Indoors and Local Positioning Systems for Interactive and Locative Audio Applications

Dominik Schlienger
99001618

A dissertation submitted to the University of the West of England, Bristol in accordance with the requirements of the degree of MSc in Audio Production. The Faculty of Environment and Technology, Bristol Institute of Technology, May 2012

Abstract

Indoors and Local Positioning Systems for Interactive and Locative Audio Applications

The human ear is able to locate a sound to its apparent location, with an accuracy of millimetres rather than centimetres. With the increased demands on audio to immerse the listener in multimedia applications this fact is more and more taken into consideration, for example by making 5.1 surround sound standard delivery format for video games.

With the developments towards interactivity in the arts and new media, and the emergence of kinetic gesture games on the mainstream market, (Nintendo's Wii for example) this research hypothesises an interest by developers of game audio, composers and musicians, the performing arts and the recording industry, in integrating real time positioning data (from sensors carried by performers or players, for example) in an interactive way.

The technological developments in local and Indoor positioning (motion tracking, radio frequency triangulation, GPS, gyro meters, etc.) present a vast array of possibilities. (Mautz et al.) However, so far no current positioning system presently applied to audio applications could be identified which is able to provide accuracy within the margins of the human ear's ability to locate sound.

The aim of this research is to identify what suitable Indoors or local positioning systems (I/LPS) there are and what possibilities and limitations for interactive and locative audio applications (iLAA) result from this in consideration of user requirements established in an online survey, conducted in partnership with Pervasive Media Studio Bristol.

The thus achieved insights are then applied to the development of a I/LPS for iLAA simulation using audible audio only.

AUTHOR’S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of the West of England, Bristol. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree.
Any views expressed in the dissertation are those of the author and in no way represent those of the University.
The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

SIGNED: .... DATE: ....

Contents

1. Introduction

1.1. Overview

The human ear is able to locate a sound (to its apparent location, see Blauert, 1997) with an accuracy of millimetres rather than centimetres. With the increased demands on audio to immerse the listener in multimedia applications (Björk & Holopainen 2004) this fact is more and more taken into consideration, for example by making 5.1 surround sound standard delivery format for video games.

With the developments towards interactivity in the arts and new media, (Flew, 2005) and the emergence of kinetic gesture games on the mainstream market, (Nintendo's Wii for example) this research hypothesises an interest by developers of game audio, composers and musicians, the performing arts and the recording industry, in integrating real time positioning data (from sensors carried by performers or players, for example) in an interactive way.

The technological developments in local and Indoor positioning (motion tracking, radio frequency triangulation, GPS, gyro meters, etc.) present a vast array of possibilities. (Mautz et al.) However, so far no current positioning system presently applied to audio applications could be identified which is able to provide accuracy within the margins of the human ear's ability to locate sound.

The aim of this research is to identify what suitable Indoors or local positioning systems (I/LPS) there are and what possibilities and limitations for interactive and locative audio applications (iLAA) result from this in consideration of user requirements established in an online survey, conducted in partnership with Pervasive Media Studio Bristol.

The thus achieved insights are then applied to the development of a I/LPS for iLAA simulation using audible audio only.

The structure of this dissertation consists of 5 chapters which are thematically distinctly separated. Firstly, After a historical overview by way of introduction, the limitations of spatial hearing are analysed in chapter 2, and how precise positioning systems for spatial audio applications need to be, based on the presumption that positioning systems do not need to be more precise than - but are at best as precise as - our ability to locate sounds.
Secondly, analysing existing positioning systems in respect to their ability to meet the above criteria in chapter 3 and thirdly comparing both the natural limitations and the existing positioning

system technology with the demands for actual applications in chapter 4. This third step will consist of an attempted taxonomy of iLAA - as different applications have different needs - and the findings from a survey into user requirements on I/LPS for iLAA. The respondents are a representative mix of developers, composers and academics in the field.

Chapter 5, describing the software development of a simulation of an audio only positioning system, will condense many of the working hypothesises and insights gained in previous chapters into the more practical project part of the thesis. The UGen++ application developed here for is contained as an appendix on disc. (Appendix I)

As the chapters are fairly self contained it was thought prudent to summarise every chapter individually with some conclusive remarks at its end, thus slimming the 6th chapter (Conclusions) where only the issues arising from all chapters together are summarised but no detailed recount of the individual chapters is given.

The survey discussed in chapter 4 remains open for the time being for further dissemination (spring 2012) and if the reader feels inclined to take part in it, he or she shall be encouraged to do so. It is available on: http://www.surveymonkey.com/s/6X86DJL

1.2 Historical Background

Spatial music is probably as old as music itself, call and response practices date back thousand of years and are evidently spatial in nature.

Historic evidence in the form of written music of spatial nature dates back to the praxis of cori spezzati, where two or more choirs are placed at separate ends of a room. In particular in Venice this polychoral style was very popular in the late Renaissance and early Baroque eras. (Zvonar 2005, Blesser & Salter 2007, Wikipedia 2012a,)

The spatial nature of a symphonic orchestra can be experienced in the many excellent auditoriums which have been built to accommodate this particular sound like the aural architecture of the Boston Symphony Hall, for example. Or the Gregorian choral music's use of the long delay times in the gothic cathedrals of the times being another historic architectural representation of the spatial quality of music. (Blesser and Salter, 2007)

Over the last hundred years the way we experience audio content has dramatically changed. (Roads,1996, Mannings 2004) With the means to electronically process audio, sound became independent of its acoustic origin, available in many different places and at the same time. This opened up new opportunities for composers and researchers with an interest in spatial music in particular.

1.2.1 The Emergence of Spatial Music Technology

The first well documented incidence of this multi-location distribution of audio content in the sense of modern technology, was the Telharmonium, an electronic instrument of importance for electronic music in general, as it was a effectively an early polyphonic, multi-timbral synthesiser applying additive synthesis.
In this context here it is of importance as the chosen way to distribute the audio content was by broadcasting alternating currents to speakers over the telephone exchange. This is not dissimilar to a modern "pod cast", but it is rather astonishing considering that we are looking at an invention which was patented the same year as the radio, in 1897 (Manning 2004).

The researcher composers of the Groupe de Recherche de Musique Concrète in Paris in the mid century developed techniques and technology which not only were crucial for the development of spatial music but for interactive spatial music in particular.

Jacques Poullin, the engineer and assistant to Pierre Schaeffer together with the latter developed a control mechanism they called Potentiomètre d'Espace (1951) which used induction coils to control the signal routing of a multichannel system feeding spatially organised loudspeakers. The Potentiomètre d'Espace was intended particularly as a means to perform musique concrète for a live audience providing real time control for fairly complex multichannel systems. It used induction of a handheld magnetic device in an arrangement of metal hoops, as illustrated in (See Fig 1) where composer researcher percussionist Pierre Henry holds the control. (Manning 2004, Zvonar 2005, Sonhors, 2012)

Figure 1(Wikipedia 2012)

A system similar in effect if different in philosophy due to it's origin in electronic music as practised at the NDWR studio Cologne, developed towards the end of the 1950's, and used in compositions by Stockhausen, consisted of a speaker on a rotating table. a tetrahedral arrangement of microphones surrounded this speaker and an operator could thus spatially control the signal flow (Manning 2004). (See Fig. 2)

Figure 2 (ComputerMusicBlog 2012)

The *interactive* nature of both these systems was probably an involuntary consequence of the limitations in technology for spatial signal control. The inter-channel activity could not yet be automatised which meant an operator was necessary who (inter-)actively controlled the system.

So in effect, this gestural interactivity was a means of control and no such *spatial* interactivity was necessary once control could be held *graphically*, i.e. by the personal computer. This might explain why spatial interactivity was not really an interest in the audio world until quite recently, with the development of adaptive sound design, pervasive media and ubiquitous technology.

Composers like Varèse, Xenakis and Stockhausen, explored the possibilities of spatial music in works which were deemed important enough to be included in the world expos of 1958 (Brussels) and 1970 (Osaka) respectively (Judd, 1961, Manning 2004, Zvonar, 2005). The newly developed ways of spatial sound distribution needed a new architecture, a link represented i*n persona* by the composer/architect Iannis Xenakis.

The Philips Pavilion at the 1958 world expo in Brussels was designed by Iannis Xenakis who worked under the architect Le Corbusier. Edgar Varèse's milestone composition *Poème électronique* was written for the pavilion, according to a detailed spatialisation scheme which made use of 425 loudspeakers, fed from a 3 track magnetic tape containing the audio material, controlled by a 15 track magnetic tape containing the data to switch the 3 audio channels to the required loudspeakers at a time within the composition. (Holmes, 2008, Judd, 1961)

For the 1970 World Expo, the same Iannis Xenakis wrote a 12 track tape work called Hibiki-Hana-Ma for the Japanese Steel Federation Pavilion, which was equipped with 800 loudspeakers.

Also at the 1970 World Expo in Osaka the German Pavilion was designed by Karl Heinz Stockhausen. It was a spherical music hall with 50 groups of loudspeakers arranged in spheres in a true surround arrangement, i.e. from all sides, including from below via a sound-permeable floor. (See Fig 3)

Figure 3 (Space Collective 2012)

This was not Stockhausens first or last work of spatial music, An early 1970 spatial Stockhausen composition, *Sternklang* is for five groups of musicians spread out in a park. The listener moves freely about in the park between the groups which are musically connected by musical "messengers" commuting between them. (D Sounds, 2012)

A later work, the *Helicopter String Quartet*, (1993) a composition "for string quartet, 4 helicopters with pilots and 4 sound technicians 4 television transmitters, 4 x 3 sound transmitters, auditorium with 4 columns of televisions and 4 columns of loudspeakers, sound projectionist with mixing console and moderator (ad lib.)" is a true example of spatial music, as the string quartet is playing from four helicopters, their sound amplified as to blend with the sound of the helicopters. In his own words: "Most of the time, the string players played tremoli which blended so well with the timbres and the rhythms of the rotor blades that the helicopters sounded like musical instruments" (Stockhausen 2012)

To summarise, Stockhausen at Cologne, and Schaeffer, Henry, Poullin and other researchers and composers at the *Groupe de Recherche Musical* in Paris produced, as far as *spatiality* is concerned, highly complex electronic and *concrète* music with intrinsically *analogue* means - *manual*, in fact - in the 1950s and early 1960s, namely the *Potentiomètre d'Espace* and

Stockhausen's *rotating speaker table* with fixed position mics at the NDWR studios in Cologne. What makes the Potentiomètre d'Espace particularly intriguing is the fact that the gestural position of the handheld device can directly and proportionally represent the controlled soundfield.

However, the 1958 Philips Pavilion's spatialisation scheme devised for Varèse's *Poème Électronic* showed automated control via a 15 track control tape. This makes it technologically interesting and is an innovative milestone for spatial audio per se but removes it from our concern for *spatially* interactive or adaptive audio applications as the automation puts the spatialisation firmly on a (linear) timeline:

the spatialisation of audio, expressed more generally in the emergence of stereophony was seen as a new, additional means of composition and production (Judd, 1961) But interactivity was not a direct concern yet for the researchers of the time, more a by-product of lack of automation.

For the *interactive* or *non-linear* aspect of spatial audio technology, there is not much evidence of developments of significance thereafter, and only now, as mentioned above, with the current drive to ubiquity of digital technology has interactivity become intrinsic to spatial audio and locative audio content.

But as far as spatial sound in general is concerned, further technical developments in the 1960s and 1970s resulted in ambisonics, based on the work by Blumlein (Malham, 1998) and binaural technology, a method using head related transfer functions (HRTF) to decode a signal with consideration of the acoustic properties of the ears' shapes and the distance between them (Blauert, 1997) and wave field synthesis. (Oellers, 2011)

Both systems are looked at in more details later, as they are of relevance for the development of locative audio applications, particularly due to the difference in nature of the two approaches and the resulting crossroads for development of iLAA.

For completeness sake, it shall be mentioned here that all these spatialisation-efforts were predated though by the development of surround sound for cinema, surprisingly initiated in pre-digital times by the Disney Studio's development of Fantasound (for the film Fantasia) in 1940 (Garity, 1941)

Besides film sound, the video - and increasingly online - game industry is the other important global player of relevance for spatial audio. And gaming can be considered the main driver behind the increased interest in interactivity in audio, along with the development of *mobile* devices. The developments of games for smart phones, with built in accelerometers and GPS make locative contents in games an obvious playground for developers.

1.2.2 Spatial Interactivity: Adaptive Sound

Despite having its beginnings as early as the 1950s, video games started without sound, *Tennis for Two*, and *Spacewar!,* developed by MIT in 1958 and 1962, respectively, were silent. Only with the development of arcade games and *Computer Space* in 1971, which featured "battle sounds" was audio introduced to computer games. (Collins, 2008)

As the early personal computers were designed for "serious applications", not much room was given to audio processing. The emerging game culture, which appropriated PCs for its uses, had to work with limited hardware, not developed for "entertainment", concerning audio it didn't come with more than the facility to *beep,* frankly.

Progressively audio was picked up more and more for computer games and via the introduction of FM synthesis and MIDI in the 1980s to today's orchestral material written specifically for game audio which, on its own can fill several CD's for just one game, (Final Fantasy IX, for example) game sound has become an industry driver for the *music* industry. (Collins, 2008)

The next and quite recent step was the development of 3D engines for games. The application of which, in first person shooter games for example, can be easily expanded to audio in a surround system. This led, in fact, to the introduction of Dolby 5.1 surround sound as the standard format for games audio on all major platforms. (Collins 2008)

The thing which makes games audio special and different to film sound, for example, is that the audio can - or should - react *adaptively* to the players interaction with the game. In respect to the audio, the audio is not following a linear timeline, and is thus described as *non-linear.*

This *adaptivity* in principle, can be understood as an actual definition of interactivity as used for computer - user interaction in respect to interactive audio applications in general.

However, Karen Collins, a leading researcher in the field states "Although the technology is now there, the tools are not and the dynamic aspects of games audio, (player participation, non-linearity and so on) have only begun to be explored." (Collins, 2008)

The current development of gaming mowing away from desktop computers (After they long left the arcades) onto mobile phones and onto the mobile internet and into the cloud, is another step towards ubiquitous and pervasive computing.

In their contribution to Collins "Pac-Man to Pop", Agnes Guerraz and Jacques Lemordant, (Collins 2008, p. 343) put the consequences of this and possibilities therein to the point: "Audio is a participatory medium, which actively engages the listener in the ongoing processing of aural information. Audio constitutes an essential part of any game, especially in helping to create a feeling of immersion. A convincing combination of sound and image is a key aspect in producing games. Spatialised sound sources, audio special effects, interactive audio and animated digital signal processing (DSP) parameters are the main ingredients for building such environments."

It is in this sense of tool finding for increased non-linearity this research into positioning systems for interactive locative audio applications (iLAAs) understands itself.

The recent developments have brought with them a terminology which probably culminates in the concepts of pervasive media and ubiquitous computing.
One of its early pioneers, Mark Weiser, in "The Computer for the 21st Century" written 1991, stresses the fact that ubiquitousness is practicable only if they are "invisible in fact as well as in metaphor" and "Already computers in light switches, thermostats, stereos and ovens help to activate the world. These machines and more will be interconnected in a ubiquitous network. As computer scientists, however, my colleagues and I have focused on devices that transmit and display information more directly. We have found two issues of crucial importance: location and scale. Little is more basic to human perception than physical juxtaposition, and so ubiquitous computers must know where they are. (Today's computers, in contrast, have no idea of their location and surroundings. [in 1991])If a computer merely knows what room it is in, it can adapt its behaviour in significant ways without requiring even a hint of artificial intelligence."(Weiser, 1991)

1.2.3 Examples of interactive Locative Audio Applications (iLAA)

Conceptually, it's easy to see that the performing arts are particularly interested in spatial interactivity between performer and music.

Feldmeier and Paradiso (2007), describe a radio frequency (RF) based system which transmits signals from sensors carried by individual dancers to several base stations, and from there to a mapping system, processing audio data according to the gathered information. This enables the use of "zoning information" "to direct the music and lighting to respond to the participants' actions in that area, localising the response to a smaller group of proximate dancers."

As the research's aim was "to create an engaging and enjoyable musical experience", the explicitly highlighted shortcomings of motion tracking devices did not form a problem. But highlight the need for further development, particularly for larger scale projects with more than individual dancers. Cost, data-communication bandwidth and the need for "highly structured and stable stage environment with tight lighting constraints" are mentioned amongst others.

Despite the fact that their system has the "ability to collect simple data reflecting each individual's motion" it can't evaluate which individual sensor sent the information, and crucially where it moved to exactly in relation to the next nearest sensor at a given moment.

Morales-Manzanares paper *An Interactive Music Composition System Using Body Movements sums up further limitations*: "There remains a distinct need for a simpler, easier to understand, and powerful coupling mechanism that mediates between sound and motion." (Morales-Manzanares and Morales , 2001)

Both these examples of recent research indicate the need to identify more suitable positioning systems or combination of such systems for interactive audio applications.

Despite the limitations, the conceptual links of spatial music and dance seem self evident and there is particular interest in the world of dance for means of translating movement to interact with sound work by Cliff Randell and Stan Wijnans, for example led to the design of a tracking system in Max/MSP for this purpose. (Randell, 2012, Stan Wijnans, 2009)

In the world of video game sound, the appeal of using interactive spatial information for audio is particularly evident in kinetic gesture games, (wii for example) where "games are designed to have players directly physically participate and respond to the sound" (Collins, 2007) However, here as well as in the other areas of relevance for audio spatial interactivity surprisingly "academic […] work into the sonic aspects of audio-visual media has neglected games" Karen Collins outlines the importance of audio in games as "adaptive sound design", i.e. as sound which “reacts appropriately to—and even anticipates— gameplay” rather than responding directly to the user (Whitmore 2003 in Collins, 2007). This further underlines the importance of spatial audio in video games and its interactive nature.

To put the importance of spatiality in gaming to the point Collins argues: "The illusion of being immersed in a three-dimensional atmosphere is greatly enhanced by the audio, particularly for newer games which may be developed in 7.1 surround sound. " And "It must be recalled that a game may take thirty to forty hours to complete […] and audio plays a crucial role in helping the player to recall places and characters, and to situate themselves in such a massive setting."

Here is another example of a non-explicit suggestion that integration of spatial interactivity for audio applications would be a valuable addition, but the explicit absence of such a suggestion shows the need for research into user requirements.

However, there are solutions which work despite the limitations:

A fairly recent example of a locative audio application is the iphone game Papa Sangre, (launched December 2010) which uses HRTFs in realtime "the first binaural real-time, 3D audio engine implemented on a handheld device" according to the developers, it uses audio only, no graphics other than game flow-control (Kiss, 2010) and the inertia - navigation aids of the iphone.

A further example, one of gestural control, where audio content is controlled by head movement of a musician is the TangibleFX Cap, which can be applied as a means of control similar to foot pedals (Pervasive, 2012).

1.2.4 Conclusive remarks on this chapter

This research into the possibilities of spatial interactivity for audio applications can be understood as part of a development towards ubiquitous computing: If computers must "know where they are", the "spatial awareness" of a computer or "computing device" can supply us with the necessary information for iLAAs: "If a computer merely knows what room it is in, it can adapt its behaviour in significant ways."

If the concept of non-linearity, as discussed in connection to game flow-control through the activity of the gamer is understood as a form of Weisers "embodied virtuality " It follows that this process needs to be, as Weiser has it "invisible in fact as well as in metaphor".

As can be seen from this, not directly a derivative of pervasiveness per se, adaptivity is none the less a crucial part of ubiquitous computing, particularly in order for computers to fade into invisibility: If visibility is understood a bit broader (or in "metaphor," rather than "fact") it could be argued that, whenever we *notice* a computer, it is visible. So if, for example, in an interactive audio application like a computer game, our actual spatial position does not correspond to our position within the game, we'll *notice* that fact as it counteracts the immersion.

As mentioned earlier, in principle, the *adaptive* nature and *non-linearity of* games audio can be understood as a definition of *interactivity* as used for computer - human interaction in respect to interactive audio applications in general.
For *spatial* or *locative* interactive audio applications, the *spatial dimension* is added. It is widely

accepted that spatial or 3D audio is essential for the immersive nature of games audio. (See Collins 2008) It simply follows then that this spatiality has to be adaptive, interactive too, which is by large realised by 3D engines in First Person shooter games, for example, where the audio reacts spatially and in surround to the players actions in the *virtual* space. As soon as the player moves around in real space, however, the need for some sort of positioning system becomes clear in order to achieve *adaptivity*

Existing interactive applications often suffer from the limitations of positioning systems for audio applications. Despite these limitations, audio *only* games like *Papa Sangre* show that it can be done. Many successful applications are successful because they don't tamper with the limitations by setting the parameters for their application within the possibilities of the positioning systems.

In short, the current literature shows a clear need for development of positioning systems for iLAAs.

In the following 3 chapters, with the aim towards a change for this situation, a threefold approach is being taken:
Firstly, by studying the limitations of spatial hearing it is analysed how precise positioning systems for spatial audio applications need to be, based on the presumption that positioning systems do not need to be more precise than - but are at best as precise as our ability to locate sounds, secondly, analysing existing positioning systems in respect to their ability to meet the above criteria and thirdly comparing both the natural limitations and the existing positioning system technology with the demands for actual applications. This third step will consist of an attempted taxonomy of iLAA - as different applications have different needs - and the findings from a survey into user requirements on I/LPS for iLAA. The respondents are a representative mix of developers, composers and academics in the field.

2. Parameter Limits based on Spatial Hearing

2.1 How hearing relates to seeing

When it comes to experience space, hearing is in many ways only a secondary means to seeing. Blauert, in his book *Spatial Hearing*, (Blauert, 1997) goes as far as defining "human beings" as "primarily visually orientated". this priority on the visual pervades, according to Blauert "scientific and technical descriptions" . He exemplifies this by quoting the German industry standard for sound, DIN 1320 which defines sound as "mechanical vibrations and waves of an elastic medium particularly in the frequency range of human hearing (16 Hz to 20 kHz)" and goes on: "This is a description of physically measurable changes of position, primarily perceived visually. What is heard, what is perceived auditorily is only implicitly included in the final phrase 'particularly in the frequency range of human hearing'"

He further analysis this definition thus: "The underlying assumption is that a normal human being generally hears something when in a medium in which vibratory or wave phenomena are occuring whose frequency is between 16 Hz and 20 kHz. This does not, however mean that the vibrations and waves of the medium are what is heard" and exemplifies this by the fact that a person with plugged ears might not hear a sound event but might perceive its waves and vibrations by visual means.

This prompts him to clarify the terminology by introducing the "auditory event" as the the "auditorily perceived sound event" rather and in contrast to "sound" as in DIN 1320.

This distinction is necessary particularly for spatial hearing as the auditory event i.e. the perception of a *sound event*, might locate the sound in an apparent location which is geographically different to the location of the physical source of the sound event. (The whole concept of stereophony relies on this!)

Interestingly, as to the question which location is the intrinsically true location of the sound Blauert writes: "The sound source and the auditory event are both sensory objects, after all. If their positions differ, it is an idle question to ask which is false." (Blauert 1997, p. 4)

By describing an experiment by Wallach of 1940 Moore shows an example of the influence of vision on auditory localisation (Moore 2008, p. 264). The subjects had their heads fixed on the vertical axis within a cylindrical screen which rotated about them. The screen was covered in vertical stripes. After a while the subjects perceived themselves to be in motion and the screen at rest. When a sound source straight ahead of the subjects was activated, the sound was perceived as coming from either above their heads or from below. This perception is the only possible consistent result of what the subjects hear *and* see and thus, as Moore has it, "the highest level of spatial representation involves an integration of information from the different senses."

Blauert (1997, p.193) uses the very simple example of the day to day occurrence - in the times of mono-TV sets, that a person watching TV hears an announcer in the location on screen and only when closing the eyes does she or he notice that the auditory event comes from the side, if that is where the loudspeaker was.

However, despite this apparent sub-ordinance of hearing to seeing, space can be experienced as an auditive phenomenon and it is hard to dispute that every sound *has* spatial attributes, attributes which we *hear.* (Blauert, 1997; Blesser & Salter, 2007; Moore, 2008)

2.2 Spatial Hearing: Experimental Considerations

What we know about human sound localisation we know from psychophysics, the study of the relation between stimulus and sensation. (The *sound event* being the stimulus of the *auditory event* which is the sensation.) In psychophysical experiments the sensation is never directly observable and we rely on the subject's account of the sensation. Blauert (1997) deals with this problematic by introducing a "black box" in the sense of a systems analysis where the sensation (the auditory event) is shown as an output which can only be observed introspectively (when the experimenter is also the subject) but the *description* of the auditory event is a quantifiable output. (See Fig 4)

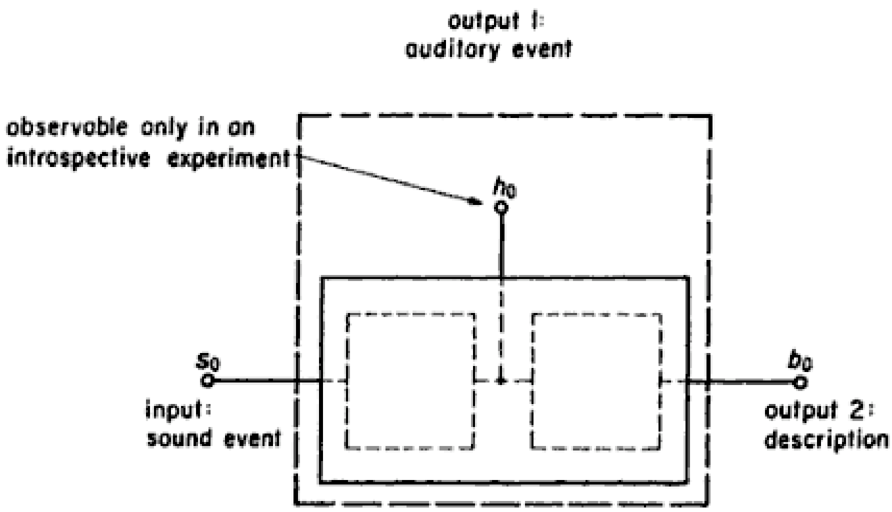


Fig 4 (Blauert, 1997)

The psychophysical experiments establish spatial hearing abilities by studying the *thresholds* of hearing a sound event in situations of interest. These observations of auditory events can be illustrated in a head-related system of coordinates on intersecting planes, namely the horizontal, frontal and median plane which are at right angles to each other and intersect in the origin between the two ears at the height of the upper margins of the ear canal.
The horizontal angle, azimuth, increases anti clockwise and the elevation angle increases upwards. The distance is usually expressed as r. (See Fig 5)

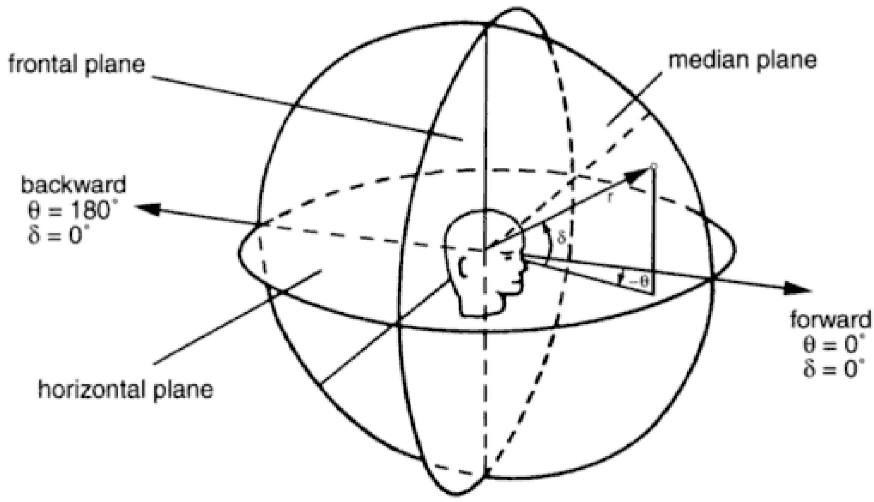


Fig 5 (Moore 2008)

As there is good evidence that something quite similar to spectral analysis (some organic variant of a Fourier transform) is happening in the neural processing of auditory events in the ear, (Moore 2008) complex sounds' components can be analysed separately for their psychophysical effects on subjects to understand the underlying principles of spatial hearing.

