INTERNATIONAL CONFERENCE ON LIVE INTERFACES

Gestural Control for Musical Interaction using Acoustic Localisation Techniques

Dominik Schlienger

University of the Arts Helsinki Sibelius Academy
Centre for Music & Technology, Finland
dominik.schlienger@uniarts.fi

Abstract. Acoustic Localisation principles for tracking technology are well researched and have many applications in medicine and industry. In creative technologies optical technologies are more prominent. For creative music technologies we know from our own research into the applicability of acoustic localisation techniques for audio in the frequency range of standard audio equipment, that acoustic localisation techniques are potentially a straightforward choice. We also know that the technology is scalable from tracking in large performance areas to smaller areas with lower latency, required for gestural tracking for real-time interaction. As a proof of concept we prototyped two implementation of the principles, using handheld microphones and standard, commercially available loudspeakers, firstly for a Theremin - like pitch control interface and secondly, a spatial trigger for percussive sounds.

Keywords: Acoustic Localisation Techniques, Theremin, Gestural Control, Interface for Musical Expression.

Introduction

Acoustic Localisation (AL) principles for tracking technology, particularly techniques using ultrasound, are well researched and have many applications in medicine and industry (Holm 2012). The principles for AL are not limited to ultrasound however, they apply to audible sound as well. This makes the use of them for tracking technology particularly interesting, namely for applications where audio equipment is already present, like in many audio, audiovisual and multimedia applications and particularly for interactive arts and new interfaces for musical expression (Schlienger and Tervo 2014). In creative technologies optical technologies are more prominent (Schlienger 2014), but there are a number of applications using AL (Rishabh, Kimber, and Adcock 2012; Filonenko, Cullen, and Carswell 2010; Janson, Schindelhauer, and Wendeberg 2010; Seob Lee and Yeo 2011).

We also know that implementation of AL principles is scalable. Tracking of performers in traditional performance areas like stages in concert halls are as feasible as tracking of small gestural movements which require very low latency and high update rates, for example for real-time interaction with virtual instruments.

For low latency applications, the authors of (Gupta et al. 2012), describe an implementation which arguably provides similar functionality to what we suggest. However our Time Difference of Arrival (TDoA) approach in lieu of Doppler, brings advantages in applications where the tracked object's identity needs to be known. For our test implementation of a Theremin like instrument, the differentiation between the left and right hand, for example, could be achieved by using separate microphones. In the Doppler scenario described in (Gupta et al. 2012) this differentiation would have to be achieved by other, more complex means.

Further, from our work with the Workshop on Music, Space & Interaction (MS&I) (Schlienger 2016a), we know about the need for simple, ubiquitous and pervasive interfaces which do not hem in the flow of gestural explorations of space. What is more, our findings from the workshop suggest that interfacing technology in form of sounding objects provides a particular engaging tool to immerse performers as well as audiences in a spatially interactive performance. We thus describe two

applications here, using AL, which answer these requirements as a proof of concept for the feasibility of the principles in low latency applications needed for interactive sounding objects.[1]

Both applications work on the basis of moving a handheld microphone but arguably the microphone could be attached to any type of moving object.

To summarise our rationale, we argue for a broader implementation of AL using mainstream audio equipment as the principle provides competitive alternatives to many other solutions as we showed using the literature in our previous work (Schlienger and Tervo 2014). Namely, by using sound in the frequency range supported by standard, main street, audio equipment, we see a possibility to improve on the performance of, for example, optical tracking systems due to the diffraction of sound around objects, allowing for tracking in non-line of sight situations. Whereas the focus of our previous work describes novel applications for larger performance areas and stages, our contribution here is to show that the AL principle can also be applied in smaller, gestural applications, for which other technologies are usually implemented: To stay with our examples, capacitance for the Theremin and optical tracking for gestural triggers. Our point is, the AL principle, using ubiquitous technology, has broader and more pervasive potential than its current state of rarity portrays.

The following sections describe our work in MS&I in more detail, give more details on AL principles followed by a description of our prototype implementations. The Matlab scripts using the Playrec Utility library are available to download and form part of this publication as an appendix (Schlienger 2016b).

Method

Workshop on Music, Space & Interaction

The workshop on MS&I runs now in its third year at the University of the Arts Helsinki, and uses Participatory Design approaches (Robertson and Simonsen 2013) to explore the affordances of technology for spatial interactivity in interactive musical applications. We use interdisciplinary, free improvisation as a method (Andean 2014) to explore existing and possibly new technologies without the restrictions of habits, genres or conventions. This approach to technology design draws on the notion of Designing Culture (Balsamo 2011) and on the notion of mess in ubiquitous computing (Dourish and Bell 2011).

