Human localization and performance measures

Dorte Hammershøi

Acoustics, Dept. of Electronic Systems, Aalborg University, DK-9220 Aalborg Ø, Denmark

Localization is for some scenarios and situations vital for the success of hearing, e.g. when listening out single sources in multi-source environments, or when navigating primarily by audible information. It is therefore of interest to know the limits of the human localization capacity, and its dependence on e.g. direction and distance. When addressed in laboratory experiments, the significance of other modalities are controlled in different ways, yet figures will inherently reflect properties of the test situation as well. The present paper will discuss the methodologies of localization experiments, generally and by examples.

INTRODUCTION

Localization is the process of linking spaces, namely that of linking the position of a given physical source, with that of the "position" of the listener's auditory event (if any). For most everyday natural situations, this is a highly meaningful part of the individual's formation of perceptual space. Localization supports navigation, facilitates communication between humans, and is believed to play an active role in attentive and possibly selective listening (e.g. the "cocktail party effect"). The localization performance in itself will rarely sufficiently describe a given human behaviour in a given situation, as localization may primarily be a supportive action.

On the other hand, successful localization may also occasionally be taken for granted in situations, where sound sources are spatially arranged, even if the test doesn't directly address the localization capacity. It is therefore a challenge to evaluate the significance of successful localization, when evaluating e.g. virtual acoustics as a supportive tool for communication and other tasks.

One aspect of this is the natural limits of the human localization capacity, incl. dependence on direction and distance. When addressed in laboratory experiments, the significance of other modalities are controlled in different ways, yet figures will inherently reflect properties of the test situation as well as the capacity of the individual.

METHODS

The experiments that give information of the human hearing localization performance can roughly be sub-divided in the following main groups, i) explorative studies on the absolute localization performance (e.g. Gardner, 1973; Oldfield and Parker, 1984; Makous and Middlebrooks, 1990; Butler and Humanski 1992), ii) just noticeable differences in direction (or distance), and iii) direct source identification experiments.

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

While many of the experiments in the first two categories have had the primary objective to explore the human hearing capacity as such, many of the experiments in the latter category serve to evaluate a given performance degradation that may inadvertently be imposed in e.g. inadequate control in recording and playback in binaural sound systems. For such experiments, it is also desirable that some of the primary parameters, the number of source locations, number of subjects, etc. represent a reasonably general range of options, for which reason the results also provide information of the localization performance more generally, although secondary to the investigation's objective.

Some of the key methodological aspects are discussed in the following, primarily based on the experience with localization experiments carried out to assess the principles of binaural recording and playback (Møller *et al.* 1996a), incl. the significance of individual differences (Møller *et al.*, 1996a; Møller *et al.*, 1996b) performance of artificial heads (Møller *et al.*, 1999; Minnaar *et al.*, 2001), and performance of binaural synthesis under "ideal" conditions (Hammershøi and Sandvad 1994).

Source representation

The number of source positions to have included and the position of these, whether physically or virtual, is always a trade-off between the primary objective and the representation required for this, and the wish to give ecologically valid surroundings for the individual during tests. If the objective is to study the capacity of distance estimation by hearing, then sources need to be placed at different distances, but this will experimentally not be possible in real life for many directions. Likewise, if the objective is accurate information on the capacity of the human directional hearing, sources needs to be represented in many different directions, with restricted options for representation of different distances, and for spatial resolution. Even with a relatively sparse representation of sources (and response options), it is possible to detect small differences in e.g. signal processing, as will be shown later in the examples.

Visibility

A separate and important aspect is the visibility of the sources, and what they represent. Nowadays most experiments are computer controlled, which enables the reproduction of the exact same audio stimuli over and over. Sounds are, for the latter reason, without doubt not perceived as authentic or ecologically valid. Even when speech signals are presented, it hardly presents a truly communicative situation, not even if the task challenges the intelligibility. With respect to visibility, source positions are either represented by the sound producing devices, the loudspeakers, or by purpose not visible, in which case subjects have to speculate to the origin of the sound, including the physical position of the source.