The such derived experimental stimuli are further grouped into periodic (tones) and non periodic (noise) sounds and various stages between them (e.g. modulated noise)

Of further importance is the duration of the stimulus: Transient sounds and periodic sounds with low repetition rates have a different effect than constant periodic sounds or constant non periodic sounds.

Further, stimuli can be observed *monaurally*, i.e. a subjects observation are about a stimulus on one ear only, (The other needs to be blocked) or it can be observed *binaurally*, i.e. the effects of a stimulus on both ears is observed.

And a stimulus can be the same for both ears or different. these stimuli are then either *diotic* (same for both ears) or *dichotic*. (different for one ear than for the other)

Another crucial factor in psychoacoustic experiments is if the experiment is conducted in a real soundfield or using headphones. There is a certain amount of controversy about the use of headphones in psycoacoustic experiments as many experiment conducted with headphones ignore the effects of the Pinna (The outer ear as we see it) on the sound perception, and, the ones that do by using head related transfer functions, (see below) often use standardised HRTFs rather than individual ones.

But, as most researchers use whichever method suits the particular experiment better and as the problematic is widely known, this controversy remains fairly philosophical and certainly doesn't interfere with the validity of our discussion here.

However, it is worth noting in view of the later discussion that even here for psycho-acoustic experiments, for technological reasons, a sound representation has to be chosen which is either happening directly on the ear or in a physical room, in a soundfield. This fundamental choice presents itself in the design of spatial audio applications too, for the same technological and or psychophysical reasons.

Ideal sound sources for psychoacoustic experiments are pulsating sound sources. (See Fig 6) They emit a sound equally in all directions and are small compared with the wavelength they generate. With increased distance the spherical wave becomes increasingly plane. For a 3m distance from a pulsating source loudspeaker, the level difference between 2 points spaced apart by approx. 9cm, (simulate the distance between the human ears), is less than 1 dB and hence ignored for most experiments. (Blauert 1997)

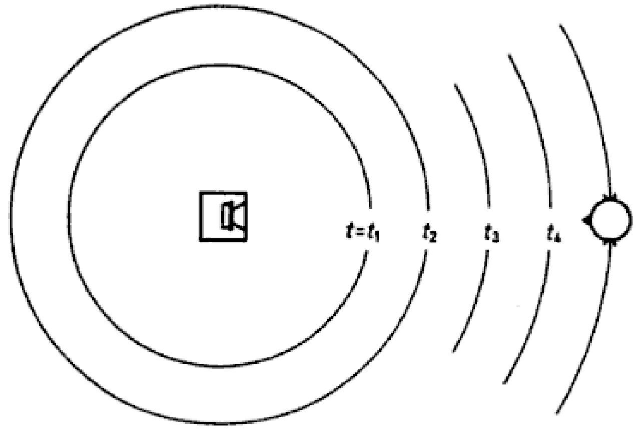


Fig 6 (Blauert, 1997)

Or in other words, the soundfield is presumed to behave according to Huygens principle according to which every point in a system of waves can be regarded as the source of a spherical elementary wave, or, as Malham puts it (Malham 1999) "a propagating wavefront may be regarded as consisting of a large (ideally infinite) number of secondary sources which add up to form the wavefront."

Keeping in mind that we ultimately want to establish the limits of spatial hearing in order to define parameters for positioning systems, we are interested in the threshold of spatial hearing, the finding of which happens to be the method used for most psycho-acoustic experiements.

One thus established factor of spatial hearing is the *localisation blur,* LB "the amount of displacement of the position of the sound source that is recognised by 50 percent of experimental subjects as a change in the position of the auditory event." (Baluert, 1997)

The other, similarly derived term used by Moore (2008), the minimum audible angle, MAA is defined as "the smallest detectable change in angular position, relative to the subject".

As both terms are expressed in degrees, they will be used here interchangeably, but in accordance with the source of the discussed findings. (MAA if referring to Moore, LB if referring to Blauert.)

The localisation blur varies with the nature of the sound source and its position in the head related system described in figure *Fig 5.* Let's look at the differences between horizontal and median plane separately and in more detail.

2.3 Spatial Hearing in the Horizontal Plane

In the horizontal plane, the main mechanism applied by the hearing system to locate sound is the interaural time difference (ITD) (See Fig. 7) which is established by measuring the MAA binauraly. However, experiments have shown that ITD as a prevalent means of locating sound relies on the located source to emit at low frequencies. (< 1500 Hz)

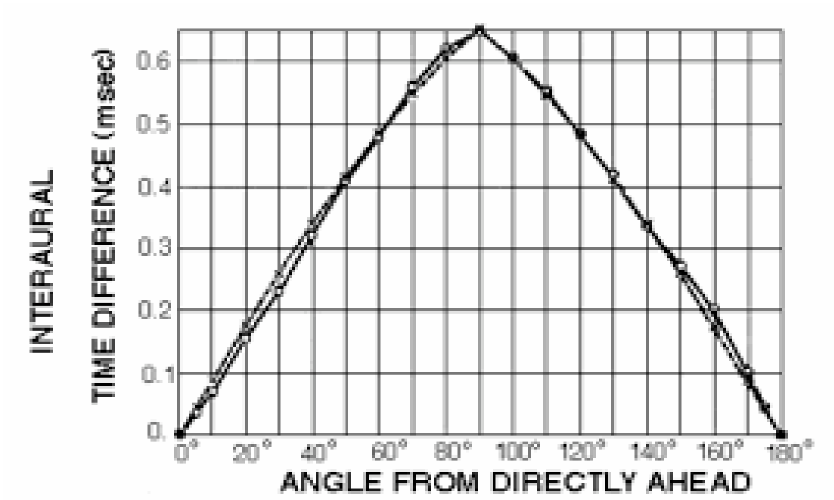


Fig 7 (Moore 2008)

The maximum time difference between the two ears (for a distant sound source) would be at an azimuth 90°. Estimating the distance between the ears as 9 cm and the speed of sound through air as 30µs per cm this would give a time delay of around 690µs, based on a spherical head. (Moore 2008. p.237)

For high Frequencies (>1500 Hz) interaural level difference (ILD) become more important cues. For transient sounds and periodic sounds with low repetition rates however, ITD is still the main mechanism even at high frequencies.

For periodic sounds of longer duration a form of progressive adaption comes into being: the onset of the periodic sound is used for positioning but not later parts. The signal needs a new trigger of slight spectrally different characteristic to re-new the localisation of that sound. (Moore. 2008)

According to a compilation of various workers' findings for various types of signals, (Blauert, 1997, p. 39) LB varies from 0.75°- 2° for Impulses (Clicks) to sinusoids between 1.1° - 4°. (See Fig. 8)

Table 2.1
A survey of measurements of localization blur $\Delta(\varphi = 0)_{min}$, i.e., for horizontal displacement of the sound source away from the forward direction. Because different measuring techniques were used, reference is made to the original works, where the techniques are described in detail.

Reference	Type of signal	Localization blur (approximate)
Klemm (1920)	Impulses (clicks)	0.75°–2°
King and Laird (1930)	Impulse (click) train	1.6°
Stevens and Newman (1936)	Sinusoids	4.4°
Schmidt et al. (1953)	Sinusoids	>1°
Sandel et al. (1955)	Sinusoids	1.1°–4.0°
Mills (1958)	Sinusoids	1.0°–3.1°
Stiller (1960)	Narrow-band noise, cos² tone bursts	1.4°–2.8°
Boerger (1965a)	Gaussian tone bursts	0.8°–3.3°
Gardner (1968a)	Speech	0.9°
Perrott (1969)	Tone bursts with differing onset and decay times and frequencies	1.8°–11.8°
Blauert (1970b)	Speech	1.5°
Haustein and Schirmer (1970)	Broadband noise	3.2°

Fig 8 (Blauert, 1997)

The LB increases with displacement of the soundsource in the horizontal plane away from the forward direction. It is best at 0° azimuth (1°-4°) worst for 90° and 270° , respectively (9-10°) , but improves for 180° to a respectable 5.5° (Fig 9)

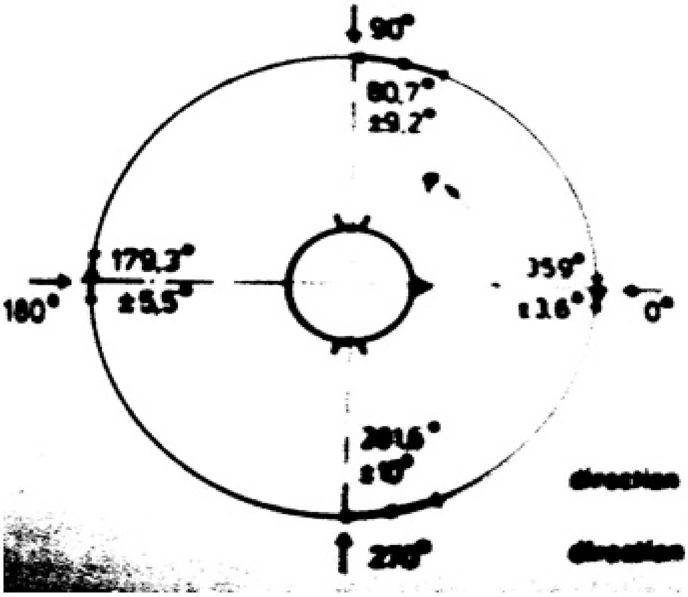


Fig 9 (Blauert, 1997)

Moore (2008) illustrates the relationship between frequency and azimuth angle for interaural *level* differences (ILD) in a table which he adapted from Feddersen *et al.* (1957)

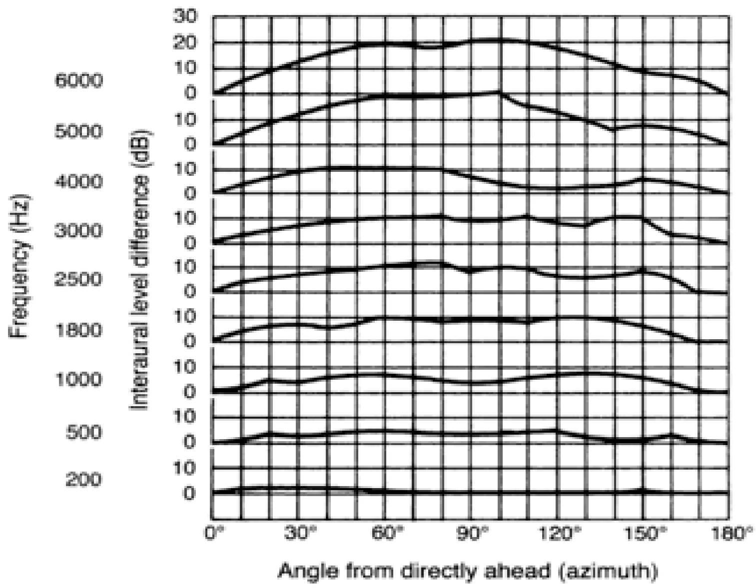


Fig 10 (Moore 2008)

In Fig 10, MAA is shown as a function of frequency for 4 different reference directions for sinusoids. This data is derived from an experiment by Mills (1958)which used loudspeakers rather than headphones, and hence generated ITD and ILD . Comparing this Fig 10 with **Fig 7** and Fig 11 shows how for frequencies up to 1500 Hz ITD's are relevant and from around 3000 Hz upwards, ILDs.

Further, the increase of MAA with azimuth difference to forward is illustrated in the starting points of the individual curves of Fig 11 straight ahead, (The 0° Curve) starts at MAA < 1°, whereas the 75°curve, for example, starts at MAA = 6.2°

Other factors than ITD do play a role for some sounds in the horizontal plane, but, their influence on LB and MAA are overwhelmingly more relevant for the majority of sounds, particularly when considering the relevance of head-movements, as discussed later.

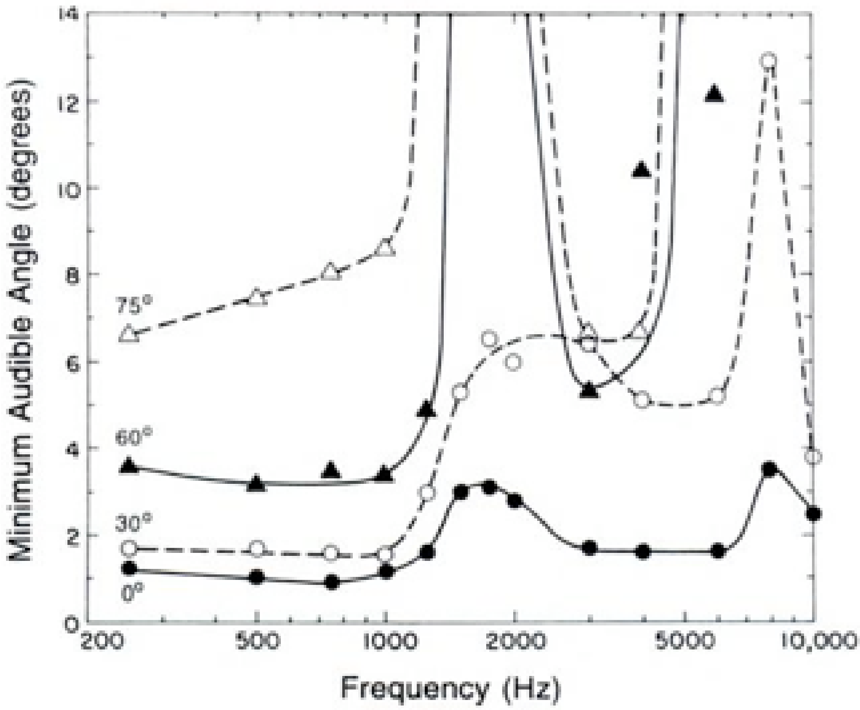


Fig 11 Moore (2008)

2.4 Spatial Hearing in the Median Plane

The fundamental difference for directional hearing in the median plane from the horizontal plane is the absence of interaural time difference information. And for sound in the median plane, subjects find it often hard to differentiate if a sound comes from directly ahead or directly behind.

Localisation blur for the elevation angle in the median plane for speech has been estabished by Blauert (1997) as 17 ° for speech of a person unfamiliar to the listener, 9 ° if the person is familiar and 4° for white noise. (Fig 12,)

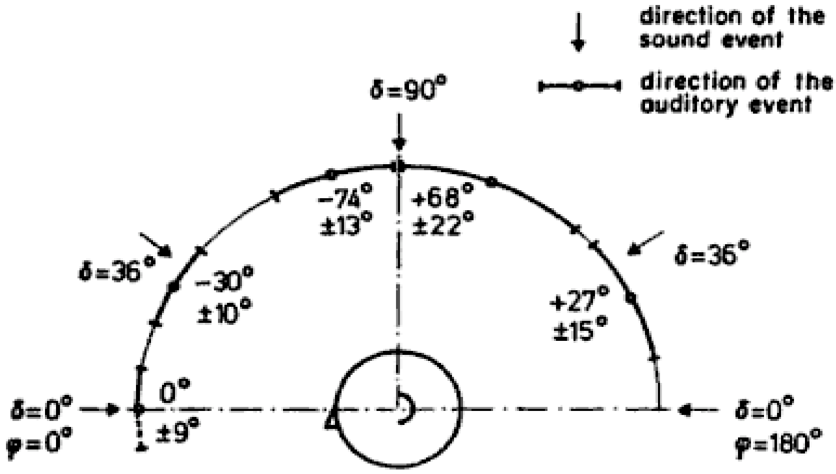


Fig 12, Blauert (1997)

In the absence of ITD cues, for vertical and front-back discrimination the frequency content of a signal becomes important for the localisation process. In particular it is the frequency response of the outer ear, the Pinna, which gives relevant clues.

The Pinna effects are not only valid in the median plane. But with the reduced availability of ITD (absent in the case of a source exactly overhead) they become more relevant. How relevant can be seen from experiments by Batteau (1967), Freedman and Fisher (1968)* described in Moore (2008) wherein microphones placed in artificial Pinnas connected to standard headphones thus provided "remote ears" eliminating the influence of the head shape (and ITD). Subjects established localisation of sounds from all directions with "reasonably accurate judgement" both vertically and horizontally.

Consequently, an impulse response measured at the entrance to the ear canal for a particular sounds spatial characteristics, can produce a so called head related transfer function (HRTF) which can then be applied to any signal, thus, if reproduced on headphones, the auditory event will be a good approximation of the one heard on the Pinna in the soundfield.

HRTFs are as good as the measurements they are based on. An illustration of 10 HRTFs with a separation of 30° (Between which a computer could interpolate the other angles) is shown in Fig 13

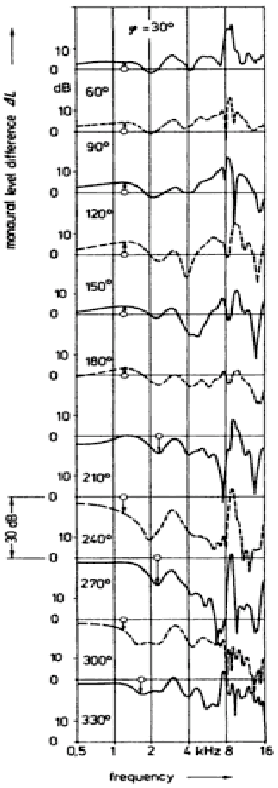


Fig 13 (Blauert 1997)

The technological limitations here are the fact that every subject's ears are slightly differently formed. Commercially available HRTFs hence are based on an average and normalised ear.

[not sure the next two paragraühs are needed, as they may complicate rather than explain?]
For signals of narrow bandwidth (< 2/3 octave) in the median plane "the direction of the auditory event depends not on the direction of the sound source but only on the frequency of the signal" (Blauert 1968b in Blauert 1997, p. 45)

This means that a a narrow band sound for which no interaural information is available , regardless to its position in the soundfield,is always heard where the peak of its HRTF would be lying. So, a sound source emitting the frequency 8 kHz lies always right above the subject, in accordance with that frequency's HRTF. (Moore 2008)

What can be a bit confusing about this experiment is that the test tone has to be diotich in order not to generate any interaural cues (ITD, ILD), hence the experiment has to happen monaurally, for example.

However, in contrast to the psycho-acoustic researcher's aim to find out what a particular restrictions effect is on spatial hearing (in order to understand the underlying principles) we are interested in the overall capability and *integrated* ability of all contributing factors in day to day spatial hearing, relevant to user requirements on positioning technology.

This is why we have to remind ourselves that these experiments were all conducted with the subjects head in a fixed position: A natural sound source in most day to day situation would only stay in the median plane *as long as we don't move our head* in relation to the sound source.

Head movement can not only put a sound source from a median position for example into a more forward or lateral position, from where the same cue can be analysed by using localisation mechanisms with lower LB or smaller MAA, but it also helps to disambiguate situations where a cue gives conflicting, rather than just inaccurate, information:

One example is the disambiguation of interaural *phase* differences. (multiples of wavelength can give "wrong" phase aspects) Another is the "cone of confusion", which expresses the fact that a signal of a particular ITD can, if ignoring Pinna effects, lie on a cone, its tip at the ear, due to the quasi spherical shape of the head. If head movement results in a equivalent shift of the auditory event, the source is perceived to be on the horizontal plane. if no shift results, the source must be on the median plane.

2.5 Distance Discrimination

Distance perception relies mainly on sound levels, and, compared with the abilities in the horizontal and meridian plane, is rather crude. Errors are often in the region of 20% for uncommon or unfamiliar sources. (Moore 2008) For familiar sounds, the attenuation of high frequencies over distance give additional cues.

Even if the distance estimation is not very precise, the ability to notice a *change* in distance of a fixed source i.e. when the listener is walking towards or away from a cue, is a lot more evolved. (Moore 2008)

So, for most practical purpose, an increase in sound level is perceived as a decrease in distance and vice versa.

A further cue might come from the reverberation of a sound. Again, it's the change in reverberation which tells us about a change in distance, not necessarily much about the exact location of an auditory event.

This discrimination of distance by changing location plus the *integration* of all aurally available information (Moore 2008) might explain phenomena like blind cyclists' abilities to cycle through New York without crashing into things, as described by Blesser and Salter (2007). But the available psychophysical data, particularly as far as distance perception is concerned, clearly can not explain such abilities directly.

Again, head-movement can often turn a change of distance into a change of azimuth. (Pure distance perception relies on the sound source being straight ahead) and a stationary sound source can be turned into a moving sound source (thus creating a change of location) by physically approaching it.

So only in a situation where no head movement is possible, it can be argued that no real accuracy can be achieved: The various adverse factors due to unfamiliarity of the to signal characteristics, the fact that distance hearing is not really possible for stationary unknown sources can be overcome: In most applications - at least applications taking place in the *soundfield* head movement *is* possible, so the localisation blur can be assumed to be the same in all directions for if the listener can move about freely.

Thus, as a working hypothesis, let the suitability of positioning systems for LAA be maximally defined by the limits of the auditory resolution imposed by our hearing apparatus, and supose that a suitable positioning system does not need to be more precise as our ability to localise sounds.

2.6 Deriving Positioning parameters from Localisation Blur

In order to define parameters for user requirements on positioning systems for iLAA,we are interested in the outer limits of our ability to locate sounds as this directly limits these requirements on the performance of localisation systems: The localisation does not have to be faster or more precise than our ability to locate sound psychophysically. As discussed, this doesn't always exclude senses other than the ear, but nonetheless, a positioning system for LAA does not need to "outperform" what we can actually perceive.

These limits can be directly derived from the thresholds of hearing, the method used in most psycho-acoustic experiements. (LB and MAA, see above)

The translation from minima and maxima of spatial perception to parameters of potential positioning systems to establish requirements is not entirely straight forward: The ability to detect movement is defined in the minimum audible *angle*, consequently the location blur increases with distance. This rmeans that a movement of a sound source by 1 cm, for example, can be detected if the sound source is close but not if it is far away.

So the resolution of a positioning system needs only be as precise as the localisation blur occurring at the position nearest to the subject which is still relevant for the application.

It has to be remembered that the localisation blur is defined as "the amount of displacement of the position of the sound source that is recognised by 50 percent of experimental subjects as a change in the position of the auditory event" (Blauert 1997) This is a common definition for threshold values in psycho-physics.

It follows then that in an experiment where 50% of subjects detected a change in position of a sound source when it was moved by the MAA, 50% can't hear a change even if there is one, or hear one when there isn't at a given MAA.

This defines the required accuracy of the positioning system: if 50% of the positions are wrong, but within the area covered by the blur, the subject wouldn't know any better as it is within the subjects spatial resolution. This should be the maximum error though, as if it is above that, the system's error becomes misleading. But If the PS error is smaller than the blur error, the PS error is negligible as the subject can't discern which one caused the error, the PS or the localisation blur.

We can relate the localisation blur to a specific distance of a sound source by the proportionality of a central angle to the distance of two points on a sphere. So for every MAA (in radians) there is a maximally axceptable error for the positioning system Err(max) (in m) at a given distance r (in m) from the subject. This could be expressed in the formula

$$Err(max) = r \quad * MAA$$

so as a numerical example, the maximally acceptable positioning error for a sinusoid (MAA = 3°) at 10 m distance would be

$$10 \times \frac{3 \times \pi}{180} = 0,52 \text{ m}$$

For a near-field application however, and using speech (MAA = 0.75°)

$$1.5 \times \frac{0.75 \times \pi}{180} = 0.02 \text{ m}$$

For the median plane the same formula for maximum Localisation blur (MAA = 17°)

$$1.5 \times \frac{17 \times \pi}{180} = 0.44 \text{ m}$$

And, as an example for large distance and large blur: for a 10m distance the 17° would result in a 3 m error.

This established maximal acceptable error is to be understood as follows: if the LB at a particular distance is 0.44 m, a subject can not be sure if an auditory event is 0.44 m more to the left or the right, but whatever she or he hears is either where it is heard or within 0.44 m of it.

The next step is to work out how to relate this to the positioning measurement error.

As the LB area for a particular point is on a perpendicular plane at a distance r from the listener but only defined as a line (between to points defined by the angle between them) we're actually interested in the projection of the measuring errors of the measured *area* surrounding the *point* of the right measurement onto the plane on which is the line covered by the LB.

But as a measurement is either right or wrong, and as the projection would contain the same amount of right and wrong data in relation to the LB, we refer to both the LB - line and the values projected onto it by the positioning systems measurements as an area of errors. (The right position fix would be on the line defined by LB, just the error might not!)

So to tune the measurement error with the localisation blur it has to be worked out what the maximum allowable error is for the positioning system. It has to be considered that potentially, the errors could accumulate.

The probability of this can be calculated based on the following two independent events:

Let's presume the maximum allowable errors for both are the same:

Event A: an auditory event heard by a listener is probably 50/100 wrong. (The listener is wrong)
EventB: The auditory event is not in its intended place 50/100 (The positioning system is wrong)

The chances of both events occurring is a joint intersection of P(A∩B) which is 1/2 * 1/2 = 1/4 or in percentages: 25 out of 100 readings are both wrongly located and wrongly heard.
This seems unsatisfactory and a bit crude. Leavitt & Payton (1996) assume, quite reasonably, for positioning systems such as GPS, the maximum allowable error for a position fix to be half the picture resolution of the acquired image.

Adapting this here we could say that the LB Error(max) as defined earlier, could be looked at as the resolution of the acquired image. (Or the projection of the image on the line defined by LB and use 1/2 of the error of the blur as the maximum of the allowable error for the positioning system.

Adjusting the probability calculation for these events, A remains the same but event B is now 25 out of 100, assuming the observations halve with the doubling of the area. (we can presume it is even less, due to measurements errors' generally assumed normal distribution)

If we looked at the resolution of LB as the acquired image, in our example, the maximum error would then be 0.22 m.

Therefore

50/100 * 25/100 = less than 12.5 % of errors are due to the positioning system.

Or more generally: Setting the maximally allowed measurement error at half the distance covered by LB at a certain point results in a maximum cumulative error of less than 12.5%.

2.7 Spatial Sound Reproduction

So far we looked at the implications of spatial hearing only from the point of view of how accurate a positioning system would need to be to reproduce this *spatiality*. But how the choice of spatial reproduction technology influences user requirements has so far only tangentially been discussed, for the special case of binaural technology.

As a working hypothesis, let's assume that the nature of an application (Discussed later) has an influence on what spatial reproduction system a developer would choose, (hence having an influence on user requirements) it seems the right place to have a look at these systems as they are closely related to the way we hear spatial audio.

There's several ways of technologically reproducing spatial sounds spatially coherently, with their strengths and lmitaions: Either to reproduce the sound as realistically as it sounded originally *in space* or to reproduce it as we hear the original *in the ear* The first way is atempted via wave front synthesis, ambisonics, holography and to an extent Dolby Surround, the latter via stereophony and binaural technology. Many technologies uses a bit of both.

Despite its seemingly tenuous connection to spatially *interactive* audio applications the spatial sound technologies developed on the back of psycho-acoustic factors crucially influence the choice of technology for developers of LAA as a means of spatial audio *reproduction*: An application in which the audio content shall only be heard by single users would not want to develop a system for ambisonics but would probably choose a reproduction system applying HRTFs as this can be done via headphones.

