Call to composers, musicians and multimedia artists to participate in a survey establishing user requirements on local positioning systems for spatially interactive applications.

Abstract

Integrating spatiality into music is not new, and historical examples are many, from the cori spezzati of the 17th Century, to marching bands or Schönbergs Helicopter String Qaurtet. Recent developments in (digital) technology in indoors and local positioning systems (I/LPS) allow for spatiality to become an interactive aspect of multimedia applications (the Nintendo wii being an example of the gaming world) and open new creative possibilities to spatial composers and multimedia artists and game developers alike. Be it to enhance immersion in video games, or as a means to track the movement of a dancer to control audio content, or interactive spatial compositions or installations, all these applications require location awareness in one way or other. Despite the fact that there are many local positioning systems using various technologies, (RSSI, WiFi fingerprinting, video tracking, etc.) there is very little data available to the engineering world as to the requirements of these systems for creative multimedia applications. As part of my MSc Thesis in Audio Production at UWE Bristol, UK, which researches the availability and suitability of I/LPS for interactive audio applications The author is calling upon composers and musicians and multimedia artists to participate in a survey to establish these user requirements.

1. Introduction

The human ear is able to locate a sound (to its apparent location, see below and in detail 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 (Wikipedia 2011a) 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 GNSS, (Global navigation satellite systems) Indoors and Local Positioning Systems (I/LPS) like motion tracking, radio signal strength triangulation, pseudolites, gyro meters on smart phones, etc. present a vast array of possibilities. (Mautz et al.) However, so far not many positioning system presently applied to audio applications could be identified which are able to provide accuracy within the margins of the human ear's ability to locate sound.

In practice this has not hindered developers and musicians to come up with a plethora of creative solutions using existing technology but the literature suggests that the limitations of available I/LPSI/ LPS, tend to transfer themselves to the creative possibilities of many applications, particularly when cost is an issue.

But in order to develop positioning systems which lend themselves more positively to these creative applications more research into the user requirements seems to be necessary, as a clear gap in the literature has been established. This was further underpinned by the findings of a focus group at Pervasive Media Studio Bristol (PMS) earlier this year which the author held in conjunction with PMS.

This focus group was a preliminary forum to formulate questions in discussion with professionals working with audio and locative media as to what positioning systems should be able to do and what limitations are acceptable. These questions are now consequently raised as part of an online survey into user requirements on positioning systems for Locative Audio Applications (LAA) and posed to a wider community of interested musicians, developers and academics.

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 practice was very popular in the late Renaissance and early Baroque eras. (Wikipedia 2012c) This polychoral style was appreciated for its spatial effects.

The spatial nature of a symphonic orchestra can be experienced in the many excellent auditoriums which have been built to accommodate this particular sound, aural architecture being another historic representation of the spatial quality of near all types of music. (Blesser and Salter, 2007)

Over the last hundred years the way we experience audio content has dramatically changed. (Roads,1996) 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.

In the 20th century 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 (Zvonar, 2005).

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. Poullin and Schaefer 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. (Zvonar 2005, Sonhors, 2012)

The Philips Pavilion at the 1958 world expo in Brussel was designed by Iannis Xenakis who worked under 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.

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, inluding from below a sound-permeable floor.

This was not Stockhausens first or last work of spatial music, An early 1970 spatial Stockhausen composition is *Sternklang* Where five groups of musicians are spread out in a park. The listener moves about in the park between the groups.

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.)" (Stockhausen 2012) 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 and the researchers at the *Groupe de Recherche de Musique Concrète* produced technologically aided spatial music with entirely analogue means and one can only wonder what works they would have written if they had the digital instrumentarium of today at their disposal.

Further technical developments of spatial sound systems resulted in ambisonics, based on the work by Blumlein (Malham, 1998) 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)

All these 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. The developments of games for smart phones, with built in accelerometers and GPS make locative contents in games an obvious playground for developers. Conceptually, it's easy to see that the performing arts rely particularly on spatial interactivity between performer and music.

3. Spatial Hearing and its technological implications

The human ability to locate sounds is used extensively for immersive purposes, historically, as outlined earlier, and technologically in stereophony, ambisonics, wave front synthesis, surround sound and binaural technology. These technological developments rely directly on the results of studying the psycho-physics of hearing.

