Audio Engineering Society
Convention Paper

Presented at the 115th Convention
2003 October 10–13 New York, New York

Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications

György Wersényi[1],
[1]Széchenyi István University, Department of Telecommunications, Győr, H-9024, Hungary.

ABSTRACT

Listening tests were carried out for investigating the localization judgments of 40 untrained subjects through equalized headphones and with HRTF synthesis. The investigation was made on the basis of the former GUIB (Graphical User Interface for Blind Persons) project in order to determine the possibilities of a 2D virtual sound screen and headphone playback. Results will be presented about the capabilities and values of typical headphone playback errors as well as minimum, maximum and average values of discrimination skills. Special localization events like left-right and up-down symmetries, missing locations in vertical localization are also discussed. The measurement method includes a special 3-categorie-forced-choice MAA report on a screen-like virtual auditory surface in front of the listeners. Test signals were presented with different spectra and movement. Conclusions are drawn both for a GUIB application as well as for the binaural synthesis about the role of the fine structure of applied HRTFs.

1. INTRODUCTION

The former GUIB (Graphical User Interface for Blind Persons) project was for finding solutions to help elderly and disabled people to use personal computers. Blind persons do not have the advantageous properties of graphical user interfaces (GUI) like MS-Windows, icons and the ability of orientation among multiple visual information [1]. Visual events on the screen, like opening files, closing windows, movement of the cursor, etc. are to be replaced only by sound events. The former results of this project related to sound reproduction are:

- a collection of sounds representing visual icons and events of the screen only by acoustical information called “Earcons” [2],
- the possibilities of different input media [3],
- and the localization blur using a multi-channel loudspeaker playback system [4].

Now, the possibilities of the headphone playback is discussed.

2. LOCALIZATION IN FREE-FIELD AND VIRTUAL ENVIRONMENTS

Sound waves reaching the eardrums are affected by directional filtering of the outer ears. This binaural filtering effect determines basically the perception of the direction of sound sources depending on the angle of incidence [5-8]. Monaural cues are responsible for the perception of elevation in the median plane, front-back directions and distance. The Interaural Time Delays (ITD) and the Interaural Level Differences (ILD) are the basic cues for the localization in the horizontal plane, which results in a much better localization performance. The directional information added by the filtering effects of the outer ears is complete at the entrance of the ear canal and this information does not vary along the cavity of the ear canal [6, 7, 9].

The transmission from a point in the free-field to the eardrums is described by the complex Head-Related Transfer Functions (HRTFs). In virtual audio environments the HRTFs have to be reproduced through headphones. We can use individual HRTFs, HRTFs from a "good localizer" or from a dummy-head. It was shown, that HRTFs from a good localizer and the use of simple methods to make them more individual results in a satisfying localization [8, 10, 11]. Other basic psychoacoustic parameters for the localization are: spectral content, bandwidth, volume, duration, adaptation and learning, a-priori knowledge, and additional visual information.

According to the statements of the *binaural technique*, if we reproduce the sound pressures at the eardrums exactly, the reproduced signal will have the full spatial information about the environment it was recorded in. For the reproduction a proper and individual headphone-equalization is required, as far as possible. This technique may contain errors as well, like front-back confusion and in-the-head localization due to headphone playback [9, 12]. In general the results from free-field measurements tend to be better than when using headphone playback [13].

2.1. Listening tests in virtual audio synthesis

Localization means finding the absolute position of the sound source. *Localization blur* is the smallest change in the direction of the sound source, which can be perceived. To measure it, we have to search for the Minimum Audible Angle (MAA) or the Just Noticeable Difference (JND), where subjects only have to compare two sound sources and identify only the change of the source direction [14, 15].

Results on this field are difficult to compare, because experimental designs and methods differ. For a direct comparison of results, similar conditions are needed. Furthermore, better results can be obtained in a MAA measurement in contrast to an absolute measurement. The application of a headphone in a virtual synthesis introduces well-known errors. These are:

- in-the-head localization (the lack of "externalization")
- front-back confusion
- sources too near
- elevation shift
- ambiguity of movements symmetrical to the median plane.

2.2. Virtual Acoustic Displays

Virtual Acoustic Displays (VAD) are widely used in several applications. VAD identifies a virtual environment, where sound sources are artificially reproduced and the listeners are able to localize and identify them.

