Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013

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

Sakari Tervo
Department of Media Technology
Aalto University
Helsinki, Finland
sakari.tervo@aalto.fi

ABSTRACT

This paper provides a rationale for choosing acoustic localisation techniques as an alternative to other principles to provide spatial positions in interactive locative audio applications (ILAA). By comparing positioning technology in existing ILAAs to the expected performance of acoustic positioning systems (APS), we can evaluate if APS would perform equivalently in a particular application. In this paper, the titles of NIME conference proceedings from 2001 to 2013 were searched for presentations on ILAA using positioning technology. Over 80 relevant articles were found. For each of the systems we evaluated if and why APS would be a contender or not. The results showed that for over 73 percent of the reviewed applications, APS could possibly provide competitive alternatives and at very low cost.

Keywords

Acoustic Source Localization, Interactive Locative Audio Applications

1. INTRODUCTION

Music, ever since its origins, has arguably always been spatially interactive. From call and response practices to marching bands, interacting through space is part of many musical practices. The advent of multichannel recording technology provided new possibilities for spatial distribution. Consequently the spatial and temporal relation between the origin of a sound and the listener fundamentally changed. Linked with the digital revolution this relation was further abstracted and new ways of dealing with this relationship became necessary and possible. The wide interest of researchers in computer human interaction and a large body of work bears witness to this.

For spatial interaction in particular, the proceedings of NIME 2001 - 2013 provide many examples of interactive locative audio applications (ILAA) corroborating the importance of the field. The systems presented apply a range of technologies to use spatial data as a a parameter in musical applications. Optical tracking principles like motion caption and infrared technology, gyro and accelerometers and hybrids of both are the largest groups of principles applied. However, to our knowledge, only very few positioning systems using acoustic source localisation have been realised for the use in ILAA [10, 6, 15], notwithstanding the principle's documented technical feasibility [17, 20]. In fact, trawling the roughly 1100 NIME proceedings for ILAA using acoustic source localisation techniques for positioning, none were found. The principle is conspicuous through its rarity; analysing its performance in applications a moot point.

Acoustic localisation is similar in principle and performance to Motion Capture (MoCap), which is predominantly used in ILAA, but it can operate without the requirement of line of sight between tracked object and camera. The effects of an object obstructing the signal path between a microphone and a speaker are a lot less detrimental to the signal than obscuring a marker in a MoCap system, due to the diffraction of sound around obstacles. Further, in our previous work [17], we compared various positioning technologies as to their suitability for ILAA. It became clear that besides well documented uses of optical tracking principles like MoCap, acoustic localisation techniques can provide a means to procure position data, particularly wherever airborne sound is part of an ILAA.

This paper shows that an acoustic positioning system (APS) could provide competitive alternatives for many applications. Based on the known performance characteristics, we evaluate for which ILAA presented at the NIME conferences from 2001 to 2013, APS could provide an equivalent or possibly even preferable positioning system.

The paper firstly summarises our previous work on the suitability of positioning systems for ILAA in general, secondly discuss early results of an ongoing online survey on the topic, and thirdly, analyses relevant NIME proceedings for comparison of positioning systems applied with the possibilities of APS. It is followed by a section on future work and conclusive remarks.

2. POSITIONING SYSTEMS

2.1 A Short Overview

Broadly, positioning systems can be typed by Systems providing absolute data in relation to a reference grid, i.e. GPS to longitude and latitude; relative data, i.e. the output of an acceleration meter; or symbolic data, for example, the fact that a mobile phone is within the reception range of a receiving mobile mast, or the statement that somebody is at home or at work.

Despite the summation of error in relative type positioning, hybrid systems using dead reckoning-principles like gyro and accelerometers, often bring good results when combined with absolute data from, say, optical tracking devices. Dead reckoning type devices require frequent updating with absolute data to be of use, which predestines them for hybrid -, rather than stand-alone systems or systems where only relative data is needed.

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

Positioning systems using radio signals to triangulate positions, are almost ubiquitous due to WLAN technology. Liu *et al.* state systems estimating positions by signal strength measurements to rarely achieve accuracy bellow two metres. Other systems which are more accurate are not as ubiquitously available, and lack robustness [9].

Acoustic source localisation techniques can be used for positioning technology using time of arrival calculations by digitally measuring the time delay of an acoustic signal between a loudspeaker and a receiving microphone. From the time delay the distance can be calculated, as the speed of sound through air is known *a priori*. By calculating this distance from several loudspeakers to a microphone, the microphones 3-D position can be obtained. These systems provide accuracy in the low decimetre range at low latency, in the milliseconds range. APS laid out as a one microphone - multiple speaker system allows participant-performers to decide if they want to be tracked or not, as they are in control of the signal receiver, the microphone. Latency issues in APS depend on the necessary buffer sizes which are determined by the time sound takes to travel through air over a distance of interest. In applications where intrinsic small gestures akin to instrument control have to be tracked, a smaller buffer size can be chosen as the distances to be measured are shorter. Thus, APS are scalable. The prerequisite for APS is the presence of multiple loudspeakers and some processor. All that needs to be introduced to the system is a microphone, ideally wireless. Further, the signal which is already part of an ILAA can be used directly as a measure signal, necessitating masked or added measure signals for silent moments only.

