

Acoustic Localisation Techniques for Interactive and Locative Audio Applications

Dom Schlienger
Sibelius Academy Helsinki
Centre for Music & Technology
email: dominik.schlienger@siba.fi

Abstract

Reviewing the literature on positioning systems using acoustic source localisation principles for Interactive and Locative Audio Applications (ILAA) it becomes clear that Acoustic Localisation and Positioning Systems (ALPS) implemented on ubiquitous devices can provide an alternative to Motion Capture systems (MoCap) wherever multiple speakers are part of an application. Providing background on and defining the notion of ILAA, this paper argues that based on comparisons of recent applications in the literature, ALPS can provide competitive alternatives to MoCap, the system prevalently used in ILAA.

1 Introduction

With the advent of multitrack recording technology and its possibilities for spatial distribution, the relation between the origin of a sound and the listener was essentially abstracted. Today, with location aware mobile technology and almost ubiquitous internet access, we can use the position of a moving object in space as a dynamic parameter and map it, arguably, to anything. In the case of musical applications these mappings can be used for musical expression, to track gestural control for hyper instruments for example, or let a dancer control musical parameters through spatial cues.

The use of these spatial interfaces for musical applications are well documented in the proceedings of the international conference on *New Interfaces for Musical Expression* (NIME), for example. The importance of digital positioning technology in general is born witness by the yearly *IEEE conference on Indoors Positioning and Indoors Navigation*, and *Ubiquitous Positioning Indoors Navigation and Location Based Service*.

In Interactive and Locative Audio Applications (ILAA), Motion Capture (MoCap) is the most commonly applied principle, using cameras to track movement. However, the question is if it is always the best choice. In our previous work [15], we found evidence that for some applications acoustic localisation techniques might perform equivalently to MoCap and often at a fraction of the cost.

In some scenarios acoustic localisation techniques could clearly outperform MoCap. MoCap requires a line of sight between the tracked object and the camera. If some-

thing obstructs this line, position data can not be obtained. An analogous situation for acoustic localisation, where an object obstructs the signal path between microphone and loudspeaker, is a lot less problematic due to the physical nature of the audio signal. In fact, the principle's feasibility is well documented [4, 8, 9, 13–15, 20], despite its rare implementation in ILAA. The implementation is particularly straightforward when multiple speakers are already part of the system, as is often the case in ILAA.

By comparing the performance of Acoustic Localisation and Positioning Systems (ALPS) with the documented performance of MoCap, we can further evaluate the principle's suitability for ILAA, and contribute to a growing body of work which advocates its implementation. ALPS are often cheaper and usually make better use of existing technology within ILAA, in accordance with the notion of ubiquitous computing, as our findings show.

The paper provides background on ILAA, followed by a discussion of technical aspect of ALPS and how they compare to MoCap and some MoCap hybrid systems. It then goes on to discuss the limits of MoCap for ILAA and why ALPS could provide alternatives, followed by a section on future work and conclusive remarks.

2 Background

2.1 Interactive and Locative Audio Applications

In the context here, ILAA are audio applications in which the position, or the change of position of an object in physical space gives a cue for some process to happen. In general, ILAA could also include audio applications where some process changes the physical position of something in the space, but for this paper the former definition shall apply.

The potential uses of positioning systems for interactive art, multimedia and interactive spatial music are well documented [24], and early concepts go back to the emergence of electroacoustic music in the 1950s. The *potentiomètre d'espace* being an example thereof [10], and the spatialisation schemes of Varèse *et al.* for the Phillips Pavilion at the Expo in Brussels 1958 [1], or Stockhausen's for *Kontakte* [18]. Admittedly, these spatialisations probably were interactive in nature merely by the absence of digital multitrack control. If the computers were available at the time to control dozens of audio track automatically, this option would have been embraced, effectively disposing of interactive control. However, the experimental spirit of the time, and an example of gestural control of spatial audio also in the modern interactive sense is visually documented by the photograph of Pierre Henry controlling the *potentiomètre d'espace* in Figure 1. Already in 1941, in a more mainstream but not less ground breaking vein, the spatialisation of the soundtrack for Disney's *Fantasia*, Fantasound [7], used several operators to control the panning of separate tracks over multiple speakers.