Ideally, if correctly instructed, subjects should relate to the auditory image and its position, disregarding any objects producing the sound or objectifying the possible positions for the source. Since localization is by definition about linking spaces, and to a great extent about finding the source, this ideal can probably not be mastered by the majority of typical listeners participating in given experiments.

There is also strong evidence for the significance of visual information on auditory perception, and its congruence to the task at hand. We are all familiar with the ventriloquist effect, where the spectator is easily fooled into believing that it is the puppet and not the puppeteer, which speaks. This reminds us that auditory perception is not only about the sound that enters our ears and our capacity for hearing it, but also about congruence to other modalities, in particular vision.

Response options

The response options are to a wide extent defined by the scope of investigation and physical setup, but yet the definition of the subject's task and his/her options for response influences the results (e.g. Perrett and Noble, 1995). In identification experiments, subjects' are typically instructed to assign the position of the physical source nearest to the position of the auditory event. The task instruction may focus on the fact that the auditory event doesn't necessarily coincide with the position of the physical source. Nowadays most subjects accept this relatively easily, since most will have heard e.g. stereo reproduction, where the image doesn't coincide with the sound producing device. Yet it influences the subjects' expectations, and it will depend on the scope, whether this is desirable.

One aspect of the response collection method relates to the congruence between auditory and visual space, and sense of self-center. Some studies (Arthur *et al.*, 2008) have discussed the significance of having either ego-centered response options (as e.g. pointing (Makous and Middlebrooks, 1990), using gaze direction (Hofman *et al.*, 1998) or calling out coordinates (Wightman and Kistler, 1989, and 1992)) versus exo-centric options (as e.g. indications on touch screens or spheres (Gilkey *et al.*, 1995), tablets (Møller *et al.*, 1996a), or paper drawings). It has been demonstrated, that although we interact seemingly effortless with objects in the physical world, there are sizeable misperceptions of spatial relationships, even in the nearby environment. Yet the effect of this will depend on instructions, the level of difficulty in the task presented, and to which extend the subject succeed in putting him/herself in the centre of the given response options.

EXAMPLE I: TEST IN STANDARD LISTENING ROOM

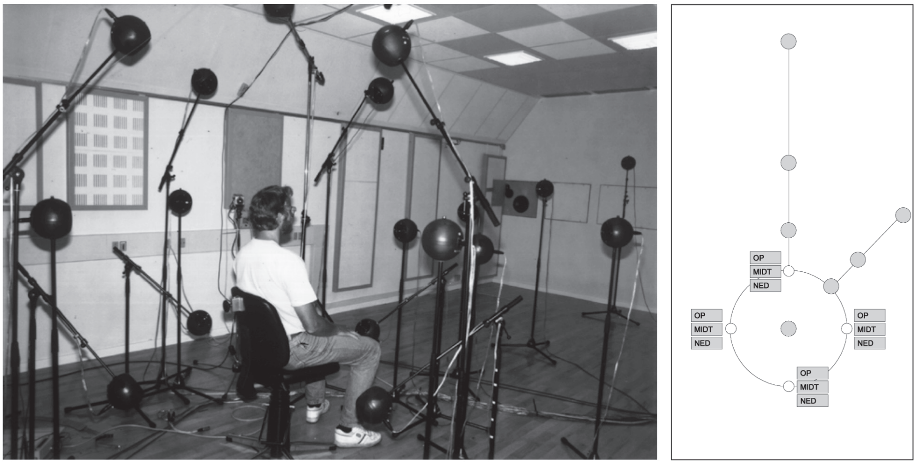

Fig. 1: Left: Photo of setup for localisation experiments with 19 loudspeakers. **Right**: Sketch appearing on tablet for response collection. Only grey zones represented valid response options. The grey boxes represented positions at different elevations, "OP" being 45º above horizontal plane, "MIDT" being in the horizontal plane, and "NED" being 45º below horizontal plane. Subjects were instructed only to look down on the tablet, when response was required, and maintain upright position during stimulus playback. This was monitored. From Møller *et al*. (1996a).