MoCap, despite being chosen by most developers, has a considerable disadvantage to APS which is inherent to the very nature of the camera and constitutes its limitation as an interface to sample spatial data. That is, 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. This provides contorted data. This contortion can of course be rectified and calculated. But APS can provide uncontorted data directly and in a much simpler set up.

2.2 User Requirements for ILAA

In our previous work, the suitability of positioning systems for ILAA in general has been studied [17]. A summary of this analysis is listed in Table 1.

In addition to analysing the suitability of the positioning systems from an objective point of view, some answers were gained through an ongoing online survey on the topic. [17]. The survey was developed in cooperation with a focus group of professionals in the field at the Pervasive Media Studio, Bristol, UK, and is aimed at developers and early adapters and asks respondents about their experiences with and expectations on positioning systems for ILAA. Besides expectations on accuracy it also enquires about respondents' opinions on cost, ubiquitousness in the sense of unobtrusiveness, implication on privacy and more. The survey has yet to reach representative proportions but shows clearly that beside optical tracking, acoustic localisation techniques would meet most requirements.

In summary the survey's answers so far suggest the preference for positioning technology to be ubiquitous, working on existing technology, virtually everywhere, virtually invisibly, unobtrusive, hence requiring as little extra devices or gadgets as possible. Thus, users would prefer not to wear backpacks, hats and goggles as suggested in [4]. And position technology should be cheap, provide accuracy in decimetre range and low privacy intrusion, for example, in public spaces.

Being tracked is a privacy-sensitive issue. Hence, having an opt-in choice, is regarded as important by respondents. This can easily be achieved with the one receiver - multiple sender model, a principle Siegel describes as an inside-out system in [18]. It means that a participant can opt in to be tracked by being in control of the receiver, i.e. the microphone.

As the survey's aim was originally to establish what *any* positioning system would be required to do to be suitable for ILAA, acoustic systems were just an option of several. Thus to say that acoustic source localisation is an ideal solution from this survey alone, would be conjecture. From the answers to questions regarding expectations on actual specifications, accuracy, for example, it is clear that optical tracking provides the best *available* solution, as APS were not known to many respondents.

In discussions with the focus group it became clear that new developments are expected to adhere to principles of ubiquitous computing. According to Weiser, ubiquitous computers must know where they are [21]. Also he proposes a computer to be invisible, the technology to stay out of the way of the task [22]. It is self evident that this also applies to interface design. In application to interface design for musical expression, this can in many way stand as a defining difference between an interface and an instrument.

3. INTERACTIVE LOCATIVE AUDIO APPLICATIONS AT NIME 2001 - 2013

1100 titles of proceedings items of NIME 2001-2013 and full texts from 2001-2012 were searched for ILAA using acoustic localisation principles, and none were found. In lieu, we searched for paper titles on ILAA with the search terms *tracking, tracker, locative, localisation, positioning, position, motion, mocap, gestural, gesture, 3D, space, spatial*. The 80 relevant papers were filtered into sub-categories of technical principles. These groups were then compared to APS' documented specifications on several criteria like precision, range, ubiquitousness, latency, cost and the presence of multiple loudspeakers. Additionally, a full text search of the proceedings 2001-2012 yielded another 51 possible contenders for APS, out of 250 relevant hits on top of the ones form the title search. To keep the reference list short the full sample is referenced, with brief comments, in a downloadable spreadsheet.[1]

Of the 80 ILAA, 28 applications use MoCap, 12 Dead Reckoning and 10 hybrids of both, 6 use other principles, one does not name a technology but discusses positioning principles in general and 7 explicitly require a positioning system to enhance or realise proposed mappings, or they present filters or other algorithmic processes, but do not name a particular principle for positioning. The closest match to an APS ever presented at NIME 2001-2012 was only identified through the full text search. [8] applies Doppler effect analysis on test signals, which, arguably, are ultrasonic.

3.1 Discounted Applications

If an ILAA was in character more of an instrument than an interface it was discounted from the list of 80 applications. The distinction here comes from the notion that an instrument has an idiosyncratic character of which the positioning technology might be intrinsic part [19]. APS, in this contrast, wants to be understood as an interface, ideally as an invisible interface [22].

