Propriosound: A spatial audio enabled proprioceptive augmentation system

Ondrej Brem
Aalto University
Espoo, Finland
ondrej.brem@aalto.fi

Thomas Bruhn
Aalto University
Espoo, Finland
thomas.bruhn@aalto.fi

Janis Heldmann
Aalto University
Espoo, Finland
janis.heldmann@aalto.fi

Michael McCrea
Aalto University
Espoo, Finland
michael.mccrea@aalto.fi

Janne Pietila
Aalto University
Espoo, Finland
janne.a.pietila@aalto.fi

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
The interplay of the auditory and proprioceptive senses show promise in the development of so-called disappearing user interfaces, especially as ubiquitous computing devices move toward eyes- and hands-free interaction. While both virtual auditory displays and gestural tracking are areas of active research, these interaction modalities have primarily served vision-driven virtual reality applications or functional user interfaces that use sound as a medium of translation (sonification) or interactive cueing (auditory menus). This project aims to enlist a variety of innate sensorial mechanisms, namely our proprioceptive and spatial auditory faculties, to reinforce the virtual presentation of parasensory phenomena in an augmented audio reality (AAR) system. In the spirit of exploration and discovery, retaining user agency through mobility is paramount, so we strive to overcome the confines of previous AAR frameworks, which are often tethered to video trackers or designated interaction areas. A framework was developed in which signals from hand-mounted electromagnetic sensing probes are presented to the user in a headphone auralization system. The hand-probe location is tracked through space, as is the movement of the head, allowing the auralization to be "spatially aware" and presented to the user as an augmentation of natural binaural hearing. A prototype was produced integrating numerous technologies. Challenges of an unmoored position tracking system were confronted and, in

response, exploratory research has charted a course for future development.

Author Keywords
Proprioception; spatial audio; sonic interaction.

Introduction
The pursuit of human enhancement might be as old as civilization itself. Humans have been constantly developing technology towards advancing their skills, communication, and knowledge of the world. One application of technology is enhancement of the human senses. While many animal species like bats, sharks or bees perceive the environment in ways that are beyond human ability, novel technologies allow us to adapt parasensorial phenomena to our natural senses [22]. The interplay of the auditory and proprioceptive senses have shown promise in human-computer interaction (HCI), especially as ubiquitous computing devices move toward voice-activated, and eyes-free control [12]. Meanwhile, gestural sensing as a means of communication, exploration, and play has advanced rapidly in gaming and virtual and augmented reality applications. Indeed the intersensory interactions of audition, vision, and proprioception are known to form mutually reinforcing supraordinate representations of the body [16]. Enlisting humans' innate intuitions built upon such representations will be crucial to successful sensorial augmentation.

In recent research, head- and hand-tracking in combination with binaural acoustics has enabled the realization of audio-augmented reality (AAR). While much work in this domain focuses on creating experiences in which synthetic or recorded sounds are inserted into a virtual environment, relatively little work has used AAR to directly experience phenomena inherent to the physical world but which is imperceptible to humans. Furthermore, AAR applications often demand a stationary setup due to external cameras or spatial anchors, so the mobility of those systems is limited. Therefore, our aim is to develop an application that is able to make a variety of parasensorial phenomena accessible to human senses while retaining explorative mobility.

In this report we document the initial work of constructing such a system. We begin by selecting the initial sensing method for study: electro-magnetic field (EM) sensing. We then cover the development of the hardware and software for tracking both hand positions and head orientation to provide "spatially aware" sensing framework. These sensing modalities are brought together through a headphone-based auralization system that delivers the sensed EM field as spatial audio. Finally, the system is made mobile by being integrated into a wearable prototype. Design decisions and initial user feedback is considered in the context of the project goals, and future development, both technical and conceptual, is discussed.

