An Experiment in the Perception of Space through Sound in Virtual World and Games

Antoine Gonot[1,2], Stéphane Natkin[1], Marc Emerit[2] and Noël Chateau[2]

[1] CNAM, CEDRIC laboratory, Paris, 75003, France, [2] France Telecom Group, Lannion, 22300, France

Abstract—**This study presents two approaches for representing space through sound, allowing being aware of what is going on out of visible screen. The first one, called *decontextualized beacon*, uses a sound indicating the azimuth of a target. The second one, called *contextualized beacon*, uses a sound indicating the shortest path toward the target. The usability of the two spatial auditory displays has been evaluated during a first-person navigation task in a virtual city. It appeared that *contextualized* beacon was more adapted than the *contextualized* one when navigation was not the major task. However, it was not as relevant as expected for navigation itself. Therefore, the *decontextualized beacons*, seems to be a better compromise between possibility of failure and effectiveness of navigation.**

Index Terms— **3D audio, Virtual world, Game, Navigation**

I. INTRODUCTION

By using a 3D Audio API, such as Microsoft's DirectSound3D®, and extensions such as Creative Labs's EAX® (see [13] for a review of modern audio technologies in games), one can create a realistic three dimensional audio world, including complex environmental effects (reverberation, reflected sound, muffling, etc.). Those technologies can be used for several purposes. Mostly, three aspects can be considered. First, it is essential to enhance the sensation of presence (related to immersion). Larson *et al.* [12], for example, has shown that "subjects in a bimodal condition experienced significantly higher presence, were more focused on the situation and enjoyed the Virtual Environment more than subjects receiving unimodal information did". Secondly, spatial cues allow increasing intelligibility when multiple sources are presented concurrently. This is typically referred to as the Cocktail Party effect (see [3] for a review). Thirdly, spatial sound can be used to enrich user interface, adding information through another medium. For example, a spatial auditory display can assist navigation in an environment either real or virtual. In this case, spatial audio is used to describe non-speech spatially presented sounds or at least non-linguistically mediated spatial content. Indeed, the works of Klatzky *et al.* [9] on spatial updating has illustrated the fact that speech adds extra cognitive processing load imposed by converting language to spatial content.

Those three aspects could affect navigability and then

usability of a virtual environment. Both the perception (at a lower level) of the user and the cognition (at a higher level) can be affected. At the lower level, the study of Larson *et al.* [12] has confirmed that consistent visual and auditory cues can enhance orientation task. In the same way, a study carried out by the present authors [6], has shown that real 3D sound (i.e. using Head Related Transfer Functions over headphone) compared to classical stereo panning improve also the effectiveness of this task. Taking the example of virtual city, it has been observed that users are able to choose a direction faster. However, acquisition of spatial knowledge does not rely simply on perception. At a higher level, learning to navigate a virtual world is referring to the formation of a cognitive map within a person's mind. This *map* "is a structure which is an internal representation of an environment which one uses as a reference when navigating to a destination" [14] (quote in [7]). According to the *image updating model* proposed by Klatsky *et al.* in [10], visual and auditory modalities differ in encoding the target locations into memory. However, the resulting spatial image functions equivalently for spatial updating across modalities. In fact, knowing whether a visual experience is a pre-requisite for image formation is still an issue.

So, the present research is a first attempt to assess what can be expected from a spatial auditory display in terms of navigability in virtual world. Researches seem to show that adding spatial auditory information consistent with visualization rather improves orientation than spatial image and global navigation. Even though Larson *et al.* [12] has observed that bimodal processing significantly improves memory, this effect has shown to be related with auditory content of a sounding object rather than with its spatial property. So, what can be the contribution of spatial auditory cues to spatial knowledge acquisition?

Essentially, two typical acquisition modes can be distinguished in game: "pedestrian mode" and "bird mode". The former is the typical situation in "first-person shooter" and most of adventure-action game in which the world is perceived from the "in-side" (even if third-person view is allowed in some cases). The later refers to the genre of strategic computer simulation, often called "god game", in which the world is perceived from the "out-side". We focus in this study on the "pedestrian mode". The present paper investigates the relative contribution of auditory modality to perceptive and cognitive spatial ability during first-person navigation in a city-like virtual world. An example of a game using auditory navigation

as part of the *gameplay* will be taken to illustrate the details of the problematic. Then two approaches for representing space through sound will be presented. The design of the virtual word used for the experiment will be introduced. Finally, the results of the experiments will be discussed. First conclusions on the contribution of spatial auditory cues to spatial knowledge will be drawn. Future directions for research will be presented.