In accordance with Blauert (1997) "sound" as in DIN 1320 shall describe "mechanical vibrations and waves of an elastic medium particularly in the frequency range of human hearing (16 Hz to 20 kHz) whereas "auditory event" shall describe what is perceived auditorily: 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: 4)

Moore (2008) summarises the results from experimental psychology of the human ability to localise sounds as a "combination of time and intensity differences at the two ears, changes in the spectral composition of sounds due to headshadow and pinna [the shape of our outer ear] effects and changes in all of these cues due to head or sound source movement".

As to the precision of the localisation: Under laboratory conditions (often dead room and single sound source) in the horizontal plane we are able to notice if a sound source has moved by less than 1 cm. In the azimuth plane (vertical movement) the localisation is somewhat coarser.
To define if a sound comes from behind or from ahead of us, head movement is often used as a means to provide extra clues (Moore 2008 : 248).

Additionally to purely auditive cues, visual clues are of very high importance to locate a sound. This can be reproduced in the laboratory as described by Moore (2008 : 264) He abstracts this sum of cues thus: "The highest level of spatial representation involves an integration of information from the different senses."

3.1 Binaural technology and stereophony

The mentioned effects of difference in time and intensity at the ears is used in stereophony by panning a sound between two speakers. By following experimentally established panning laws, (Wikipedia 2012a) the sound on the two speakers can create an auditory event with an apparent location somewhere between the speakers.

The effects of the pinna and headshadow combined with time and intensity difference at the two ears owing to the physical distance between them are expressed in the head related transfer functions (HRTF's): The ratio of the spectral analysis of a sound at the sound source and that of the same sound at the eardrum.

By convolving a sound with the for its position angle relevant HRTF a so called binaural reproduction can be created with great immersive effect over earphones.

In an ideal world every listener would have his/her own HRTFs measured and calculated. The available HRTFs are only approximations of an "average" head and pinna.

However, there's two ways of technologically reproducing spatial sounds "spatially correctly", each with their strengths and limitations: Firstly, as described above, we can try to create or recreate an auditory event as we hear it. this can be done by excluding the physical room as much as possible and create a virtual room in the ears of the listener. These systems as near as always apply headphones to avoid intrusion of the "real" room into the virtual room. On the other hand or secondly, we can create a sound event in the real physical room which creates the virtual room within the existing one: Not suffering from the fact that whenever the listener turns her/his head the whole virtual world turns with him/her, the limitations are that this virtual world is shared with everybody else within the space and its implementation needs a considerable amount of infrastructure:

3.2 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).

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.

In this sense ambisonics can be viewed as a form of wave front synthesis (Daniel et al. 2003).

3.3 How this relates to locative audio applications

Despite its seemingly tenuous connection to spatially interactive audio applications the psycho-acoustic factors and the technologies developed on the back of them crucially influence the choice of technology for developers of LAA and hence the user requirements on positioning systems for LAA .

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 (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. 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.

Thus the suitability of positioning systems for LAA are maximally defined by the limits of the auditory resolution imposed by our hearing apparatus, i.e. a suitable positioning system does not need to be more precise as our ability to localise sounds.

The question as to the minimum requirements albeit remains as they are specific to particular applications. Yet another pointer towards the need for a survey establishing user requirements from the developers of those applications.

4. Available systems

The actual existing positioning systems available today are many and of varying limitations in affordability and accuracy, these two factors often in inverse proportionality.

A complete taxonomy of positioning system would probably stretch this paper out of proportion but a rough overview would split the field into the categories of table 1

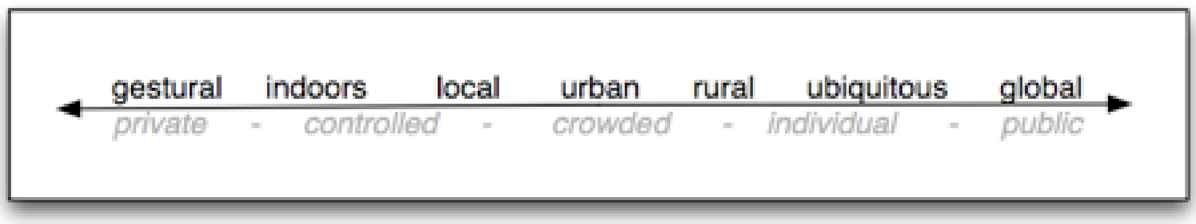
Global navigation satellite systems	(GNSS, GPS one of them)
Ultra Wide Band positioning systems	(a form of radar)
Radio signal strength system	(RSSI) using WLANm for example
Radio Frequency Indication	WLAN Fingerprinting, for example
Acoustic Positioning	Measuring time delays of a known signal on sender and receiver
Magnetic Field systems	Creating an as hoc north and using compasses, so to speak
Induction	Like the *potentiomètre d'espace*
Motion Tracking	Using cameras
Infrared tracking	using special cameras
Inertial Navigation	Using accelerometers and gyro's as in Papa Sangre
Appropriation of existing entertainment electronic	Hacking wii controls and Kinec