Gathering Requirements
With Propriosound users can explore properties of physical environment that are naturally inaccessible by the human senses. This interaction is enabled by hand-mounted sensors and perceived as a spatial auditory representation of the phenomena. This work builds upon research that has been done on head- and hand- tracking in combination with and binaural synthesis to realize audio-augmented reality (AAR). In other works such as the augmented tangible sonic interaction (ATSI) [25] the user can augment physical objects with spatial sounds and listen to the sound attached to the object while moving it around. Some spatial sound design systems depend on tracking objects by sensors and therefore requiring that objects be designed in advance. Others allow the user to manipulate sound in the physical room detached from any object [3, 13].

Figure 1: Two variations for mounting the EM sensor; on the fingertip (left) and flush with the fingertips (middle). Various sensors may be used, including "contact" vibration sensors (right).

In contrast to these works, with Propriosound the user does not choose the sound that she aims to attach to the explored object, but the object itself 'emits' the sound on behalf of its physical origin, such as its electro-magnetic field. Hence, compared to the sonification approach the sound is not processed to benefit a certain understanding, but provides a direct representation with only minimal filtering, adapting pleasant aesthetics to the user. It also does not demand an outside positioned depth camera as in ATSI, but solely depends on the head-orientation tracking, the hand-orientation tracking and the auralization process. Therefore, the exploration is not limited by a predefined area, but can be done anywhere. Also the objects don't have to be modified beforehand, offering a non-intrusive natural AR experience.

Relating to the two dimensions *embodiment* and *metaphor* as proposed in the taxonomy of Fishkin et. al for tangible user interfaces (TUIs) [9], Propriosound can be called fully-embodied due to its first person perspective. It can also be called "noun" as it represents a certain quality of the explored objects and "verb-like" as the user takes action in using their own hands to explore the quality of the objects and therefore generating the audio representation. On a bigger scale, Propriosund draws aspiration from the development of disappearing user interfaces (DUI) as proposed by Hui and Sherratt [11]. This area of research is based on the interaction with computers through all kinds of enhanced human senses to allow intuitive accessible and ubiquitous computing as proposed by Weiser[30] for future computers being invisibly enmeshed in our day to day life. While Propriosound does not offer interaction with a computer, the insights gathered through its means of interaction can contribute to the field of DUI development.

Design & Methods

Parasensory listening

As a sensory augmentation framework, Propriosound enables a person to experience phenomena that would otherwise be beyond the reach of our innate sensory faculties. We adopt the term parasensory phenomena to indicate physical phenomena which lay just outside our perception which nevertheless surrounds us everyday, and which may provide for a rich experience of exploration and discovery [4]. Examples include ultrasonic sound, infrared light, radio frequency signals, and vibrations that are too high in frequency or too low in amplitude to perceive tactilely.

Other dynamic phenomena are further from reach, such as gravitational waves, cosmic rays or electromagnetic (EM) fields. We chose the latter for integration into the first iteration of Propriosound for numerous reasons. First, electromagnetism pervades our environment through the prevalence of electronic devices we use daily. Second, given our chosen method of sensing, via a coil "pickup" which allows transduction of EM fields directly to sound, the dynamic behavior of EM emitted from electronics—at times chaotic, at other times rhythmic and syncopated—reveals the highly individualized character of each object being explored. Third, many sources emit very localized EM fields, making them well-suited for tactile exploration and an audible synchrony with the movement and position of the hands, showing potential for a tight link to one's proprioceptive awareness.

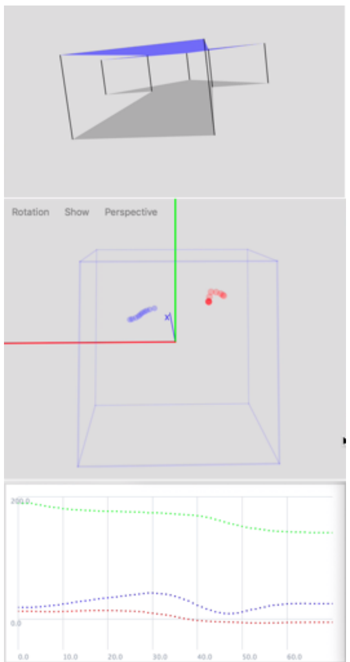

Figure 2: Development view of head tracking (top) and hand position tracking control signals (middle, bottom) in SuperCollider.