To realize a VAD two independent questions have to be answered. First, which sounds correspond the best to the visual and deeper meaning of the object to be reproduced? In other words, what is the best mapping between sounds and events on the screen? Second, what is the localization blur like through headphone playback? In principle, three-dimensional VADs can be realized by reproducing *depth* or *distance* information as well, e.g. by an object approaching the listener or by overlapping windows. State-of-the-art multimedia computers and applications nowadays allow full auralization and orientation in a virtual reality. Only the last decade made it possible to handle huge amount of computation data, real-time filtering of HRTFs, reverberation and head movements [16].

3. LOCALIZATION BLUR IN A 2D VIRTUAL AUDIO SYSTEM

The listening test was for investigating the localization blur using the headphone playback method. The Earcons are short pure tones or special noisy-like sound events. Therefore, we decided to use 300 ms long sound events of broadband and filtered version of broadband noise (signal A, B, C) to match and model in a generic but not too specific way the possible real application of the Earcons.

Furthermore, the measurement has novel methods like the 3-categorie-forced-choice in order to determine the "uncertainty domain" of the subjects during the localization judgments. A two-directional discrimination will be applied to determine the localization blur independent of the direction of a

moving source. Instead of the commonly used method of constant source distance (in a circle around the head), a "virtual rectangle screen" is simulated (Fig. 1.).

3.1. Measurement method

Our virtual sound screen is a 2D square surface in the front of the listener. The mapping from a visual screen (PC monitor) is better to a "screen-like" 2D virtual sound screen for the orientation with the mouse (see Fig. 1.). The maximal range of simulated sources is ±60° horizontal and vertical. We assume that the listener in a real life application would be able to adjust volume, so the parameter "depth" is neglected.

The measurement setup is based on a PC with the Beachtron DSP board. Real-time convolution of the mono input signal and the HRTFs is made in the time-domain (16 bit; 44,1 kHz). The system is precisely equalized for the circumaural, open-dynamic Sennheiser HD540 headphone. The HRTFs originate from a good localizer in a measurement of *Wightman and Kistler* [17-19]. 72 measured HRTFs are available in a form of 75-point minimum-phase-FIR-filter set in 30° spatial resolution. Duration and volume of the test signals were determined during a pre-test with 7 subjects. The main test was made with 40 untrained subjects, all with normal hearing. The individual setting of the HRTFs corresponds to measure the size of the head (distance of the ear canal entrances). To reduce the parameters we work with constant signal volume and duration.

Excitation signals for the MAA-measurement are 300 ms noise burst impulse-pairs: white noise (signal A), 1500 Hz low-pass (signal B) and 7000 Hz high-pass filtered version of the white noise (signal C).

Novelties and general conditions in our measurement:

1. Use of a 2D virtual sound screen in the front of the listener. Sources can move only in the horizontal (left and right) and in the median plane (up and down) from the origin in 1° resolution. The source distance is not constant and the source is *not moving around the head* as usual.
2. Subjects have to report in a 3-categorie-forced-choice (MAA): "no difference between the sources", "different sound sources" and "I'm not sure".
3. Source-pairs have to be discriminated first as the second source is moving away from the static reference source, then as it moves toward the reference point. We are looking for the nearest point to the reference, where (from both directions) the subject is able to discriminate the sources with certainty. If we determine the localization blur from both direction of moving, we will get a direction-independent localization performance.

The first impulse of the burst-pair is always a fixed reference point, and the second is moving first away, then toward the reference point. During the MAA measurement subjects were asked to report in a 3-categorie-forced-choice. Possible answers are: "no difference" if the subject is not able to discriminate the sources and they seem to come from the same direction. „Different sound sources" means that he is able to distinguish between the signals. He may have the possibility to choose „uncertain" as the answer, if he is not sure which is the case. At the beginning, the reference point is always in the origin. The second source is moving away from the reference point to the left. After the subjects have reported "different sound sources" the moving source moves backward. The nearest point where the subject in both direction of moving was able to distinguish the sources will be selected as the new reference point.

In [20] a similar method was used, but only in a 2-alternative forced choice as the subject's response was used to initiate the next trial. In [21] the subjects had also to report in a forced-choice using pulse-pairs and they had the possibility to be uncertain. But this was not investigated deeply.