This was confirmed in discussion with professionals in the field in the focus group at PMS Bristol: the realisation of many LAA relies directly on our perception of locative audio content which, intrinsically is spatial.

From a technological point of view and in technological reproduction of locative audio, it is hence often necessary, sensible or beneficial to recreate the spatiality of audio content in the for the particular application more appropriate form. (by means of stereophony, binaural technologies or ambisonics for example).

The requirement onto the precision of this spatiality is dependent on the *nature* of the application too. But if an application's required locative resolution is high the limiting factor for the audio content is the limits of the spatial resolution of our hearing.

Application specific requirements are looked at closer in the chapter on User Requirements, But the two main technologies with a bearing on iLAA shall be shortly highlighted here:

2.7.1 Implications for Binaural Technology

The inherent limitations of a binaural system when using headphones are the conflicts of spatial coherence arising from the fact that the "virtual" world, or the audio world portrayed over the headphones moves when the head moves, or indeed if the listener moves about. (Malham, 1999) As Malham already stated in 1999, the head movement could be tracked, and the HRTFs interpolated in realtime, so that headphones could actually give the illusion of a static soundfield outside the headphones.

However, it gets more complicated to keep the congruence intact as soon as a person wearing headphones moves about within the "virtual" soundfield portrayed via the headphones: It would probably be easier to correct the signal in a way that the "real" sound source is spatially signal processed in accordance with the "virtual" audio event. A simple example of such a processing system is suggested in form of a VST plug-in for example in Schlienger (2011): The plug-in's Cartesian co-ordinate system allows a sound source to be three dimensionally panned. A positioning system could provide these parameters from the position of the listener directly to the plug-in.

Malham (1999) gives reasons why Dolby surround can be seen as coherrent, despite - or in fact due to - the spatial warping that happens due to Dolby systems being basically stereo *plus more speakers*: The two dimensional screen is a spatial simplification which is coherently portrayed with the Dolby technology. The problems start when these systems are used for non-cinematic content.

2.7.2 Ambisonics (and Wave Front Synthesis)

Ambisonics, as a method of reproducing recorded soundfields with an arrangement of more than two speakers is based on technological developments which started in the early 20th Century, with Alan Blumlein and the American RCA studios, amongst others, but Michael Gerzon has to be credited for the development of most theoretical and practical aspects (Furse, 2011, Malham, 2009 Wikipedia 2012b).   The basic idea behind it is, that, building on Blumleins ideas, two near coincident perpendicular figure of eight microphones give an astonishingly appropriate representation of the soundfield surrounding them. If reproduced on speakers on a sphere of known proportion, the signal of the recording can be decoded according to relatively simple mathematical formulas as to the relative angle aspects of the four directions of the signal (X ,Y, Z, W) This type of ambisonics is referred to as B format or first order ambisonics (Malham 1998).
In this sense ambisonics can be viewed as a form of wave front synthesis (Daniel et al. 2003).

The near incidental arrangement of the microphones for recording the soundfield for ambisonics stands in direct conflict with the understanding that the shape of our head and in particular the pinna influences our spatial experience of what we hear (Blauert 1997).
Over the last hundred years, schools of thoughts which rate one method over the other have developed (Martin, 2006) but recently technologies are being developed which combine the two approaches (Noisternig et al.).

What is an undisputed advantage of ambisonics over binaural or stereophonic systems is the reduced reliance of the positioning of the head: As binaural systems rely on headphones "the world" turns with the movement of the head. Stereophony (and to some extent Dolby surround systems) rely on the listener to be sat in an ideal spot. Ambisonics recreate a soundfield as it was recorded, hence moving around within it, does not distort it, the position of the listener within just changes according to her/his movements.

2.8 Concluding summary of this chapter

In this chapter, aspects of spatial hearing relevant for the development of positioning systems for iLAA have been discussed based on the literature on the psychophysics of spatial hearing. The limitations derived from localisation blur LB and minimum audible angle MAA have been discussed and it was found that the limitations of the horizontal plane can be assumed to be the limitations a positioning system would need to represent. This despite the fact that in the median plane the limitations are more severe than in the horizontal plane as head movement can move an auditory event from one plane to another. Similarly distance discrimination is assumed to be relative to the position and mobility of the listener, thus the azimuth blur is taken as the limiting parameter too.

For binaural technology, using HRTF, the same blur is being assumed, as the ITD are part of the technology but it is suggested that further technological improvements like head direction tracking for increased spatial coherence could improve the technology's potential for spatially interactive technology.

A formula to roughly indicate the expected blur at a specific place in space in relation to a listener is established by $LB = r * MAA$, where r is the distance between auditory event and the listener in meters, MAA the blur angle in radians.

This thus established blur error can be used a direct indicator as to the resolution a positioning system must minimally have to be able to accommodate spatial reproduction convincingly for a listener at a given distance.

The practice of using the half of an observed area as the

At last, the different technical approaches towards spatial sound reproduction or distribution are broadly discussed as to their suitability and a link to the nature of the application emerges resulting in the need to further analyse what a specific interactive audio application is and needs.

This is discussed later in chapter 4. The next chapter however, deals with existing positioning systems as they can now be compared to the minimal requirements established through the limits of spatial hearing: Systems which don't potentially fulfil these requirements can be filtered out.

3. Positioning Systems

This chapter looks at the most common positioning technologies. Many audio applications will not need an accuracy exceeding the psycho-physical limitations of human hearing discussed in chapter 2. However, technologies which are not providing the required accuracy within these limits but which are expect to considerably improve in the near future, as is the case with the new generation of GNSS, for example, are included here, (Rizos et al., 2010) as are technologies exceeding these limits as they might provide intuitive, immersive an novel means of control. (gestural, for example, as described in Packi et al, 2010)

To provide an overview, the technologies are grouped by conceptual similarity rather than by technological implementation. So, technologies based on "dead reckoning methods" like inertia measurements are grouped together regardless if developed for stand-alone hardware or for smart phones, for example.

Most of the systems of interest for iLAA are indoors and local positioning systems (I/LPS). However with the recent improvements in GNSS technology and in the aim - out of the drive for ubiquitousness, of indoor systems to be integrated into the global navigational grid, virtually *all* positioning systems become of interest.

One area of interest for developers of positioning systems and tracking devices is micro tracking used for medical and surgical applications in particular. These systems are of little relevance for iLAA as discussed here and are hence not included in this overview.

Most of the systems discussed here, have been presented at one of the two International Conferences on Indoors Positioning and Indoor Navigation (IPIN), 2010 in Zürich (ETH 2010) and 2011 in Portugal (IPIN 2011) or at the UPINLBS 2010 in Kirkkonummi, Finland (UPINLBS 2010)

3.1 RF systems

3.1.1 Global Navigation Satellite System (GNSS)

Currently there are two GNSS systems active, the Russian GLONASS and the American GPS.

The European system Gallileo, and the Chinese Compass and the Indian IRNSS systems are not active yet. The reason why these systems are being developed beside the existing ones is, that the American Government, who owns the GPS system and opened it for (restricted) civilian use in the mid 1990, reserves the right to reduce accuracy or even shut down the system in "hostile situations" (Global Positioning System, 2012)

The artificial reduction of accuracy, the "selective availability" the original GPS came with, was discontinued in 2000 and accuracy down to centimetre level is available to civilian users. However, this accuracy depends on the quality of the receiver, and for navigational use for cars in civilian traffic inaccuracies of up to 20m have to be expected, not accounting for errors induced by indoors situation or urban canyons.

At that level of error GPS is not of interest for iLAA, however, there are developments which make better use of the available information, even on low cost units available in handsets and mobile phones.

Differential GPS (DGPS) , in Europe pinoneered by the Finnish and Swedish maritime administrations in order to improve safety for boats sailing in the archipelago between the two countries, but now omnipresent throughout, uses fixed, ground-based reference stations to broadcast the (time) difference between the positions indicated by the satellite systems and the known fixed positions.

In the absence of multipath errors (reflections on buildings, trees or other objects) and shadowing (The signal is weakend as it cannot go through an object of high density) DGP can reach stable accuracies up to 10 cm.

Thus, At least for outdoors iLAA, GPS with optimum use of DGPS is a possibility.

The other improvement, currently available commercially only in expensive surveying equipment allowing accuracy down to centimetre level (In the absence of shadow and mutli-path) is Carrier-Phase Enhanced GPS, or CPGPS.

Normal, or code based GPS compares a satellite specific code it receives with the code it expects to synchronise itself wit the satellite's atomic clock. However, the (current) code's phase lengths are long enough to cause errors up to 2-3 meters, without falling out of synch. Most modern receivers deal with this by optimising the synchronisation. The other way, however, is to use the carrier signals phase shift. At 1.57 GHz phase locking will have a lot more precision than on the current code, which modulates at a bit rate of 1MHz. This is what carrier phase enhanced GPs does, after acquiring rough time and position data from the "normal" code sequence and then fine tune to the carrier signal. (Global Positioning 2012)

As the market drivers behind this technology are massive, improvements, which are planned and anounced can be expected to filter down into the mass market.

Summary of GPS Error Sources

Typical Error in Meters (per satellites)	**Standard GPS**	**Differential GPS**
Satellite Clocks	1.5	0
Orbit Errors	2.5	0
Ionosphere	5.0	0.4
Troposphere	0.5	0.2
Receiver Noise	0.3	0.3
Multipath	0.6	0.6

Fig 14 (Global Positioning 2012)

The remaining limitations on the system are multipath and errors induced by the Ionosphere. A list of errors and a comparison of GPS and DGPS is shown in table (Fig. 14)

3.1.2 Pseudollite

Pseudolites, standing for pseudo satellites, conceptually stands for ground based fixed satellites not dissimilar to the ground based stations used for DGPS. The European GNNS system Gallileo, for example is being developed with the help of pseudolite technology.

Pseudolites are of interest here as many workers suggest pseudolites for indoor use, again similar to DGPS providing GPS connection in areas where LOS cannot be achieved. With the presence of pseudolites, GPS could go ubiquitous.

A system working on similar principles is Locata, which provides very accurate positioning with the help of algorithmic multipath correction by separating the direct signal from the rest. Systems like this need exact calibration, but once installed they can be used fro many purposes other than the primarily intended ones, thus, if a municipality for example decided to locally extend the GPS system this way, location-based services (LBS) could pick up the advantages. (Rizos et al, 2010)

The concept of pseudollites is to provide centimetre level accuracy where satellites can not reach. This means they do this absolutely independent of any satellites, but fit seamlessly into the same grid as, for example, GPS.

3.1.3 Ultra Wide Band UWB

Ultra wide band (UWB) positioning system use electromagnetic waves' scattering off objects in their path to establish the distance of an object from the known velocity of the waves propagation and the time it takes the scattered signal to travel back to the origin of the wave.

The bandwidth of UWB is defined to be > 500 MHz thus in theory allowing multiple use of the same bandwidth.

The - in effect same as - radar principle of UWB is mostly applied in short range applications (maximally a few 10 metres): Ultra wide band technologies are limited by regulation on transmitted signals' spectral power density (ECC in Europe and FCC in America, with slightly different bands excluded) to avoid conflicts with other radio services, which restricts the application broadly to (low power) short range applications (< 10m) (Herrmann *et al*., 2010)
However, There are way round this, Robert *et al.*, (2010), for example, use low duty cycles (LDC) which are sufficient for indoors situations (around 1Hz) which reduces power levels considerably allowing them to place their bandwidth between 3.4 and 4.8 GHz, as the ECC restriction on LDC emissions at the applied power levels exclude this band.

One of the challenges for UWB positioning is the situation that the UWB emitted radio waves do scatter as well as penetrate: The frequency range is in fact in the (low powered) microwaves - range, so UWB fairly easily travels through walls, furniture and so on. For positioning though a returning scatter path is needed. This means a lot has to be known about the scattering characteristics of an object to be localised. The system needs to differentiate between scatter patterns evoked by diferrent type of objects.

Are the scattering characteristics known, UWB with sub centimetre accuracy can be used to monitor breathing by the varying scatter pattern of a human body, so uses for example in emergency services (search for earthquake victims for example) are a probably more likely than their use for iLAA.

Further limitations stem from the required data-streams at sub-decameter solutions, which can easily reach gigabits per second - order. Not necessarily a problem in laboratory conditions but limiting for small mobile devices.

The feasibility for localisation in a home entertainment applications, where the possibilities of tracking a listener by UWB was conducted by Zetik *et al* (2010) with the conclusion "that without a proper
background subtraction, data validation and tracking algorithm, the precision of the location estimation is usually very inaccurate."

There is a successfully applied commercial example for UWB for iLAA, Ubisense is used in an audio guide application for the Coronation Hall in Aachen, Germany. Visitors are handed a device which prompts the system to provide contextual audio content in relation to the position of the visitor within the room.

The company's own publication (Ubisense 2011) admits that calibration was difficult and called it a "very tough and special environment condition". This causes pause for thought, coming from a company providing positioning systems which successfully guide multimillion jets in and out of hangars.

UWB is being used with radio frequency identification (RFID) tags as well. UWB's ability to overcome multipath is an advantage here. The commercial system Dart UWB is an example used for industrial asset tracking. (Zebra, 2012)

3.1.4 WLAN

Wireless local area networks, electromagnetic communication networks ranging from peer to peer connections to outdoor "hot-spots" allowing wireless internet access in parks and public places, are probably as ubiquitous as GPS and hence utilised for data transfer in indoors positioning applications as well as as actual positioning systems.

WLAN adhere to the IEEE 802.11 (a, b, g and n) standards and WiFi is one of the commercial trade names using the standard.

The bandwidth for 802.11 WLAN is 20 MHz for each version (20 or 40MHz for 802.11n) at a frequency of 2.4 (3.7 GHz under FCC regulations allowing 5000 m range at higher power levels)

The indoor range is, depending on protocol, between 30 - 70 m indoors and 100 - 250 m outdoors.

There are various systems using WLAN wireless networks. Roughly they can be grouped into three categories: They either use the radio signal strength indication (RSSI) to work out the distance of a device by measuring the amplitude of the signal or time of flight (TOF) measuring the time a signal takes to travel from a sender to a receiver, and thirdly, time difference (or delay) of arrival TDOA, a variation of TOF: The correlation of signals received at multiple receivers are compared to each other to calculate the lag (delay) between them to solve for distance.i
TOF and TDOA measurements rely on (known) synchronisation of the signals. So does a further variation of TOF, round trip time of flight (RTOF) which needs slightly less demanding timing than TOF, as it uses a form of radar-principle. (Liu et al, 2007)

Beside those categories, radio frequency identity (RFID) and fingerprinting, two approaches which will be looked at separately, often use WLAN as a distribution network. They are not necessarily using any of the characteristics of the WLAN signal itself though as a means of positioning.

WPAN, or wireless personal area networks a term coined for networks smaller than WLAN (providing connections for wireless computer mouse, mobile phone head-sets, etc.), are potentially of interest for gestural control, the principles, limitations and advantages are the same as for WLAN if based on the 802.11 standard.

WLAN systems using RSS only, vary in accuracy between 3 - 30 meters. The RSS systems can be vastly improved up to about 2 - 3 m accuracy with algorithms compensating for multipath by modelling know errors. (Liu et al, 2007)

Despite the immense improvements achieved through error estimation and other optimisations most RSS based systems using WLAN (and similar technologies) can not improve accuracy to anything better than 2.5 - 3m (Gansemer et al (2010), Larranaga et al (2010), Barsocchi et al (2010).

As theoretical values in simulations can achieve half - meter accuracy, maybe RSS systems can improve with further development. (Bosisio, 2011)

However, the problematic of the use of WLAN, even if used for fingerprinting lies in the nature of the RSS signal. Described as "almost random" by some, (Ogunjemilua et al, 2009), due to its dependence on environmental factors. (Multipath, variations due to characteristics of materials, as some are more permeable than others, and so on.)

By the introduction of multiple in and multiple out (MIMO) antennae, the situation is further complicated, as the multiple frequency use this brings with it makes the RSS even less predictable and propagation predictions made for the older protocols can not be transferred to the newer ones. (Ogunjemilua et al, 2009)

The intrinsic problem for position technology using the 802.11 standard is that the standard is simply not designed to give a stable signal at a particular strength to indicate distance by its attenuation, but to provide a signal which is as strong as possible at any point of time in order to allow communication, an aim in many ways juxtaposed to the aims of positioning.

So optimally, these systems either assure line of sight (LOS) between sender and receivers or albeit means of dealing with multipath problems (filtering, modelling, and so on) due to the nature of radio signals.

3.1.5 WLAN Fingerprinting

Fingerprinting has two phases. In the first (offline or calibration) phase, measurements are taken of a device's signal strength at a receiver of known location. The such acquired data is compared to the devices actual (absolute) position and entered into a database noting the devices RSS and its relating position. In the second phase, the device's signal strength is compared to the data base entry for that device and the absolute position is interpolated from the data held for this device. (Hansen et al, 2010)

The upside is that the accuracy is as good as the data base, the downside is, the RSS is not stable and adversely influenced by shadowing and multipath and a manual calibration for every type of device is necessary. However, due to the ubiquitousness of devices that could be tracked in this way on the near ubiquitous 802.11 network, this method is being used for a huge number of location based services (LBS) and applications.

Research into fingerprinting concentrates thus on finding ways of improving the RSSI readings, be it through automated updates through algorithmic strategies (Hansen et al 2010; Segou et al, 2010) or shortening RSS paths and filling gaps in the database with ad-hoc peer to peer connections of already calibrated devices to non-calibrated devices as described by Della Rosa *et al*., 2010, for example.

Further, the design and algorithmic improvement of the radio map can increase accuracy and usefulness of these systems. (Gallagher et al, 2010, Leppäkoski et al., 2010)

But despite the evident workability of fingerprinting location methods using RSS, the issue remains that the radio signal's propagation in relation to signal strength is poorly understood. (Kaemarungsi and Krishnamurthy, 2004)

3.1.6 WSN

Wireless sensor networks, by definition a collection of networked transducer - nodes of any kind, are used for positioning by monitoring an area via locative nodes, motion sensors, for example.

If there are enough nodes, the proximity of an object to that node can be expressed as the object's location.

WSN rely on communication via a wireless RF network but they don't necessarily use the signal of the carrier network for positioning. As they thus don't rely on the nature of the radio signal per se, their positioning accuracy relies on the technology specific to the nodes function, which often is working in the sense of relative (or symbolic) positioning: The position of an object is defined by its proximity to a sensor node which has been triggered by the presence of the object. Relative positioning is inherently stable. (Kaseva et al, 2010)

This principle - which is the underlying logic of RFID - can in this sense be understood as a 1 bit RSS system: the signal strength is either 1 or 0. As such it is evident that these nodes need very little processing power other than being able to broadcast their state and ID to some processor or monitor.

The more nodes there are, the better becomes the RSS reading, as the RSS reading improves with closeness to the node. (Wei et al, 2010)

By attaching a node onto a moving object, WSN can be useful for many complex localisation tasks as long as other, responding nodes are densely deployed.

In fact, many WSN use a combination of these approaches: RSS is used to find the distance to the nearest access point (AP) node and the AP node position is known. Additionally fingerprinting (see above) is integrated too, if an offline calibration is possible. (Kaseva, et al)

WSN nodes are typically designed as low power devices as the idea of high density deployment makes manual battery changes difficult. Therefore the development aims for power efficiency and low cost and small size.

A further strategy applied to economise power consumption is daisy chaining the nodes, reducing the required reach of the RF signal to the next node rather than the whole system.

3.1.7 Bluetooth

Bluetooth, a very ubiquitous standard protocol using electromagnetic radio waves at the same 2.4 GHz frequency as WLAN 802.11 has been suggested by many as an option for I/LPS, suggesting the use of RSS measurements for distance estimation. Again, the challenges due to the nature of the signal are the same as for all systems using RSS.
These are heightened though by the fact that from standard v3.0 onwards, the Bluetooth protocol uses the (built in) RSS indication (RSSI) value of a Bluetooth device as a means to know when to increase the signal power based on its reading. Not an ideal condition if, what really is required, a *stable* signal which stays so over time and attenuates evenly.

The new v4.0 protocol introduces further means of dynamic power management. The ad - hoc nature of Bluetooth connections basically makes them very unwieldy for positioning tasks based on RSS. (Bluetooth, 2012)

3.1.8 RFID

Radio frequency identification, RFID, uses principles discussed in connection with WSN, in fact, WSN are often using RFID tags, as are many UWB based technologies. Passive RFID tags are very small and ubiquitous: They fit into book covers, clothes, pets and are very cheap.

RFID, in their most simple form, are 1-bit signal strength indicator: The signal is either present (state 1) or not (state 0) By giving a mobile tag a particular identity code, and making the nodal network dense enough RFID systems can be used for applications way below meter level. Again, with probabilistic algorithms and filters, the performance can be vastly improved. (Uchitomi et al., 2010)

Many RFID systems use RSS as well to improve on the relative or symbolic nature of a position.

3.2 Infrared

Besides expensive motion tracking systems for film productions (IR LEDS worn under clothes can still be recorded by IR cameras) there is a fairly ubiquitous IR system available at comparatively low cost:

The Nintendo wii console game controller wii remote is a wireless handheld device with an accelerometer and an infrared camera. In its original form, (the wii is being used for many non-nintendo applications after reinstalling the firmware with Wiibrew, for example (Wiibrew, 2012)) the wii remotes infrared camera calculates its distance to the "sensor bar" (confusingly named so, as it does not contain *sensors* but IR LEDs), by comparing the distance between the two groups of LEDs at the ends of the bars with the distance seen on the camera. (Wii Remote) The handheld controller connects to the game via Bluetooth connection and up to 10 players can connect simultaneously. (wii Remote, 2012)

The camera has a 1024 x 768 resolution and a IR LED can be tracked as one pixel with (x / y) coordinates. (Tas et al 2010)

This allows tracking of movements in the azimuth in realtime, for forward backwards moves the accelerometer reacts faster. The wii remote can be calibrated to other IR sources than the one provided on the sensor bar. (Yildiz et al, 2011) This allows for positioning in whichever plane the LEDs are set.

As the wii remote is fairly cheap, whole systems are being developed around it beyond gaiming applications: Tas et al, (2009) for example, install the wii remote on a robot which orients itself this way in relation to ceiling mounted IR LEDs. There is practically no error for these systems as long as line of sight is guaranteed. to at least 2 LEDs.

In fact Microsofts Kinect motion sensing input device uses IR too, it contains an IR projector and an IR camera. The IR projector together with a monochrome complimentary metal-oxide semiconductor enables depth discrimination.

IR can be used for LEDs worn under clothes, but shielding a LED from a camera with hand or other body parts results in a non-LOS situation. This is why both wii and Kinect don't rely on IR as the only means of positioning.

3.3 Inertial Systems

Inertia systems, in their pure form, use dead reckoning methods for navigation, whereby no absolute information is being added, only speed, heading direction, and estimated errors (like leeway) are calculated: They provide estimated positions (EP) not fixes.

This is why inertial positioning methods are often used when an absolute position is (temporarily) not available.

The inclusion of accelerometers, gyroscopes and magnetic compasses in smart phones, in combination with the built-in GPS make them powerful navigational tools and integrated systems can be deviced, together with fingerprinted WLAN positioning which makes ubiquitous positioning a reality, at least in well equipped, mostly urban, areas. (Lukianto et al., 2010)

Applications using inertia only systems, like step counters used for fitness purposes, for example, often only need relative information, be it because the speed information is enough as the absolute orientation is provided by a map, or like in some gaming applications where the absolute position is not actually required, just the heading, as the "absolute" positioning is virtual.

Conceptually, inertial navigation systems, (INS) are the solution to practically all short comings of other technologies. However, INS are unfortunately notoriously inaccurate as error in INS, intrinsically accumulate. This problem is not helped by the erratic ways pedestrians move. (One way to ascertain the inherent inaccuracy of human propulsionexperimantally is by trying to walk in a straight line over a snow field or a large beach and compare the actually walked track visible from the footprints with the intended route.)

Foot mounted inertial measurement units (IMU) take advantage of the zero velocity moment when the foot is at a stance, to overcome the problem that an accelerometer measures just the acceleration, (and deceleration) but not the (constant) speed. Non- foot mounted system like the one described by Goyal et al.(2010), consisting of a device which is worn like a belt, consisting of a tri-axial gyroscope, a tri-axial magnetometer and a tri- axial accelerometer, need "more sophisticated algorithms".

The difficulty in developing algorithms for these devices is, that, if one element is erroneous (say the magnetometer) but the gyroscope is closer to the actual path, the information of the magnetometer on its own is not only redundant but plainly misleading. However, with the application of an extended Kalman filter (EKF), which establishes averaged positions in the process of the trajectory based on probabilities, the corrected results are a lot more promising.
(More on Kalman filters in the section on optimisation below)

Indication on accuracy for inertia only INS have to be expressed as error over distance, as they are initially spot on. Accuracy in INS is the factor modulating the proportionality of error to distance travelled. INS estimation error is thus usually expressed as a percentage of deviation from the truth track, called *relative distance error* (RDE). In the example of Goyal et al, for one experiment the magnetometer had a 11.53% relative distance error, the gyroscope on its own 4.51 % and the combined track using EKF 4%. Due to the constant updates happening in the EKF, the average error at any point of the trajectory did not exceed those 4%, resulting in a distinctly improved comparison to the truth path. (*Fig 15)*

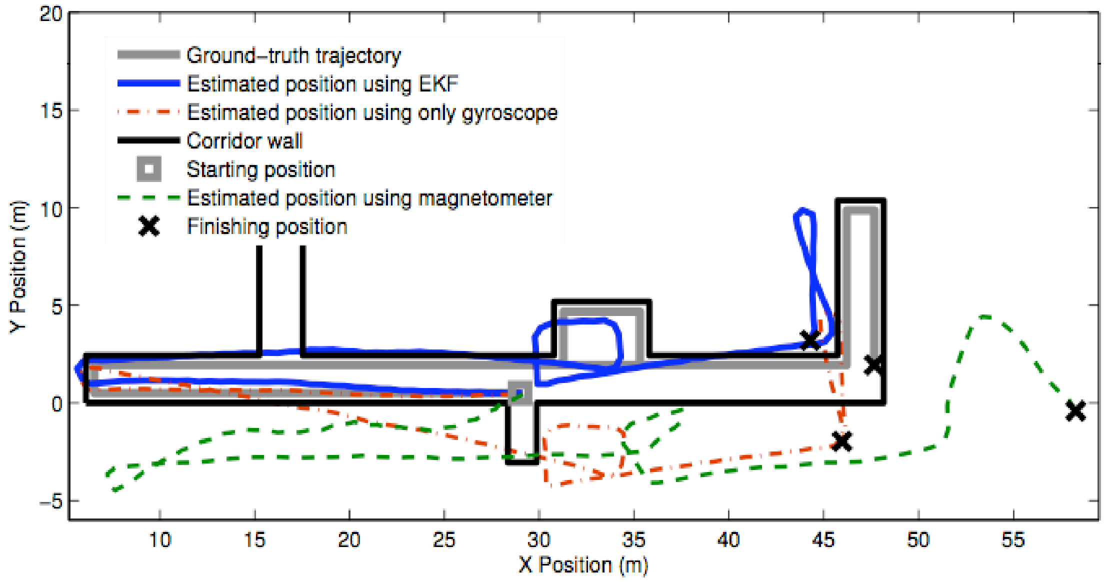


Fig 15 (Goyal et al 2010)

For game controllers where the control is gestural, IMU are extremely useful, as they can account for quick and jerky movements. Updating to an absolute position is no problem as devices like the Sixaxis game controller (Playstation), or the wii remote use IR sensors too, with centimetre accuracy.(Purkayastha, 2010)

An interesting addition to the instrumentarium of IMU is a barometer to add height information to the measurement, as suggested - and implemented by Kröger *et al* (2010)

3.4 Optical Systems

Motion tracking, as used in he film industry, records devices attached to moving actors on (multiple) camera and the path of the device is then digitalised. This information can then be applied to any picture, like a monster or a cartoon character and so on. The technology is very precise, very time intensive and very expensive.