(table 1)

All of these systems and more have been described in detail and repeatedly at several conferences, namely the International Conference on Indoor Positioning and Indoor Navigation (IPIN) conferences 2010, and 2011, but also the International Conference and Exhibition on Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS) 2011 in Kirkonummi (2012 in Helsinki)

There are clearly plenty of possibilities for development, however, the prioritised calls for papers on user requirements at these conferences highlight an evident gap in the literature this research here is hoping to help filling.

The suitability of positioning systems for LAA is, as outlined earlier very much dependent on the nature of the application. What the weighting of different applications as a factor for generally defined user requirement should be will hopefully be established or approximated by the survey. As a un-weighted generalisation however, each LAA seems to find itself conceptually somewhere on a continuum between gestural motion tracking and global localisation. (See Ill. 1)



(Illustration 1)

5. Examples of locative audio applications

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."

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

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.

An further example, this one using 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).

Another very promising approach is a system using smart phones wherein exterior ambient sound is used as a signal to triangulate the distance to various synchronised receiver. Only the standard and low cost microphones of the mobile phones are being used and a timecode to time the signals arrival at a receiving phone (Janson et. al. 2010). It further suggests, as an aim for further research, that signals sent from known points (Wifi senders) could provide similar incidents of sounds from which a distance could be triangulated as well as positioning without time code.

6. Concluding remarks

The spatial nature of music has been explored and used as a creative element for hundred of years. With the development of new technologies the possibilities of integrating locative content into musical practices has been seized upon by the gaming industry in particular for its immersive character.

The possibilities of the technology are evident for a wide band of creative applications but a limitation in the availability of positioning technology can be established both explicitly and in-explicitly on the side of the developers of audio content, a lack of user requirements for locative audio applications on the side of the developers of positioning systems on the other.

Hence in the interest of bringing both aspects of the development together productively, this author would like to call all interested parties to participate in the online survey on

http://www.surveymonkey.com/s/6X86DJL

which forms the next step of this research to understand what positioning systems would be suitable for locative audio applications and of course with the ultimate aim to develop such applications and systems.

7. References

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

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

Edition. MIT, Massachusetts

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

Collins, Karen 2007*. "An Introduction to the Participatory and Non-Linear Aspects of Video Games Audio." Eds. Stan Hawkins and John Richardson. Essays on Sound and Vision.* Helsinki: Helsinki University Press. pp. 263-298

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]

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 and Humphreys (2005) "Games: Technology, Industry, Culture" in Terry Flew, New Media: an Introduction (second edition), Oxford University Press: South Melbourne

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]

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]

IPIN 2011 (2011) *2011 International Conference On Indoor Positioning and Indoor Navigation(IPIN)* website. [online] available on [http://ipin2011.dsi.uminho.pt/cfp.php] accessed [02/04/2011]

Janson,T.; Schindelhauer, C.; Wendeberg, J. (2010) Self-Localization Application for iPhone using only Ambient Sound Signals in Mautz, R., Kunz, M. and Ingensand, H. (eds.)(2010) *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* IEEE Computer Society

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]

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]

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]

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]

Blauert, J. (1997) *Spatial Hearing - the Psychophysics of Human Sound Localization* Revised
Moore, B.C.J. (2008) *Psychology of Hearing* 5th Edition Emerald, Bingley

Noisternig et al (2003) Ambisonic based Binaural Sound Reproduction *AES 24th international conference proceedings* [online] available from: decoy.iki.fi/dsound/ambisonic/.../3d-ambisonic-based-binaural.pdf [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]

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

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

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

Stockhausen (2012) *IntroductionHELICOPTER STRING QUARTET (1992/93*) [online] Available from: www.stockhausen.org/helicopter_intro.htmll

Wikipedia 2012a *Pan Law* [online] available from: http://en.wikipedia.org/wiki/Pan_law [Accessed 01. February 2012]

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

Wikipedia 2012c *Venetian Polychoral Style* [online] available from: http://en.wikipedia.org/wiki/Venetian_polychoral_style [Accessed 01. February 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]