3.2. Capability and errors in headphone playback

A *preliminary test* was also made with 25 subjects in order to find well-known headphone playback errors, like in-the-head localization, elevation shift and front-back confusion [22]. During this test, special sound source locations and movements were generated and specific questions had to be answered. According to former results poor localization performance was observed:

- only 20% of the subjects were able to "externalize" the sound source and avoid in-the-head-localization.
- Front and back directions were mostly confused as the source was in the front (69%); only one third was able to localize the source at its correct position.
- 58% reported elevation shift.

This test proved that even a carefully made headphone-equalization, the use of HRTFs of a good localizer and individual setting of the size of the head are maybe insufficient. All of the subjects were easily influenced and they reported all kinds of answers by the same signal reproduction, which suggests low quality localization in the median plane. In-the-head localization and front-back confusion are more significant than elevation shifts.

3.3. Localisation blur and discrimination skills

The main test includes listening tests using noise impulse pairs in the horizontal and in the median plane in order to determine the localization blur. 40 untrained subjects all with normal hearing participated in the test, and results are presented below showing average (AVG), maximal (MAX) and minimal (MIN) values of measured data. The test was carried out in the anechoic room.

Results were found to be independent of age and computer skills, but little improvement in the localization performance was found by subjects using headphones often.

Figure 2 shows the average values from the *median* plane for all signals up and down as well. The average resolution for signal A is about 15-17°, 19-24° for signal B and 18-23° for signal C.

In the *horizontal* plane signal A is localized the best with an average resolution of 7-9°, signal B with 9-11° and signal C with 8-10° (Fig. 3.). In general we can support the finding that broadband sources are localized the best as well as signals with lots of high-frequency information, but the differences in our measurements are relatively low: the results of signal A are only 1-2° better than results of signal C.

It is interesting that the resolution (the difference between nearby source locations) is almost constant.

Figure 4 shows the average localization of signals with different spectra in the horizontal plane (only left side).

The possible source locations are shown on Fig. 5 in the median plane and horizontal plane respectively (on average).

3.4. Left-right and up-down symmetry

Other studies reported asymmetries on the left and right sides of the hearing system in connection with right or left-handed persons [23, 24]. Our results also showed systematic asymmetry but we had only right-handed subjects. Sources from the left side were typically harder to localize. The results show 2-4° average differences that correspond to a difference of 20-40%. Further measurements are suggested to find regularities on this field.

3.5. Vertical localization

In the median plane the localization is only made based on the HRTFs because no interaural differences are present. This results in a decreased localization performance in contrast to horizontal plane localization. This fact is supported by our results as well: first, the spatial resolution is poorer, second, some were not able at all to localize the sources. Only 67% of the subjects reported correct localization but 33% made the MAA-judgments only why the impulses „sound different“ (based on the spectral distortions of the applied HRTFs) The same observed *Mills*: subjects reported that the difference between the stimuli seemed to be in the loudness or quality of the sound rather than its location [25].

3.6. Missing locations

Subjects had to discriminate new source locations (reference points) within a domain of ±60°. The number of possible source locations is limited: maximal 6 in the horizontal plane and maximal 2 in the median plane. Subjects, who can not determine so many different source locations, have poor localization performance (“missing locations”). In the *horizontal* plane only the half of the subjects could discriminate 6 sources for all signals. In *vertical* directions 70% of females and 62% of males were able to detect 2 new sources. This shows a bit poorer performance of males.

3.7. Uncertainty in discrimination skills

The subjects reported in a 3-categorie-forced-choice, so they determined a domain in which they were uncertain. By some of the subjects this domain is quite large: by 57% it reached 3-5° or more independent from the signal. 43% of all subjects reported only “different sources” and “no difference”. The uncertainty by them is about 1°.

4. SUMMARY

Minimum Audible Angle measurements were made in order to determine the localization blur for signals with different spectral containment. 40 untrained subjects reported in a 3-categorie forced choice using headphone playback and synthesized HRTFs. The goal was to determine how many virtual sound sources can be placed in the horizontal and in the median plane respectively and in which spatial resolution. The summarized findings are:

- the localization is poorer in the median plane than in the horizontal plane,
- the lack of individual HRTFs and head movements cause in-the-head localization, front-back reversals and elevation shift. (The last is not very significant in our measurement.)
- in the median plane, one third of the subjects could not localize the sources at all.
- Source movements symmetrical to the median plane are confusing and hard to

perceive, sources are often localized only in the back hemisphere.
- Age, sex and computer skills do not influence the localization, but subjects wearing often headphones delivered better results.
- Broadband signals are to localize the best, followed by high frequency stimulus and low frequency tones at last.
- The hearing system is not symmetrical: different resolution can be measured on the left and the right side as well as up and down.