Optical positioning systems achieve sub millimetre accuracy, which extends its uses to applications which require this type of precision. (Mautz and Tilch, 2011)
High precision laser based positioning like iGPS (owned by Nikon) are being used in precision industry and for robot technology. The iGPS system maximal path tracking errors are up to 0.3 mm

at speeds of up to 3 m/s (Depenthal , 2010) Even higher accuracies are reported, for example, for the DAEDALUS system, in the high end bracket of optical systems, with tracking accuracy up to 0.04 mm. (Mautz and Tilch, 2011)

As far as ubiquitousness is concerned it is again, the gaming industries consoles' tracking technology which is probably most interesting for iLAA: As long as LOS is not an issue, these optical systems are very accurate, particularly in audio visual applications where a 2 dimensional screen is involved there is probably no system which can outperform the accuracy of (optical) motion tracking.

The "motion sensing input device" Kinect uses - as mentioned in the Infrared section, an IR projector and camera, a monochrome CMOS (complimentary metal-oxide semiconductor) and a VGA Color camera which records red green and blue and is hence called an RGB camera.Positioning systems using Kinect are very precise and do not need any mobile devices like tags, or handheld gadgets, it tracks gestural movements. (Smisek et al, 2011)

Besides the cameras it contains a microphone array of 4 microphones which makes sound localisation possible, based on principles further discussed in the section on acoustic tracking.

Again, the system is near ideal for applications where LOS is not an issue, i.e. the majority of audio visual iLAA based on 2 dimensional screens.

The Playstation Move, uses RGB technology similar to Kinect. A handheld device with a ball shaped top can be moved in front of the Playstation Eye camera. The ball shaped top, containing the RGB LEDs choses which colour to send based on best contrast to the background.

There are not as many Move "hacks" to be found as for Kinect or wii, but the product is still fairly new, in comparision to its competitors. (launched March 2010)

Overall, the accuracy of optical systems lies between approximately 2 μm and decametre, in volumes ranging from 4m^2 to "arbitrary scalable systems". And the costs are expected to come down for components, opening the technology further to the mass market. (Mautz and Tilch, 2011) The limitations of these systems are within their optical nature: Occlusions will have to be overcome by non-optical means.

3.5 Magnetic

Magnetic field systems, as discussed in Blankenbach & Norrdine (2010) have the following conceptual advantages: Magnetic fields don't suffer mutli-path, they don't need active sensors (no sender -receiver principle necessary) and, surprisingly, they don't interfere with radio frequency signals as long as the field's signal is emitted at very low frequencies. (Hz range)
As a passive system, it allows as many mobile users as can be physically present in the observed volume. (Blankenbach and Norrdine. 2010)

The Blankenbach and Norrdine (2010) system uses a layout not dissimilar to RF systems, with a mobile magnetic field sensor and magnetic coil reference stations which emit the magnetic field at low frequencies. This system achieves - with "hand made coils" precision within half-metre accuracy.

The magnetic field can be expressed in vectors which has the additional advantage of supplying orientation as well as distance, which, theoretically reduces the amount of needed reference stations.

There are two commercial providers for magnetic field sensing tracking systems discussed by Blankenbach and Norrdine (2010), Polhemus and Ascension, which claim accuracy in the sub cm range and and 0.5 ° for orientation. Low cost development of such systems are feasible for many tasks.
However, the technology is not ubiquitous and would need installation and calibration. But, the underlying technology is simple and affordable.
Other magnetic systems on similar principles are proposed by Sherman et al (2007), Paperno et al (2001) and others.

These systems are not exactly new, Raab et al. proposed one in 1979 (Raab et al., 1979) this system was based on alternating currents field, others use pulsed fields. which resulted in the need for error correction in the presence of metallic objects. (Blankenbach and Norrdine 2010)

All these systems have to be calibrated for the existing magnetic fields (the earth's, for example) and electromagnetic inference. As long as these fields are stable and localised, calibration can account for them.

As far as iLAA are concerned, the presence of potentially strong magnetic fields of speaker systems might be an issue, on the other hand more research could possibly establish if the fields of magnetic coils in loudspeakers could actually provide the necessary field.

Electromagnetic induction would be a conceptually possible magnetic application too which could be used for tracking or positioning, and the Potentiomètre d'Espace discussed in chapter 1 was built on these principles. However, no recent literature on any such system could be found.

3.6 Sonic Positioning Systems

3.6.1 Ultrasonic

The borders between sonic and ultrasonic are difficult to draw, and, around the upper threshold of hearing, are possibly arbitrary to the developers choice. Filonenko et al. (2010) for example, describe their system as ultrasonic, using smart-phones as sources for "ultrasonic" sound, in the frequency band of 20KHz -22KHz, whereas Packi et al. (2010) describe a system for "telepresence" which they describe as acoustic, rather than ultrasonic, as the signals are just above the hearing threshold, i.e. between 20KHz and 50KHz. (Filonenko et al, 2010)

What is undisputed is that ultrasound is of the same physical property as sound, ultrasound is just too high in frequency for the human ear to hear it. Bats, dogs, cats and other mammals can.

In localisation systems. TOF, TDOA and angle of arrival (AOA) of ultrasound signals are being used. Ultrasound has to deal with multipath errors in indoor environments but the acoustic signal is well known and its propagation predictable.

It is hence not surprising that there are successful commercial systems using ultrasound, namely the Bat and Cricket, the latter using radio signals as well as ultrasound, but both using a grid of ceiling mounted sensors (microphones). Both systems can track mobile devices with accuracy below 3 cm.

Filonenko et al, (2010) successfully showed that the necessary ultra sound tones (in the 20 KHz to 22KHz band) can be emitted by standard mobile phones, by analysing the spectral graphs of recordings of such produced tones.

Another upside of ultrasound technology lies in the nature of possible errors: Most acoustic noise errors are normally distributed, which makes them good candidate for error corrections using stochastic corrections based on Gaussian noise distribution like Kalman filters, etc.

This all sounds promising. The downside of these systems, including, in principle, the Bat and the

Cricket system, is that microphones don't necessarily record ultrasound *only*, raising concerns for privacy issues.

Moreover, it has to be considered that at ultrasonic frequencies the attenuation of the signal happens fast over distance. (Filonenko et al 2010) this suggests the use of audible signals, which, coming back to iLAA might not be the worst idea if audio signals are present anyway.

3.6.2 Acoustic Tracking

Non-ultrasound acoustic tracking is not as common but very promising in many ways. Packi et al. (2010) demonstrate with a device developed for "telepresence", how, with a microphone array and speakers sending signals just above the hearing thresholds sub - meter results can be achieved.

It is feasible to expect similar systems could be developed using audible signals *within* the audible range. So far very little literature could be found employing acoustic signals in the audible range. However, doing just this are described in Filonenko et al (2010) and Janson et al (2010) which uses ambient noise for a very clever and intriguing localisation system using nothing but networked and time-synched mobile phones.
The authors' suggestion of moving their research onto non-synchronised devices, using just trilateration of acoustic transients of sound sources on spaced devices inspires further research as outlined in Chapter 5. (Acoustic Only System - A Suggestion)

The Kinect input device has a built in microphone array featuring four microphone capsules and operates with each channel processing 16-bit audio at a sampling rate of 16 kHz. The makers claim to achieve "sound localisation" that way, but as the software is proprietary no further information how this is used in standard kinect applications could be found. (Kinect, 2012)

For audio applications, intrinsically when the aim of the application is to locate an actual sound-source, the audible signal itself could be observed from various microphones and transients' DTOA on geographically spaced microphones compared and the sound-source thus be located by trilateration. Same as Janson et al (2010) suggest in their section of further research, no time synchronisation would be needed if the correlation of the signals would be used to work out the distance between the sound source and the microphones. This is the principle idea behind the suggested system in chapter 5.

The possible resolution, even at standard sample times at 44.1 kHz provides at 8 mm per samples, certainly an encouraging amount of precision, remains to be seen how multiple path error and timing issues can be dealt with.

It shall be mentioned here again that, same as for ultrasonic systems, the noise distribution is Gaussian for audio signals, rather than for electromagnetic waves which do not adhere entirely to normal distribution. (Ogunjemilua, 2009).

Beep, a system developed by and described in Mandal et al (2005) uses audible signals from a mobile phone and wall mounteed sensors (microphones). it measures TOA of the signal at the known position of the sensors. The authors plan is to develop a mass market compatible hardware solution for the wall mounted sensors in the near future.

The afore mentioned system developed by Janson et al. (2010) which aims for a smart phone application which uses only environmental sounds, currently needs to synchronise clocks between the participating devices. In principle, the sound triggers a time stamp, and the time-stamp is then compared to the clock shared by all participating devices. The positions could thus be calculated by trilateration.

3.7 Optimisations

A series of optimisations are used in for positioning systems, the most typical one being the Kalman filter and the extended Kalman filter (EKF) which are based on a recursive Baesian estimation.

The Kalman filter is a real time filter which recursively compares the state of a stream with the state of an imminent predecessor. Assuming noise to be present it makes predictions based on a physical model (normal noise distribution, for example) and then compares this prediction with the measurement of the next step.

In the extended version (EKF) the underlying state transition need not to be linear as in the simple version, but differentiable functions, in fact non-linear stochastic difference functions. (Welch & Bishop 2001)

Other filters exist, but the Kalman filter is practically omnipresent in localisation technology, as it is evidently effective. This was illustrated also in Fig (Goyal et al.2010) in the section on INS.

The code for a simple Kalman filter in is appended as APPENDIX., A Kalman filter was considered for implementation in chapter 5 for the suggested audio only positioning system discussed there.

3.8 Discussion and Conclusive Remarks on this Chapter

Overview of Positioning Systems

principle	*system*	*accuracy in m*	coverage	cost to user	availability	Ubiquitous	initial cost
RF	GNNS	5 - 10	global	low	market	yes	very high
	Pseudolites	0.05	Local	low	planned	no	high
	Ultra Wide Band	0.012 - 0.1	indoors	high	market	no	medium
	WLAN & WLAN FP	2 - 5	local	very low	market	yes	low
	Wireless Sensor Net	0.5 - 2	scaleable	low	market/DIY	no	low
	Bluetooth	5-10 m	20 m	very low	DIY	yes	low
Infrared	wii	0.005	scaleable	low	market/DIY	no	low
INS	inertial Meas. Units	0.5%-20%	1 - 100 m	low	market/DIY	yes	low
Optical	motion track.hi- end	0.0005	scaleable	high	market	no	high
	moti'n track. low-end	0.05	scaleable	medium	market	no	low
Magnetic	magnetic field	0.004	1-20 m	medium	market	no	medium
	induction	?	?	?	no	no	?
Sonic	Ultrasonic	0.001	scaleable	medium	market	no	medium
	Acoustic tracking	0.001	scaleable	low	DIY	yes	low

Fig 16

The available positioning systems as discussed in this chapter show varied characteristics of which some self-evidently present themselves for iLAA, be it by the potentially shared infrastructure of the iLAA and the positioning system as could be the case for video games or acoustic tracking, for example.

From this we can tell that the nature of the ILAA finally, is decisive which system can be used. This aspect will be discussed in chapter 4, based on the findings of a survey.
At this point, however, we can summarise principles of positioning which might have a bearing on their usefulness for iLAA.

An overarching conceptual difference between the discussed systems is where the information of the position lies, with the sender or the receiver. The difference can be easily shown by the difference between the wii remote and the Kinect system: The wii remote has the camera (receiver) which works out its position in relation to IR LEDs (sender), the position information is (initially) with the remote.

In the Kinect system, the camera (receiver) is stationary in relation to the moving object which is being traced, so the information is initially with the stationary part of the system. These differences, in connection to the technological aspects - other aspects of these principle will be discussed in connection with privacy issues and requirements for specific iLAA - have the following implications: A system like Kinect needs to identify the object to be tracked by means of information about the object. The system needs to "tag" the object, so it can be identified. The wii remote in contrast, is the moving part of the system and it works out its own position as the receiver lies within.

Occasionally these systems like wii Remote are called passive, as no information about the object to be tracked needs to be obtained. This might be confusing in this context as it has been used in this chapter, in line with the discussed literature, to define if a tag, for example, emitted a RF signal *actively* as opposed to magnetic field localisation methods where the tracked object does not need to propagate a signal and hence was described as *passive*. From this point onwards it seems more appropriate to define systems based on a sender - receiver model rather than terming them active or passive when referring to where the position information lies: The sensor shall be defined as the point where the position information data is *initially* gathered, (The camera, the microphone, the IMU, the RF receiver); the sender as the part which emits a signal which the sensor uses for location.

N.B. the use of *initially* in connection with data gathering at the receiver: The information is often passed on immediately to other parts of the system, the *processing* of the data does not necessarily happen at the place of the receiver! Nor does the emitted signal need to be emitted by the object. All the system needs is some data containing information about the position of an object to compare a second set of position data in relation to it.

It follows that all systems process some data gathered on some sensor to work out where an object is in relation to another set of data. Both sets of data have to be available to the system. Where within the system these sets are, is basically the difference of one system to another.

Moreover, it follows that, if the receiver is not identical to or in the same place as the object to be tracked, the object needs a tag, a means of identification for the receiver. This tag is then a sender, by sending a RF signal, as in RFID for example, or by sending information data in form of a pixel position of a LED, as in IR positioning. A mobile phone, for example in this model is a transceiver: With the GPS it it uses satellites as senders to obtain a position (The information lies with the phone) but through its connection over the mobile network, its position can be obtained from its signal's strength at the nearest mobile network mast (the information lies initially with the mast, i.e. the mobile network provider).

(An example of a sender receiver in the same "object" is a person with a map, for example, comparing the set of data of the landscape with the second set of data on the map.)

The one system where this crucially does not apply is for inertial systems. By definition the inertial sensor is *always* within the moving part of the system.

From this discussion can be seen that consequentially, a system wherein the object to be tracked is not identical with the receiver, tagging is necessary. The nature of this tag can be RF, IR, magnetic, optical or sonic. Which tag to chose optimally depends on the nature of the application.

To summarise, The various existing systems cover a vast range of possibilities in accuracies as low as sub-millimetre for some systems, but of variable scale, reach and real-time compatibility. Moreover, the systems differ in their demands on infrastructure, which make some system more feasible than others. Once a large scale system is installed, there is no reason why it couldn't be used for applications other than its intended original use, as was the case for GPS which was originally intended as a military system only.

The use of existing systems for other means can be fraught with systemic difficulties as encountered when using existing communication networks like Bluetooth and WLAN for positioning, as they are designed to provide strong signals for communication, not a stable metric for positioning.

Systems developed for video games like wii Remote, Kinect and others provide a low cost means of new uses for positioning. As the intended use of these systems lies thematically close to iLAA already, it seems evident that these system play an important role. The reliance on LOS for these systems make them less likely candidates to be a one-fit-all solution.

For many systems which provide the required accuracy as discussed in chapter 2, like UWB and ultrasonic, sonic, magnetic, optical - given LOS and other systems it remains to be seen what the user requirement survey's findings favour, but it looks likely that factors similar to the ones discussed for positioning in general like accuracy, scalability and, of course, cost as well as ubiquity will play a role in what system is being chosen for an application. Optimisation and in particular algorithmic optimisation will play a central role in I/LPS for iLAA, like in all positioning systems.

The overview of these general factors are shown in the comparison table (Fig 16)

4. User Requirements

From the previous chapter it is fairly clear what positioning systems can and can not do. In chapter 2 we looked at the ability of the human ear to locate sound but the question as to the minimum requirements for particular applications remains however, as they are specific. This chapter will give an overview of possible iLAA and discusses the findings of an online survey wherein developers and musicians in the field have been asked a in a questionnaire what their expectations on positioning systems are when developing iLAA.

As iLAA are, generally speaking, very much a concept rather than a day to day reality, actual examples are not as many as there are concepts and ideas, many of which have not been realised as no I/LPS could be found to develop them. This was one of the outcomes of a focus group forum held at Pervasive Media Studio (PMS) Bristol, a preliminary to the online survey whose findings are discussed in this chapter.

Further, it was noted in the forum, that due to the secondary nature of hearing to seeing (see also chapter 2) often the visual clue within an application is so clear that the spatial position of the sound becomes less relevant. This holds true for audiovisual iLAA only.

The most evident outcome of the forum was, however that the vast variations and differences between various iLAA needs to be categorised in order to be able to generalise requirement parameters.

4.1 iLAA Taxonomy

In order to obtain user requirements for I/LPS for iLAA, it has to be defined what iLAA are and how they differ from non-interactive locative audio applications and non locative audio applications likewise.

At the focus group meeting at PMS Bristol, the diversity of iLAA was noted. In fact, the concern was voiced that these variations make it near impossible to have a general set of parameters which can apply to all of them.

A taxonomy of the various applications was suggested in the aftermath of the meeting to see if it was possible to group applications in order to supply generalised requirement parameters which can be applied to more than just one possible application.

The thus devised taxonomy tries to follow principles of taxonomic ranking and understands itself as a subsystem of the Association for Computing Machinery (ACM) Computing Classification System (CCS) (ACM, 2002) somewhere in J.9, possibly under J.9.a (Location-dependent and sensitive) but with relevance to J.9.e and J.9.d. (see Fig 17)
(This explains why in this context *audio applications* is a subsection of *locative applications*.)

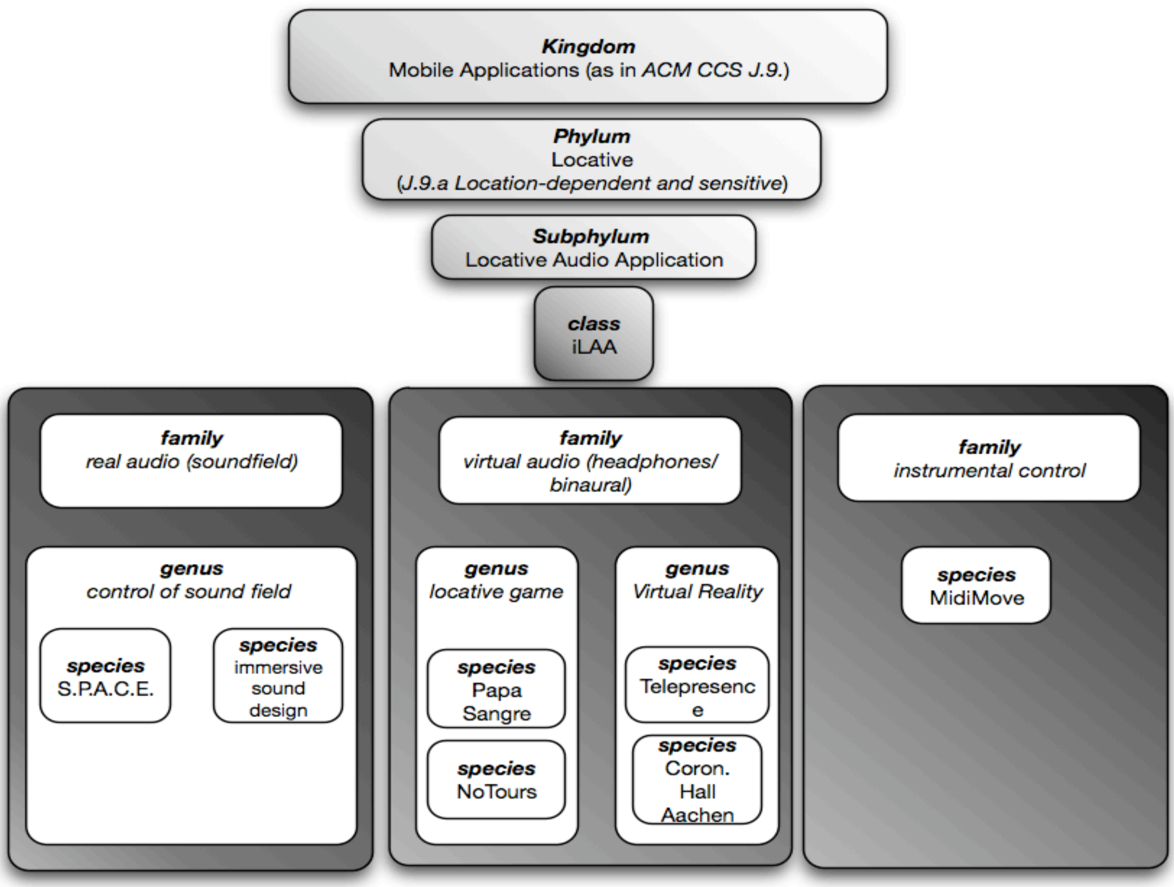


Fig. 17

The applications mentioned here are used for their exemplary nature:

The MidiMove, an application wherein movement of the head can control audio content, (TangibleFX, 2012)despite not actually *affecting* sound spatially, is still very much an iLAA, but of the Control genus. Papa Sangre, on the other hand, is of the binaural family as it uses HRTFs as means of spatially distributing audio.

Binaural systems often use headphones, so the best positioning system would be one which gives information about the directional position of the head. As this applications is of *virtual* family, the absolute position would not matter so much, for example.

There are applications of course which could be understood as virtual or as real, like an interactive museum audio guide which uses the real room to evoke a "virtual world" of times gone by.

Virtual is understood here in relation to the space in which the audio is happening: If the evoked space is congruent with the real space, i.e. the "virtual reality" is happening in the soundfield, the application is considered real audio. Headphones are an obvious means to create a virtual audio space in this sense, but not the only one: stereo imaging could be described as virtual, for example, according to this definition.

The third family listed here, instrumental control, includes all gestural controls which don't translate directly into proportional spatial control of either virtual or real audio worlds. So the Potentiomètre d'Espace for example, *is a real audio control*, as the gesture translates to a change of the spatiality of the sound field, whereas the MidiMove controls the output of an instrument. Typically, the gestural movement controls a process other than spatial sound processing.

S.P.A.C.E., the example - species used for the genus "control of sound field" is an i/LAA developed by the author for his final year project at UWE for a BSc in Creative Music Technology in 2010, presented at the faculties final year project presentations of that year. The abbreviation stands for Spatial Position Audio Control Environment. The idea behind it was that a mobile device moving around in a room controls the audio directly by its position. The positioning system developed for it used RSSI of Bluetooth devices to estimate distances in relation to receivers of known position. (The encountered limitations of the technology inspired this research.) The simulation of a possible audio-only system discussed in chapter 5 can be seen as a further development of the same idea.

Besides categorisation by nature of the application, iLAA can further - or otherwise - be classified in respect of the 3 aspects of their dimensions:

1. reach
2. density
3. latency (speed)

These 3 aspects give direct information as to the suitability of a positioning system for a particular iLAA:

The reach of an iLAA defines the physical size of the whole system, (local, global, indoors, urban etc.) which would have to be covered by a positioning system and in this context here would also express the demands on the system as to how many degrees of freedom need to be tracked and in which dimensions. (Two-or three dimensionally.)

The necessary density, or density resolution, describes how many tracked instances per area are necessary.

The third aspect describes how quickly a system needs to be able to react depending on how fast the tracked object/subject moves.

So, taking into consideration the psycho-physical limits of hearing, as discussed in chapter 2 and the place of an application in the suggested classification system under consideration of its dimensions, what remains to be taken account of is what actual real world users and developers think a positioning system needs to be able to do to be of use for iLAA.

4.2. Survey

4.2.1 Methodology

The surrvey conducted for this research consisted of a preliminary forum with professionals working in the field at PMS Bristol in which questions were formulated for an online questionnaire which went online in the beginning of 2012 and which is still open for further respondents at the time of writing.

An approach using an enquiry cycle model with three stages of 1. requirement documentation, 2. requirement discussion and 3. requirement evolution has been used and iterated twice by the initial documentation based on the authors view which was then challenged in discussion with the focus group, the such evolved documentation was used as the base of the questionnaire which in its turn was challenged by the respondents leading to further evolution of the documentation. This approach was formulated for user requirement analysis by Potts et al (1994) for software development.

The focus group, consisting of early adapters and developers of iLAA was considered to be relevant based on the evaluation of the literature, from where people in this group were identified as primary stakeholders (See 4.2.2)

The questionnaire was designed such that in the case of poor respondent numbers some qualitative answers could be obtained even if the sample size did not reach representative proportions.
In particular the findings of the focused interview, which carried out a situational analysis based on the views and experiences of the group of professionals, provided the in-depth qualitative insight necessary to evaluate the answers to the questionnaire. This was done in the understanding that "real world research" needs often to combine approaches in order to obtain valid results. (Robson, 1993)

As an online survey - in theory open to anyone, the questionnaire was exposed to the usual pitfalls of this type of survey, namely low response rate, misunderstanding of and ambiguities in the questions and sample errors due to answers by non-relevant respondents. (Robson 1993)

These problems were addressed as follows: Presenting the survey on PMS Bristol's website, on the authors website dedicated to creative music technology and by emailing university students on relevant courses along with inviting known professionals and academics in the field to respond personally by email and in conversation, it was felt that the specific population addressed consisted entirely of stakeholders. Moreover, by the fairly technical nature of the questions it was considered unlikely that respondents without a stake in the subject would actually summon the effort to respond in full.

The other complication arising with online questionnaire is that random sampling is impractical as the respondents can not be chosen. (For *representative* surveys random sampling is an intrinsic requirement of the method, see Robson (1993) for details) So the respondents represent at best the part of the population (as outlined in the stakeholder analysis) which happened to come across the survey and had time and motivation to fill it out. This is clearly not representative of the whole population we're interested in. However, this part of the population is not the *wrong* sample either, as the choice of filling out the survey per se shows an interest in the subject, thus qualifying the respondent as a typical stakeholder.

The population size is based on an educated guess. Basing it on the video-game industry as the best known and probably largest industry with a stake in iLAA, we can roughly interpolate world wide how big the population of developers for interactive audio could be.

In the UK, 7000 people work in the Gaming Industry and 950 people (13%) in technical development, and 2% (140) in sound/music/audio (Creative Skill Set 2012)

Let's assume that worldwide, the section of society interested in I/LPS for iLAA exceeds the size represented by the 2% of audio developers for games, but is not smaller: The 2% shall represent the smallest probable population.

In the 13% of technical developers there are bound to be people with an interest in i/LAA in addition to the 2% of audio-people, as developers of consoles, controls, and locative media have a stake in i/LAA too, even if this is not their main field.

There are, of course, stakeholders in other industries too, e.g. mobile phone developers. . Let's therefore presume 1/2 of the 13% (50%, the statistic average of "all of them" and "none of them") are part of the population. If this still seems a high number in comparison with the 2% of the audio workers, let's remember that other non-game industry stakeholders for which we don't have a number need to be accounted for somewhere too. So, the population probably lies somewhere in the area of a percentage represented by between 2% and 6.5% of the number of people working in game development worldwide.

In the US, the market leader, 32000 people work in the gaming industry in total. (Siwek, 2010) How many of those are stakeholders in iLAA is guessed here by adopting the percentages used for the UK industry as Siwek (2010) does not provide any employment numbers other than for the whole sector i.e. 2% and 6.5%, so, between 640 and 2080 people, respectively.

Interpolating from the market sizes as illustrated by De Prato et al (2010) wherein EMEA (Europe, Middle East and Africa) cover roughly one third, the Americas another and Asia Pacific the last third we are looking at about a global population of people with a stake in iLAA of between 1920 and 6240.