II. USING SOUND FOR CONTROLING NAVIGATION-BASED CHALENGES IN GAME

A. An example: Eye

The Eye video game (http://eye.maratis3d.com) is a second year project designed by a group of students (Matthew Tomkinson, Olivier Adelh: Game Design, David Elahee, Benoît Vimont: Programming, Johan Spielmann, Anaël Seghezzi: Graphic Design, Timothée Pauleve: Sound Design, Julien Bourbonnais: Usability, Vivien Chazel: Production) from the Graduate School of Games (ENJMIN: www.enjmin.fr). This video game is based on the classical "blind man's buff game". The player character, hero of the game, Vincent Grach, has an extraordinary power: he is able to visit others' people memories. Travelling physically in a lunatic asylum and in the memory of all patients, he tries to save his wife. During this journey, he is confronted to the anguishes of the other characters. To save his mental health, he must close his eyes and progress in an almost dark world where only the circle of the strong lights appears. In this state, he must also protect himself from numerous dangers like falling from a barge or into a fire. His progression relies on his memory of the space and on the location of sound sources.

As a consequence, an original and complex sound world is one of the main features of Eye. It was designed using ISACT™ from Creative Labs©. It relies on a real time 3D localisation of sound sources, using the OpenAL® library which was integrated in the game engine "Maratis". This localisation can be heard trough a 5.1 system using the Sound Blaster® technology. Two other effects are used to help the player when Vincent Grach's eyes are closed. Firstly, the decay of the attenuation curve of sound objects are accentuated (i.e. the "Roll Off" parameter is higher), then The "Eiffel Tower effect"[1] mutes the sounds which are not related to dangers or which do not help for localisation.

B. Auditory navigation in games

Except for audio games for visually impaired (for example, GMA Games's *Shades of Doom*® or Pin Interactive's *Terraformers*®, etc.) or games revolving around a musical experience (Sega's Rez®, Nana On-sha's Vib Ribbon®), only few games use sounds as part of their *gameplay*. However, as point out by Stockburger in [17], although a game can be seen as a larger genre, it sometime uses sound in an innovating way.

For example, sound is an important element of *gameplay* in stealth intrusion game like Konami's *Metal Gear Solid 2 Sons of Liberty*® (MSG2), where, most of the time, the player can not see his opponent. Indeed, according to Begault [4], spatial auditory displays are very effective for conveying alarm and warning messages in high-stress environments. One aspect of the *acousmatic*[2] situation of the player in MGS2 is referred to as *sound awareness*, similar to what can be experienced by a pilot in an airline cockpit. In the same game, another type of *acousmatic* situation occurs when the player has to use a directional microphone to locate a specific hostage in the environment. Similar challenge is also encountered in *Eye,* except that, visual cues are only partially available. According to the *image updating model* [10], those two games illustrate two complementary challenges involved by an auditory navigation-based *gameplay*.

The first one is rather based on the initial encoding of target(s). In *Eye*, because visual perception can be frequently interrupted, auditory modality has to be strongly involved in the encoding of the immediate surrounding scene. For example, a fast memorization of the exit's topology of a room in more of the objects it contains turns out to be fundamental. This is part of the *game balance* when considering the navigability of the virtual world the player is exploring. In the case of MSG2, because it is rather visual, the encoding of the surrounding is not a problem *a priori*. The *game balance* mostly relies on another kind of challenge, related to spatial updating. Finding a character by using the directional microphone is similar to find a given street in a town by using a compass. If we assume that the player didn't know the environment, the effectiveness of such task depends mainly on the complexity of the road (or corridor) pattern. For example, the pattern in Figure 1.a can be considered more complex than the pattern in Figure 1.b, referred to as raster pattern by Alexander *et al.* [1] (quote in [19]).
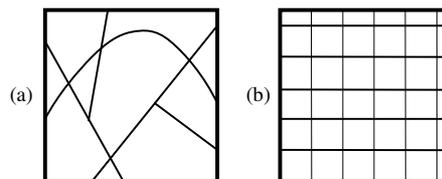
Fig. 1. Example of two different complexities for road-network. (a) is supposed to be more complex than (b)