The coil pickup is also well-suited to the form factor of the application: the sensing surface of the coil pickup roughly matches the size of a fingertip. A glove conveniently allows mounting of both the coil pickup and the hand-tracking sensor in close proximity (so that the auralized EM source could be positioned back at its spatial origin). Two variations on mounting the coil, as seen in Figure 1, were tested. The first approach was to mount the coil pickup directly to the tip of the index finger, thereby extending it. This offered both intuitive and fine motor control of the sensor.

We note that a similar solution was devised for mounting of a contact microphone to the sensing glove (See Figure 1, right), in order to explore surface and material vibrations, though because fewer sources of vibration were readily available for testing, we focused on the EM coil pickup for the first iteration.

Hand position tracking

In order to reinforce the character of the object being explored via its EM field, it is important that an intuitive connection be established between the audible sound of the field, and where that field is originating. The hope is that hearing an object "through" one's hands reveals another side to the nature of electronic objects and, by extension, in aggregate with the many devices that comprise much of our environment, the nature of our everyday environment. Two of our spatial faculties are enlisted to for this purpose: our proprioceptive sense—feeling of our hands' locations and movements in space (often reinforced through vision) [16]—and our spatial auditory awareness.

Exploratory research at the outset of the project showed that video skeleton tracking offered high accuracy [7] and stability in specific scenarios such as gait-tracking [15], but is impeded by low throughput rates, less than ideal depth tracking in the desired field of view, and in some cases the need for multiple cameras [27]. Infrared time-of-flight tracking offers high spatial fidelity [29], but requires fixed external sensing hubs and static calibration. However, a primary aspiration of the Propriosound project is unimpeded spatial exploration, so inbuilt portability is paramount. Therefore, we sought a solution which was entirely body-mounted so the motion sensing frame moves with the body.

The chosen tracking implementation is path integration, also known in electronic navigation as dead reckoning; a process by which position is inferred from movement from a previous position. While the method is known to be error-prone do to accumulated errors in sensing and integration [18] [14], the method is feasible in contexts where the behavior being tracked can be anticipated an incorporated into the design [8] [1]. Because portability is paramount, we saw

[1] Gait-trackingsite:https://x-io.co.uk/gait-tracking-with-x-imu/

the inherent drawbacks of the method as welcomed challenges to be addressed by compensatory design decisions, which are discussed in the section Discussion & Future Work.

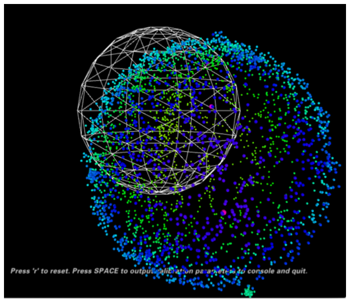

Figure 3: Output from IMU orientation calibration to remove distorting effects of the headphones from the magnetometer data. [5]

The implementation of dead reckoning was achieved through the use of an inertial measurement unit (IMU) mounted on the back of each hand. The IMU has nine degrees of freedom among three sensors: an accelerometer, gyroscope, and magnetometer. We chose the Sparkfun Razor IMU for its notable accuracy and commercial availability, as well as prior use for related studies [21]. Importantly, it features inbuilt sensor fusion which provides on-chip sensor calibration and yields quaternion outputs required for stable orientation calculations [14]. The algorithm for dead reckoning can be summarized simply: 1) using the IMU's orientation, the axes on which acceleration axes are rotated to the world frame (in which gravity points downward), 2) the world-oriented accelerations are then integrated over time once to yield the IMU's velocity, and finally 3) the velocity measurement is integrated to yield position. The code implementation was largely derived from [2].