Adopting these numbers for our survey, this results in a required sample size at a confidence interval of 10% and a confidence level of 90% of 66 people in the case of the smaller population, or 67 for the 6240 people in order for the survey to be representative. To establish trends this is considered to be adequate, as the sample is of a fairly homogeneous population and no cross section of society as a whole is needed. Further, in the vein of a combination of approaches as suggested in Robson (1993), the findings of the survey are being used to test assumptions and working hypotheses developed in previous chapters. For these tests the findings don't need to be statistically representative, but they help to confirm already existing evidence.

The length of the questionnaire was kept within the usual length for online surveys, in keeping with the online-survey providers recommendations (Surveymonkey, 2012) and shouldn't have taken more than maximally 20 minutes for most respondents and averagely around 12 minutes.

For all questions multiple answers were allowed as the questions were often of the nature that developers who are working on more than one iLAA would answer differently for various projects they have in mind when filling out the questionnaire. Or many developers are indeed musicians too and answer some questions with one hat on and some with the other. Again, it was thought that in this situation multiple answers give a more illustrative view of what stakeholders need than if only single answers were possible thus leading to frustration with the process and drop outs before completion, beside showing a limited aspect of a particular stakeholder's activity.

There are of course instances where these multiple answers seemingly distort the sample size. In the evaluation of the findings it will be highlighted whenever multiple answers have led to a virtually increased respondent number.

To illustrate this, the multiple choice question "Which positioning systems have you used?" might have a number of answers exceeding the sample size. The absolute number of answers is still valid though, if the interpretation of the answer is "X % of respondents have used GPS for iLAA" and the percentage can refer to the percentage of all answers, rather than the sample size, as somebody might have used other systems too.

If the question has a rating though, for example "which of the following systems did you like best in a scale from dissatisfied, satisfied, to fully satisfied" if most respondents give GPS a satisfied but one respondent gives WLAN RSS a fully satisfied, but it is the only "fully satisfied" in all answers, it is sensible to refer to the one person as a percentage of the whole sample. This allows the interpretation of the answer, for example, "Only a small percentage of respondents were fully satisfied with the systems used" but to claim that WLAN RSS was the most satisfying system would be misleading.

As the questions were directed at developers it had to be considered that respondents might feel reluctant to forward ideas which they consider to be their intellectual property. Accordingly, the questions were formulated in an abstract way, asking about aspects only which influence the design of the positioning system not the iLAA itself. Reading through the question it becomes apparent that it is nigh impossible to reverse engineer any iLAA a respondent is referring to from the answers.

The open ended questions, kept to less than 20 % of all questions, were intended to provide Respondents with the opportunity to push the survey into another direction, adding parameters which might have been forgotten, voice disagreement about possible assumption and to disambiguate.

4.2.2 Definition of the "user" *(stakeholder analysis)*

Due to the emergent nature of iLAA the "users" of iLAA are almost entirely developers and early adapters. There is academic interest in these development from the sides of music technology, dance, pervasive and ubiquitous computing, video games and sound design. There are, however, not many end users who could help in the process of establishing user requirements as not enough iLAA have been launched so far. All information has therefore to come from developers and early adapters.

Stakeholders in iLAA have thus been identified in groups which show an interest in iLAA through academic literature. Here is a short list of relevant literature for a given field.

Spatial composition (Bates, E. et al. 2008)
Dance (Wijnans, 2009)
Immersive video games (Collins, 2007, Collins, 2008)
Interactive Music (Szinger, 1993)
Installation art (Ouzuonian, 2008)
Education (Brown et al., 2003)
Recording Technology (Churnside et al., 2011)
Assistive technology (Wersenyi, 2003)
Robotics (Trifa et al., 2007)
Ubiquitous and Pervasive Computing (Tuters, 2008)

To broaden the scope of considered stakeholders outside of literature in order to make the questionnaire as relevant as possible to as many stakeholders as possible the afore mentioned focus group has been formed to discuss matters casually with professionals active in this field.

One of the things to consider within the developers and early adapters - section of the population is that within the world of creative technology, "early adapters" by definition hear of a technology first and then think of creative uses for it as a consequence. This "hunter gatherer" approach makes requirement acquisition harder as what a technology *can do* is intuitively more interesting to work with than what it can *not* do in a creative process.
Nevertheless, people have expectation as to what should be possible *conceptually* with a technology, i.e. the question is, e.g. "what would you do if precise positioning was ubiquitous" And from the answers to this we can infer what the current technological limits are, as probably the expectation would be met, exceeded or failed by the existing technology.

4.2.3 Focus Group

This stage aimed to establish quantifiable parameters for localisation systems in order to compare user requirements with existing technologies in discussion with professionals working in the field. It was introduced to the group as an aim to inform thoughts on designing a dedicated system for iLAA. The forum was the first step to formulate relevant questions for the survey.

As the author is a stakeholder himself he presented himself to the group as an involved person and stated the aim for this forum to "objectify" his personal views by comparing them with other people's (subjective) ideas to broaden the analysis and discuss scenarios and identifying and formulating actual questions for a questionnaire designed for a yet broader group of people in an online survey. (In a combined approach as described in Robson, (1993))

With the aim to get the group to help in identifying quantifiable parameters for relevant stakeholders the author presented his own views as a starting point. An initial list of stakeholders and scenarios was then extended with the help of the group. The focus group confirmed the following stakeholders:

- Gaming
- Spatial Music
- Interactive Art
- Multimedia
- Education
- Assistive Technology
- Recording Technology
- Home Theatre
- Performance Art
- Manufacturing *(New)*

In respect as to the spatial reach of iLAA and the relevant scenarios, the following list was collated: iLAA can be

- Global
- Local
- Indoors
- Urban
- Rural

- Crowded
- Controlled
- In Motion *(New)*

As abstract scenarios beyond spatiality but with huge impact on the way how iLAA are developed, the following scenarios were added as suggestions by the group:

- Public *(New)*
- Private *(New)*
- Commercial *(New)*
- Non-Commercial *(New)*

One aspect which clearly was a concern for a majority of people in the group was privacy: Many tracking devices consist of cameras and or microphones which could record sensitive data. How privacy in the presence of ubiquitous microphones can be ensured and how involuntary broadcast of audio - or video content can be managed has to be considered in the development of I/LPS for iLAA.

In fact the question of "tracking" in itself can be seen as controversial in consideration that in the wrong hands the knowledge of the whereabouts of a person, can be misused. (From rather banal situations where a burglar could use the information of somebody being not at home, for example, to more sinister situations of Orwellian nature.)

To summarise, the focus group established *stakeholders* and possible scenarios' relevant quantitative *parameters* for a positioning system for iLAA. This happened using qualitative methods.

A focus group explores subjective views on a matter (Robson, 1994) but as this happened in the interest of objectifying the analysis from one stakeholder's ideas (The author's) to several, (the forum's) this method was deemed justified. It is hoped that the process of distinguishing quantifying parameters with a group of stakeholders has increased the relevance of the questionnaire and hence the survey's findings.

4.2.4 Survey: Online Questionnaire

The survey had 10 questions divided into 3 sections. the first one asks about the professional background of the respondent, the second about what the respondents (potential or existing) iLAA is or could be and the third about the respondents stance on privacy issues which arise with some positioning systems.

The survey did not ask for sensitive data and it was made clear that the results are for scholarly purposes only, followed by assurance that no record will be kept of respondents details other than what has been disclosed in the survey.

A short summary of each question will be listed in the following section with the findings, the full wording of the survey is attached to this report in the appendix.

The results as per 01 May, 2012 can be accessed on http://creativemusictechnology.org/ILPSforiLAAsurveyresults.html but the survey will remain open for the time being on http://www.surveymonkey.com/s/6X86DJL

4.3 Findings of Survey

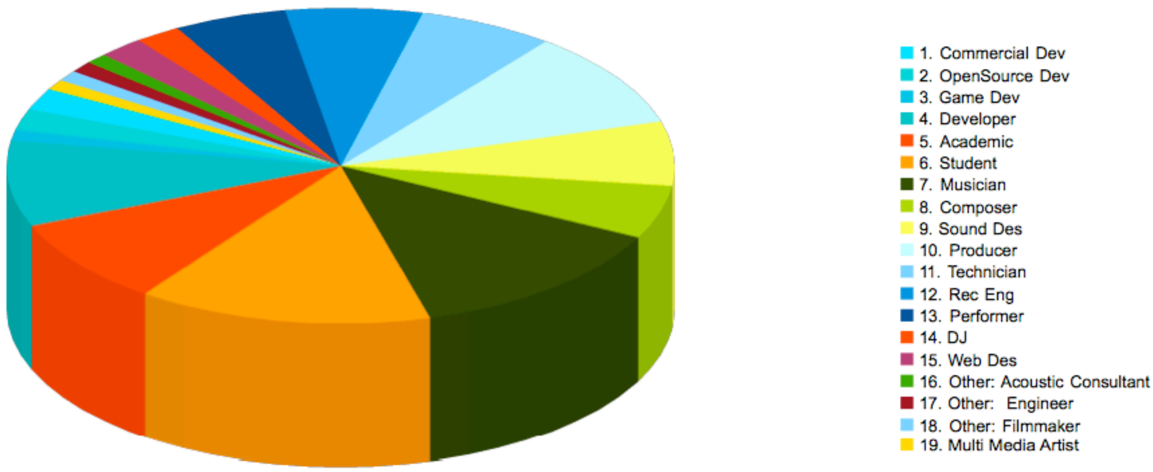
By 15th April 2012, 30 respondents answered the survey. (Less than 50% of the required sample size, alas.)

4.3.1 Question 1

The 1. question asked about the respondent's professional position. Multiple answers were possible. The aim of the question was to gain an insight into the sample structure in order to know what generalisations are possible.

Multiple responses in this question can be understood as indicating a multitude of requirements for more than one application per respondent, so a hypothetical expansion of the sample size as discussed in 4.2.2 could be assumed, as our interest is in requirements per *iLAA*, (existing or proposed) but not per *respondent*.

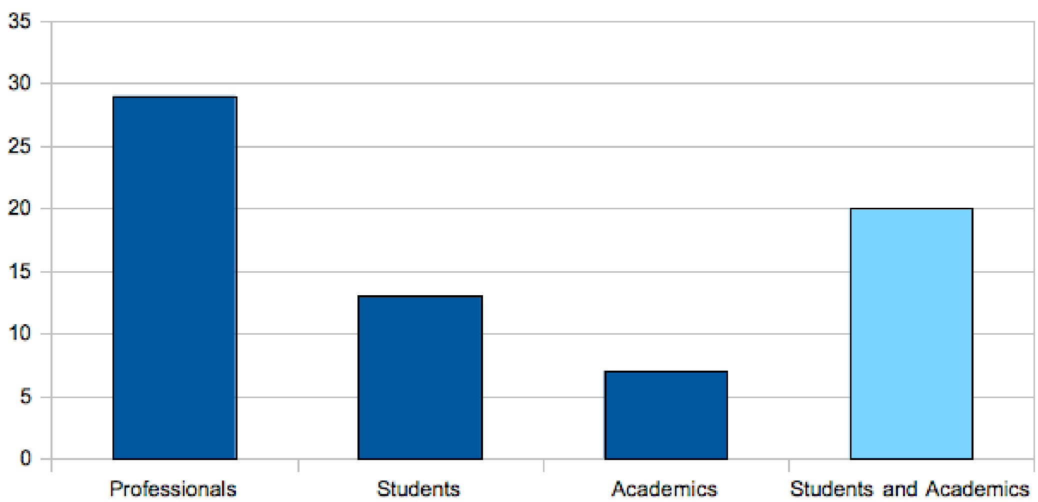
The chart showing all responses thus shows a rich mix of stakeholders. (See Fig. 18)



(Fig 18)

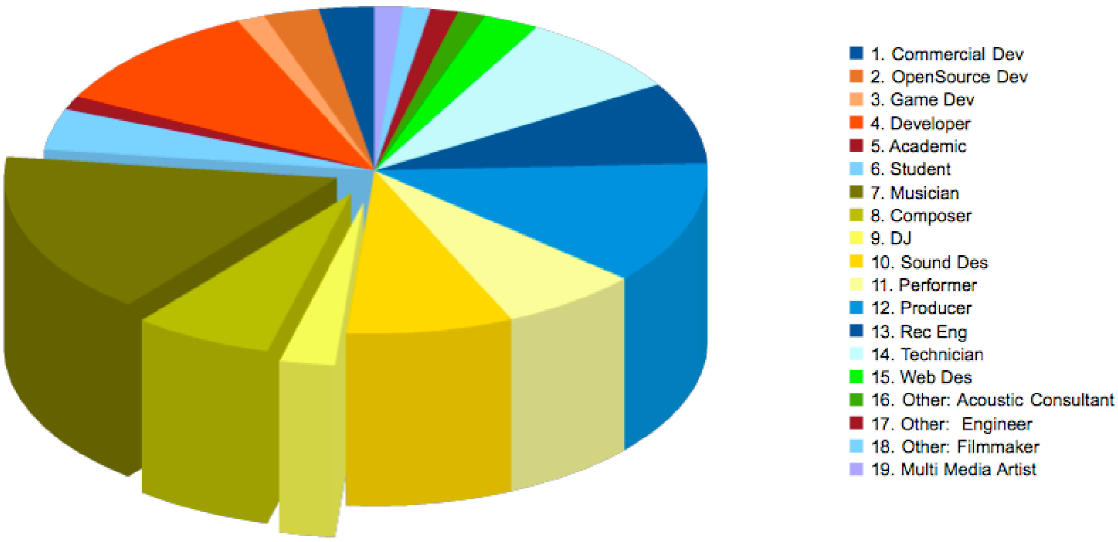
This hypothetically enlarged number has to be moderated downwards though by the multifunction of students and academics, who all answered to a second field, by the exception of 3 students and 1 academic.

A high percentage of academics and students in the sample was expected due to the nature of the study and its academic environment, however, splitting up the academics from the industry professionals, by "allowing" only the latter to be active in more than one field, showed students and academics to be in a clear minority. Professionals which are academics too were counted as academics. Thus, no bias towards academia seems to be evident. But as discussed in the stakeholder analysis in section 4.2.2, academics hold an important stake in the development of I/LPS for iLAA and should be well represented in the sample. See Fig 19 for the percentage of academics/students in relation to professionals.



(Fig 19)

To illustrate the respondents interests beyond academia, in Fig 20, only students and academics who stated these options as their sole occupation are shown. This should illustrate the spread over the fields of interests of the sample more generally and gives information of the mix of respondents with a more technical, versus a more musical/artistic background. Also form Fig 20 it can be seen that respondents with an interest in musical content make up a quarter, (musicians, composers and DJs) but they are not an overwhelming majority.



(Fig 20)

In summary, the results from question 1 thus present a fairly even distributed sample of the general population of stakeholders we are interested in.

4.3.2 Question 2

Question 2 asked respondents about their attitude towards development in general, in order to place the sample in the content of respondents associations to the open source development, for example, respondents attitude to hi-tech and cost, and a general idea where the I/LPS for iLAA developers stand in respect to these issues.

Question 2 was:

It is important to me that the positioning system I use for my application...

(strongly disagree, disagree, agree, strongly agree)

...uses day to day technology

...is state of the art

...is ubiquitous

...uses open source software

...is cheap

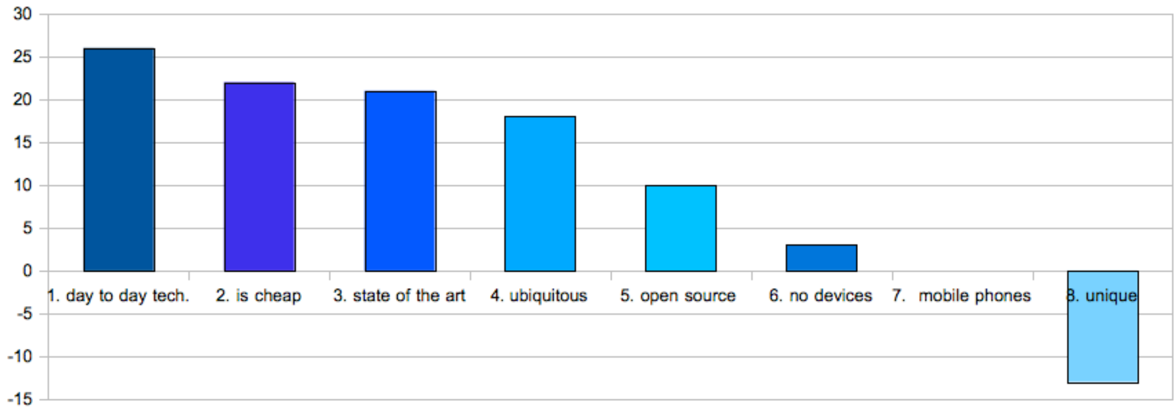
...doesn't require any devices

...is unique to my application

...works on mobile phones

As respondents were left free to decide how they fill this question out, it was deemed best to spread the answers from -2 for strongly disagree to +2 for strongly agree. An answer left open would then be a neutral score. (0) respondents who filled out all options will thus be equally represented in their opinion as are people who only expressed agreement for example but didn't voice a negative opinion.

However, the most divisive issues are thus slightly misrepresented as they show a near neutral or neutral score even if opinions were strong, but divided. This was namely the case for the question about open source, where the (weighted) difference between the negative score and positive score was as high as 32. (See Fig 21)



(Fig 21)

Interestingly, the respondents value affordability as much as state of the art. The expectations on technology are clearly high, and demanding, *and* day to day availability is a clear a priority. In the question design it was assumed that respondents would either favour state of the art *or* affordability. However a majority decided to chose state of the art *and* affordability ("..is cheap) as a prerequisite.

(This is an example how multiple answers can bring forward unexpected but relevant insights.)

Further, carrying a device is not seen as a problem, but it is evident that this device can be something else than a mobile phone. Uniqueness is not necessary.

A highly contested - as mentioned, but still seen as positive over all, is if the I/LPS uses or is based on open source software.

Some respondents favour a no device approach, but devices are clearly an option respondents are happy to deal with.

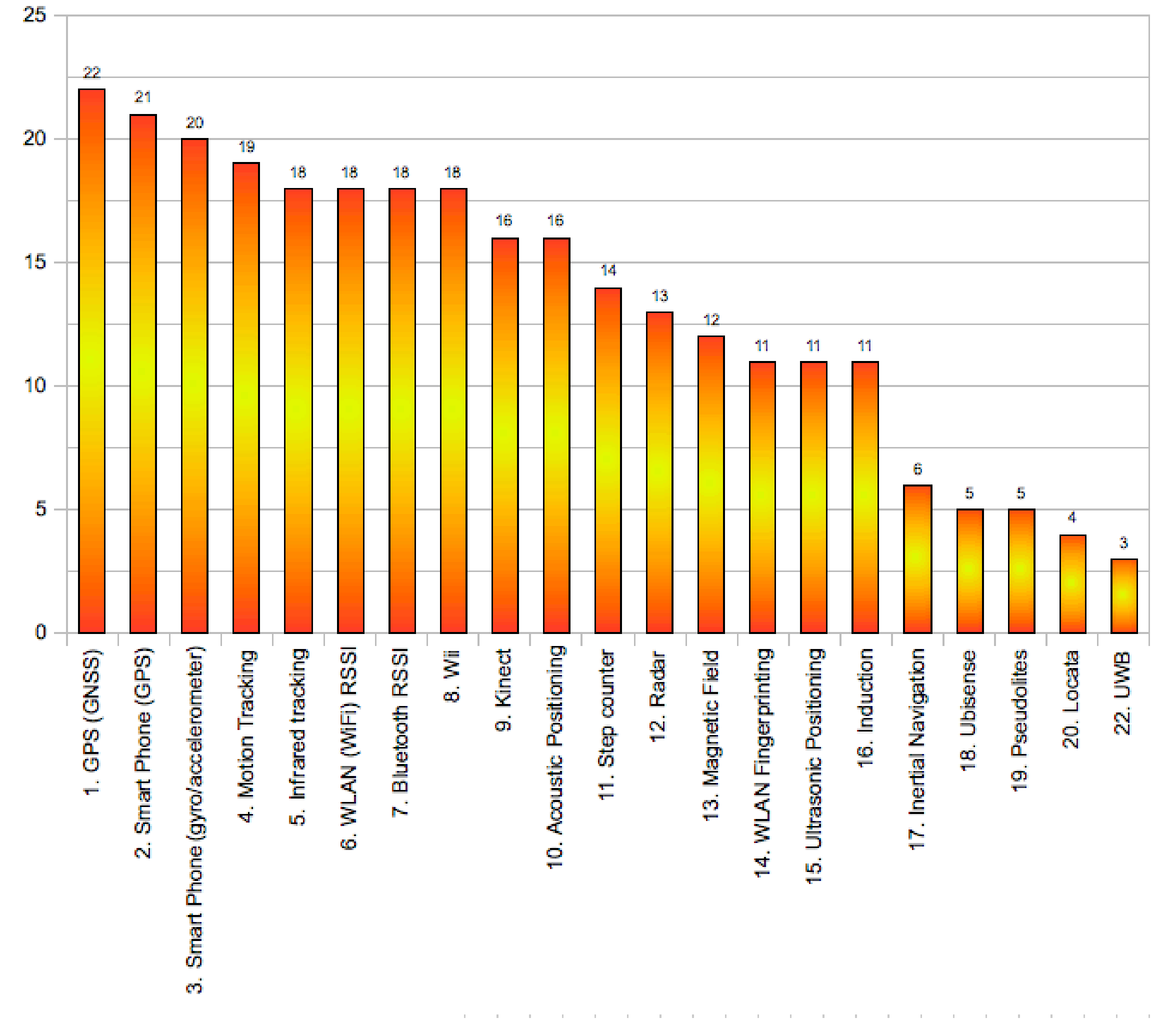
The question after uniqueness is interesting as it indirectly indicates the willingness of the respondents to develop their own positioning system: A clear indication is given that a system which works for the respondents would be adopted in consideration of the other factors like price, etc. but unique it doesn't need to be.

In summary, respondents want positioning technology to be day to day technology, state of the art at low cost, and available everywhere.

4.3.3 Question 3

The first section of the 3rd question asked about what systems respondents have heard of. The answers yielded evidence for the ubiquitousness of GPS and GPS on smart phones. Interestingly despite the general knowledge about gyro and accelerometer in smart phones, the term "inertial navigation" seems not to be familiar to everybody, ranking 17th, despite the gyro-/accelerometer smart phones ranking 3rd.

For an overview, the charted answers probably speak best for themselves. in Fig 22.



(Fig 22)

The question about which systems respondents are interested in was aimed at finding out which of the systems developers have heard of *and* are considered to be of interest to them.
To find out what motivates them to that interest can not be conclusively said from the available data.

The answers might be compromised by a perceived ambiguity in terminology: It seems that respondents either ignored "inertial systems" because they already ticked the gyro/accelerometer field or they did not know what it was supposed to describe. There is no way of knowing which is the case.

Therefore, as far as the interests of respondents are concerned the (shared) underlying technologies shall be grouped together, where a respondent has not already done so. (respondents who are interested in smart-phone GPS and "normal" GPS are given one "vote" for GPS.) The raw data shall be shown in Fig 23, in order to compare to other sections of question 3 and for transparency's sake. Fig 24 shows the weighted table.

position	*system*	*votes*
1	GPS	19
2	motion tracking	8
3	Kinect	7
3	acoustic positioning infrared smartphone GPS smartphone gyro/accelerometer	6
4	ultrasonic magnetic radar induction	5
5	inertial navigation wii WLAN RSSI	4
6	step counter WLAN fingerprinting Bluetooth RSSI Pseudolites	3
7	UWB Locata Ubisense	1

(Fig 23)

position	*technology*	*votes*
1	GPS	13
2	sonic (acoustic and ultrasonic)	11
3	IR	10
4	motion tracking (optical) Inertial	8
5	magnetic	7
6	radar	5
7	WLAN/WPAN	4
8	UWB Loacata Ubisense	1

(Fig 24)

As IR and optical systems both need cameras, they could be grouped together. This would bring them out top, putting GPS second and sonic systems third. From the data it is not clear if the respondents distinguish between "optical" and IR or not.

A further limitation which has to be mentioned here, is that it can not be clear from the data if the omission of ID-tag systems in the list has had a confusing influence. It was thought helpful to leave it out as the tags can be representative of a multitude of technologies, e.g. ultrasonic, RF; WLAN, etc. On hindsight this might have had an adverse effect. However, no respondent added tag-systems in the "other" - field either, so maybe the omission was if not fully justified at least irrelevant.

Unfortunately there is no way to disambiguate from the data if respondents are aware of the multi-functionality of the wii as a combination of IR *and* INS. An arbitrary decision needed to be made here, and the wii was considered to be primarily IR. The adjusted table could be altered however to account for this by counting the wii twice, for IR *and* for INS. this puts *inertial technologies* ahead of IR and *sonic technologies* into 2nd place with a vote of 12. This is in as much misleading as the success of wii (and Kinect, for that matter, but with added acoustic positioning) stems from their *hybrid* nature.

What can be said with confidence, however, is that respondents follow with clear interest the developments in GPS (GNNS) technology, optical *and* sonic systems, and there is some indication that the interest in systems like Kinect and wii can be understood as an interest in hybrid systems in general.

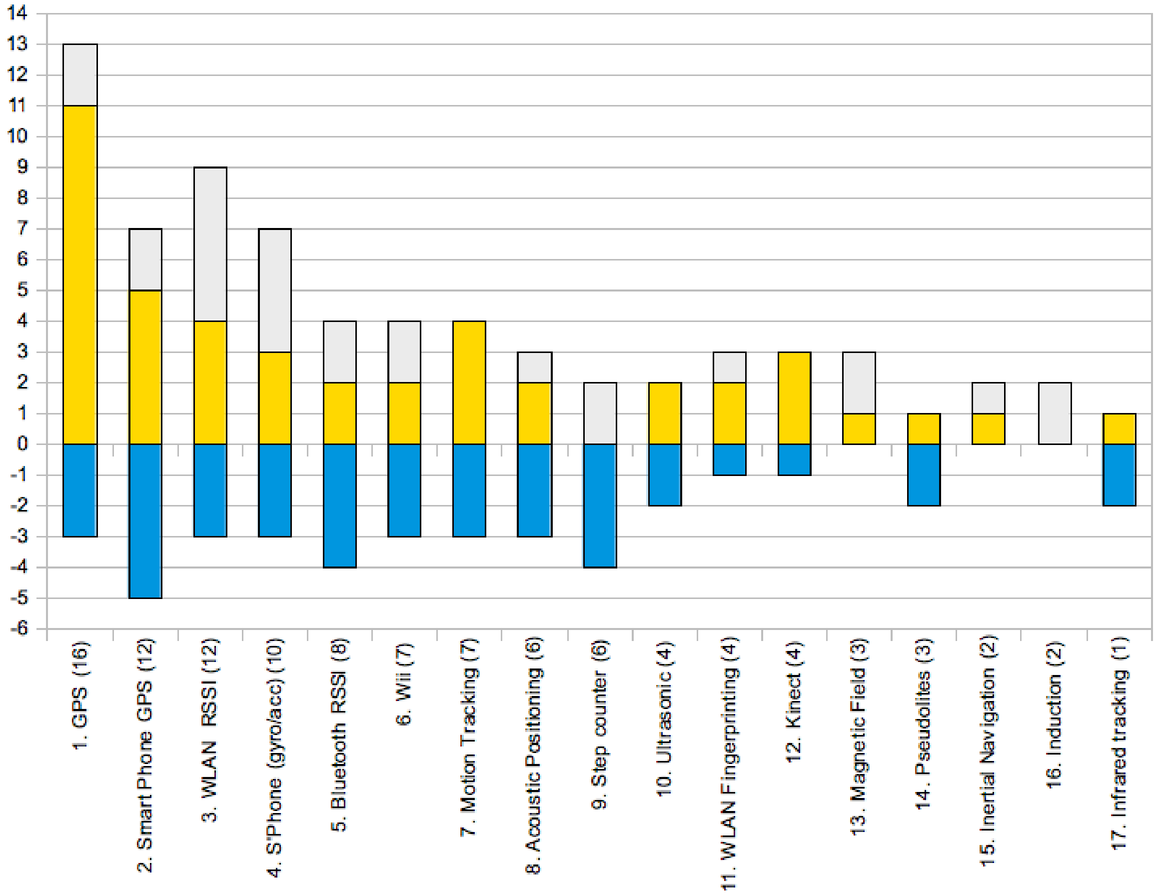
The next table shows the answers to the remaining sections of question 3, posed to respondents who have worked with any of the listed systems and asked if they found the systems adequate for their needs or limiting. (Fig 25)

This question is problematic as the answer depends of course on the nature of the application. The answers can only state abstractly what the respondent thought. No generalisation can be derived really other than that in some application some system was perceived as limiting or adequate, respectively. If, however, a particular system shone through as throughout adequate, it would indeed show that a fully scaleable, versatile and yet accurate system has been found. (The working hypothesis as outlined in section 4.2 is that this does not (yet) exist.)