A more recent example for ILAA in the mainstream are interactive video game controllers, like Wii and Kinect, where spatial interactivity provides increased immersiveness, also through surround sound [2, 3, 15] to create a virtual space. The virtual



Figure 1: Pierre Henry at the *potentiomètre d'espace* in a concert at the *Salle de L'Ancien Conservatoire*, Paris, 1952

space shall be defined here similar to Normandeau's definition in [12], as the audio space created for the listener by the composer or in paraphrase for the gaming environment, the developer. Virtual audio space can but does not have to be congruent with a real existing sound field in space. For the discussion on suitability of positioning technology for ILAA, this generalisation is helpful, as it provides a denominator for space as a concept in electroacoustic music, video games and spatial sound reproduction.

The very recent presentations held by Normendeau and Siegel at Sibelius Academy, both as guest performers at the 2013 *SibA MuTe Fest* [5] highlight exemplarily the interest of electroacoustic composers in ILAA. Compositions like *Movements in Possible Histories or a Composition for 24 Windows* [11], also at Sibelius Academy, in the surroundings of our research, as well as research conducted at SARC, Belfast; Locus Sonus Laboratoire Art Audio, Aix-En-Provence; Zentrum für Kunst und Medientechnologie Karlsruhe, the activities of NIME, just to name a few, show the importance of the field for composers and musicians.

Electroacoustic spatial composition is just one area of ILAA where positioning technology can provide interaction. The ubiquitousness of mobile technology facilitates locative games like *Papa Sangre* [23], for example. Spatial audio interaction is a growing area in music technology, and is becoming increasingly pervasive. Various ongoing projects at institutions like Pervasive Media Studio Bristol, are examples therefore [19].

2.2 Acoustic and Local Positioning Systems

Using the principle of acoustic source localisation we can obtain the distance between a speaker and microphone by measuring the time delay of the acoustic signal, as we know the speed of sound through air *a priori*. If we know the distances for multiple speakers, a 3-D position can be triangulated. The accuracy of such a system lies within the low decimetre range and its latency is the same as for audio recording systems, in the low millisecond range. The latency derives from the buffer lengths applied, which means that at the cost of covered area, by shortening the buffer size, the latency can be reduced. This is useful for applications where small gestural movements need to be tracked, for example, in case of instrument control. The system is thus scalable. The presence of multiple loudspeakers is a prerequisite, which in most ILAA is given. Other than that, all that is required is a microphone. This makes the principle cheap for ILAA.

ALPS have not often been implemented to date. Rishab *et al.* presented a system which uses controlled ambient sound, i.e. pseudo random noise, as a signal to measure the time delays [14]. Random noise can be problematic for the use in ILAA, as the noise would need to be masked at all time by the audible signal. The system described in 2010 in [8] compares arrival times of a distinct signal on networked devices, applying a multiple receiver principle. The systems in [9] and [4] measure signals outside the limits of audible sounds, making them effectively ultrasound. Our suggestion is for a one receiver multiple sender model, using the audible airborne sound which is already part of an ILAA directly as a measuring signal, as described in [15]. ALPS are rarely implemented like this, particularly not by measuring the time delays of the signals carrying the audio content itself of an audio application, despite their documented feasibility in [4, 8, 9, 13–15, 20]

Our previous work, about the suitability of positioning systems for ILAA in general, compared specifications in the literature to user requirements obtained from early findings of an ongoing survey.¹ The survey shows that optical tracking in form of MoCap provides good solutions as long as the requirement for line of sight between tracking device and tracked object does not cause issues. When this indeed does create issues, respondents revert to hybrid systems, where MoCap is combined with auxiliary systems using dead reckoning principles. Dead reckoning systems predominantly consist of inertia meters. In the literature the term inertial navigation system is thus often used. However, dead reckoning methods also include the use of compasses, providing a more general term. Dead reckoning systems usually require frequent updating with absolute data in ILAA, which predestines them for hybrid systems. On their own, they are usually only applied where absolute position is not necessary.