The data gained from the workshop is ethnographic in nature, it consists of notes and participants written contributions along with some audio and video documentation. Some insight can be gained from the workshops blog the participants are encouraged to contribute to (Schlienger 2016a).

The applications we describe in this paper are based on our early findings from the workshop which provided us with the idea of spatially interactive sounding objects: These are things in a space which might be actual physical objects but also virtual objects which can only be heard in a particular position, rather than seen. On this notion we developed the idea of a spatially controlled pitch - object, probably most descriptively described as a Theremin - like instrument and a percussive - object, whereby at distinct positions percussive sounds can be triggered.

[1] We describe a proof of concept for larger scale applications in an article accepted for presentation at the NIME2016 conference

Figure 1. Workshop on MS&I, (Courtesy of Timo Pyhälä, 2015)

Acoustic Localisation

As we are only interested in AL in respect to sound which can be produced by standard audio equipment, we refer to AL in the rest of the text in reference to the frequency range form roughly 20 Hz - 30 kHz. Further, we are interested in the Time Difference of Arrival (TDoA) technique specifically, as this is the approach taken for the implementation we are working on for tracking in larger spaces. Compared to other AL approach, Doppler, for example, TDoA techniques lend themselves better to applications where the identity of the tracked object needs to be known, as the position estimation happens actively for a sender or a receiver's own position, Doppler or also echo-location relies on an estimation based on an indirect measurement. The principle of TDoA measures the time difference between the sending of a signal and its recording at one (or several) receiving microphones. From the correlation of the recorded and the original signal the time delay can be calculated directly. As the speed of sound through air is known *a priori* the distance of the receiving microphone from the sound source (the loudspeaker in this case) can thus be derived from the time difference. Using several such measurements, a position estimate can be trilaterated. The technique thus works in principle for 3 dimensional localisation.

For the prototypes discussed here we use the most simplest of principles, namely one single distance reading. *Nota bene*, this limitation to a single dimension is not a limitation of the AL principle! Trilateration from 4 distance measurements can be used to estimate an absolute 3D position. However, for the applications we discuss here, simple distance measurements between one sender and one receiver are, indeed, sufficient. We would like to stress that although a single distance reading is not enough for a 3 dimensional, absolute *position*, the application can still be spatial in character, as the distance reading is available *radially* from a fixed point of the receiver or sender which allows for 3 dimensional interaction with the object.

Experimental set up

To demonstrate the applicability of AL for applications which are latency sensitive we prototyped two musical applications. The first application is a Theremin like instrument wherein pitch can be gesturally controlled. For the second application, we implemented a gestural trigger mechanism for percussive sounds. Both implementations require to move a microphone in front of a loudspeaker emitting a high pitched measuring signal just above the frequencies audible to the human ear and within the frequency range of standard audio equipment. (Depending on the type of loudspeaker, 17-30 kHz.) The loudspeaker used for our tests was of type Alesis M1 MK II, the microphone of type Sure SM58.

The applications were implemented in Matlab R 2013a, using the Playrec (Humphrey 2011) utility. Playrec allows for non-blocking soundcard access and thus continuous and simultaneous play and record. This could have been achieved natively in newer versions of Matlab, but Playrec provides insight into the sample-by-sample workings which was considered helpful for prototyping. A simple Max Patch received the measurements from Matlab via udp and dealt with the content audio.

Patches and Playrec Script are available online (Schlienger 2016b). The processor running Matlab was an 11-inch, Mid 2011 MacBook Air, 1.6 GHz Intel Core i5, running OSX 10.8.5, the soundcard a DigiRack 002.[2]

The room in which we tested the applications is a typical living room without particular acoustic treatment, with a reverberation time below 0.4 seconds, 6 by 3.5 meters with 2.6 height.

Both applications are latency sensitive: For musical interaction, in order to play within an ensemble as well as to be able to play an overdub for a multitrack recording it is crucial that the performer can monitor her or his playing in real-time or in a very close approximation to real-time. To define the criteria of what shall be considered a close enough approximation, the following thoughts were decisive:

- Latency up to a length of 10 ms is generally tolerated by musicians in performance situations as well as in the recording studio.[3]
- Just over 10 ms at 48 kHz sampling rate can be achieved with a buffer size of 512 samples. This also happens to be the lower limit at which our current set-up runs stably.
- The actual, overall latency between is somewhat higher: We can calculate this by measuring the time it takes a signal to arrive at a microphone at a millimetre distance from a loudspeaker. Possibly due to hardware restrictions and processing power of our set-up, our best achievable latency is often as high as four times the buffer size.
- For many mainstream sound cards 20 ms were considered acceptable until fairly recently.
- We further estimate typical gestures for these application to stem from arm movements of a stationary person, so we scaled the functionality of both applications to a range of 1 m.
- The trajectories through air which can be expressed in 512 samples at 48 kHz represent a distance of 3.66 metres. For gestural interfaces we consider this adequate.