Head orientation tracking
It is well-established that the sense of realism and immersion in virtual and augmented reality environments is greatly enhanced by head-tracked audio [1]. In virtual auditory displays, sound sources placed virtually in the environment can be made to appear fixed in space by manipulating the virtual sound scene in response to head movement (See §Auralization). To achieve this, a third IMU was mounted to the top of the headphone band to follow the head's changing orientation (Figure 4).

Figure 4: The headphone-mounted IMU sensor for tracking head orientation.

Mounting the IMU to headphones introduces the unique requirement to filter out the electromagnetic field disturbance introduced by the speaker coils in the headphones. The inbuilt sensor fusion could not alone be used to provide accurate results. The orientation calculation was improved by a modification of sensor firmware [3] using a Direction Cosine Matrix algorithm [24] for improved sensor fusion. This firmware requires a one-time calibration procedure which measures and generates compensation filtering for hard-iron disturbances magnetometer readings [4,5]. This filtering is possible because the headphones are fixed to the frame of the sensor and their adverse impact can therefore be minimized through calibration (See Figure 3). The origin of the movement frame was set to the location of the head so that all three IMUs could be brought into a common coordinate system for auralization system.

Auralization
The Propriosound auralization system consists of a signal processing network that encodes the audio signal from the EM sensors into a spatial audio format which aims to create an augmented audio environment. This is done through first-order ambisonic encoding in which the sound sources are virtually positioned in the auditory scene at the location from which it was generated; at the hand positions. The virtual sound scene is then decoded through head-related transfer functions (HRTFs)—which recreate directional cues produced by the anthropometric features of the head—and presented to the user via headphones. [10]

[2] https://github.com/xioTechnologies/Gait-Tracking-With-x-IMU

[3] https://github.com/lebarsfa/razor-9dof-ahrs/tree/v1.5.7.1-ptrbrtz-PR

[4] https://www.fierceelectronics.com/components/compensating-for-tilt-hard-iron-and-soft-iron-effects

[5] https://github.com/Razor-AHRS/razor-9dof-ahrs/wiki/Tutorial#sensor-calibration

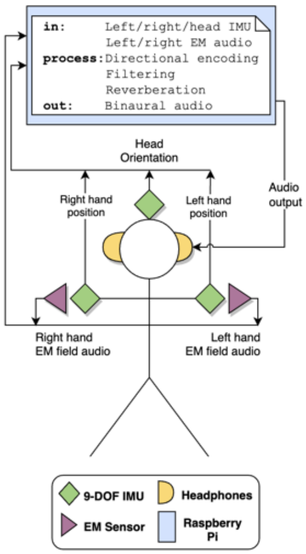

Figure 5: Hardware and software processing diagram. IMU position and orientation inputs drive the audio spatialization processing of the EM sensors.

Further processing is added to the signal processing network to increase the realism. Basic ambisonic encoding treats the sound source as a monopole, i.e. a single point from which sound radiates. Real-world objects, however, radiate sound from across a broader surface, which is to say real-world objects have spatial *extent*, lending to a subtle but important spectral radiation characteristics [20]. To emulate this behavior, spectral dispersion filters are applied to the incoming signal, which allow for parametric control of the apparent extent of a sound source [19]. To further improve the realism of the audio scene, reverberation was added, tuned to the audible size of a moderately-sized indoor space, to aid in the externalization of the auditory images [5, 1].

Finally, in response to early feedback to the aesthetic treatment of the audio generated from the EM sensors (some listeners deemed the sound to be "harsh" and thought it may become annoying over time), musical filtering was added as an optional feature to the auralization system. To this end, banks of resonant filters were applied to the input signal, which can be tuned to variable musical scales. The result is that the EM signal excites the musical resonance, adding a performative aspect to the system, transforming the interaction modality from one of exploration to instrumental performance and gestural metaphor [28], which is not further explored in the current work.