Further, the survey showed respondents' concerns about tracking as a privacy-sensitive issue. To have an opt-in choice is regarded as important. This can easily be achieved with the one receiver - multiple sender model. The control over privacy stays with the person holding the mobile device, the microphone.

¹http://creativemusictechnology.org/research_&_development_EN.html

3 Comparing MoCap to ALPS

In the discussion on suitability of ALPS for ILAA one does not get around the fact that the currently predominantly used optical tracking systems are generally considered to be satisfactory. Not many people who use MoCap criticise it as there is not much to compare it to. Thus, it is in many cases the best available option.

However, the requirement of line of sight between a tracked device and a camera is an issue for many conceptual possibilities. Particularly when the tracked devices are to be on members of a crowd [21].

Also for the systems described by Normandeau for the Klangdom using ZirkOsc [6], positioning information from ALPS could open new possibilities. In the examples discussed in Siegel [17] on interactive applications, where dance controls music, ALPS could provide interesting and dynamic possibilities.

In view of audio-mobility, the dependence of MoCap on cameras means that ad-hoc networks of ubiquitous devices using MoCap are unrealistic. The environment needs to be controlled and the tracked object defined by colour or shape. Even if line of sight could be established between mobile devices, in a multi user environment this would require some form of choreography. Despite its popularity, MoCap has a considerable disadvantage to ALPS which is inherent to the very nature of the camera and constitutes its limitation as an interface to gather spatial data. The camera provides a 2-dimensional view of space and the further away from the camera an object is, the less information we gain about it.

The following example does not want to be misunderstood as a criticism on the artistic quality of the work in question. The tracking system applied by Siegel for *Two Hands, (Not Clapping)* [16], was originally developed as an interface for dancers to interact with music, wherein a movement in the field of the camera is digitally registered. But interaction through this interface requires the dancer to adapt performance to the interface's characteristics. Due to the optical perspective of the camera, a moving object close to the camera causes more change in pixels than an object further away. That is, it provides a warped perspective. Thus the interface's resolution for a movement in space is not linear but proportional to the distance. As innovative this is for the performance of *Two Hands, (Not Clapping)*, the fact that the interface weighs a particular amount of movement in one area of the performance space differently to the same amount in a different area makes it a poor interface for 3-D interaction. As a spatial interface, a lot of movement should be mappable to a *pro rata* equivalent. If this is not the case, additional information is needed, to differentiate between small cow very close, and large cow far away. ALPS could provide better results. N.B., not for *Two Hands, (Not Clapping)*, where the warped perspective is idiosyncratic part of the composition, but for a more general spatial human computer interface. This is particularly poignant if the performance space in question does want to be understood as 3-dimensional, which is usually the case in ILAA, especially in implementations in ad-hoc networks.

By using multiple cameras, depth information is added in some MoCap systems. For ILAA, this is a cost, which, in the presence of multiple loudspeakers, could be avoided

by using ALPS.

It is evident that every medium influences the message by its characteristic [12]. This might not necessarily be negative, but it is understood that this influence can warp the data, as exemplified above. If the system is supposed to be a good interface according to the notion of ubiquitous computing, the influence of its characteristic, its visibility, has to be reduced [22].

4 Future Work and Conclusive Remarks

Looking at the existing literature and the early findings of our own research, the impression prevails that ALPS provide a competitive alternative for many spatially interactive applications, where audio is diffused over multiple speakers. The possibilities arising from ALPS implementation in ad-hoc networks stand in contrast with what is achievable with MoCap. MoCap's limitation as an ubiquitous interface for ILAA are further compounded by the intrinsic distortion of spatial data through the 2-dimensional depiction of space by a camera.