This last point might need some more elaboration: The length of the buffer we iterate when calculating the time delay between two signals sets the limit of the longest time delay we can actually measure. We cannot measure time delays which are longer than the window that the buffer provides. As we translate the time delays to distance covered by sound, the window sizes also stand for maximum distances that can be estimated within a window. These relations between window sizes and time delays and distances covered by sound, respectively, are further dependent on sampling rate. So with an increased sampling rate we need to higher the window size to cover the same distances. The higher sampling rate does not only provide a higher update rate, but also allows for measurement signals at higher frequencies.

	44.1 kHZ	**48 kHz**	**96 kHz**	**192kHz**
Measurement signal top frequency (Nyquist)	22.05 kHz	24 kHz	48 kHz	96 kHz
1 sample	0.0244 ms	0.0208 ms	0.0104 ms	0.0052 ms
1024 sample window	23.2199 ms	21.3333 ms	10.6666 ms	5.3333 ms
2048 sample window	46.4399 ms	42,6666 ms	21.3333 ms	10.6666 ms
Maximum distance represented by window 1024	7.9644 m	7.3173 m	3.6586 m	1.8293 m
Maximum distance represented by window 2048	15.9288 m	14.6346 m	7.3173 m	3.6586 m
Distance in one sample	0.0077777 m	0.0071458 m	0.0035729 m	0.0017864 m
Samples to travel 1 m	128.5714	139.9416	279.8833	599.7667

speed of sound $= 343\ m/s$
time for sound to travel 1m $= 2.9154\ ms$

Table 1: Overview of Relations between Sample Rates and Distances Covered by Sound

[2] For the video demonstration we used Meyer Sound MM-4XP, the Motu 16A soundcard and AKG c417 Omni Lavalier Microphones, and the same processor.

[3] We differentiate here between what is tolerable and what is noticeable: even latency as short as 1-3 ms can be noticed, for example as comb filtering by singers monitoring themselves on headphones.

Proof of Concept for Low Latency Applications

We chose the two applications as proof of concept for low latency applications for the reason that they show conceptual differences in slightly different requirements. Thus, we look at the applications separately:

Pitch Control

Gestural pitch control with a Theremin is a skill which needs to be acquired through practice. Similarly, our take on the concept is not meant to simplify musical playing: With a stable and controlled hand it should allow for stable, controlled pitches and the (skilled) playing of melodic material.

The pitch range covered within 1 m we set too 80 - 6000 Hz, as we believe this to be roughly representative of the frequency range of interest for most musical applications. We experimented primarily with sine waves, but this was an arbitrary decision, and other sound material could, of course, be used instead.

Percussive Control

We generally associate percussion playing with hitting an object with our hand, or with a stick, or similar. Our implementation of gestural percussive control is thus somehow quite abstract, as moving the microphone through a particular distance to a loudspeaker triggers a sound. There is a commercial implementation doing a very similar thing though, Aerodrums (Aerodrums 2016), using optical tracking. Similarly to our scenario, the lower latency limit is set by the system buffer size, however, their implementation allows for settings as low as 128 samples per buffer, albeit at 44.1 kHz sampling rate.

Results & Discussion

We achieved latencies for both application scenarios we considered sufficiently low for a proof of concept. (Typically around 10 - 20 ms) We are also confident that these numbers can still be improved on with further development.

For the pitch control application, we found that our current, averaging, smoothing algorithm[4], makes it difficult to know if the gestural position responds to the proper pitch location or if the algorithm is averaging out under the influence of a few wrong readings. For pitch sensitive application averaging filters don't seem to be the best choice.

One peculiarity we can report for the percussive control application: Despite the latency being quite high at 512 or more samples per second (48kHz sampling rate) we didn't really notice this at first, as we seemed to just make a mental note where it was that the sound triggered when we tried it out the first time. As we knew that there was some latency we then investigated and came to the following insight: As it happens, the fact that in these air-drum type applications the performer is not actually hitting a physical object, the latency, (as long as constant and not varying) has the effect of moving the virtual object further away from the place where the performer thinks she or he will hit it. In this sense it is actually easier to live with the latency of such a virtual sound object than it would be to live with the latency of a physical object triggering a virtual sound source. The lack of haptic response means that, for our scenario, the virtual object is just further away than the performer would initially presume if we knew at what distance measurement the sound is being triggered.

The need for a mobile device to be held by the performer, even as a clip on wireless microphone, remains an obstacle towards totally transparent interfacing. Yet, compared to Doppler techniques which would allow device - free tracking of gestural movements, we see great advantages in the mobile device approach as the identity of the tracked device is known to the system.