Thus, controlling the reliance on physical space can be a critical issue for *game balance*. It depends on the relative importance of navigation during the different phases of the game and more generally on the player activity. For example, if the player has to fight an enemy, navigation becomes a secondary task, and should be achieved with the minimum of cognitive load. Indicating the shortest path to the target could then annihilate the complexity of a pattern, removing any

---

[1] It refers to the study of Roland Barthes on "Eiffel Tower and other mythologies". He believed that the tower was so popular because a person looking out over Paris felt they could master the city's complexity.

[2] According to M. Chion [18], the term "Acousmatic" is used to describe the listening situation of someone hearing a sound without seeing the object which produced it.

challenge in navigating. In the contrary, if the goal is to collect equipments (typically, weapon, armour or ammunition) then, navigation could be more challenging. The following illustrates how to make the most of the potential of 3D audio for this particular balance.

C. Two approaches for representing space through sound

In the domain of sonification, the term *beacon* has been introduced by G. Kramer [11] to describe a category of sound used as a reference for auditory exploration and analyse of complex, multivariate data set. *Beacons* do not have intrinsically spatial property, but has been naturally adapted to navigation by Walker and Lindsay [20]. This concept is very close to the concept of landmark used in the domain of urban planning.

As Johnson and Wiles point out in [8] that it is preferable that the interface remains the most *transparent* as possible. They hypothesise that "the focus on, and lack of distraction from, the major task contribute to the facilitation of the flow". For example, Lionhead Studio*'s Black & White®*, have been released with the interface virtually absent during *gameplay*. Such design rule can be transposed to auditory modality, considering its strong ability to facilitate player's selective attention. For example, Andresen [2], creating a blind-accessible game, has installed noisy air conditioning vents into the centre of each hallway to indicate the location of the exits. The importance of this type of *beacons* indicating location in the immediate surrounding has been well illustrated by *Eye*. Then, the previous guideline will be applied to *beacons* which indicate a distant location.

Let's consider the environment shown in Figure 2, presenting many adjacent rooms with opening, communicating each other. The auditory situation the player experiences in *MSG2* and *Eye* involves direct path from the source to the listener. In this case, the use of the information conveyed by sound is similar to the use of a compass. This approach, described as *decontextualized* (a), is common since sound engine only recently support complex environmental effects (*i.e.* take into account the interaction of sound with physical space).
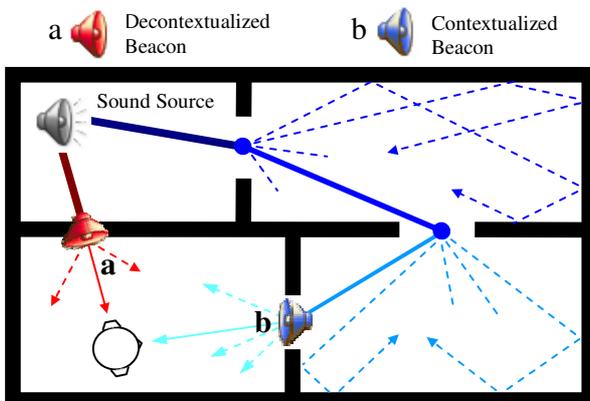
Fig. 2. Two approaches for auditory representation of distant location through sound

Let's now consider a propagation model of acoustic waves from a sound source to a listener. By extrapolating the exclusion phenomenon, the apparent position of the source is the position of the opening. Thus, such approach, described as *contextualized* (b), defines a *beacon* indicating both, exit location in the immediate surrounding and a path to the distant location of an object.

For describing clearly those two *beacons*, two terms have to be defined: the *target* is a location to reach and the *beacon* is a sound, whose spatial cues indicate this location.

The usability of those two types of beacons has been evaluated in a city-like virtual world, enabling one to create complex environment, constraining visual modality to local perception of space.