Embedded processing
Propriosound has been conceived primarily as a framework for spatial and sonic exploration. As such, it was important that the user not be tethered to external tracking systems or power sources, and could be free to move about the environment following wherever her ears, hands, and curiosity might lead. Therefore, all sensor and audio processing was ported to a Raspberry Pi computer [6], mounted with a high-quality audio card [7] and portable battery pack. The data and audio processing processing was performed with SuperCollider [8], a programming language built for real-time audio synthesis, which processed the serial data from the IMUs as control signals driving the parameters of the ambisonic encoding, transformation and decoding, which was facilitated through the Ambisonic Toolkit library [9].

Wearability
To allow good experience of spatial sound while exploring the environment, wearability of the whole system had to be considered. The whole system basically consisted of three parts. Headphones with an orientation-tracking IMU, position-sensing gloves fitted with EM sensors, and a central processing unit worn around the neck. The processing unit and its power source evolved from being belt-mounted to a small bag hanging from the neck. The belt-mounted version was less intrusive but the wiring of the prototype became cumbersome once all the sensors were integrated, so a hanging enclosure was more convenient during the final tests.

Results
A user tested the system for the sake of demonstration and critique, revealing number of usability issues. Regarding wearable aspects of the system, the user reported that the complex wiring caused fitting of the system to be too complicated. This was to be expected, but after some adjustment, the user found the system "*surprisingly comfy*" and "*didn't feel limitation in movement*".

[6] https://www.raspberrypi.org
[7] https://www.hifiberry.com/shop/boards/hifiberry-dac-adc/
[8] https://supercollider.github.io
[9] http://www.ambisonictoolkit.net

Figure 6: A user explores the EM fields of his environment.

It was clear that the depth of the EM sensor forced a distance between the hand and the object being sensed, limiting the otherwise beneficial sense of touch, as the user reported. A glove was selected as the wearable sensor mount, as it facilitated wire routing and offered more sensor mounting surfaces, including between the tips of the index and middle fingers, allowing it to be flush with the fingertips. While this configuration slightly hinders the dexterity of both fingers, it allows the whole hand to engage with the observed object. Both mounting options were facilitated by a simple velcro surface which allows a user to choose and alter the mounting to their preference (See Figure 1). The user reported that after a while the glove grew slightly uncomfortable on account of a small fit.

Regarding the impact of the sound of the sensed EM fields, the user was surprised by how distinct the sounds generated by different objects were. The user described the sounds as "*alien but fascinating*". He made the suggestion that if the sound was more closely studied and characterized, the system may serve the work of electrical engineers. Further study is required to determine what, if any, filtering might mitigate apparent harshness of the electronic sounds, or encourage other uses, such as musical uses of the system.

Importantly, the spatial behavior of the sound was not pronounced enough to give a tangible proprioceptive feedback. While movement was perceptible, the spatial accuracy had not reached sufficient accuracy to convince the user of added "reality" through auralization. Over time he noted an isolated feeling, on account of the headphones.

A number of these observations were anticipated during development and attempts were made to preemptively mitigate them. For example, because the IMUs accumulate positional drift over time, a known strategy was used to zero out the velocity once movement fell under a calibrated threshold [8]. Even still, it was necessary to apply spatial filtering to limit the drift of the hand tracking signals.

In an effort to improve accuracy through interaction methods, a simple gesture was added. The user can reset the positioning by a simple double-tap of the hands together (detected through the acceleration of the IMUs) in front of the chest, with arms bent at 90 degrees, and gaze of the head directed forward. This provided a natural and easily-repeatable position from which an approximate coordinate origin could be inferred and reset. Feedback is provided to the user as a non-obtrusive bell sound to ensure the gesture was successfully registered.