In the next steps of the research, it is considered of paramount importance to find ways of establishing requirements directly from the potential uses in musical practices. As ALPS are rarely implemented for ILAA, not much material is available to document precisely how the system will be used by musicians or any other early adapters. So insight needs to be gained experimentally. As very little music of spatially interactive nature exists, free improvisation suggests itself as a means to study interaction of space and audio. To use free improvisation as a methodology to explore the relation of organised sound and space is in itself a very interesting but vast field, which, as a means to establish user requirements, shall be explored in the future work.

References

- [1] Spatialisation et image : la mise en espace des sons concrets et électroniques. <http://sonhors.free.fr/panorama/sonhors13.htm>, 2014.
- [2] K. Collins. *An introduction to the participatory and non liner aspects of video games audio*. Helsinki University Press, 2007.
- [3] K. Collins, editor. *From Pac-Man to Pop Music: Interactive Audio in Games and New Media*. Ashgate, Aldreshot, 2008.
- [4] V. Filonenko, C. Cullen, and J. Carswell. Investigating ultrasonic positioning on mobile phones. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.
- [5] Centre for Music and Sibelius Academy Helsinki Technology. Mute fest 2013. <http://tinyurl.com/mqz16nx>, 2013.

- [6] Zentrum für Kunst und Medientechnologie Karlsruhe. Zirkonium. <http://www.zkm.de/zirkonium>, 2014.
- [7] WM. E. Garity. Fantasound. *Journal of the Society of Motion Picture Engineers*, 1941. <http://www.widescreenmuseum.com/sound/fantasound1.htm>.
- [8] T. Janson, C. Schindelhauer, and J. Wendeberg. Self-localization application for iphone using only ambient sound signals. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.
- [9] A. Mandal, C.V. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi. Beep: 3d indoor positioning using audible sound. In *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pages 348–353, Jan 2005.
- [10] P. Manning. *Electronic and Computer Music*. Oxford University Press, New York, revised and expanded edition, 2004.
- [11] L. Muredu and M. Mela. Movements in possible histories or a composition for 24 windows. <http://www.youtube.com/watch?v=1tqQVrVoHyE>, 2011.
- [12] Robert Normandeau. Timbre spatialisation: The medium is the space. *Organised Sound*, 14:277–285, 12 2009.
- [13] V. C. Raykar, I. V. Kozintsev, and Lienhart R. Position calibration of microphones and loudspeakers in distributed computing platforms. *Speech and Audio Processing, IEEE Transactions on*, 13(1):70–83, 2005.
- [14] I. Rishabh, D. Kimber, and J. Adcock. Indoor localization using controlled ambient sounds. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, Nov 2012.
- [15] Dominik Schlienger. Indoors and local positioning systems for interactive and locative audio applications. <http://creativemusictechnology.org/MScAudioProduction99001618TextNoApcs.pdf>, 2012.
- [16] Wayne Siegel. Two hands (not clapping). http://waynesiegel.dk/?page_id=24, 2010.
- [17] Wayne Siegel. *Dancing the Music*. Oxford University Press, Oxford, 2011.
- [18] Karl Heinz Stockhausen. Kontakte. <http://www.youtube.com/watch?v=aNt6a5xF0nE>, 1959.
- [19] Pervasive Media Studio. Pervasive media studio. <http://www.watershed.co.uk/pmstudio/>, 2014.
- [20] S. Tervo. *Localization and tracing of early acoustic reflections*. PhD thesis, Aalto University School of Science, January 2012.

- [21] Ryan Ulyate and David Bianciardi. The interactive dance club: Avoiding chaos in a multi participant environment. In Ivan Poupyrev, Michael J. Lyons, Sidney Fels, and Tina Blaine, editors, *NIME*, 2001.
- [22] Mark Weiser. Building invisible interfaces. Presentation Slides, Novemeber, 2nd 1994.
- [23] Wikipedia. Papa sangre — Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Papa_Sangre, 2014.
- [24] R. Zvonar. A history of spatial music. http://cec.concordia.ca/econtact/Multichannel/spatial_music.html, 2005.