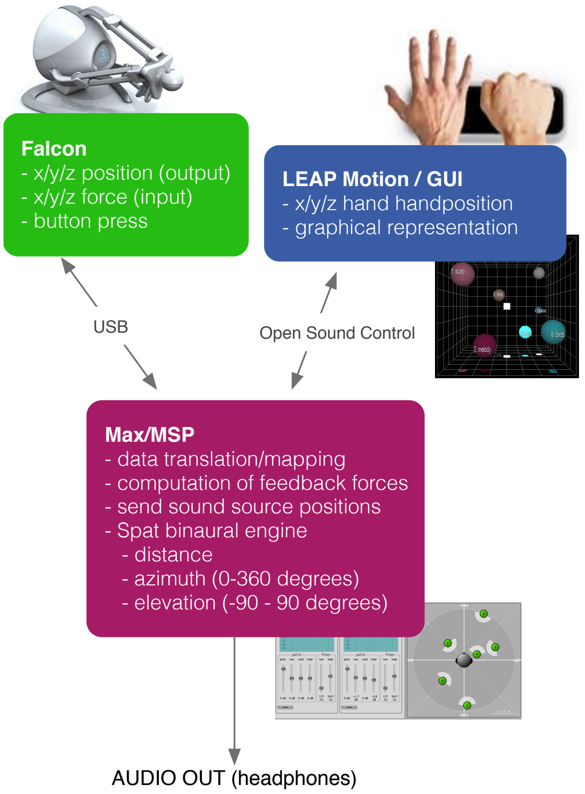

Figure 4: Overview of the overall system architecture.

the audio engine and the main communication hub of the overall system—see Figure 4.

Regarding the force feedback, three different overall functionalities have been implemented: center pull, snapping and sound source targeting. The center pull is simply implemented as a spring/damping system constantly exciting a subtle force on the controller in the opposite direction of its position relative to the center. Here the spring stiffness was set to 35 N/m with a damping of 0.1 N/(m/s).

The snapping has been implemented by emulating a subtle stick/slip motion similar to that of plucking a string (using the *contact-pluck˜* abstraction in the Open Haptics library). Here we have used a stiffness of 130 N/m, damping of 0.5 N/(m/s) and a minimum displacement difference for contact of 6 mm. A simple thresholding makes sure that the stick/slip emulation is only active when the position of the cursor is actually near enough to the position of the sound source to select it.

Finally, the automatic targeting of sound sources has been implemented using the same approach as the center pull. When the user hits a number key (for instance number "3" on the computer keyboard), we induce a strong force (stiffness of 500N/m, damping of 8 N/(m/s)) pulling the controller towards the position of audio channel number 3. In order to avoid the Falcon overshooting, which would result in a strong force pulling it back in the opposite direction, which would then lead to oscillation, we do not set the center of gravity of the force to the desired position immediately. Instead we slowly interpolate between the position of the cursor and the position of the audio source over a timespan of 400 ms. This creates a smooth motion ending at the position of the selected audio source.

5. EVALUATION

An exploratory pilot study was conducted in order to gather initial feedback from users as to the potential of the proposed control schemes. The goal was to understand whether the introduction of haptic feedback would (1) augment the feeling of the 3D space, (2) improve the speed and precision of the selection and positioning of the sound sources and (3) lower the dependency on visual feedback. Three test subjects (both with more than 8 years of mixing experience) were asked to try all six different control schemes in an exploratory fashion. For each they were given the task to imagine that they had just acquired the interface and were trying it out for the first time exploring its capabilities. As part of the exploration they were also asked to create a rough mix of the 6 audio channels. After having explored each control scheme they were asked to comment on their experience. Finally, after having tried all schemes they were asked to compare them in terms of the three goals mentioned above. Detailed notes were taken during the test sessions. The notes were analysed using a critical incidents approach where critical events relevant to the overall purpose of the evaluation were identified. These were then categorised into different themes and used to compare the different interfaces.