Eight presentations were on such actual idiosyncratic instruments or tools using positioning data, both absolute or

[1] `http://tinyurl.com/nngwcoa`

Table 1: Positioning Systems Performance in Overview

Principle	System	Accuracy	Area	Cost to user	Availability	Ubi.	Cost to installer
RF	Satelite navigation	low	global	low	market	yes	NA
	Pseudolites	medium	local	low	planned	no	high
	Ultra wide band	high	indoors	high	market	no	medium
	WLAN	low	local	very low	market	yes	low
	Wireless sensor net	medium-low	scaleable	low	market/DIY	no	low
	Bluetooth	low	20 m	very low	DIY	yes	low
Inertial	Gyro/Accelerometer	0.5% - 20%*	1-100 m	low	market/DIY	yes	low
Optical	Infrared, wii	medium	scaleable	low	market/DIY	yes	low
	MoCap hi-end	very high	scaleable	high	market	no	high
	MoCap lo-end	medium	scaleable	medium	market	no	low
Magnetic	Magnetic field	high	1-20 m	medium	market	no	medium
	Induction	NA	NA	NA	no	no	NA
Sonic	Ultrasonic	high	scaleable	medium	market	no	medium
	Acoustic Tracking	high	scaleable	very low	DIY	yes	low

*No absolute measure. DIY:Do It Yourself. WLAN:Wireless Local Area Network, NA:Not available, Ubi.:ubiquitousness

relative. As an instrument, the idiomatic way of playing with e.g., an optical interface [1] or floor pads [7], can be essential part of the instrument design, and using APS would change the character completely. Systems with a clear haptic idiom, like [12], were also discounted, as APS can not provide haptic feedback. Further, one system is excluded as it applies positioning data primarily as a global or symbolic parameter [11]. Three applications do not require positioning data at all, as they are spatialisation schemes where the position data applied is not actually spatial.

Of the 28 systems using optical tracking, 7 can be discounted as their specification explicitly states, that the gestural tracking of movement is to happen without any sensors attached to the body or fingers as in [5]. Similarly, where the object to be tracked is a person *passing by* in a public space [1], APS can not provide alternatives, as the activity of *picking up a device* can not implicitly be expected as being part of *passing by.* Except when mobile phone technology is mentioned and its presence can be considered to be ubiquitous, then APS is an option.

Two system which cover a table size area in 2D were considered to be not typical for ILAA, and [13] uses optical sensors in mm scale for bow tracking for violin-family instruments. APS is not able to measure in necessary detail.

Systems of gestural control were included if the tracked gesture could be expressed as a Cartesian position in space. This excluded some systems where the gesture was more symbolic in nature, similar to a fader or joystick movement. For obvious reasons, systems tracking facial expressions and contour tracking applications were excluded too.

Systems which use haptic information as integral part of their workings were not included, as haptic feedback is intrinsically absent in APS. If an auxiliary positioning system was explicitly mentioned as part of a haptic system, and its performance could be improved by APS, it was included.

3.2 APS provides alternatives

The above exceptions still leave us with 59 systems out of 80, (73.75 percent) wherein APS could provide comparable performance. Including the additional yield from the full text search, there are 110 contenders for APS out of 358 ILAA using positioning technology (30.72 percent). One common factor to all 110 systems is that airborne sound is explicitly part of the application, i.e. multiple loudspeakers are already part of the system, mostly in form of surround sound [3].

The most common principle of positioning applied in ILAA is optical tracking and specifically MoCap. The larger part of the applications in this group tracks performers in a room or performers area, at a precision level in the low decimetre range. Most MoCap systems use multiple cameras, 8 in the case of [2]. In these ILAAs APS could possibly provide a competitive alternative as it does not rely on line of sight to the camera. And by replacing 8 digital cameras with one analogue microphone the reduction both in processing power and calibration effort will be reflected in the cost.

To provide alternatives for MoCap in applications were a small area is being tracked APS can be scaled to higher accuracy at the cost of range by shortening the buffer size. Particularly conducting-type applications would profit, as they mostly rely on single camera systems or handheld devices. Evidently a handheld microphone could easily replace the IR functionality of a WiiMote as a baton [14].

The second largest group uses dead reckoning methods. The group can be further split into systems using wearable devices, wristbands or gloves, for example, and ones using handheld devices. The fact that this group relies on a device being on the performer means that replacing it with a small microphone is certainly not a step back and an improvement in precision can be achieved due to the availability of absolute position data.

In the group of hybrids are quite a few idiosyncratic tools like data gloves which work very well for their intended purpose. In the nature of their multi modality, adding a small mic would be rather in keeping. In many cases the camera - depending part of a hybrid system, i.e. MoCap or infrared, could be replaced with APS [16].