Discussion & Future Work

The prototype of the augmentation framework is in a stage where the auditory interface—the spatialized auralization of the sensor data—can now be tested with people. Observations during-, and interviews after experimenting with the prototype could hold valuable information for future iterations. Since the sound was described as unpleasant during a feedback session, further investigation through user tests could provide a better understanding of how different filters for the sound affect the experience. Future research could also explore the vast landscape of sensors; to discover appealing parasensory phenomena for users, or new potential use cases for the augmentation framework. In the team we discussed many options to extend the interaction paradigm. One promising possibility is to ''anchor" discovered sounds in the room at the location in which they were found [2, 12]. This would allow the sensed phenomena to be mapped and revisited, documented, compared, spatially rearranged and composed.

The prototype of our proposed augmentation framework has many opportunities to improve from a technical standpoint. Position tracking, for example, can be improved with more advanced algorithms [14], or with the addition of more inertial sensors—as few as two per arm—with consideration of the state-space of limb movement [6]. As reported from the user, the cabling of the system is cumbersome, which could be addressed through wireless sensing and transmission solutions, such as Bluetooth, and future iterations would strive for seamless integration of a so-called Disappearing User Interface [11].

Numerous improvements are also possible to improve the realism of the virtual auditory scene [23]. These include sound encoding into higher-order ambisonics (HOA) for increased spatial resolution or parametric directional enhancement [17] Important to the goals of the project is to approach the sensation natural hearing, to avoid the "isolated" feeling reported by the user. Further research should be done on strategies to externalize the encoded sounds, likely with the help of a binaural microphones to create "hear-through" headphones where sound of external environment can pass through to the listener for a more realistic fusing of the virtual sound in the real auditory environment [26].

Conclusion

The goal of this work was to explore the potential of our innate proprioceptive awareness and highly sensitive auditory spatial perception as a means to experience a new sensory field. We developed a proof of concept for a spatial audio-enabled sensory augmentation system. The resulting device allows its wearer to explore EM fields transmitted by electronics in their surroundings – supported by spatial auditory awareness. While an initial test showed the potential awe and curiosity of users, imprecise translation of position in space prevented tangible proprioceptive feedback.

We received valuable feedback with which we were able to chart a course for future research, to improve spatial fidelity of both the movement tracking and auralization, as well as explore new interaction modes for which this framework may be uniquely suited.

Additional Resources

Code used to produce this project can be found at https://github.com/mtmccrea/propriosound.

REFERENCES

[1] Durand R Begault, San Jo, and San Jo. 2000. Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. In *Proc. Audio Engineering Society 108th Conv.*

[2] Jamie Bullock, Tychonas Michailidis, and Matthieu Poyade. 2016. Towards a Live Interface for Direct Manipulation of Spatial Audio. In *Proc. of the International Conference on Live Interfaces*. 134–141.

[3] Daniel Dobler, Michael Haller, and Philipp Stampfl. 2002. ASR: augmented sound reality. In *ACM SIGGRAPH 2002 conference abstracts and applications*. 148–148.

[4] Gershon Dublon and Joseph A. Paradiso. 2014. Extra Sensory Perception. 311, 1 (2014), 36–41.

[5] N. I. Durlach, A. Rigopulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel. 1992. On the Externalization of Auditory Images. 1, 2 (1992), 251–257.

[6] M. El-Gohary, L. Holmstrom, J. Huisinga, E. King, J. McNames, and F. Horak. 2011. Upper limb joint angle tracking with inertial sensors. In *2011 Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 5629–5632.

[7] A. El-Sawah, N.D. Georganas, and E.M. Petriu. 2008. A Prototype for 3-D Hand Tracking and Posture Estimation. 57, 8 (2008), 1627–1636.

[8] Mostafa Elhoushi, Jacques Georgy, Michael Korenberg, and Aboelmagd Noureldin. 2014. Robust motion mode recognition for portable navigation independent on device usage. In *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*. IEEE, 158–163.

[9] Kenneth P Fishkin. 2004. A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous computing* 8, 5 (2004), 347–358.