The test paradigm used in Møller *et al*. (1996a, 1996b, and 1999) is an example of localisation tests, where the test is carried out in acoustically "normal" conditions, with sound source positions at different directions and distances, with a fairly simple task for the test person, but with options for detecting even mild deteriorations of the sound reproduction. Figure 1 illustrates the test scenario. For details, please consult the original publications.

The test paradigm illustrated in Fig. 1 was used for various tests of the "authenticity" of binaural reproduction. In summary, the following was tested: Whether individual (the person's own binaural recordings) could provide a localisation performance similar to real life, whether non-individual (binaural recordings from other subjects) could, whether artificial head recordings could, whether the headphone reproduction needed individual (personal) equalisation, and more.

An example of accumulated test results for the given test paradigm is given in Fig. 2.

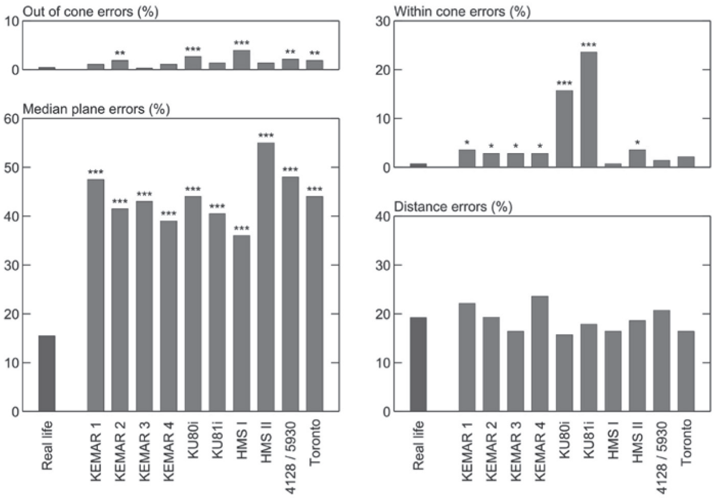

Fig. 2: Accumulated responses (percentage “correct”) from localisation tests with various artificial recordings vs. real life listening. Data from Møller *et al*. (1999).

Figure 2 summarizes the results for listening tests with different artificial heads vs. the localisation performance of the same listeners in the relevant real life situation (with sound played back over loudspeakers in the same setup).

From top left panel it can be seen, that a few of the artificial heads provide significantly more out of cone errors. This is a relatively severe error, since the cones represented in the setup is 45º apart (in horizontal plane), and the errors thus represent a confusion of sources relatively far apart. Such confusion would normally indicate that the arrival time of sound at left vs. right ear is incorrect, which could suggest that the artificial head has an inappropriate geometry.

From left lower panel it can be seen that the artificial heads without exception give more median plane errors than the corresponding real life test. These errors represent confusions between sources in the median plane, which would indicate that spectral fingerprint of the signals do not well match what the listener normally hears. This is to a great extent controlled by the detailed geometry of the outer ear, but could also be due to imperfect headphone equalisation (if not individually designed).

From the top right panel it can be seen that some artificial heads has a high number of “within cone” errors. These errors represent confusions between sources on the cone that extends out from the listeners ears’ at 45º elevation angle, e.g. the “left low” and “left high” direction. One can again speculate to the origin of these confusions, and it is remarkable that it is the artificial heads without torso that have the most of this type of error. This suggests that the torso and the related shoulder reflections are important for sound localisation for certain directions.

From the lower right panel it can be seen that all heads provided a near real distance perception. In view of the magnitude of other types of errors this would seem to indicate that distance perception is not controlled by features of the head, torso or ear, but most probably by the acoustics of the room.

In summary, the test scenario proved useful in detecting even small differences in processing, including the significance of individual vs. non-individual recording, and individual vs. non individual headphone equalisation. The latter is normally considered one of the weaker compromises to make, but as results from Møller *et al.* (1996b) showed, the difference was significant when tested.