III. THE EXPERIMENT: MOTIVATION, DESIGN AND DISCUSSION

A. Presentation of the experiment

1) Motivation and hypotheses

As pointed by Rollings and Adams in [16], "as action games became more complex, the play area began to span multiple screens of action, although the player still needed to be aware of what was going on in the game-world not visible onscreen". There are several common configurations used to achieve this goal. The original one, used in Williams Electronics's *Defender®*, shows the entire game world to the player. Another configuration offers a zoom-out view of the area surrounding the player. At last, a third configuration, rather used in strategy and war game, presents a map building up as the player explore. So, the amount and the quality of the spatial information displayed by this mean can vary, depending on the *gameplay* and the required *challenge*. That way, the *decontextualized* and *contextualized beacons* can be considered as two different approaches of *auditory minimap*. Taking into account the interaction of sound with physical space (*i.e. to use contextualized beacons*) is not just an improvement of sound realism. It offers contrasted, not to say complementary, auditory representations of the virtual world. Thus, the aim of the experiment presented here is to assess the relative contribution of these two *beacons* to spatial knowledge acquisition, even when visual perception is involved. Then, discerning the navigation task (*i.e.* way finding) from the orientation task (*i.e.* choosing a direction when bifurcation occurs), the following hypotheses are ventured for each type of *beacons*.

For *contextualized beacons*, initial encoding of the beacon is facilitated by consistency between visual and auditory cues. This improvement of initial encoding should be reflected in a more effective orientation. Moreover, because the spatial configurations of targets and *beacons* do not fit, auditory modality can not contribute to the formation of the mental map. The later rather relies on local visual perception. However, this also means an effortless navigation that should be reflected by a lower cognitive load. Respectively, using *decontextualized* beacons requires greater focus of auditory attention to the spatial configuration in order to correct bearing (*i.e.* angle between the target and the direction imposed by the

road-network). This should involve a higher cognitive load but the formation of a better cognitive map.

2) Task and data

The goal is to find, as quickly as possible, nine "streets" represented by "real"w sound (for example "fireworks", "fanfare", etc.). They are equally distributed in three zones (marked on the floor by three different colors), the corresponding sound is always audible, and they are sought one after the other. When the player reaches the target (i.e. enters the right street), the program presents the next one. Interaction log are recorded during the game and subjective evaluation are achieved at the end of a session, including the recall of the street's location on the map, auto-evaluation of cognitive load (NASA-TLX) and impressions questionnaire. There are two experimental factors: *contextualized* versus *decontextualized* beacons and stereo vs. binaural rendering (using non individualized Head Related Transfer Functions).

B. The design of the Virtual World

Here are presented the design choices that have been made, in order to guarantee *believability* of the virtual world and a control of dependant variables. One can refer to [6] for a complete description of the setup.

1) The foundations of the game

Navigation in a city can be seen as a succession of choices of directions to take. Thus, as shown in Figure 3, the 3D model of the road-network has been simplified drastically, so navigation is like moving from square to square on a chessboard (excepted for corners).
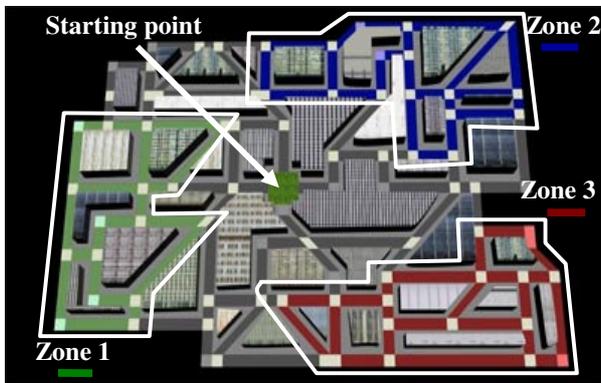
Fig. 3. Road-network of the Virtual City

The game has been designed using Virtools©, and sound sources has been spatialized using algorithms developed at France Telecom Group©.