6. RESULTS & DISCUSSION

Since there were only three participants in the study, the following results are only indicative. They serve to provide an initial idea of the potential of different haptic interaction schemes. A future study will have to be conducted in order to properly assess this potential.

In regards to the LEAP motion control scheme all participants found the interaction reasonably natural and easy to use with minor comments including it being a little too sensitive and not as ergonomically pleasant as the Falcon. However, when beginning the first interaction scheme implementing the Falcon all participants were surprised at how much easier, faster and more accurate they were able to select and position the sources. After having explored the control scheme with no active haptic feedback, one participant stated:

> *"Wow, this is 10 times better.. I am much faster and more secure"*

The difference experienced was not due to any imprecisions of the LEAP motion controller. In fact the LEAP motion works at a frame rate of 200 fps and with an accuracy of below 0.2 mm [19], and it was actually perceived as having less latency than the Novint Falcon. The experience of the LEAP motion being less accurate in use was most likely due to the nature of the freehand gestures having no haptic feedback, thus making the hand less steady.

The control scheme only implementing passive haptic feedback was preferred over both the center pull and the snapping control scheme, because participants felt more natural and free in their movements. The center pull haptic feedback felt counter-intuitive for all three participants. One participant said that it actually influenced his mix. Another stated that

> *"It feels like I have to fight against the controls a bit."*

The snapping to sound sources control scheme was not appreciated either. Here subjects reported that they liked the idea of getting help with fast selection, but that the snapping actually was in the way for their free movement. One stated that

> *"It was sort of in the way - it was pulling me away from what I was doing"*

Two of the three participants stated that they could imagine this feature working if the implementation would have somehow been less distracting.

The control scheme implementing key presses for automatic translation to a sound source worked well but was not preferred over the simple Falcon implementation with no active haptic feedback. However, all three participants stated that they would probably be able to work fast and accurate with it after some practice. It was observed that participants used a combination of relying on the visual feedback moving the cursor towards the sound source they wanted to target while also hitting the key for that sound source.

Participants all strongly preferred the last control scheme implementing the Falcon with key press to target sound sources **with no visual feedback.** This was surprisingly apparent even when just observing the participants. They would gaze into the distance or close their eyes to really listen to the audio content in a completely different way than with visual feedback. All participants were really surprised about the difference it made having no visuals and how intuitive and natural the controls felt. One stated that

> *"I was really worried at first, whether I would be able to locate the sounds or use the controls properly with no visual reference, but I was much better than before - it felt like I was much more forced to use my senses. It just felt totally different."*

Another participant who seemed very surprised about how intuitive the controls were stated:

> *"Really really cool.. it was almost scary how intuitive it felt. I have never tried anything like it. It felt like I was almost inside the sound."*

Other comments compared the feeling to how a physical mixing console can perform compared to using a Digital Audio Workstation (DAW), where the lack of visual feedback for the console forces one to listen more. The same phenomenon was also observed in an earlier study by the authors when testing tangible tabletop mixing interfaces [6].

7. CONCLUSION

We have presented a new interface for 3D spatial audio mixing, which implements different forms of haptic feedback using a Novint Falcon device. Different haptic interaction schemes were compared to a mid-air freehand gesture control scheme using the LEAP Motion in an informal evaluation. Haptic control schemes included (1) center pull, for augmentation of the perception of 3D space, (2) snapping to objects, for faster sound source selection and (3) automatic translation to selected targets, in order to reduce dependency on visual feedback. Although the evaluation was very informal, we strongly believe that there is a great potential in using a haptic controller like the Falcon for 3D audio mixing, because it achieves many of the same benefits as found on a more traditional mixing console. These include tangible feedback for a sense of precision and especially the reduced need to look at a screen, which leads to the users focussing more on the task at hand—namely listening. Interestingly, participants did not appreciate the active haptic feedback involved with center pull and snapping, stating that it felt counter-intuitive. Future studies will include a performance test of speed and accuracy testing improved schemes as the ones tested above using a methodology similar to the one used in [18].