The 2D virtual acoustic display is suited for replacing the screen and visual information for blind and elderly people in case of proper mapping between acoustic and visual information, so these results can be the basis for further GUIB applications and investigations.

- Average resolution of 7-11° and 15-24° were measured in the horizontal plane and median plane respectively dependent on the spectral content of the signals.
- White noise is to localize the best, low frequency filtered noise the least. It is also suggested for a GUIB application to use broadband noisy like sound events and/or tones with more high frequency content.

Earcons are already available based on the decisions of blind people. Based on these results, for a GUIB-based simulation it is recommended

- not to use vertical displacement of simulated objects, because one third of the users are not able at all to localize virtual sound sources in the median plane. One possible solution could be timbre or pitch modulation based on psychoacoustic observations: signals having higher frequency components are „above“; signals with lower frequency elements are „below“,
- to partitioning in the horizontal plane for maximal 9 source position in a resolution of 10 degrees.

The Beachtron system is suited for listening tests and for low-cost solutions for everyday users: it offers real-time filtering of HRTFs, user-friendly applications and programming, headphone equalization and even individual settings of the HRTFs through the measurement of the head diameter. We found this system suitable for GUIB applications.

On the other hand, the preliminary test showed that low-cost real-time system with many efforts to a correct binaural reproduction have all kinds of headphone playback errors. This assumes that the problem of insufficient localization is not due the „quality”, fine structure or overall accuracy of the applied HRTFs.

5. REFERENCES

[1] D. Burger, C. Mazurier, S. Cesarano, J. Sagot, “The design of interactive auditory learning tools,” Non-visual Human-Computer Interactions, vol. 228, pp. 97-114, (1993).

[2] M. M. Blattner, D. A. Sumikawa, R. M. Greenberg, “Earcons and Icons: their structure and common design principles,” Human-Computer Interaction, vol. 4(1), pp. 11-44, (1989).

[3] K. Crispien, K. Fellbaum, “Use of Acoustic Information in Screen Reader Programs for Blind Computer Users: Results from the TIDE Project GUIB,” The European Context for Assistive Technology (I. Porrero, R. Bellacasa), IOS Press Amsterdam, (1995).

[4] K. Crispien, H. Petrie, “Providing Access to GUI’s Using Multimedia System – Based on Spatial Audio Representation,” J. Audio Eng. Soc. 95th Convention Preprint, New York, (1993).

[5] J. Blauert, “Spatial Hearing” The MIT Press, MA, (1983).

[6] H. Møller, M. F. Sorensen, D. Hammershøi, C. B. Jensen, “Head-Related Transfer Functions of human subjects,” J. Audio Eng. Soc., vol. 43(5), pp. 300-321, (1995).

[7] D. Hammershøi, H. Møller, “Sound transmission to and within the human ear canal,” J. Acoust. Soc. Am., vol. 100(1), pp. 408-427, (1996).

[8] H. Møller, M. F. Sorensen, C. B. Jensen, D. Hammershøi, “Binaural Technique: Do We Need Individual Recordings?” J. Audio Eng. Soc., vol. 44(6), pp. 451-469, (1996).

[9] H. Møller, “Fundamentals of binaural technology,” Applied Acoustics, vol. 36, pp. 171-218, (1992).

[10] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman, “Localization using nonindividualized head-related transfer functions,” J. Acoust. Soc. Am., vol. 94(1), pp. 111-123, (1993).

[11] J. C. Middlebrooks, “Individual differences in external-ear transfer functions reduced by scaling in frequency,” J. Acoust. Soc. Am., vol. 106(3), pp. 1480-1491, (1999).

[12] J. Kawaura, Y. Suzuki, F. Asano, T. Sone, “Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear,” J. Acoust. Soc. Japan E 12, pp. 203-215, (1991).