2) Avoiding cross-modal influence

As pointed by Pellegrini [15], "when assessing psychoacoustic features within an AVE (auditory virtual environment), the auditory test setup needs to be designed with care to minimize unwanted cross-modal influences". Thus the textures were choosing for their banality, avoiding that a building serves as visual landmark. Moreover, as can be seen in Figure 4.a, the Virtual World is not illuminated. Only few spotlights are used to allow the visualization of the directly reachable nodes. By this way, local visual perception is under control and reduced to desired cues: the direction choices, the distance and the azimuth of next nodes. The Figure 4.b shows a screenshot of the first-person view.
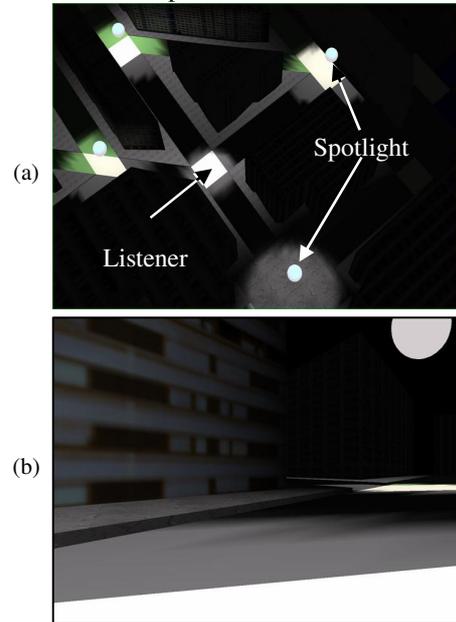
Fig. 4. Lightning (a) and first-person view of the Virtual World (b)

3) The design of contextualized beacons

Researches on usability have shown that "the design goals for Auditory Virtual Environments shift from "reproducing the physical behaviour of a real environment as accurate as possible" to "stimulating the desired perception directly"" [15]. Then, it is recommended to rather reproduce required features for a given specific application. Consequently, implementing *contextualized beacons* does not necessarily require modelling the exclusion phenomenon. However the *beacon* needs to exhibit its main characteristics, that is:

- The sound of the *beacon* is coming from a particular exit. This is implemented by calculating the shortest path toward the location each time a new node is reached. Then, the azimuth of the sound source is given by the first node of the path.
- The sound must reflect the effect of wave's propagation. Only the effect of distance (length of the shortest path) on sound level has been included.
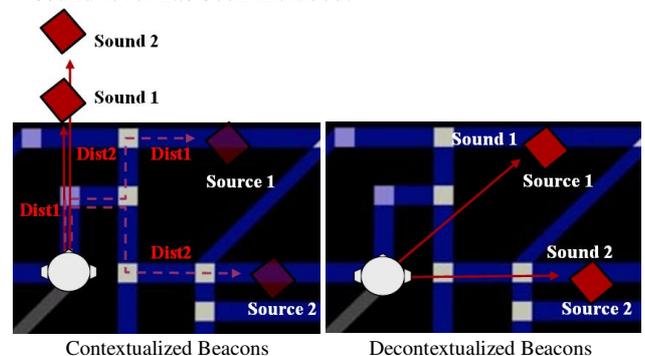
Fig. 5. Implementation of the two approaches for auditory representation of distant location through sound

For smooth changes, the position of the source between two nodes is determined by linear interpolation.

C. Results and discussion

As was expected, the *beacon*'s type has significantly affected orientation task. The time for initial encoding and choosing directions was significantly lower for *contextualized* beacons (using binaural or stereo). It was correlated with a lower cognitive load (see [6] for details on the statistical analysis). Those results have confirmed that such *beacon* could be more effective when navigation is not the major task. However, no significant effect has been observed on navigation. Indeed, the two *beacons* have allowed effective navigation. For both, the covered distance to reach the target was close to the shortest path and the perceived performance was high. So, it seems that the reliance on physical space was not as strong as expected for this environment. Even if a comparative measure of complexity with other environments encountered in FPS game should be achieved to rigorously conclude, it shows that *contextualized beacons* were not necessarily relevant for navigation. According to Johnson and Wiles [8], "during *gameplay*, the joy of success is dependant upon the possibility of failure". From this point of view, *decontextualized beacons* could offer a good compromise. Finally, the angular and absolute distance errors when recalling locations on the map did not differ. So the mental map seems not to be better with *decontextualized* beacons. Once again, nothing can be concluded, since the recall task was maybe not a good evaluation of mental map for first-person navigation in virtual city (*i.e.* local visual perception). However, if it was, this result could mean that *decontextualized beacons* could not really improve mental map constructed by visual experience.