As GPS can only really apply to applications of outdoor nature, it shows that developers working in that field are generally happy with the way the system works.

If the positive experiences are compared to the negative ones as ratios of success of a system - which has to be done with caution, as for example the only respondent to rate IR systems found it adequate for some things (giving it an adequate rating) but totally inadequate for others. To say that

2/3 of respondents found IR systems inadequate would be misleading - beyond GPS are no real contenders which would shine above the rest.



(Fig 25T: grey = no opinion, yellow = adequate, blue = limiting. numbers in brackets = total respondents who worked with system.)

4.3.4 Questions 4 - 6

Question 4 aimed to find out if respondents support the suggested iLAA classifications, and if there are any categories of more relevance than others.

Two fields had no respondents (manufacturing and surveillance) and two fields were added. (Acoustics and storytelling, sustainability/Climate change) the latter two probably could be included in education and / or performance art.

The responses are fairly evenly spread around education, spatial music, gaming, interctive art and multimedia, with a slight bias towards performance art and distinct bias towards recording technology. Assistive technology does not seem unimportant, manufacturing and surveillance seem to be irrelevant, however.

1.	Recording technology	13
2.	Performance Art	11
3.	Education	9
4.	Spatial Music	8
	Interactive Art	8
5.	Gaming	7
	Multimedia	7
6.	Assistive technology	4
7.	Home theater	1
	Other: Acoustics	1
	Other: Storytelling, sustainability/Climate change	1
	~~Manufacturing~~	0
	~~Surveillance~~	0

Questions 5 and 6 were about the nature of the respondents' applications scenarios and particulatly asked about the spatial reach of the application. Calculating the correlations to the answers of Question 4 show that local and indoors scenarios apply to most iLAA. And immersiveness seems to be a generally apt description for the majority of applications.

The geographical scenarios will be compared to the absolute parameters stated in question No 8 for respondents who filled out both, question 5 and 8. This is of interest as this can give us an indication as what respondents mean when they say "global" or "local", and will be discussed in the section on Question 8.

Purely audio and *audio contro*l feature high in preference for iLAA, just behind immersiveness. This seems to indicate a general interest by the repondents in applications where the resulting media is of acoustic nature.

A large proportion describes the application as ad-hoc. This has of course direct implications as to what position technologies are applicable. (Fig 26)

Geographical Scenarios

1. Local	15
2. Indoors	14
3. Urban	11
4. Global	7
5. Rural	6

Scenarios

1. One-off (ad hoc, for a show, a gig)	13
2. Public	12
3. In Motion	10
4. permanent (architectural, etc.)	10
5. Private	9
6. Non-Commercial	9
7. Crowded	8
8. Commercial	5
9. Ubiquitous	5
10. Controlled	3

more on iLAA

1. Immersive	13
2. Audio control	12
3. Purely audio	11
4. Narrative	9
5. Interactive	9
6. Dynamic	7
7. Static (in a museum)	6
8. Visual, audio only assisting	6
9. Gestural Control	3
10. Assistive technology	3
11. Gestural	2
12. Other: Mainly audio, but image and moving used as well	1

(Fig 26)

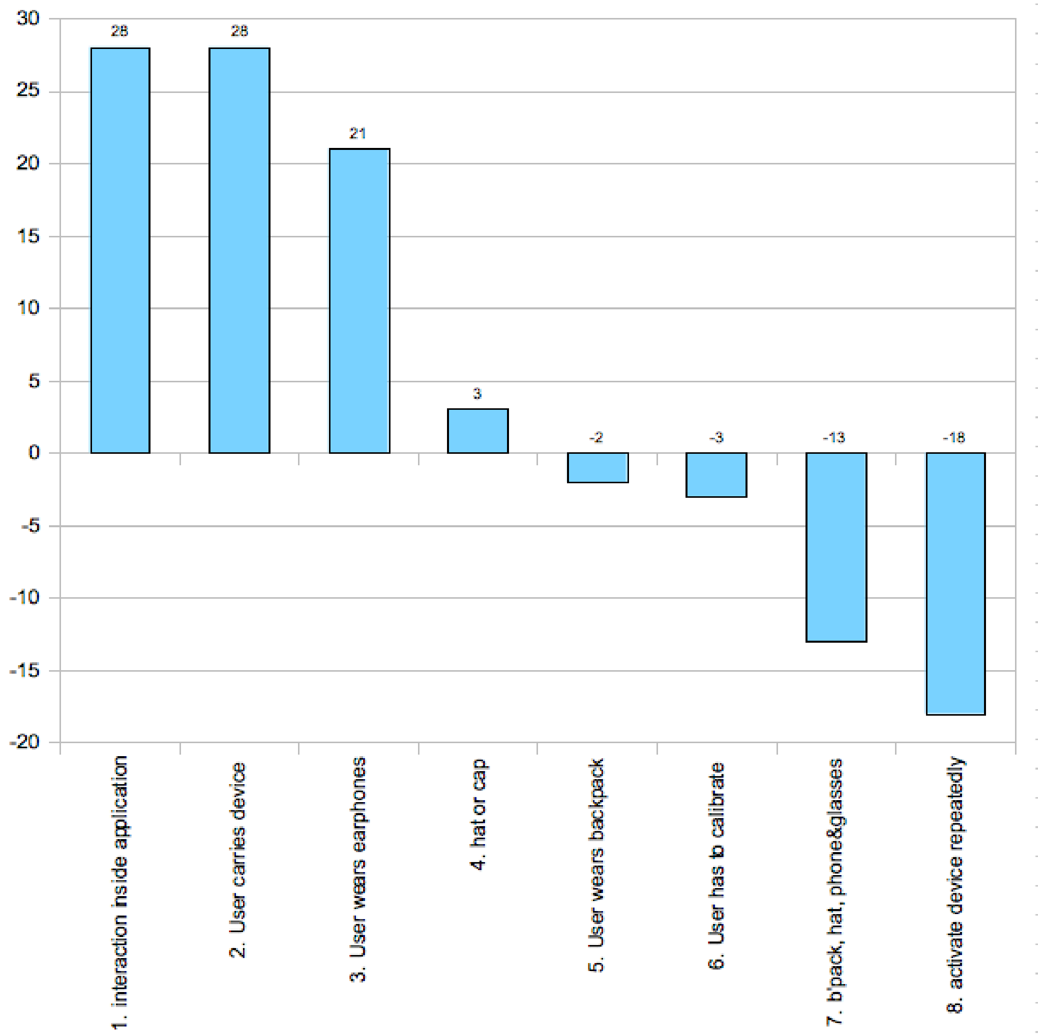
4.3.5 Question 7

Question 7 enquires about the acceptability of user interaction with the positioning system. This informs user requirements directly. (Fig 27)

Same as in Question 2, a respondents non-opinion was expressed as a 0 score, totally unacceptable as a -2, unacceptable as a -1 acceptable as a 1 and totally acceptable as a 2.

The answers were nearly unanimously in agreement that carrying a device is acceptable and that interaction could be part of the application. Calibration before application start is experienced as a negative aspect, and interaction with a device which is not part of the application seen as unacceptable. So are multiple devices like backpacks etc. A hat or cap seems to be met with moderate approval. Interestingly, the second least popular option, the wearing of backpack, glasses, head phones and a hat, is indeed an option suggested by developers of telepresence systems like Packi et al. (2010)

In this context it has to be considered that we happily don glasses to view 3D films, for example. But it seems that the accumulation of devices is the main stumbling stone here. A backpack on its own, for example, only meets small resistance by the respondents.



(Fig 27 left to right: from acceptable to unacceptable)

4.3.6 Question 8

Question 8 asked directly for parameters' maximum or minimum values. As these parameters could directly guide decisions on I/LPS development it is important to view this in connection with respective iLAA scenarios. Crucially, the averages in Fig 28 are *illustrative* only. It is quite evident that there are several trends expressed in these values: One part of respondents requires accuracies in the meter range but incidence of trackable devices up to 180 instances, others need sub-metre range but a lot more modest incidences of traceable items. A system which lies somewhere between these limits would possibly be useless to both groups.

Further, if an iLAA takes place in an area of 10m^2 it is hardly feasible that 45 instances of traceable items are within this area if the items are hand-held devices carried by the same number of people.

Absolute Parameters	*averages*	*range*
1. Accuracy > 1m (m)	4.75 m	1 - 20 m
2. Accuracy < 1m (mm)	0.35 m	0.001 - 1.0 m
in metres for < and> 1 m	3.36 m	0.001 - 1 m
3. Acc. vertical > 1m (m)	3.28 m	1 - 9 m
4. Acc. vertical < 1m (mm)	0.1 m	0.003 - 0.3 m
in metres for < and> 1 m	2.5 m	0.003 - 9.0 m
5. latency (s)	6.9 seconds	1 - 30 seconds
6. User cost (£/€)	10.6 €/£	1 - 500 €/£
7. Whole System cost (£/€)	4009 €/£	1 - 25000 €/£
8. Area	658 m^2	10 - 2400 m^2
9. incidence of tracked instances	45 incidences	1 - 180 incidences

(Fig 28)

From the more detailed Fig 29, showing relation to iLAA scenario using values of individual respondents, specific disambiguation is possible. However, the evidence is too anecdotal really to formulate more than trends based on the available findings: No numerically evident correlations between respondents parameters to particular scenarios can be found which would allow to group them. It is hoped, as the survey remains open, that more respondents could provide more data for future research.

Respondents No	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
scenarios:													
1. Global			X			X	X				X		
2. Local	X				X		X	X	X		X	X	X
3. Indoors	X	X	X				X	X	X	X	X	X	
4. Urban				X	X		X			X	X	X	
5. Rural							X				X		X
6. Crowded	X				X				X		X		
8. In Motion	X		X								X	X	X
13. One-off (ad hoc)	X	X	X	X			X	X			X		
Absolute Parameters													
1. Accuracy > 1m (m)		2	20	1	5	10	5	1	2	2	5	1	3
2. Accuracy < 1m (mm)	400		500	1000	100			100	3				
3. Acc. vertical > 1m (m)	1		9	1	1			1		2			8
4. Acc. vertical < 1m (mm)			300	1	100			100					
5. latency (s)	3		1	10	1	1	1	1	30	1	2	2	30
6. User cost (£/€)	5	100	2	1	10	5	3	5	2		3	1	1
7. Whole System cost	500	5000	500	1	100	500		10	10000		25000	10000	500
8. Area	400	10	50	25	100	1000		10	2000	500	2400	1000	400
9. tracked instances	8		180	10	50	1		2	100	40			20

(Fig 29)

The parameter range derived from the data provided by question 8 can be used with more confidence, as it is based on more readings than for correlation of a sub-sections and is based on sample sizes in solid double digits.

As far as the start of sub-metre range is concerned, the one respondent's value setting the minimal value at 3 mm might be an outlier, or the victim of a possible metric conversion error, as the same respondent describes the iLAA as "*urban*" and "*local*", rather than *gestural contro*l, the only type of iLAA which would require such an extreme minimum. The bulk of iLAA which require sub-metre accuracy lie around the 0.3 m mark. which is probably more realistic.

The parameter maxima/minima derived from this question give an insight into various requirements, however, for conclusive generalisations the data is a bit too scarce and more respondents would be needed for a more concise picture. The data does support all working hypothesises reached in previous sections, and certainly not contradict them: the findings do not contradict the taxonomy of iLAA, the experiences respondents had with existing systems are congruent with the technical discussion of them in chapter 3.

4.3.7 Question 9 and 10 (Privacy Issues)

Some positioning systems actively broadcast the position of a user. This potentially creates privacy issues as the location information could be passed on to a third party and/or be used for surveillance.

An alternative to this are the implementation of passive systems where the location-information stays with the user and is not broadcast. (The most simplest example being a map)

Hence respondents were asked as to their stance towards these issues.
Multiple choice answers, in a weighting in four stages, were:

1. This is not a problem as for my application the user explicitly wants to be tracked and has given consent to this.

2. This is a problem as for my application I can't ask the consent of the user.

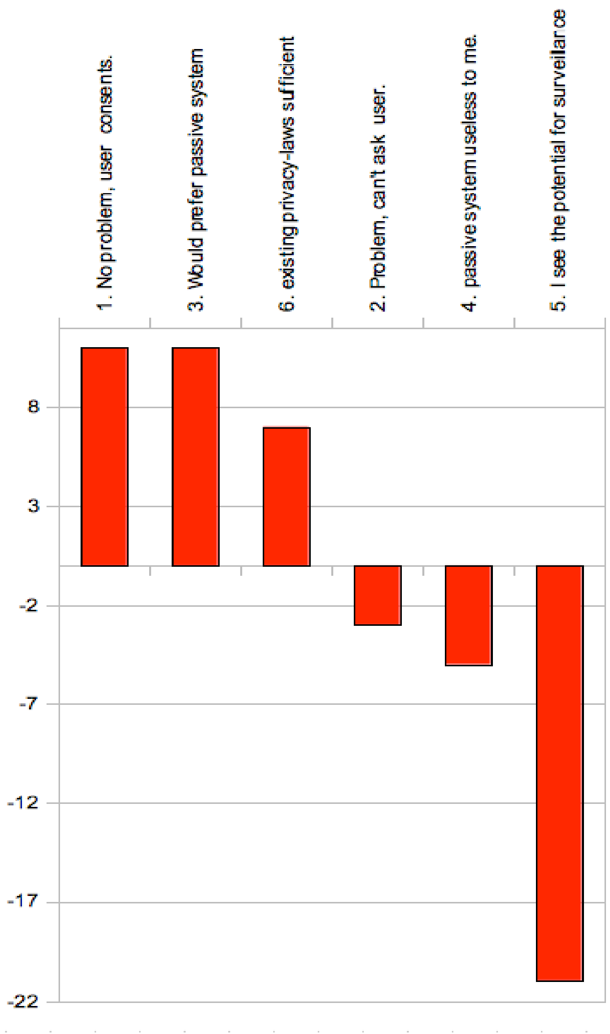
3. This is why I would prefer to work with a passive system

4. The whole point of my application is that the users position is broadcast, a passive system would be useless to me.

5. I see the potential of surveillance for public security and hence welcome it.

6. As long as existing privacy-laws and ethical guidelines are respected, I'm not worried about privacy

The following chart shows the weighted answers. (Fig 30)



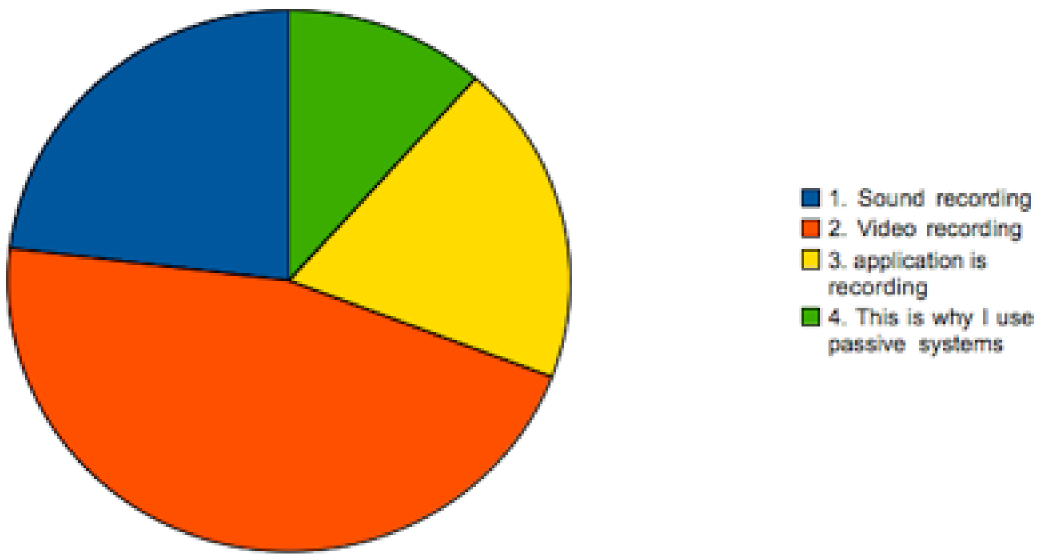
(Fig 30)

About half of the respondents agree that the current legislation and best practices - if adhered to - are sufficient but respondents clearly distance themselves from practices using positioning systems for surveillance, even in respond to the explicitly cautious wording used in order not to appear to be leading.

However, more respondents have no issue with the fact that a user can not consent with being tracked than do have an issue. This might be related to the nature of the application the respondents had in mind, but it has to be looked at individually how a particular system deals with this non-consentual aspect.

On the other hand a lot of privacy issues can probably be dealt with in the way suggested by a respondent in response to question 9: "[...]prefer 'passive' systems - or systems which generate data which can be of no value or interest to any third party e.g. MIDI data" This is of course a valid point, but if the information in the end intrinsically contains information about the *position of somebody*, the data will remain sensitive.

Some positioning systems use cameras, some use microphones. From the point of view that a malevolent third party could get hold of the recordings, respondents were asked in question 10 which would worry them more to get into the hands of the wrong person, a sound recording or a video recording. The alternative answers were to state either "Not relevant as my application IS in fact a means of recording audio or video" or "This is why I use passive systems only, where the location information is not broadcast." or tick "I prefer not to answer".



(Fig 31)

Besides the answers contributing to the pie (Fig 31), the following two remarks were added under "other":

"I prefer 'passive' systems - or systems which generate data which can be of no value or interest to any third party e.g. MIDI data"

"Different for different projects I'm working on. Ethics related to specific context and users is strongly important always."

In respect to the discussion in the forum at PMS this is an interesting result: Within the group opinion was voiced by several participants that audio recordings can contain a lot more sensitive "content" as the abstract content of information encoded as language can be recorded (and potentially misappropriated)

The findings don't support that this impression is shared by respondents.

4.3.8 Conclusive Remarks On This Chapter

To summarise, The respondents answers to Question 1 show a fairly broadly spread sample of the relevant stakeholders. Question 2 shows that day to day technology, ubiquitousness and affordability inform the choice of I/LPS. In Question 3, GPS turns out to be the most succesful system for people who have worked with it, however, the results show too that beyond the applicability of GPS (i.e. for indoors applications) no system shows a distinctly positive track record.

Questions 4 - 6 enquired about the scenarios for iLAA envisaged by developers, showing in response to Question 4 that *recording technology, performance art, education, spatial music* and *interactive art* being are the best represented scenarios by the responding stakeholders. Responses to Questions 5 and 6 show that *local, indoors one-off (ad hoc) immersive* and *audio control* were the other contenders named most often as possible scenarios.

In Question 7 it was established beyond doubt that one device to be carried around is acceptable but multiple devices is not and that interaction with the system should happen as integrated part of the application, not as a separate action in form of calibration etc.

Answers to Question 8 in correlation to answers to Question 6 gave some direct insights into absolute values on accuracy, reach, density, response time and cost in the expectation of the respondents. They show that indoors systems should have an accuracy bellow around 0.3 metres and allow multiple incidents of traced devices, a reach between 10 and 100 meters (depending on the application) a latency of approximately 5 seconds and cost around £5.00 per user and £500 in total.

Answers to Questions 9 and 10, asking respondents about privacy issues, show clearly that the stakeholders are aware of the implications on privacy when using tracking devices for iLAA, particularly when these technologies use cameras or microphones which could be used to record sensitive information.
However, in general respondents believe the existing legislation and best practice if adhered to, provides sufficient protection.

5. An Acoustic Only I/LPS for iLAA - A Suggestion

This chapter describes the suggested system and a software simulation applying correlation to audible audio signals as a means for trilateration of a position of a mobile device within a soundfield.

The simulation uses some simple filter techniques and employs a simple model to introduce first reflections and white noise to test the theoretical possibilities of the system.

5.1 Potential of acoustic-only systems:

In Janson et al. (2010) Packi et al. (2010) Mandal et al. (2010) Wendeberg et al. (2011) we see systems which clearly suggest that audible sound, in the form of coded signals or as environmental sound, can provide a means for localisation. In fact, many principles of ultrasound technology could theoretically be applied to the audible range.

Moreover, principles used in RF technology, like frequency identification, can be applied to audio signals as can signals strength and TDOA, TOF and RTOA. and trilateration.

The advantage of audio signals is that we know audio signals' propagation paths in the near field rather well. We know too how the environment affects them (temp, pressure etc.).

The downside of the application of audio in the audible range is of course the fact that in many environment the audio signal could be perceived as a noise. As we are looking at audio applications, however, the presence of audio signals is self evident, as is the necessary infrastructure. The use of audible range audio for use of positioning in iLAA is undoubtedly a conceptual possibility in the vein of ubiquitousness. But its possible uses where acoustic disturbance is not an issue probably reach beyond iLAA.

5.2 Application Scenario and Simulation

The following scenario was envisaged:

In a room of small venue proportions, i.e. approximately 12 * 12 m or similar, there is a surround sound system consisting of loudspeakers which are roughly spherically distributed. The exact number of loudspeakers is not relevant to the principle, but 3 speakers were chosen according to the reasoning outlined in section 5.6.

For the purpose of demonstrating the systems three dimensional possibilities and for its mathematical simplicity, the former, simpler, layout is being used here.

The audio content distributed on those speakers shall be used to provide the information necessary for a mobile device to calculate its position in relation to the speakers. The audio content is available to the mobile device at the same time as it is available to the speakers. (Synchronously so.)

Additional microphones beside the one on the mobile device are an option but in the interest of keeping infrastructure minimal, shall be, if possible, avoided. (This is as well in the interest of minimising privacy issues discussed in previous chapters) For the simulation the microphones were left out all together, to reduce influences of hardware - limitations to the system at this stage. Instead a sound file was delayed in accordance to the speed of sound formula (FORMULA) for each of the three (simulated) speaker positions.

To be of value for iLAA, the tracking of the mobile device needs to be accurate enough to meet the maxima and minima parameters derived from our ability to hear spatially and the parameters derived from the user requirement survey in previous chapters.

The mobile device for a real system would ideally be a commercially available wired or wireless audio microphone (lavalier, clip on, hand-held, etc.) or a mobile phone or laptop for example. The mobile device either has on-board processing power to conduct multiple realtime convolutions/correlations or is connected in real-time to a device with the required processing power.

The parameters derived in chapter 4 shall apply, and a definition of the systems requirement is summarised in table (TABLE), The development of the suggested system simulation can thus be regarded as an experimental test of the findings so far.

5.3 Three Theoretical and Conceptual Approaches

When looking at the options of developing audio - only systems initially three conceptual approaches were identified:

1.) Correlate a known signal with its delayed spatially distributed copy to calculate distance
2.) De-convolve a known signal with HRTFs to gain locative information (binaural, needs 2mics)
3.) De-convolve a known signal as to its axis content when recorded on a soundfield mic

The third system is (most likely) a variation of how the Kinect gains acoustic - locative information. However, Kinect does use proprietary software and no more information could be found which explains the mechanisms further.

If the signal was in an ambisonic format, and the microphone a soundfield, it should be possible to work out the position of a soundsource by decoding from B-format. The positions of the speakers would have to be known. Near coincidental tetrahedral microphones, which is what a soundfield mic is, could certainly be built out of cheap mass produced components, but ubiquitous, they are not.

The second system would require a precise range of HRTFs. It would model the ear in many ways, as it relies on "recognition" of spectral sounds and works out the distances thereby like the ear would, through computation of frequency shifts, and DToA based on HRTF of a listener model, or, in a simpler model, by comparing ToA at various physical locations as ITD.

A more complex system based on the same principle would compare (real-time) FFT's of a sound-source recorded at various physically spaced microphones and work out the position of the source by comparing the spectral analysis. This analysis could possibly be used to identify and isolate various sound-sources in a complex signal and track their individual movement.

Both these approaches might be interesting enough to follow up in future research, but by its simplicity the first approach appealed most and hence was followed in order to create a prototypical, software based simulation with the view to progress to real world experimentation.

5.4 Correlation

Cross correlation is the correlation of two signals as a function of the time deplacement T in the formula

$$Rab\ (T) = Sum\ a(t)\ b(t + T)dt$$

In Randall (1987) this formula is given for all transient signal and a derived form for stationary signals. As we are interested in transient signals rather than stationary signals, no other formula was considered necessary.

The similarity of the formula to the one for Fourrier transform is evident, and indeed, cross correlation can be derived from the cross spectrum of the inverse Fourier transform. (Randall 1987) The aspect of cross correlation we are interested in here is the time delay between 2 identical transient signals. This special case of cross correlation is called auto correlation, which is the case when a(t) = b(t).

$$Raa\ (T) = Sum\ a(t)\ a(t + T)dt$$

If the cross correlation of two signals set apart by a time delay is being processed, the resulting function of time will show a slope to a peak and a descent there after. This peak, which, in case of auto correlation equals 1 and generally represents the moment of highest correlation conveniently shows the delay-time of the two signals.

By measuring this time delay we thus have the time delay of two signals, which, if we know that the delay stands for the difference of time of arrival of a signal broadcast from a speaker at a distance can be used to calculate that distance as we know the speed of sound in air.

A further advantage of cross correlation is its robustness to multipath and white noise. An early reflection adds a peak within the cross correlation function, but it would be, per definition, smaller than the direct signal due to the attenuation associated with early reflections. The robustness to white noise is such that auto correlation is due to the fact that the *correlation* of the two signals is observed, noise would be providing non- correlated values, returning 0 in the correlation function.

According to Randal (1987) the auto correlation of broadband signals return a lot clearer results than auto correlation of narrow band signals. He further suggests the use of the envelope of the *analytic* signal to achieve clearer peaks. The analytic signal is the absolute of the sample values in the frequency domain. (As this would have involved an extra step within the signal process used for the model simulation, this was noted as an option but not implemented.)

In the realtime implementation to model a positioning system based on these principles, a windowed correlation is necessary, which introduces certain problems as the window size will not necessarily correspond to crossing points at 0 values. A certain jumpiness will have to be expected and possibly clipping-glitches could occur.

The way the windowed correlation works, the cross-correlation function has (input samples A + input samples B) -1 answers. Or: If 16 samples are compared with 16 samples, there are 31 points of comparison. (example: The five fingertips of one hand can meet the 5 fingertips of the other 9 times if correlated in the order as they appear from left to right or right to left respectively.)

In summary, it is important here to acknowledge the limitations imposed by choosing to ignore stationary signals and narrow band signals as an option to obtain values from. Even if they are not the main characteristics of musical (and as it happens speech-) sounds, they do ocur and possibly introduce error to the system. However, even if there is more than one peak in a correlated signal of a window, (due to noise induced by multipath or similarities in the signals from other speakers), it should still be the largest one which expresses the relevant time delay. Finding the maximum amplitude of the correlated signal of a window should provide the right clue.