[13] R. L. McKinley, M. A. Ericson, “Digital synthesis of binaural auditory localization azimuth

cues using headphones," J. Acoust. Soc. Am., vol. 83, S18, (1988).
[14] W. M. Hartmann, B. Rakerd, "On the minimum audible angle – A decision theory approach," J. Acoust. Soc. Am., vol. 85, pp. 2031-2041, (1989).
[15] D. R. Perrott, A. D. Musicant, "Minimum auditory movement angle: binaural localization of moving sources," J. Acoust. Soc. Am., vol. 62, pp. 1463-1466, (1977).
[16] D. R. Begault, "3-D Sound for Virtual Reality and Multimedia" Academic Press, London, UK, (1994).
[17] S. H. Foster, E. M. Wenzel, "Virtual Acoustic Environments: The Convolvotron," Demo system presentation at SIGGRAPH'91, 18th ACM Conference on Computer Graphics and Interactive Techniques, Las Vegas, NV (ACM Press, New York), (1991).
[18] Crystal River Engineering, Inc.: BEACHTRON – Technical Manual, Rev.C., (1993).
[19] F. L. Wightman, D. J. Kistler, "Headphone Simulation of Free-Field Listening I.-II.," J. Acoust. Soc. Am., vol. 85, pp. 858-878, (1989).
[20] D. R. Perrott, J. Tucker, "Minimum Audible Movement angle as a function of signal frequency and the velocity of the source," J. Acoust. Soc. Am., vol. 83, pp. 1522-1527, (1988).
[21] W. Mills, "On the minimum audible angle," J. Acoust. Soc. Am., vol. 30, pp. 237-246, (1958).
[22] Gy. Wersényi, "Acoustic Signal Processing for Listening Tests in Virtual Audio," 2001 Polish-Czech-Hungarian Workshop on Circuit Theory, Signal Processing, and Telecommunication Networks, Budapest, pp. 175-181, (2001).
[23] K. Hartung, „Modellalgorithmen zum Richtungshören, basierend auf den Ergebnissen psychoakustischer und neurophysiologoscher Experimente mit virtuellen Scallquellen" Dissertation, Ruhr-Universität, Bochum, (1997). (Shaker Verlag, Aachen, 1999)
[24] S. M. Abel, C. Giguere, A. Consoli, B. C. Papsin, "Front/Back Mirror Image Reversal Errors and Left/Right Asymmetry in Sound Localization," Acoustica, vol. 85, pp. 378-389, (1999).
[25] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am., vol. 62, pp. 157-167, (1977).