EXAMPLE II: ANECHOIC TEST

Another localisation test scenario is illustrated in Fig. 3.

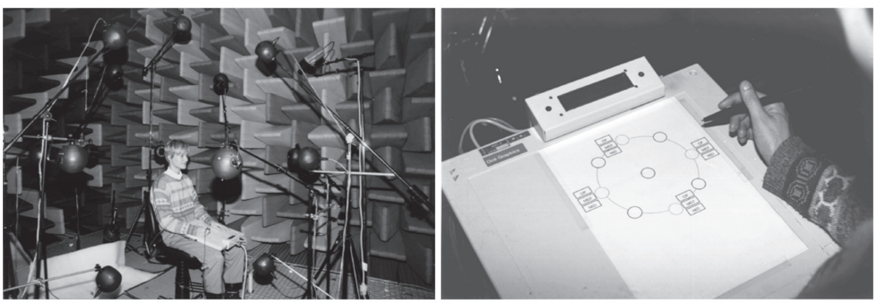

Fig. 3: Left: Photo of setup for localisation experiments with binaural synthesis. **Right**: Sketch appearing on tablet for response collection. Only grey zones represented valid response options. The grey boxes represented positions at different elevations, "OP" being 45º above horizontal plane, "MIDT" being in the horizontal plane, and "NED" being 45º below horizontal plane. Subjects were instructed only to look down on the tablet, when response was required, and maintain upright position during stimulus playback. This was monitored. From Hammershøi and Sandvad (1994).

The test scenario presented in Fig. 3, was used to assess the performance of "the best possible" binaural synthesis. This is theoretically obtained using the individual's own head-related transfer functions (HRTFs) in the synthesis, and using individual headphone equalisation.

To avoid the possible shortcomings of numerical room modelling, synthesis was carried out assuming only the direct sound transmission path from loudspeaker to listener, and did not include any representation of reflections. This has the consequence that the synthesis effectively simulates an anechoic environment, which is unnatural to most listeners, both from an acoustical and visual point of view.

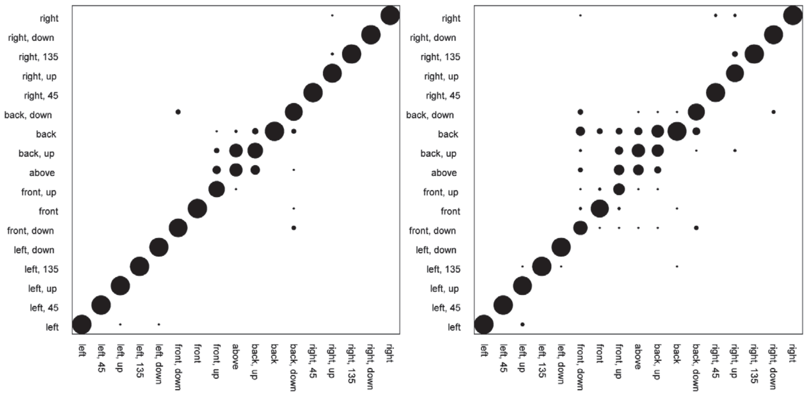

Fig. 4: Results from listening tests in anechoic chamber. Left: Stimulus vs. response for the real life playback situation. Right: Stimulus vs. response for playback of individual binaural synthesis. The area of each circle is proportional to the number of responses it holds. From Hammershøi and Sandvad (1994).

Experiments included listening to binaural signals reproduced over headphones, and to the real life setup for two types of stimuli, noise and speech.

The results of the localisation tests in anechoic chamber (Fig. 4) indicate that more errors are made with binaural synthesis, than in the corresponding real life situation.

Most errors are generally made between source directions that are within the cones of confusion represented in the setup. In both situations most confusion exists between directions in the upper hemisphere. This can be explained by the fact that the head-related transfer functions are quite similar in this region, thus the hearing has only few cues available for the localisation process.