8. ACKNOWLEDGEMENTS

We would like to thank Dannie Korsgaard for helping interface with the LEAP Motion controller and Francesco Grani for initial thoughts on the research. Finally, we would like to thank the participants who helped evaluate the interface.

9. REFERENCES

[1] E Berdahl, A Kontogeorgakopoulos, and D Overholt. Hsp v2: Haptic signal processing with extensions for physical modeling. In *5th International Workshop on Haptic and Audio Interaction Design-HAID, Copenhagen*, pages 61–62, 2010.

[2] Edgar Berdahl and Alexandros Kontogeorgakopoulos. The firefader design: Simple, open-source, and reconfigurable haptics for musicians. In *Proceedings of the Sound and Music Computing Conference*, pages 90–98, 2012.

[3] Edgar Berdahl, Günter Niemeyer, and Julius O Smith. Using haptics to assist performers in making gestures to a musical instrument. In *Proceedings of New Interfaces for Musical Expression*, 2009.

[4] Nicolas d'Alessandro, Joëlle Tilmanne, Ambroise Moreau, and Antonin Puleo. Airpiano: A multi-touch keyboard with hovering control. In *Proceedings of New Interfaces for Musical Expression*, 2015.

[5] Steven Gelineck and Dannie Korsgaard. An exploratory evaluation of user interfaces for 3d audio mixing. In *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

[6] Steven Gelineck, Dan Overholt, Morten Büchert, and Jesper Andersen. Towards an interface for music mixing based on smart tangibles and multitouch. In

Proceedings of New Interfaces for Musical Expression, 2013.

[7] Jürgen Herre, Johannes Hilpert, Achim Kuntz, and Jan Plogsties. Mpeg-h audio—the new standard for universal spatial/3d audio coding. *Journal of the Audio Engineering Society*, 62(12):821–830, 2015.

[8] Michal Lech and Bozena Kostek. Testing a novel gesture-based mixing interface. *Journal of the Audio Engineering Society*, 61(5):301–313, 2013.

[9] Mark T Marshall, Joseph Malloch, and Marcelo M Wanderley. Non-conscious control of sound spatialization. In *Proc. of the International Conference on Enactive Interfaces*, 2007.

[10] Justin Mathew, Stéphane Huot, and Alan Blum. A morphological analysis of audio objects and their control methods for 3d audio. In *Proceedings of New Interfaces for Musical Expression*, 2014.

[11] Frank Melchior, Chris Pike, Matthew Brooks, and Stuart Grace. On the use of a haptic feedback device for sound source control in spatial audio systems. In *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.

[12] Frank Melchior and Sascha Spors. Spatial audio reproduction: from theory to production. In *Audio Engineering Convention 129*, 2010.

[13] William Moss and Bryan Cunitz. Haptic theremin: Developing a haptic musical controller using the sensable phantom omni. In *Proceedings of the International Computer Music Conference*, pages 275–277, 2005.

[14] Jörg Müller, Matthias Geier, Christina Dicke, and Sascha Spors. The boomroom: mid-air direct interaction with virtual sound sources. In *Proc. of the ACM conference on Human factors in Computing Systems*, 2014.

[15] Nils Peters, Georgios Marentakis, and Stephen McAdams. Current technologies and compositional practices for spatialization: A qualitative and quantitative analysis. *Computer Music Journal*, 35(1):10–27, 2011.

[16] Jarrod Ratcliffe. Hand motion-controlled audio mixing interface. In *Proceedings of New Interfaces for Musical Expression*, 2014.

[17] Stephen Sinclair. *Force-feedback hand controllers for musical interaction*. PhD thesis, McGill University, 2007.

[18] Robert Tubb and Simon Dixon. An evaluation of multidimensional controllers for sound design tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 47–56. ACM, 2015.

[19] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, 2013.