IV. CONCLUSION

This study has tried to investigate what could be the relative contribution of auditory modality to spatial ability during navigation with a first-person view. It appears that auditory modality can improve the orientation task. However, the experiment failed to exhibit a real improvement of the mental map constructed by local visual perception. Indeed, it seems that memorizing locations when finding his way as fast as possible is a very difficult task. Memorization could have been better without time pressure. However, pre-test had shown that without this pressure, participant sometimes went all the way trying to better remembering locations. Thus, by achieving a non-effective navigation, no local effect on orientation could have been observed.

Finally, assuming that spatial auditory cues are not predominant in auditory scene analysis, as pointed out by Bregman [5], it has been concluded that spatial auditory display should be reconsidered in a more ecological way. The spatial abilities of the auditory system have been surely overestimated. It seems more relevant to study the player's responsiveness to audio cues more closely related to *diegetic* space (*i.e.* the space figured on screen). Future works will focus on games offering multiple out-side views of the world, as in strategic computer simulation or "god game". This will lead us to the study of audio-visual interactions in a multiresolution interface, introducing the notion of level-of-detail (LOD) for spatial auditory display.

REFERENCES

[1] Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I. and Angel, S. "A Pattern Language". Oxford University Press, New York, 1977.
[2] Andresen, G., "Playing by Ear: Creating Blind-accessible Game". Gamasutra article, May 20, 2002. URL: http://www.gamasutra.com/resource_guide/20020520/andersen_01.htm
[3] Arons, B.. "A review of the Cocktail Party". Effect. Journal of America Voice I/O Society, 1992.
[4] Begault D. R.,"3-D Sound for Virtual Reality and Multimedia". Cambridge, MA: Academic Press Professional, 2004.
[5] Bregman A.S., "Auditory Scene analysis: The perceptual organization of sound", Cambridge, Mass., MIT Press, 1992
[6] Gonot, A., Chateau, N., Emerit, M., "Usability of 3-D Sound for Navigation in a Constrained Virtual Environment", 120th AES Convention, Paris, France, 2006.
[7] Ingram, R., Benford, S., Bowers, J., "Building Virtual Cities: applying urban planning principles to the design of virtual environments". Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST96), pp 83-91, 1996.
[8] Johnson, D., & Wiles, J., "Effective affective user interface design in games". International Conference on Affective Human Factors Design, Singapore, June 27-29, 2001.
[9] Klatzky, R. L., Lippa, Y., Loomis, J. M. and Golledge, R. G., "Learning directions of objects specified by vision, spatial audition, or auditory spatial language". Learning & Memory, 9, 364-367. 2002.
[10] Klatzky, R. L., Lippa, Y., Loomis, J. M. and Golledge, R. G., "Encoding, learning and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision." Experimental Brain Research, 149, 48-61. 2003.
[11] Kramer, G., "Some Organizing Principle for Representing Data with Sound". Auditory Display: Sonification , Audification and Auditory Interface, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, USA, pp 202-208. 1994.
[12] Larsson, P, Västfjäll, D, & Kleiner, M., "Ecological Acoustics and the multi-modal perception of rooms: Real and unreal experiences of auditory-visual virtual environments". In Proceedings of ICAD, Helsinki, 2001
[13] Menshikov A.,. "Modern Audio Technologies in Games", article based on the presentation given at the Game Developers Conference in Moscow in 2003. URL: http://www.digit-life.com/articles2/sound-technology/index.html
[14] Passini, R., "Wayfinding in Architecture", Van Nostrand Reinhold, 1992
[15] Pellegrini, R. S., Quality Assessment of Auditory Virtual Environments, In Proceedings of ICAD, Espoo, Finland, 2001
[16] Rollings, A., Adams, E., "Andrew Rollings and Ernest Adams on Game Design". New Riders Publishing, May 05, 2003.
[17] Stockburger, A., "The Game Environment from an Auditive Perspective", Level Up (Utrecht Universiteit: DIGRA 2003)
[18] M. Chion, "L'audio-vision, Son et Image au cinéma", 2nd edition, Nathan Cinéma. 1997.
[19] Sun, J., Baciu, G., Xiaobo, Y., Green, M. "Template-Based Generation of Road Networks for Virtual City Modeling". VRST'02, Hong Kong, November 11-13. 2002.
[20] Walker, B. N., Lindsay, J., "Effect of Beacon Sounds on Navigation Performance in a Virtual Reality Environment". Proceedings of the International Conference on Auditory Display, Boston, MA (6-9 July) pp 204-207. 2003