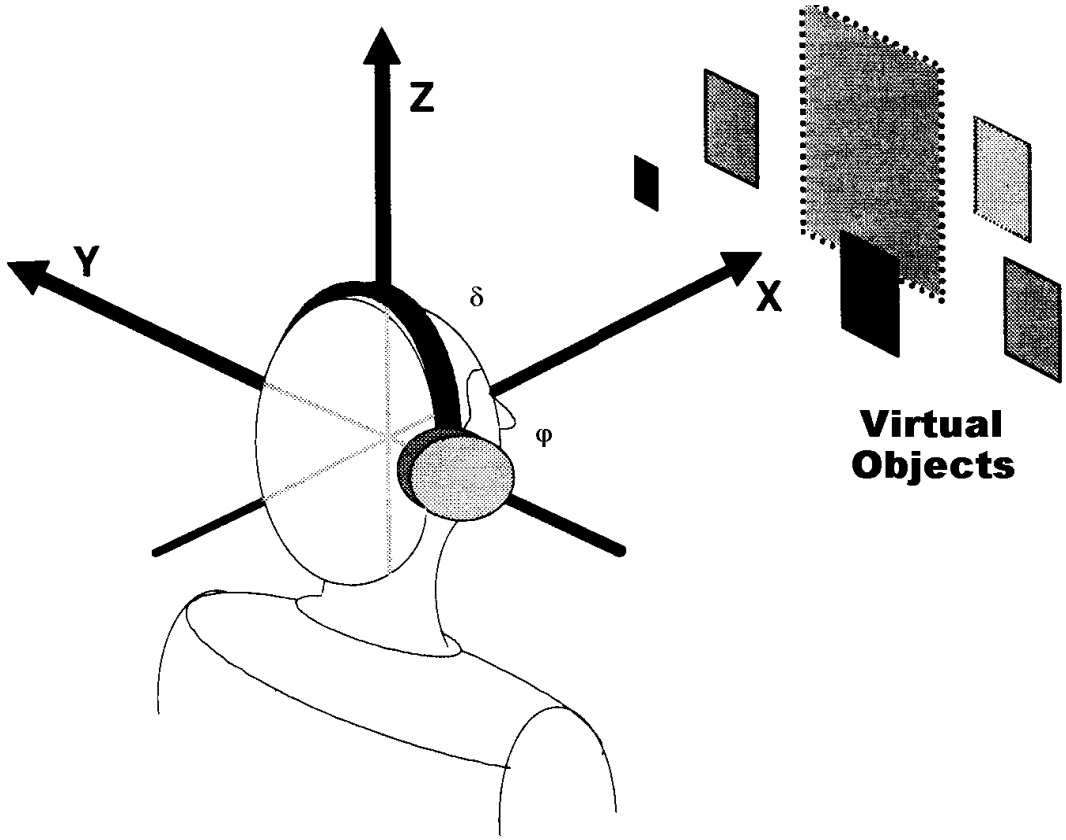
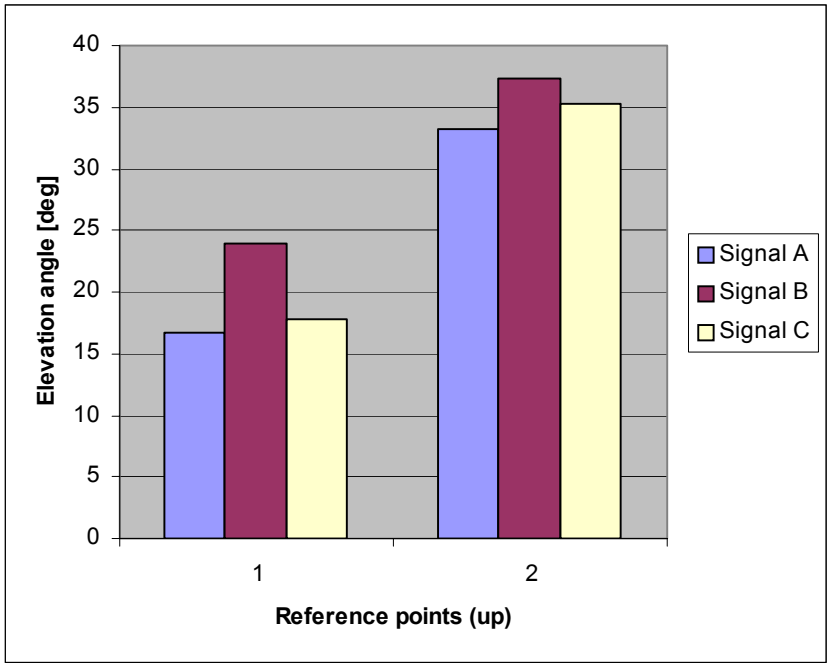


Fig.1. Illustration of virtual sources in a 2D representation [4]. The virtual acoustic surface is parallel with the Z-Y-plane. The origin is in the front of the listener: φ=δ=0°. Virtual objects move during the measurement parallel with the Y or the Z-axis in the horizontal or median plane respectively.