4. FUTURE WORK

These findings inform our ongoing development and implementation of an APS for ILAA which uses the same airborne audio signal as the one which carries the content of the application (Music, speech, sound) to measure the time delays of the signal on one microphone in relation to multiple speakers. This development project is further iguided by the notion of the invisible interface, and thus to use ready available technology in the typical set-up of an audio application. No further technology like cameras or wearable devices other than a microphone, which might already be part of the ILAA, shall be introduced. [22]

Based on the broad and documented interest of the community of developers, musicians and performers in spatial interactivity, we find it of utmost importance to include early adapters in the development of this system. We are implementing a workshop exploring musical spatial interaction to inform what the *system* needs to do, rather than presenting users with a system asking them *to do something with it*. The methodical implications of this approach shall be disseminated in a future paper.

More short term, a prototype APS shall be presented, implemented for use in a professional live - sound enforcement or recording environment, where musical performers' position on a stage can be tracked. This position data can then be used, for example, to automate the monitor mix depending on the performers position on stage.

5. CONCLUSION

The body of previous and related work shows clearly to what specification various positioning systems perform and give a clear notion of what these systems need to be able to do to match user requirements in ILAA. As very little literature exists on how the principles of APS would perform in ILAA, literature on *existing* ILAA using other principles was reviewed with the aim to identify the examples for which APS could provide an alternative with comparable performance. The astonishing result of this review is that none of the 1100 presentations, posters or papers included in the NIME proceedings from 2001 to 2012 presented a system using acoustic localisation techniques, despite the ubiquitousness of multi-track speaker arrangement in almost all of the 80 applications reviewed in detail. Peculiarly astonishing considering that the position data could be gathered cost free in case of mobile phone based applications or at the price of a simple omni directional microphone in others. Even for implementation in a professional audio environment using wireless microphone technology the costs pale into insignificance compared to professional MoCap systems.

As to the question if APS can *always* provide alternatives to optical tracking, the limitations are clear: APS can not provide an alternative for face- or contour tracking, which are typical visual interface tasks, nor can it track objects onto which a microphone can not be attached.

But, given the presence of multiple loudspeakers, wherever a performer's position needs to be tracked within a performance space of room size dimensions, APS can provide similar or equivalent results to many systems currently in use. APS are very easy to implement and use a modest amount of processing power compared to other systems.

Last not least, due to the possibility of opt-in positioning, APS could provide an alternative to some public space-installation where the presence of a camera might be perceived as an infliction on privacy issues.

6. REFERENCES

[1] G. Beyer and M. Meier. Music Interfaces for Novice Users : Composing Music on a Public Display with Hand Gestures. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 507–510, 2011.

[2] M. Ciglar. "3rd. Pole" - A Composition Performed via Gestural Cues. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 203–206, 2008.

[3] A. W. Colin Honigman and A. Kapur. The Third Room: A 3D Virtual Music Framework . In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 29–34, 2013.

[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] E. Guaus, T. Ozaslan, E. Palacios, and J. L. Arcos. A Left Hand Gesture Caption System for Guitar Based on Capacitive Sensors. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 238–243, 2010.

[6] 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.

[7] N. H. Keating. The Lambent Reactive : An Audiovisual Environment for Kinesthetic Playforms. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 338–343, 2007.

[8] J.-s. Lee and W. S. Yeo. Sonicstrument : A Musical Interface with Stereotypical Acoustic Transducers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 24–27, 2011.

[9] H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, Nov 2007.

[10] A. Mandal, C. 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.

[11] C. McKinney and N. Collins. An Interactive 3D Network Music Space. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 400–405, 2013.

[12] T. Mitchell and I. Heap. SoundGrasp : A Gestural Interface for the Performance of Live Music. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 465–468, 2011.

[13] L. Pardue and A. McPherso. Near-field optical reflective sensing for bow tracking. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 363–368, 2013.

[14] L. Peng and D. Gerhard. A Wii-Based Gestural Interface for Computer Conducting Systems. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 155–156, 2009.

[15] 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.

[16] J. C. Schacher. Motion To Gesture To Sound : Mapping For Interactive Dance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 250–254, 2010.

[17] D. Schlienger. Indoors and local positioning systems for interactive and locative audio applications, 2012. http://creativemusictechnology.org/MScAudioProduction 99001618TextNoApcs.pdf.

[18] W. Siegel. *Dancing the Music*. Oxford University Press, Oxford, 2011.

[19] A. Tanaka and P. Gemeinboeck. A Framework for Spatial Interaction in Locative Media. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 26–30, 2006.

[20] D. B. Ward, E. A. Lehmann, and R. C. Williamson. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *Speech and Audio Processing, IEEE Transactions on*, 11(6):826–836, 2003.

[21] M. Weiser. The computer for the 21st century. *Scientific American*, 265(3):66–75, January 1991.

[22] M. Weiser. Building invisible interfaces. Presentation Slides, Novemeber, 2nd 1994.