5.5 Possibilities and Limits of Acoustic Positioning using Correlation

The limitations of localisation using correlation of audio signals lie within the sampling rate of the correlation. Here is, in more detail, the principle behind the idea:

If in the scenario outlined in 5.2 a pre-recorded (thus "known") signal to which the processing device has access is being played over (to start with) a single speaker, a microphone on the same device can record the signal coming from the speaker and compare the recording to the original. There will be a time-delay on the recording in comparison to the pre-recorded original.

To work out the delay-time, the two signals, (the pre-recorded one and the one recorded on the device after distribution through air via speaker) are being correlated. The resulting correlation function's first peak will be, if looked at in the time domain, at the point where the correlation is maximum. This will be after the delay-time Δt.

Using the speed of sound through air we can work out the distance d from Δt. Doing this for multiple speakers sending separate signals we can trilaterate the position from the device. This works even if the position of the loudspeakers are not known if there are 4 or more loudspeakers, if the positions of the loudspeakers are known, 3 loudspeakers will be sufficient. Even easier it is if the speakers happen to lie on known points of an x, y and z axis.

So, in summary, here is another description of the system with the layout we assume for the following explorations:

1 microphone on a mobile device
3 speakers (at known points on a x, y and z axis)

- A 3-track recording of a complex signal (music, for example) which is played over the speakers.
- The mobile device records the signal from the 3 speakers as a mono track.
- The signal of that monotrack is analysed as to its cross corelation to the 3 original tracks.
- This should result in 3 separate delaytimes of the mono signal in relation to the 3 original tracks.
- The device's position can then be calculated from the delay-times based on the speed of sound through air.

The relevant numbers at a sampling rate of 44.1 kHz are:

during 1 sample of 0.0000227s sound travels 0.007718 m
t * v = d 0.0000227 s * 340 m/s2 = 0.007718 m (distance travelled per sample)

This, as a - theoretically at least - sub-centimetre value, is considered adequate for iLAA purposes and 44.1 kHz will be used in the following simulations and experiments.

As mentioned above, to work out the delay-time, the two signals, (the pre-recorded one and the one recorded on the device after distribution through air via speaker) are being correlated. This happens in "windowed" real-time, i.e. a buffer sized window is taken and the correlation for that window is done sample-by sample, and then when this window is done, the next buffer size window is correlated.

This window size is thus another of limiting factor to consider: the distance sound travels within one window is the maximum distance we can calculate.

in 1 window of 1024 samples sound travels 7.9 m which gives us the
maximum range: 7.9 m (1024 * 0.007718 = 7,9)

To find out how far an amount of delay is in meter (at 44100 Hz and window size 1024)
we multiply the samples by the distance travelled per sample

e.g 512 samples * 0.007718 = 3.95

To find out how many samples a certain distance is (at 44100 Hz and window size 1024)
we divide the samples by the distance travelled per metre.

e.g. 7.9 m / 0.007718 = 1024

So, for a distance of a sound source from the microphone on the mobile device, we can expect sub-centimetre accuracy at 44.1 kHz, and a range of 7.9 m at a sample window size of 1024. For these experiments and simulations, Apple Macintosh's vDSP_conv() function which is part of the accelerate library is being used, which has a maximum length of 1024 samples. The range of 7.9 meters was hence accepted as the maximum range, not least as larger distances were not available in a real world set up to experimentally test anyway.

How changing the sampling rate influences range, resolution and latency is summarised in Fig 32

	44.1 kHz	96 kHz	192 kHz
1 sample in ms	0.0227	0.0104	0.0052 ms
speed of sound	340 mm/ms		
sample window of 1024 samples =>	23.24 ms	10.64 ms	5.32 ms
time needed for sound to travel 1m	2.94 ms		
distance of sound travelled in 1 sample	7.718 mm	3.536 mm	1.768 mm
max.distance represented by window	7.9 m	3.6 m	1.81 m
processing power	constant		

(Fig 32)

So by halving the window size when doubling the sampling rate, the distance represented by the window stays constant.

To extend the range of the system (increase the distance represented by the window) but keeping the processing power the same, the option is a larger window size at the expense of lower sampling frequency. How this influences the accuracy of the correlation remains to be tested experimentally. With the current set up using Apples vDSP_conv() function, the maximum window size is set to 1024 and these experiments have to be deferred to future research.

The increased sampling rate however increases the accuracy (at the cost of range) which potentially extends the uses of the system to other applications. At 192 kHz and theoretical accuracy below 2 mm, uses in asset tracking and other logistical applications are certainly worth considering.

5.6 The Principles of Trilateration Applied in the Simulation

with 3 loudspeakers at known positions, knowing the Euclidean distance (FORMULA) between the measuring device and the sound source is sufficient to calculate a 3d- position. To start with it is assumed the 3 positions of the senders are on the end points of an x,y and z axis of a right handed Cartesian co-ordinates system. Further, it is assumed that the volume in question is entirely in the positive quadrant. And the x-y plane is horizontal, corresponding to azimuth, the z axis vertical and corresponding to elevation.

As it might not actually be possible for a real system to set up the speakers according to this layout, a fixed offset for each speaker could be calculated and the measurement corrected accordingly.
This makes particularly sense when setting up a system in a room where the maximum distance for z is not practical. (7.9 metres)

Basically, as long as the positions of the speakers are known in relation to a Cartesian system, the position of a point in the system can be derived from the distances to that point.

Of course other layouts are possible. For example a horizontal two dimensional system where the speakers are nodes of a network could work over distances exceeding the maximal 8 metres, as long as the device is within reach of at least two loudspeakers, and the system has a way of recognising the loudspeakers identity.

5.6.1 Tracking a person moving at walking speed

At walking speed of approx. 1.4 m/s (= 1.4 mm/ms) with a sample window of 1024, the update would happen every 23.24 ms, (1024 samples = 23.24 ms at 44.1 kHz) which corresponds to a displacement of 1.4 m/s * 23.24 ms = 32.536 mm . A person moving at this speed in the direction which could cause the biggest change, i.e. moving at 0° or 180° direction to or from the sound source, would cause a delay in the next window of maximally 4.2 samples, as sound needs (0.032536m / 340 m/s) to travel the 32.536 mm, which is 0.0957 ms or 4.2 samples based on:

1 ms = 44.1 samples
0.0957 ms = 44.1 * 0.0957 = 4.2 Samples

So jumps per window larger than 4.2 samples can be considered to be from something or somebody moving faster than a walking person, or, crucially for error filtering, from a reading error for example, if we know only walking people are being tracked in the application.

5.6.2 (Gestural) Tracking at Higher Speeds

For applications which need to trace faster movements, like a dancer, or a game participant wielding a sabre, we'll establish the maximum speed we can expect from a persons gestural movements. The numbers found for a fast cricket bowler have been taken as an appropriate maximum.

A fast bowlers max. hand movement speed is named in (ABC 2012) as 43 m/s and shall be our limit. In the time window of 1024 samples (23.24 ms) he/she moves

d = v*t = 43 * 23.24 ms = 0.99 m

Sound needs (0.999/340 m/s) to travel the 0.999 m, which takes 2.93 ms or samples based on

2.93 ms = 44.1 * 2.93 = 129.5 samples. Conveniently this is just what the sample window of 1024 can cover! Anything moving faster than our bowler will need a shorter window, or a higher sampling rate!

Thus, jumps in delay readings higher than 129.5 samples are definitely not human made, and can, in all cases, be excluded as error.
Moreover, this presumes as well that the bowler managed to accelerate to this speed in the time-span of one window.
From

a = Δv/Δt; 43 /0.02324 = 1850 m/S^2

we know this is not likely to be the case.

So even if a jump of 129 samples over a window is possible, the acceleration would mean that predating windows to the jumping ones must have had increasing values.

Let's take our fast bowler again, and presume 5.5 g, or 53.9 m/s2 .

From

a = Δv/Δt; we get Δt = Δv/a;

so

(0 -> 43)/53.9 = 0.8 s

so the time taken to accelerate to 43 m/s is 0.8 s (800 ms)

From these relations we can assume that not many applications will need to be able to trace accelerations of more than 53.9 m/s2.

The maximum velocity difference between two consecutive windows we can expect are

$$\Delta v = a * \Delta t = 53.9 m/s2 * 0.02324\ s = 1.2526\ m/s.$$

In terms of time delay this means:

at Δv = 1.2526 m/s within one window of 23.24 ms length the object travels d = v*t = 29.1 mm Sound needs t = d/v (0.0291m /340 m/s) to travel the distance, which is 0.855 ms, corresponding to a sample delay of

$$44.1Hz * 0.855\ ms = 37.7\ samples$$

So the increase in jumps over the length of one window can not exceed 37.7 samples if the movement is presumed to be from a person moving.

If the movement is the movement of a piano players finger (gestural control of an instrument) the following things have to be considered: Small movements like finger movements can do fairly high speeds in very short times. For example a piano players finger moves 20 mm in approx. 0.8 ms. that's a speed of 25 m/s. the acceleration would be a = $\Delta v/\Delta t$; 25/0.8 = 31.25 m/s2. (Fast, but well bellow the bowler!)

For the piano player the same calculations as done for the bowler result in the maximum velocity difference between two consecutive windows we can expect to be

$$\Delta v = a * \Delta t = 31.25\ m/s2 * 0.02324\ s = 0.72625\ m/s.$$

In terms of time delay this means:
at Δv = 0.72625 m/s within one window of 23.24 ms length the object travels d = v*t = 16.9 mm Sound needs t = d/v (0.0169m / 340 m/s) to travel the distance, which is 0.497 ms, corresponding to a sample delay of

$$44.1Hz * 0.497\ ms = 21.97\ samples$$

So the increase in jumps can not exceed 21.91 samples over the timespan of a window of 1024 samples.

What's more, it is to note that in these sort of movement the reached speed is not usually being sustained for any length of time.

Obviously, deceleration can happen more abruptly. (The sabre of the above example might hit something) For this initial simulation it was deemed acceptable to ignore this fact in the interest of simplicity.

5.7 The Simulation (SPACE V2 simulation)

As a continuation of the authors attempts at developing a spatial position audio - control environment, (SPACE) the following section describes the development of the SPACE V2 simulation, a UGen++ application which uses the principles discussed above to obtain three dimensional positioning information from correlating delayed signals with the un-delayed original.

The user interface consists of 3 faders for the delay-times representing the three speakers of the layout discussed above. The values shown to the left of the faders are the metric distances the fader values represent.

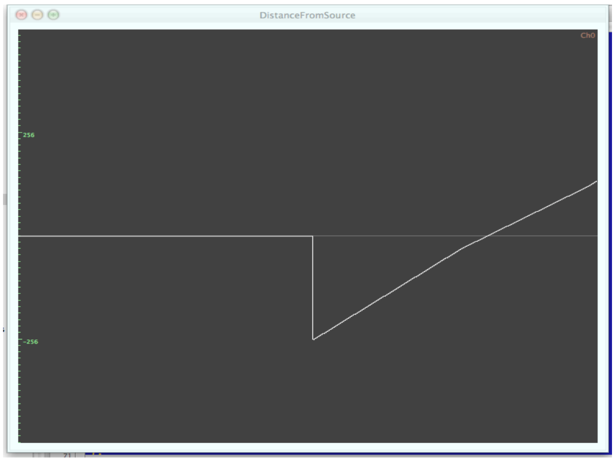
The set of three faders bellow control the amplitudes of the input signals. The three scope components show the values of the correlation between 0 and 1023, as in the signal process. Above the scopes the readings for x, y and z are the calculated values from the correlation. This allows for direct comparison of the measured (output) values with the theoretical (input) values in the display next to the first set of faders.

The 2 faders bellow the scopes control the amount of white noise and a first reflection which can be added to the correlation signal. They are mixed in with the un-delayed signal, to simulate the presence of room noise an application in a real sound-field would encounter.

Next up, some details of the code shall be looked at:

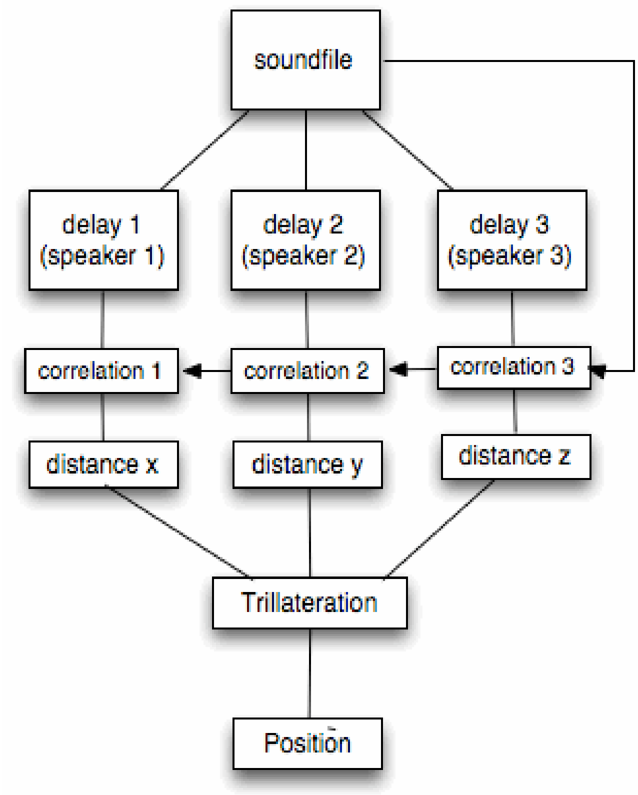
The Apple Mac OSX vDSP_conv function of the Accelerate Framework was used for the correlation.
As the correlation is windowed, the correlation functions signal is somewhat jumpy, but testing it with a square wave resulted - roughly - in the expected sawtooth form of the correlation function.

The error occurring at the window end is clearly visible in Fig 33.



(Fig 33)

The flowchart Fig 34 shows the signal flow from the audio wav file and how the distance is being derived. The filters are left out in this flowchart for clarity and shown separately on flowchart (Fig35). Included in flowchart 1 are (in brackets) the elements which the simulation represents in relation to the real application or prototype.



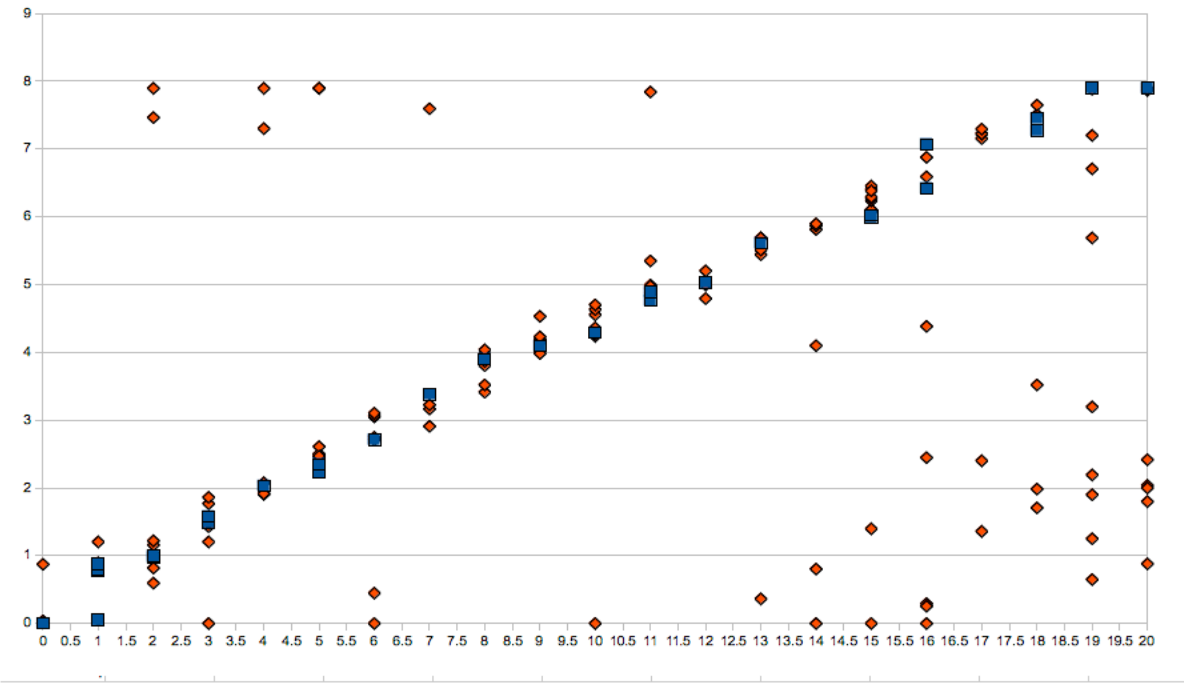
(Fig 34)

For details, the code in Appendix 1 (CD) is commented throughout, but the major steps shall be highlighted here:

In a first step directly after the correlation of the two signals, the sample with the highest value of the corelation has to be picked. This happens initially by storing the index number of the sample with the the highest absolute value. (One maximum per window)

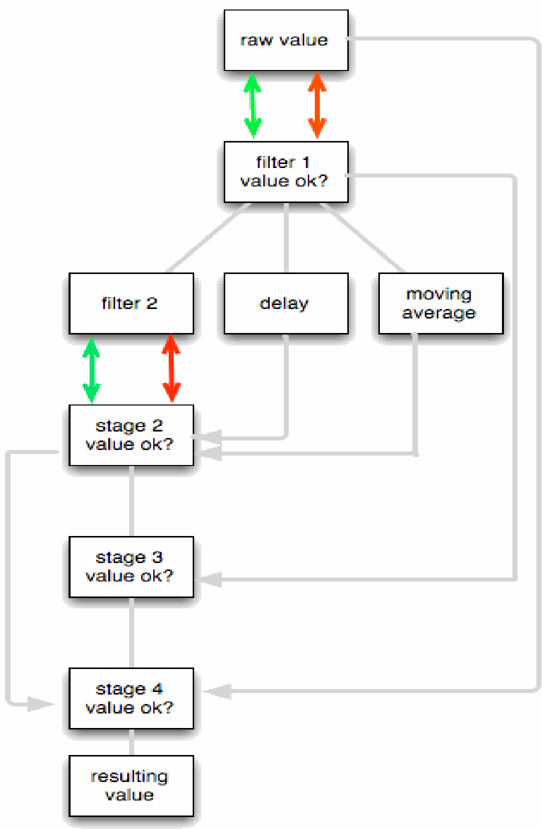
The index numbers are then fed back into a buffer. This buffer on its own, displayed as an audio rate signal on the scopes components show the expected values, but also some random values which are clearly erroneous. (See Illustration RAW) It has not been possible to find out what causes these errors. It should be possible to obtain a better signal to start with: As the signal processing at this stage is entirely digital and the the signal previously known (a .wav file) no error should occur at all.

However, a series of filters have been applied. A first attempt to employ a simple Kalman Filter (Appendix 2) has not provided much improvement, as the errors were clearly not of white noise nature, which is assumed for most straight forward applications of the filter, but outliers produced by some malfunction in the correlation.



(Fig 35)

Moreover, the simple Kalman filter is only effective on linear signals and the implementation of the extended kalman filter was beyond the scope of this project and deemed possibly ineffective due to the above mentioned non - white noise characteristics of the errors observed.



(Fig 36)

The code of the filters is part of Appendix 1 which is commented throughout, and summarised in Fig 36 but the main point are as follows:

In f*ilter 1* all values which are isolated deviations from an established value (i.e. values which have occured at least once already) are ignored and the previous value is used instead. The fact that a correction happened is noted in a negative credit score, as the used value is a presumption not a measurement. (The score is then later used to compare corrective values, as will be seen)

Exempt from *filter 1* are all value - changes which are smaller than the change equivalent to the maximally possible acceleration (*moveMax* in the code), as discussed in section 5.4.1 of this chapter: Changes of this size are possible even if they are isolated jumps, and thus plausible.

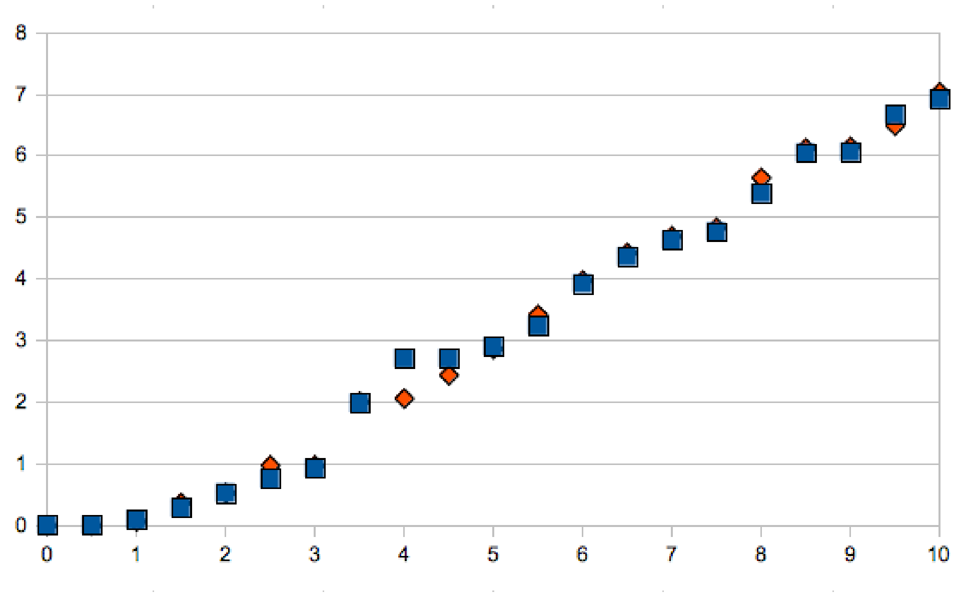
In *flter 2* (gate) initially the same rule is applied, but this time, *all* values exceeding moveMax are ignored. Now this on its own would result in hangers out of which it would be difficult to escape. (*filter 1* can not "hang" as a succession of 2 identical values passes it without activating it. *Filter 1* would only "hang" if a series of isolated value changes which happen to negate each other were to occur. (e.g. 5 10 5 10 5 10 etc.) This pattern was not observed at all and does not seem to cause any issues.)

So in order to use filter 2, exceptions have to be made to decide if a value is trustworthy or a "hanger."
Thus the *moving average* of values from *filter 1* is used to compare the results from the first phase of *filter 2* with. But to ascertain that a value is in fact likely to be a hanger, it is compared to an older value. If the older value (*delayed* in the code) is the same as the most recent one, meaning there was no change for a certain time, the current value *could* be (but doesn't *have to* be) a hanger. Hence in the next step the gate is opened, and the current value compared to the moving average. (Averages are *always* wrong, but always a little right too, so if in doubt, provide an option with possibly more credibility.) In the next two steps the current value is either increased or decreased respectively until the value of the *moving average* is being reached.

The *moving average* is not trusted blindly though: If the moving average has shown erratic behaviour between 2 steps, its credit score is low. At the point where the moving average is used as a reference fro the current value its credit score is thus being checked. if it is low, the value from *filter 1* is considered and If f*ilter 1*'s credit score is positive, the value from *filter 1* is being used instead as the new current value.
If both the credit scores of filter 1 and the moving average are negative, none of the available values is reliable. In that case the raw value is chosen instead, as a last option, allowing some fresh information into the system.

This results in fairly stable reading values which follow steady movements fairly accurately. The time lag however is notable and gestural control would need either faster filters or a more accurate signal to start with.

Tests showed that an object moving through the full range of 7.9 meters over 10 seconds results in an accuracy still bellow half meter values, but better accuracy can be achieved in slower movements. (Fig 37)



(Fig 37)

The time lag for movements bellow moveMax is around 2 - 3 seconds, but the escape from hangers take considerably longer. It is possible though, to jump to any value and let the system catch up, and eventually it does: In that sense the values are rarely wrong but not often on time.

However, the accuracy of the simulation lies above the theoretically possible values, which should be in the sub-centimetre range. On the other hand the values lie within the margins expected by users and developers according to the survey in chapter 4, at bellow 0.5 metres.

As discussed in the section on correlation, the correlation of periodic signals like sine waves for example is more difficult to spot and the observed errors in the here implementation of vDSP_conv are enough to throw these values enough to just anecdotal accuracy. However, other non-musical signals like square waves and pseudo random noise resulted in readings as accurate as for musical signals, the latter, conveniently, being the ones the application is aimed for.

What is promising is the results of the simulation under the influence of noise and first reflections. The UGen++ Application features a means to introduce noise and an extra delayed signal to the system via the 2 sliders on the bottom of the display. The effect on the resulting position is fairly minimal, for white noise Reflection and both in white noise *and* reflection.

It seems that the combined error of late reflection and white noise has the worst influence.
But even under the combined negative influence of first reflection and white noise (at roughly equal RMS power as the signal), the accuracy stays well bellow the half meter mark.

As far as latency is concerned, from Fig 37 one can make out that preferred walking speed (1.4m/s) is roughly the limit before the latency increases. The values per se are not necessarily wrong though, but they are delayed. (they show where the device just was, rather than where it is)

As the nature of the applied filters are rather crude, more sophisticated stochastic designs which work on *predictions* rather than comparison to old values, could certainly improve the response time of the system.

In summary, three approaches, one using correlation, the other using binaural technology and a third one employing ambisonics have been considered as possible candidates for development of a simulation of an I/LPS for iLAA and the first system, which correlates a known signal with its delayed spatially distributed copy to calculate the distance to the sound source was chosen. The systems possible technical limits were established to be of sub-centimetre accuracy nased on a sampling rate of 44.1 kHz and correlation of a window size of 1024 samples. Considerations as to various possible layouts were discussed and a layout of 3 speakers in known positions chosen as they provide a simple three dimensional volume where positions of a mobile receiving device (a microphone) can be calculated by Euclidean between two points in a Cartesian co-ordinate system.

Further limitations and requirements were discussed based on the required speeds for various applications in particular the maximum possible movement to be effected by a human being (around 5 times gravitational force) but in consideration of the necessary acceleration to achieve such speeds. For walking humans, a likely object to be tracked in an i/LAA and at the rather moderate speed of 1.4 m/s was assumed to be a lower limit to consider. Faster speeds as do occur in interactive gaming action, could *theoretically* still easily be traced, though.

The software implementation of a simulation of SPACE, the authors spatial position audio control environment, showed that some of the requirements can be met, for example accuracy bellow half meter errors, and fairly stable behaviour under white noise and first reflection - multipath issues.

However, the response times necessary for fast interactive gaming action could not be achieved in the simulation model. It was notable too that the digital-only nature of the simulation should provide no-error correlation as all signals are known. The reasons for the errors which occurred none the less could not be established and are awaiting further research.

5.8 Experiments With Signal Through Air and Further Research

In addition to the simulation, a prototypical application was developed which was devised to work with real through-air audio signals. However, The results were very poor and deemed insufficient to warrant recording beyond consideration for future research. Possible reasons for this shall be discussed here, but in view of the promising results of the simulation it is very much the point that further research *is indeed* necessary, and to be welcomed, and no conclusive answer has been found why the prototype is not working anywhere near as well as the simulation. However, the following paragraphs are not unlikely to provide some explanations for both the possibilities of what is to come and the limitations of the existing prototype.

Both the simulation SPACE_simulation and the prototype SPACE V2 application were developed for Apple Mac OSX *Leopard* and the tests were run on a laptop running OSX 10.5.8 (1.67 GHz PPC G4 with 2GB DDR2 SDRAM) This hardware was deemed sufficient for the simulation, however, there are indications that the next step, a realisation of an actual prototype would have been possible on a newer processor. Some tests towards an actual prototype have been conducted, and a prototype version which might or might not work on newer hardware is appended to this dissertation. (appendix IV) But no recognisable correlation could be achieved between the signal through the air and the signal on the processor using the available hardware: Glitches in the playback over speakers suggest that the synchrony between playback and processor is not sufficient. To overcome this, further tests using 2 microphones rather, (one at the position of the speaker, one on the device) were run. But still no actual correlation could be established.