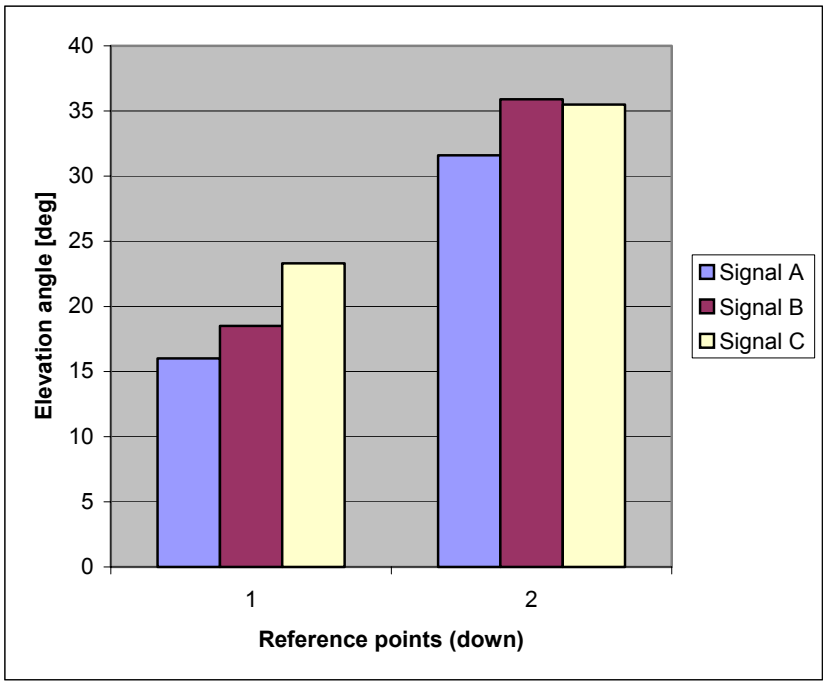
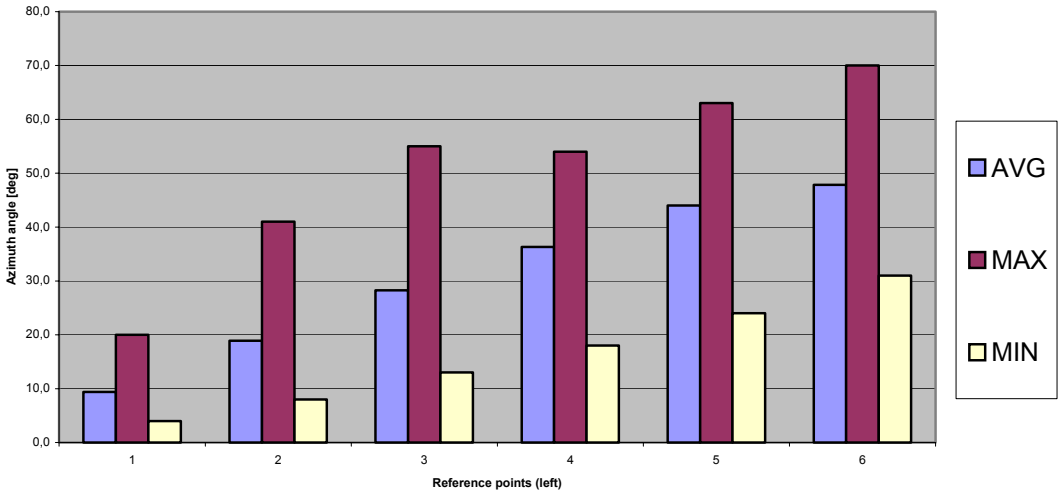
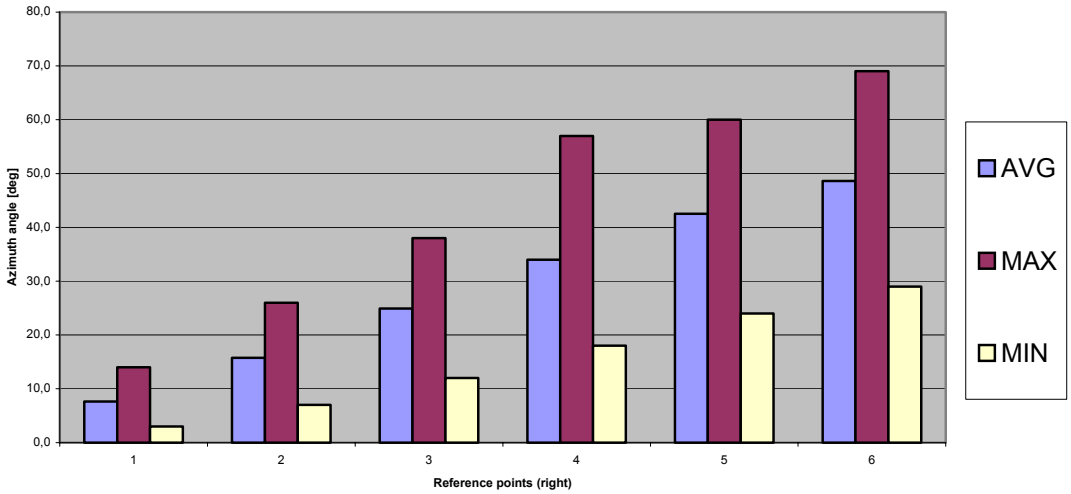


Fig.2. AVG values from the median plane for all signals.



(a)



(b)

Fig.3. MAX, MIN and AVG values for new reference points (signal A): (a) left side, (b) right side.