[10] Huseyin Hacihabiboglu, Enzo De Sena, Zoran Cvetkovic, James Johnston, and Julius O. Smith III. 2017. Perceptual Spatial Audio Recording, Simulation, and Rendering: An overview of spatial-audio techniques based on psychoacoustics. 34, 3 (2017), 36–54.

[11] Terence K. L. Hui and R. Simon Sherratt. 2017. Towards disappearing user interfaces for ubiquitous computing: human enhancement from sixth sense to super senses. 8, 3 (2017), 449–465.

[12] Raine Kajastila. 2013. Interaction with eyes-free and gestural interfaces. (2013).

[13] Hirokazu Kato, Mark Billinghurst, Ivan Poupyrev, Kenji Imamoto, and Keihachiro Tachibana. 2000. Virtual object manipulation on a table-top AR environment. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*. Ieee, 111–119.

[14] Manon Kok, Jeroen D. Hol, and Thomas B. Schön. 2017. Using Inertial Sensors for Position and Orientation Estimation. 11, 1 (2017), 1–153.

[15] Chandramouli Krishnan, Edward P. Washabaugh, and Yogesh Seetharaman. 2015. A low cost real-time motion tracking approach using webcam technology. 48, 3 (2015), 544–548.

[16] Jr Lackner and B Shenker. 1985. Proprioceptive influences on auditory and visual spatial localization. 5, 3 (1985), 579–583.

[17] Mikko-Ville Laitinen, Tapani Pihlajamäki, Cumhur Erkut, and Ville Pulkki. 2012. Parametric time-frequency representation of spatial sound in virtual worlds. 9, 2 (2012), 1–20.

[18] Johnny Leporcq. 2018. Position Estimation Using an Inertial Measurement Unit Without Tracking System. (2018).

[19] Trond Lossius and Joseph Anderson. 2014. ATK Reaper: Ambisonic Toolkit as JSFX plugins. In *The joint 40th International Computer Music Conference & 11th Sound and Music Computing Conference*.

[20] Dylan Menzies. 2007. Ambisonic Synthesis of Complex Sources. 55, 10 (2007).

[21] Dylan Menzies and Filippo Maria Fazi. 2018-09. A Complex Panning Method for Near-Field Imaging. 26, 9 (2018-09), 1539–1548.

[22] Scott Novich and David Eagleman. 2015. Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Experimental brain research* 233 (06 2015). DOI: http://dx.doi.org/10.1007/s00221-015-4346-1

[23] Makoto Otani, Haruki Shigetani, Masataka Mitsuishi, and Ryo Matsuda. 2020. Binaural Ambisonics: Its optimization and applications for auralization. 41, 1 (2020), 142–150.

[24] William Premerlani and Paul Bizard. 2009. Direction Cosine Matrix IMU: Theory. (2009). https://www.researchgate.net/publication/265755808_Direction_Cosine_Matrix_IMU_Theory

[25] Roberto Pugliese, Archontis Politis, and Tapio Takala. 2015. ATSI: augmented and tangible sonic interaction. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. 97–104.

[26] Rishabh Ranjan and Woon-Seng Gan. 2015. Natural Listening over Headphones in Augmented Reality Using Adaptive Filtering Techniques. 23, 11 (2015), 1988–2002.

[27] Markus Schlattman and Reinhard Klein. 2007. Simultaneous 4 gestures 6 DOF real-time two-hand tracking without any markers. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology - VRST '07*. ACM Press.

[28] John Sullivan, Alexandra Tibbitts, Brice Gatinet, and Marcelo M. Wanderley. 2018. Gestural Control of Augmented Instrumental Performance: A Case Study of the Concert Harp. In *Proceedings of the 5th International Conference on Movement and Computing*. ACM, 1–8.

[29] Michael Van den Bergh and Luc Van Gool. 2011. Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In *2011 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 66–72.

[30] Mark Weiser. 1999. The computer for the 21st century. *ACM SIGMOBILE mobile computing and communications review* 3, 3 (1999), 3–11.