There is also a slight over-representation of errors going from front hemisphere positions to rear hemisphere positions, again dominated by upper hemisphere confusions. Whether this generally describes the human hearing, or whether it relates to the specific setup and task is harder to determine. If the subject doesn't see him/herself in the centre of the setup distortion can occur.

In the quest for perfection of binaural synthesis (the original motivation for the study), explanations for the difference in number of errors in the two situations are also called for.

One methodological aspect relates to the "perfectly" dry simulation. With the binaural synthesis, there are really no reflections from the room, whereas in real life, any anechoic chamber will have a minimum of reflections from the setup, floor, etc. With no room-related information at all to support that the source is positioned "out there", it is possible that some sounds were perceived within the head of the listener. This could explain the (few) responses, which shifted more than 45° horizontally.

In the design of response options, it was considered, whether the subject should have the option of indicating that he/she heard the sound within the head. One reason for not including this option anyway was that there is little ecological validity in the localisation process, if the subject is left with such unnatural options for possible source positions.

This illustrates very well the most difficult challenge in the design of localisation experiments: On one hand you investigate the success with which the subject successfully links the physical world with the perceptual world. On the other hand you want the uncensored report of what the subject hears (characteristics of the auditory event) in given situations, but just by asking you bias perception.

ACKNOWLEDGEMENTS

The author would like to acknowledge the many fruitful discussions with colleagues at Aalborg University on the subject of localisation experiments, this includes in particular Henrik Møller, Michael Friis Sørensen, Clemen Boje Larsen (former Jensen), and Jesper Sandvad.

REFERENCES

Arthur, J. C., Philbeck, J. W., Sargent, J., and Dopkins, S. (**2008**). "Misperception of exocentric directions in auditory space," Acta Psychologica **129**, 72-82.

Butler, R. A., and Humanski, R. A. (**1992**). "Localization of sound in the vertical plane with and without high-frequency spectral cues," Perception and Psychophysics **51**, 182-186.

Gardner, M. B. (**1973**). "Some monaural and binaural facets of median plane localization," J. Acoust. Soc. Am. **54**, 1489-1495.

Gilkey, R. H., Good, M. D., Ericson, M. A., Brinkman, J., and Stewart, J. M. (**1995**). "A pointing technique for rapidly collecting localization responses in auditory research," Behavior Research Methods, Instruments, and Computers **27**, 1-11.

Hammershøi, D., and Sandvad, J. (**1994**). "Binaural auralization. Simulating free field conditions by headphones," Proc. 96th Audio Eng. Soc. Conv. Amsterdam, Feb. 26 – Mar. 1, 1994, preprint 3863, 1-19. Abstract in J. Audio Eng. Soc. **42**, 395.

Hofman, P. M., van Riswick, J. G. A., and van Opstal, J. (**1998**). "Relearning sound localization with new ears," Nature Neuroscience **1**, 417-421.

Oldfield, S., and Parker, S. (**1984**). "Acuity of sound localisation: A topography of auditory space: I. Normal hearing conditions" Perception **13**, 581-600.

Makous, J. C., and Middlebrooks, J. C. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2188-2200.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996a**). "Binaural technique: Do we need individual recordings ?" J. Audio Eng. Soc. **44**, 451-469.

Møller, H., Jensen, C. B., Hammershøi, D., and Sørensen, M. F. (**1996b**). "Using a typical human subject for binaural recording," Proc. 100th Audio Eng. Soc. Conv. Copenhagen, May 11-14, 1996, preprint 4157, 1-18. Abstract in J. Audio Eng. Soc. **44**, 632.

Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F. (**1999**). "Evaluation of artificial heads in listening tests," J. Audio Eng. Soc. **47**, 83-100.

Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. (**2001**). "Localization with binaural recordings from artificial and human heads," J. Audio Eng. Soc. **49**, 323-336.

Perrett, S., and Noble, W. (**1995**). "Available response choices affect localization of sound," Perception and Psychophysics **57**, 150-158.

Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868-878.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of lowfrequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648-1661.