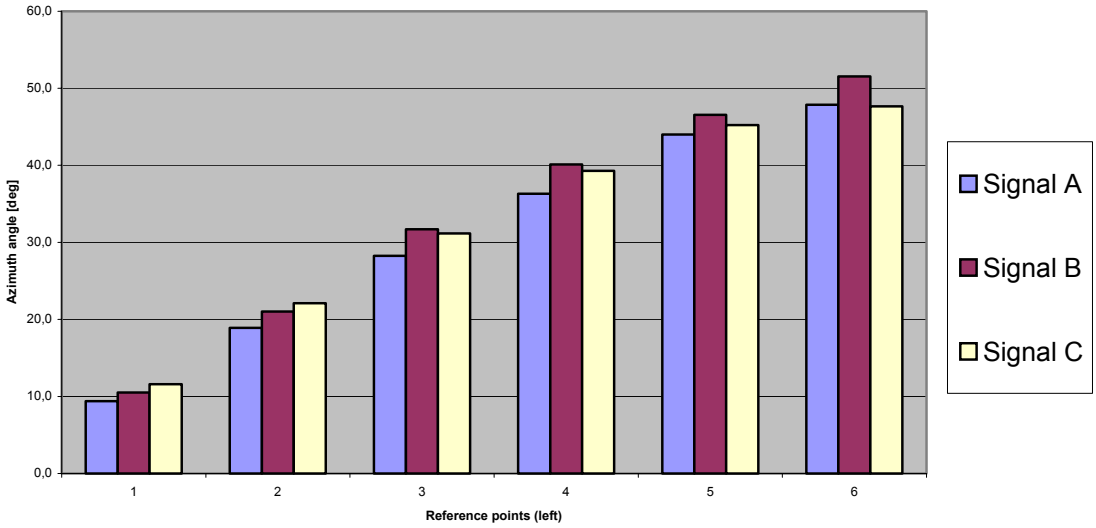


Fig.4. Localization of signals with different spectra (AVG values, left side).

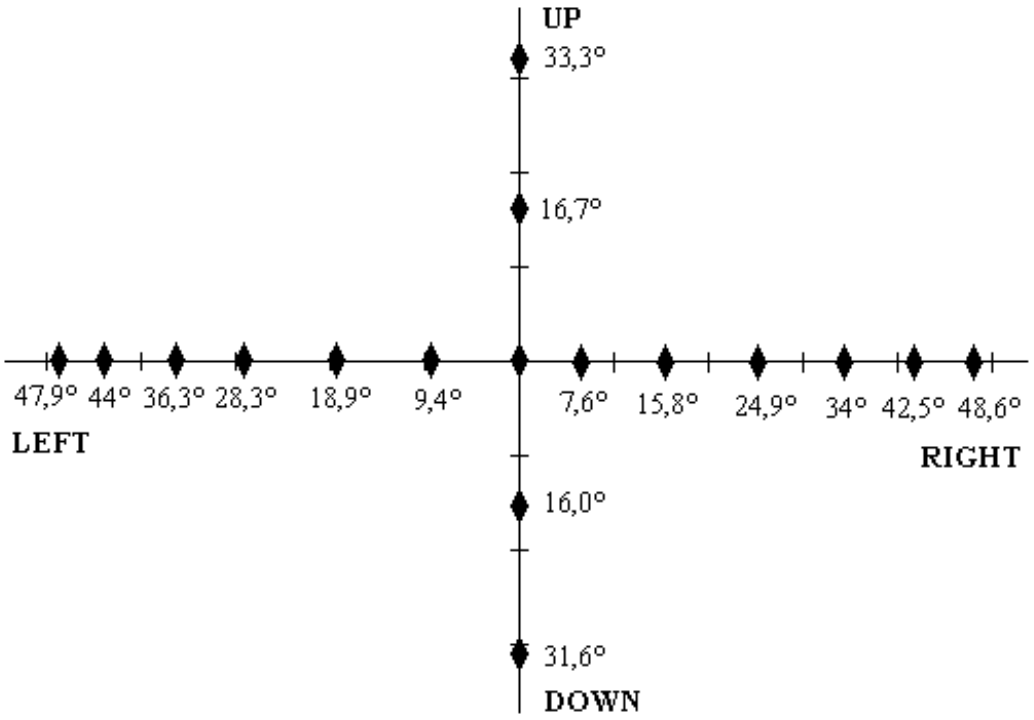


Fig. 5. Average values as possible source locations for signal A.