The differences between the simulation and the prototype were evident. In fact, to obtain the level of non-correlation observed in the prototype experiments, the amount of white noise and first reflection which had to be added to the musical signal, made it near un-distinguishable to the ear.

It is assumed that some hardware error or error other than what could be simulated is responsible for the non correlation of the signals observed in the prototype.

Instead of running further experiments with the current prototype it was thus considered more beneficial to discuss the potential shown by the simulation. And in consideration that a more accurate correlation can probably be achieved, further experiments are necessary and shall possibly benefit from newer hardware. Or at least, with tests on newer hardware, and in controlled laboratory surroundings (dead rooms etc...) various sources of error could be eliminated.

Beside the obvious limitations and the need to solve these issues, there are several fields of interest which could be addressed once a working prototype is available:

It would be interesting for example, to see how speech signals correlate and to what accuracy positioning would be possible there. Furherl, the values in the simulation show that the signal does not have to be particularly loud to correlate. More experiments with correlation of signals of low amplitude would be interesting.

The problem of multipath, despite its notable absence in the results when it was purposely introduced to the simulation, could probably be further addressed by taking impulse responses of points of interest, i.e. at the speakers and at the device, and done in a separate step to the correlation and in the knowledge of the current time lag. (It would not be possible to do this in the same step, as the de-correlation of the signal with the IP would lose the time lag information. But as a separate step it could be beneficial.)

What is more, in the current simulation application, the resulting values are displayed separately. A further filter could be readily devised using all three values in relation with each other. The trilateration of a moving device follows rules based on the geometry of the system which can be used for further filters for error reduction: It is impossible for example, that all 3 axis values get shorter or longer at the same time within the observed quadrant of the Cartesian system, least the device explodes or implodes, so some further predictions will be possible based on the sum of the distances etc. At that stage, Extended Kalman Filters would of course help considerably.

6. Conclusive Remarks

This dissertation's aim was to establish user requirements for the development of I/LPS for iLAA. It has looked at various aspects which might influence these requirements. Tangible parameters and both their maximal and minimal limits where defined from various angles, namely from the nature of possible and existing applications' scenarios, from the limitation of our ability to locate sound and by directly asking developers what they expected from the relevant existing and evolving technologies.

Early adapters and developers input through the survey showed that for global applications the existing GPS technology is fairly adequate, but WLAN - based systems do not seem to be able to provide the improvement necessary to fulfil the needs in more local scenarios, and from chapter 3 we know that radio frequency systems relying on RSSI are notoriously temperamental.

From these various sources and fields of research it became evident that not many systems are able to provide the necessary accuracy and if they do, they come with requirements which might not lend themselves naturally to the application in question. From the user requirements survey in chapter 4 we know that the developers community is quite clear on the need for unobtrusive technology, in a physical sense and in context with privacy issues.

The technologies described in chapter 3 which fulfil the requirements' needs on accuracy, like some of the optical methods, depend on LOS, which for audio applications seems counter intuitive. From this angle, and based on the discussed background, acoustic systems, using audible sound which might form part of an applications content, were identified to conceptually and technically fulfil all needs, as well in accuracy, potential ubiquitousness, unobtrusiveness, and cost: the suggested systems would rely on the infrastructure in place for the audio aspect of the application. A simulation in form of a computer model was thus devised, and some encouraging results were obtained. The results were not as good as the theoretically possible outcome, but they showed clearly the possibilities of audible audio-only position technology for iLAA and clearly showed development in this direction to be worthwhile of further research.

7. References

ABC Forum (2003) "*How fast can the human hand move*" Available from: URL http://www2b.abc.net.au/science/k2/stn/archives/archive51/newposts/324/topic324975.shtm [Accessed 02 May 2012]

ACM (2002) "Association of Computer Machinery Computing Classification System" in *IEEE Computer Society - Applications* IEEE Computer Society [online] Available from: http://www.computer.org/portal/web/publications/acmapplications#J9 [Accessed 02.03.2012]

Al Nuaimi, K.; Kamel, H.; (2011) , "A survey of indoor positioning systems and algorithms," I*nnovations in Information Technology (IIT*), 2011 International Conference on , vol., no., pp.185-190, 25-27 April 2011
doi: 10.1109/INNOVATIONS.2011.5893813
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5893813&isnumber=5893793

Barsocchi, P.; Furfari, F.; Nepa, P.; Potorti, F.; , "RSSI localisation with sensors placed on the user," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-6, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5647868
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647868&isnumber=5645818

Bates, E.; Furlong, D.; Dennehy, D., (2008) "Adapting Polyphonic Pickup Technology for Spatial Music Performance.pdf" *International Computer Music Conference, S.A.R.C. Belfast, August 2008* [online] Avavilable from: http://www.endabates.net/EndaBatesICMC08.pdf [Accessed 16 April 2012]

- Enda Bates, Dermot Furlong and Donnacha Dennehy

- International Computer Music Conference, S.A.R.C. Belfast, August 2008.

Bosisio, A.V.; , "Performances of an RSSI-based positioning and tracking algorithm," Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on , vol., no., pp.1-8, 21-23 Sept. 2011
doi: 10.1109/IPIN.2011.6071952
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6071952&isnumber=6071904

Brown, D.R.; Dunn, D.B.; (2011) "Classification schemes of positioning technologies for indoor navigation," *Southeastcon, 2011 Proceedings of IEEE* , vol., no., pp.125-130, 17-20 March 2011
doi: 10.1109/SECON.2011.5752919
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5752919&isnumber=5752891

Brown, B., MacColl, I., Chalmers, M. , Galani, A., Randell, C., and Steed, A. (2003) "Lessons from the lighthouse: collaboration in a shared mixed reality system." In: *Conference on Human Factors in Computing Systems, 5-10 April 2003*, Fort Lauderdale, Florida. [online] Available from: http://eprints.gla.ac.uk/3428/1/CHIlighthousepaper.pdf [Accessed 16 April 2012]

Björk, S.; Holopainen, J. (2004). *Patterns In Game Design*. Newton Center: Charles River Media. pp. 423

Blauert, J. (1997) *Spatial Hearing- the Psychophysics of Human Sound Localization* Revised Edition. Massachusetts: MIT.

Blesser, B & Salter, L. R. (2007) *Spaces Speak, Are You Listening?: Experiencing Aural Architecture* Massachusetts: MIT.

Bluetooth (2012) *Core Specifications* website. [online] Available from: [https://www.bluetooth.org/Building/HowTechnologyWorks/CoreSpecifications.htm] [Accessed 04 April 2012]

Churnside, A.; Pike, C.; Leonard, M., "Musical Movements - Gesture Based Audio Interfaces" *AES Convention:131 (October 2011) Paper Number:8496* [online] available from http://www.aes.org/e-lib/browse.cfm?elib=16022 [Accessed on 16 April 2012]

ComputerMusicBlog (2012) *Stockhausen* [online] Available from: http://computermusicblog.com/blog/wp-content/uploads/2009/06/stockhausen.jpg [Accessed 02 May 2012]

Collins, K. (2007) An introduction to the participatory and non liner aspects of video games audio in Richardson, J. & Hawkins, S. (eds) *Essays on Sound and Vision* Helsinki: Helsinki University Press. pp. 263 298

Collins, K. (ed) 2008 *From Pac-Man to Pop Music: Interactive Audio in Games and New Media* Aldershot: Ashgate.

Creasey, D. (2011) *Audio Processing*, *UFIEUA-40-2* Book 2, 2010-2011 Version, Bristol: University of the West of England

Creative Skill Set (2012) *Who works in computer games?* Available from: http://www.creativeskillset.org/games/industry/article_2194_1.asp [Accessed 12 April 2012]

D sounds (2012) *Karlheinz Stockhausen: STERNKLANG* website. [online] Available from http://dsounds.netzwerkneuemusik.de/ausgabe-3/stockhausen-sternklang.html [accessed 01 April 2012]

Daniel et al. (2003) *AES 24th international conference proceedings* [online] available from: http://pcfarina.eng.unipr.it/Public/AES-114/00048.pdf [Accessed 01. February 2012]

De Prado, G.; Feijóo, C;. Nepelski, D.; Bogdanowicz, M.; Simon, J.P.; (2010) *Born Digital / Grown Digital: Assessing the Future Competitiveness of the EU Video Games Software Industry* [online] EUR 24555 EN - 2010 JRC Scientific and Technical Reports. Available from http://ftp.jrc.es/EURdoc/JRC60711.pdf [Accessed 12 April 2012]

Della Rosa, F.; Leppäkoski, H.; Biancullo, S.; Nurmi, J.; , "Ad-hoc networks aiding indoor calibrations of heterogeneous devices for Fingerprinting applications," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-6, 15-17 Sept. 2010 [online]
doi: 10.1109/IPIN.2010.5647827
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647827&isnumber=5645818
[Accessed 04/04/2012]

ETH (Eidgenössische Technische Hochschule Zürich) (2010) "*2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*" Available from: http://www.geometh.ethz.ch/ipin/ [Accessed 31 December 2011]

Feldmeier, M. & Paradiso, J. A. (2007) An Interactive Music Environment for Large Groups with Giveaway Wireless Motion Sensors *Computer Music Journal*, Volume 31, Number 1, Spring 2007, pp. 50-67 (Article)

Flew, T. & Humphreys, S. (2005) "*Games: Technology, Industry, Culture* Oxford University Press, South Melbourne (second edition), South Melbourne: Oxford University Press.

Filonenko, V.; Cullen, C.; Carswell, J.; (2010), "Investigating ultrasonic positioning on mobile phones," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-8, 15-17 Sept. 2010 IEEE Xplore[online].
doi: 10.1109/IPIN.2010.5648235

URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5648235&isnumber=5645818
[Accessed 31 December 2011]

Furse, R. (2011) *First and Second Order Ambisonic Decoding Equations* [online].
Available from [http://www.muse.demon.co.uk/ref/speakers.html] Accessed [23 March 2011]

Gansemer, S.; Grossmann, U.; Hakobyan, S.; , "RSSI-based Euclidean Distance algorithm for indoor positioning adapted for the use in dynamically changing WLAN environments and multi-level buildings," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-6, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5648247
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5648247&isnumber=5645818

Garity, WM. E. (1941), *Fantasound* [online] Journal of the Society of Motion Picture Engineers.
available from [http://www.widescreenmuseum.com/sound/fantasound1.htm] accessed [04/04/2011]

Global Positioning System. (2012) Wikipedia [online]. 2 April. Available from:
http://en.wikipedia.org/wiki/Global_Positioning_System [Accessed 02 April, 2012].

Hansen, R.; Wind, R.; Jensen, C.S.; Thomsen, B.; (2010), "Algorithmic strategies for adapting to environmental changes in 802.11 location fingerprinting," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-10, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5648270
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5648270&isnumber=5645818

Herrmann, R.; Sachs, J.; Bonitz, F.; , "On benefits and challenges of person localization using ultra-wideband sensors," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-7, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5647896
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647896&isnumber=5645818

Holmes, T. (2008) *Electronic and experimental music: technology, music, and culture* New York: Routledge.

IPIN 2011 (2011) "*2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*" Available from: http://ipin2011.dsi.uminho.pt/
[Accessed 31 December 2011]

Janson, T.; Schindelhauer, C.; Wendeberg, J.; , "Self-localization application for iPhone using only ambient sound signals," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-10, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5648094
[online] available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5648094&isnumber=5645818
[Accessed 29 February 2012]

Judd, F. C. (1961) *Electronic Music and Music Concrète* London: Neville Spearman

Kaemarungsi, K. and Krishnamurthy, P. (2004) "Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting" *Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04)* [online] Available from [http://edge.cs.drexel.edu/regli/Classes/CS680/Papers/Localization/01331706.pdf] [Accessed 04/04/2012]

Kaseva, V.; Hämäläinen, T.D.; Hännikäinen, M.; (2010), "Positioning in resource-constrained ultra low-power Wireless Sensor Networks," *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS)*, 2010 , vol., no., pp.1-11, 14-15 Oct. 2010
doi: 10.1109/UPINLBS.2010.5654345
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?

tp=&arnumber=5654345&isnumber=5648783

Kietlinski-Zaleski, J.; Yamazato, T.; , "UWB positioning using known indoor features - environment comparison," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-9, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5648089
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5648089&isnumber=5645818

Kiss, J. (2010) *Papa Sangre: The sonic iPhone horror game you've been looking for* guardian.co.uk [online] 20 December. Available from:
http://www.guardian.co.uk/technology/pda/2010/dec/20/papa-sangre-game-audio [Accessed 31 January 2012]

Kroger, T.; Yuwei Chen; Ling Pei; Tenhunen, T.; Kuusniemi, H.; Ruizhi Chen; Wei Chen; (2010), "Method of pedestrian dead reckoning using speed recognition," Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010 , vol., no., pp.1-8, 14-15 Oct. 2010
doi: 10.1109/UPINLBS.2010.5653994 [online] Available from
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5653994&isnumber=5648783
[Accessed 04 April 2011]

Larranaga, J.; Muguira, L.; Lopez-Garde, J.-M.; Vazquez, J.-I.; , "An environment adaptive ZigBee-based indoor positioning algorithm," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-8, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5647828
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647828&isnumber=5645818

Leppakoski, H.; Tikkinen, S.; Takala, J.; , "Optimizing radio map for WLAN fingerprinting," *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010* , vol., no., pp.1-8, 14-15 Oct. 2010 [online]
doi: 10.1109/UPINLBS.2010.5654332 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5654332&isnumber=5648783 [Accessed 04/04/2012]

Liu, H.; Darabi, H.; Banerjee, P.; Jing Liu; (2007) , "Survey of Wireless Indoor Positioning Techniques and Systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* , vol.37, no.6, pp.1067-1080, Nov. 2007
doi: 10.1109/TSMCC.2007.905750
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4343996&isnumber=4343957

Lukianto, C.; Hönniger, C.; Sternberg, H. (2010); , "Pedestrian smartphone-based indoor navigation using ultra portable sensory equipment," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-5, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5646697 [online] Available from:
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5646697&isnumber=5645818
[Accessed 04 April 2011]

Malham, D.G. (1998) *Spatial Hearing Mechanisms and Sound Reproduction* [online] available from [http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm] [Accessed 04 April 2011]

Malham, D.G. (2000) *Homogenous and Nonohomogenous Surround Sound Systems* [online] Available from : http://www.york.ac.uk/inst/mustech/3d_audio/homogeneous.htm#N_3_ [Accessed 02. 03. 2012]

Mandal, A.; Lopes, C.V.; Givargis, T.; Haghighat, A.; Jurdak, R.; Baldi, P.; (2005), "Beep: 3D indoor positioning using audible sound," Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE , vol., no., pp. 348- 353, 3-6 Jan. 2005
doi: 10.1109/CCNC.2005.1405195 [online] Available from
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1405195&isnumber=30469
[Accessed, 05 April 2012]

Manning, P. (2004) *Electronic and Computer Music* Revised and Expanded Edition, New York: Oxford University Press

Mautz, R.; Tilch, S.; (2011) , "Survey of optical indoor positioning systems," *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on* , vol., no., pp.1-7, 21-23 Sept. 2011 doi: 10.1109/IPIN.2011.6071925
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6071925&isnumber=6071904

Mautz, R., Kunz, M. and Ingensand, H. (eds.)(2010) *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, [online] available from [http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?reload=true&punumber=5637226] [Accessed 23 March 2011]

Mautz, R.; Tilch, S.(2011) , "Survey of optical indoor positioning systems," *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on* , vol., no., pp.1-7, 21-23 Sept. 2011 doi: 10.1109/IPIN.2011.6071925 [online]
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6071925&isnumber=6071904 [Accessed 04 April 2012]

Morales-Manzanares, R; Morales, E.F.; Dannenberg, R.; Berger, J. (2001) SICIB: An Interactive Music Composition System Using Body Movements, *Computer Music Journal*, Volume 25, Number 2, Summer 2001, pp. 25-36 (Article)

Malham, D. (2009) *Sound in Space* [online]. Available from [http://www.york.ac.uk/inst/mustech/3d_audio/welcome.html] [Accessed 23 March 2011]

Martin (2006) *Why Ambisonics Can Not Work* [online]. Available from [http://www.tonmeister.ca/main/textbook/node826.html] Accessed [23 March 2011]

Moore, B. C. J (2008) *Psychology of Hearing* 5th Edition Bingley: Emerald.

Noisternig et al (2003) Ambisonic based Binaural Sound Reproduction *AES 24th international conference proceedings* [online] available from: http://www.aes.org/e-lib/browse.cfm?elib=12314 [Accessed 01. February 2012]

Oellers, H. (2011) *Die Kopie des räumlichen Schallfeldes* [online] available from [http://www.syntheticwave.de/WFS-Holophony_ger.htm] [Accessed 04/04/2011]

Ogunjemilua , K., Davies, J.N., Grout, V. & Picking, R., (2009) “An Investigation into Signal Strength of 802.11n WLAN”, *Proceedings of the Fifth Collaborative Research Symposium on Security, E-Learning, Internet and Networking (SEIN 2009*), Darmstadt, Germany, 26-27 November 2009, pp191-204. [online] Available from http://www.glyndwr.ac.uk/computing/cards/pubs/SEIN_ODGP.pdf [Accessed 04 April 2012]

Ouzounian, Gascia. (2008). *Sound art and spatial practices : situating sound installation art since 1958*. UC San Diego: b6636093. [online] Available from: http://escholarship.org/uc/item/4d50k2fp [Accessed 16 April 2012]

Packi, F.; Beutler, F.; Hanebeck, U.D.; (2010), "*Wireless acoustic tracking for extended range telepresence,*" *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-9, 15-17 Sept. 2010 IEEE Xplore[online].
doi: 10.1109/IPIN.2010.5647416 Available from
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647416&isnumber=5645818 [Accessed 28 December 2011]

Pervasive Media Studio (2012) *TangeableFX* [online] available from: http://www.pmstudio.co.uk/collaborator/tangiblefx [Accessed 01. February 2012]

Peters, N.; Marentakis, G.; McAdams, S (2011) *Current Technologies and Compositional Practices for Spatialization : A Qualitative and Quantitative Analysis Computer Music Journal* Volume 35, Number 1, Spring 2011 pp. 10-27

Piras, M.; Cina, A.; , "Indoor positioning using low cost GPS receivers: Tests and statistical analyses," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-7, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5647559
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647559&isnumber=5645818
[Accessed 11 April 2012]

Potts, C.; Takahashi, K.; Anton, A.I.; , "*Inquiry-based requirements analysis*," Software, IEEE , vol.11, no.2, pp.21-32, Mar 1994 [online]
doi: 10.1109/52.268952
Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=268952&isnumber=670] [Accessed 11 April 2012]

Purkayastha, S.N.; Eckenstein, N.; Byrne, M.D.; O'Malley, M.K (2010) , "Analysis and comparison of low cost gaming controllers for motion analysis," *Advanced Intelligent Mechatronics (AIM), 2010 IEEE/ASME International Conference on* , vol., no., pp.353-360, 6-9 July 2010
doi: 10.1109/AIM.2010.5695830 [online] Available from:
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5695830&isnumber=5695712
[Accessed 04 April 2012]

Randall, R.B. (1987) *Frequency Analysis* 3rd Edition, first print.Brüel & Kjaer, Naerum

Randell, C., 2012 *Cliff Randell*, [online] Available from http://www.cs.bris.ac.uk/~cliff/
[Accessed 11 April 2012]

Roads, C. (1996) *Computer Music Tutorial* Massachusetts: MIT.

Robert, C.; Tomé, P.; Botteron, C.; Farine, P.; Merz, R.; Blatter, A.; , "Low power ASIC transmitter for UWB-IR radio communication and positioning," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-7, 15-17 Sept. 2010
doi: 10.1109/IPIN.2010.5647856
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647856&isnumber=5645818
[Accessed 11 April 2012]

Robson, C. (1993) *Real World Research* Oxford: Blackwell.

Rizos, C.; Roberts, G.; Barnes, J.; Gambale, N.; (2010), "Experimental results of Locata: A high accuracy indoor positioning system," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-7, 15-17 Sept. 2010 IEEE Xplore[online].
doi: 10.1109/IPIN.2010.5647717
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647717&isnumber=5645818
[Accessed 31 December 2011]

Segou, O.E.; Mitilineos, S.A.; Thomopoulos, S.C.A.; , "DALE: A range-free, adaptive indoor localization method enhanced by limited fingerprinting," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-8, 15-17 Sept. 2010 [online]
doi: 10.1109/IPIN.2010.5647933
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647933&isnumber=5645818
[Accessed 04/04/2012]

Schlienger, D. E. (2011) *MOASpan User Manual* Dom Schlienger: Creative Music Technology [online] Available from: http://creativemusictechnology.org/MOASpanManual.pdf [Accessed 02.03.2012]

Siwek, S. (2012) *Video Games in the 21st Century, the 2010 report* [online] Available from: http://www.theesa.com/facts/pdfs/VideoGames21stCentury_2010.pdf [Accessed 12 April

2012]

Smisek, J.; Jancosek, M.; Pajdla, T.(2011) , "3D with Kinect," *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* , vol., no., pp.1154-1160, 6-13 Nov. 2011 doi: 10.1109/ICCVW.2011.6130380 [online] Available from:
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6130380&isnumber=6130192
[Accessed on 04 April 2012]

Sonhors 2012 *Spatialisation et image : la mise en espace des sons concrets et électroniques* Sonhors [online] Available from: http://sonhors.free.fr/panorama/sonhors13.htm [Accesed 31 January 2012]

Space Collective (2012) [online] *Osaka Dome by Stockhausen* Available from: http://spacecollective.org/soCinematic/6726/Spatial-Sound-Stockhausen-to-Stanford
[Accessed 02 May 2012]

Stockhausen (2012) *IntroductionHELICOPTER STRING QUARTET (1992/93*) [online] Available from: http://www.stockhausen.org/helicopter_intro.html [Accessed 29. 02. 2012]

Szinger, J. (1993) [online] *On Composing Interactive Music* Available from: http://zingman.com/spew/CompIntMusic.php [Accessed 16 April 2012]

Tas, B.; Altiparmak, N.; Tosun, A.S.; (2009), "Low cost indoor location management system using infrared leds and Wii Remote Controller," Performance Computing and Communications Conference (IPCCC), 2009 IEEE 28th International , vol., no., pp.280-288, 14-16 Dec. 2009 [online] doi: 10.1109/PCCC.2009.5403833 Available from:
[URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5403833&isnumber=5403797]
[Accessed 04 April 2012]

Trifa Vlad M. Trifa, V. M.; Koene, A.; Mor, J.; Cheng, G.;"Real-time acoustic source localization in noisy environments for human-robot multimodal interaction" *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on* [online] Available from http://www.vs.inf.ethz.ch/publ/papers/trifam_07_roman.pdf [Accessed 16 April 2012]

Tuters, M. (2008) "*The Locative Utopia*" [online] Available from:
http://www.rixc.lv/reader/txt/txt.php?id=186&l=en&raw=1 [Accessed 16 April, 2012]

Uchitomi, N.; Inada, A.; Fujimoto, M.; Wada, T.; Mutsuura, K.; Okada, H.; (2010), "Accurate indoor position estimation by Swift-Communication Range Recognition (S-CRR) method in passive RFID systems," Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on , vol., no., pp.1-7, 15-17 Sept. 2010 [online]
doi: 10.1109/IPIN.2010.5646819
Available from: URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5646819&isnumber=5645818 [Accessed 04. April 2012]

UPINLBS (Ubiquituos Positioniong Indoors Navigation and Location Based Services) (2010) "*The International Conference on Ubiquitous Positioning, Indoor Navigation and Location-Based Service UPINLBS 2010*" Available from: http://217.152.180.26/upinlbs/?q=node/91
[Accessed 31 December 2011]

Wan, E.A.; Paul, A.S.; (2010) , "A tag-free solution to unobtrusive indoor tracking using wall-mounted ultrasonic transducers," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-10, 15-17 Sept. 2010 IEEE Xplore[online].
doi: 10.1109/IPIN.2010.5648178
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5648178&isnumber=5645818

[Accessed 28 December 2011]

Wei, B.; Chen, W.; Ding, X.; (2010), "Advanced MDS based localization algorithm for location based services in Wireless Sensor Network," *Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS)*, 2010 , vol., no., pp.1-8, 14-15 Oct. 2010
doi: 10.1109/UPINLBS.2010.5654301
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5654301&isnumber=5648783 [Accessed 04/04/2012]

Weiser, M. (1991)"*The Computer for the 21st Century*" - Scientific American Special Issue on Communications, Computers, and Networks, September, 1991 [online] Available from: http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html [Accessed 29. 02. 2012]

Wendeberg, J.; Hoflinger, F.; Schindelhauer, C.; Reindl, L.; , "Anchor-free TDOA self-localization," *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on* , vol., no., pp.1-10, 21-23 Sept. 2011
doi: 10.1109/IPIN.2011.6071909 [online] Available from
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6071909&isnumber=6071904
[Accessed 12 April 2012]

Wersényi, G. (2003) "Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications" *AES 115th Convention, New York, New York, 2003 October 10-13* [online] available from http://www.aes.org/e-lib/browse.cfm?elib=12408 [Accessed 16 April 2012]

Wijnans, S. (2009) A Choreography of a Spatial Sonic Disembodiment in Motje Wolf (Ed.) *Proceedings of Sound, Sight, Space and Play 2009* Postgraduate Symposium for the Creative Sonic Arts De Montfort University Leicester, United Kingdom, 6-8 May 2009
Available from [http://www.mti.dmu.ac.uk/events-conferences/sssp2009/proceedings/01_SSSP09_Wijnans.pdf]

Wikipedia 2012b *Ambisonics* [online] available from: http://en.wikipedia.org/wiki/Ambisonics [Accessed 01. February 2012]

Wikipedia 2012a [online] available from: http://en.wikipedia.org/wiki/Venetian_polychoral_style [Accessed 01. February 2012]

Wii Remote. (2012) Wikipedia [online]. 2 April. Available from: http://en.wikipedia.org/wiki/Wii_Remote [Accessed 04 April, 2012].

Wiibrew, 2012 [online] available from: http://wiibrew.org/wiki/Main_Page [Accessed on 04 April 2012]

Wikipedia, (2012)*File:Henry260.jpg* [online] available from: http://en.wikipedia.org/wiki/File:Henry260.jpg [Accessed 02 May 2012]

Yildiz, A.; Akay, A.; Akgul, Y.S.; , "Wii Remote calibration using the sensor bar," *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011* IEEE Computer Society Conference on , vol., no., pp.7-12, 20-25 June 2011 [online]
doi: 10.1109/CVPRW.2011.5981683 Available from:
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5981683&isnumber=5981671
[Accessed 04 April 2012]

Zebra (2012)*Dart UWB Technology* [online] Available from: [http://www.zebra.com/id/zebra/na/en/documentlibrary/product_brochures/dart_uwb_technology.File.tmp/ds_dart_UWB_tech.pdf] [Accessed 10 April 2012)

Zetik, R.; Guowei Shen; Thomä, R.S.; (2010) , "Evaluation of requirements for UWB localization systems in home-entertainment applications," *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* , vol., no., pp.1-8, 15-17 Sept. 2010 [online]
doi: 10.1109/IPIN.2010.5647483
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5647483&isnumber=5645818
[Accessed 04/04/2012]

Zvonar, R. (2005) *A history of Spatial Music* [online] available from http://cec.concordia.ca/econtact/Multichannel/spatial_music.html [Accessed 04/04/2011]