[4] The algorithm calculates averages over time, the smoothing creates a lag and the interpolation of positions result in slurring of the pitch

A further caveat applies to both Doppler as described by (Gupta et al. 2012) and our (current) TDoA approach: With measurement signals in the frequency range between 18 to 30 kHz we lose the advantage of acoustic tracking over optical tracking to a certain extent: The much praised advantage of AL in view of the requirement of line of sight between tracked object and a camera in optical tracking systems, is much reduced at high frequencies, as the corresponding wavelength do not diffract around obstacles but reflect if the obstacle is wider than the wavelength in question. AL still works in the dark, certainly an advantage over some optical tracking systems. We are thus very interested into further research about the possibilities of using the audible sound of the content in an audio application as a measurement signal, or certain frequency bands within the content audio.

The resulting proof of concept - applications are being demonstrated in the appended video clips available online via the following link: http://creativemusictechnology.org/lowlatencyapps.html

Conclusions and Future Work

We showed with a simple implementation that AL techniques are feasible for pitch control and percussive triggering of sounds in musical applications. We also showed that more research is necessary, and that the current implementation can not be considered the state of the art of what is possible: With a more advanced implementation, a direct comparison with commercial systems on the market using optical tracking, for example, will provide an actual evaluation, which we can not sensibly provide yet with the current, rudimentary, example code. We also point out that these limitations are not due the principles of AL but due to the basic nature of the prototyped implementation. To summarise, we think we have a couple of interesting virtual sounding objects whose affordances we will explore much further in the workshop on MS&I. we would also like to invite interested parties to look at the code appended to this paper, available online (Schlienger 2016b). The whole project is intended to be open source, and an implementation as a Pd and/or Max MSP object is planned.

Acknowledgements. We would like to thank all the participants past present and future of the workshop on Music, Space & Interaction. This research is made possible by a Kone Foundation Researchers Grant.

References

Aerodrums. 2016. Aerodrums: The best Drumset You've Never Seen. http://aerodrums.com/aerodrums/.

Andean, James. 2014. Research Group in Interdisciplinary Improvisation: Goals, Perspectives, and Practice. http://www.academia.edu/4158856/Research_Group_in_Interdisciplinary_Improvisation_Goals_Perspectives_and_Practice.

Balsamo, Anne. 2011. *Designing Culture: The Technological Imagination at Work*. Durham, NC, USA: Duke University Press.

Dourish, Paul, and Genevieve Bell. 2011. *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. Cambridge, MA: MIT Press.

Filonenko, V., C. Cullen, and J. Carswell. 2010. "Investigating ultrasonic positioning on mobile phones*." International Conference on Indoor Positioning and Indoor Navigation (IPIN).*

Gupta, Sidhant, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. "SoundWave: Using the Doppler Effect to Sense Gestures." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*. New York, NY, USA: ACM, 1911–1914.

Holm, S. 2012, Nov. "Ultrasound positioning based on time-of-flight and signal strength." *International Conference on Indoor Positioning and Indoor Navigation (IPIN).* 1–6.

Humphrey, R. 2011. Playrec. http://www.playrec.co.uk/index.php.

Janson, T., C. Schindelhauer, and J. Wendeberg. 2010, Sept. “Self-localization application for iPhone using only ambient sound signals.” *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*.

Rishabh, I., D. Kimber, and J. Adcock. 2012. “Indoor localization using controlled ambient sounds.” *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*.

Robertson, Toni, and Jesper Simonsen. 2013. *Routledge international handbook of participatory design.* Edited by Jesper Simonsen and Toni Robertson. Routledge New York.

Schlienger, Dominik. 2016a. Workshop on Music, Space & Interaction Participants’ Blog. http://creativemusictechnology.org/MS&P.html.

Schlienger, Dominik. 2014. “Acoustic Localisation Techniques for Interactive and Locative Audio Applications.” *Locus Sonus Symposium #8 on Audio Mobilité*, Proceedings.

Schlienger, Dominik. 2016b. Gestural Control for Musical Interaction using Acoustic Localisation Techniques (ICLI2016). http://creativemusictechnology.org/lowlatencyapps.html.

Schlienger, Dominik, and Sakari Tervo. 2014, June 30 – July 03. “Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013.” Edited by Baptiste Caramiaux, Koray Tahiroglu, Rebecca Fiebrink, and Atau Tanaka, *Proceedings of the International Conference on New Interfaces for Musical Expression.* Goldsmiths, University of London, 439–442.

Seob Lee, Jeong, and Woon Seung Yeo. 2011. “Sonicstrument: A Musical Interface with Stereotypical Acoustic Transducers.” *NIME*. nime.org, 24–27.