

Investigation of multisensory spatial hearing: from the sense of audition to multisensory interactions

Klaus A J Riederer

Helsinki University of Technology, Laboratory of Computational Engineering, Cognitive Science & Technology
P.O. Box 9400, FIN-02015 HUT, Finland, E-mail: Klaus.Riederer@hut.fi, URL: <http://www.lce.hut.fi/~kar>

Background of author

MSc EE Klaus Riederer has worked as a research scientist on audio, acoustics and specifically on human spatial hearing for more than five years at the Helsinki University of Technology (HUT). He has a strong background of 10 years in professional level, and non-professional level of 15 years in technical sciences. He has gained a wide-scaled theoretical and practical know-how in various fields of engineering, mechanics, craftsmanship and photography. The emphasis of his academic research is on various aspects of human spatial hearing; theory, practice and analysis of acoustical head-related transfer function (HRTF) measurements. He has measured ca. 160 subjects' HRTFs (from 252 sound incidents) with the high-quality automatized measurement system, he has devised. He is the only person of this basic research field in Finland. He has published various papers on these issues and currently he is devising perceptual experiments. A long-time close collaboration is running with his former affiliation Laboratory of Acoustics and Audio Signal Processing (HUT), Brain Research Unit at Low Temperature Laboratory (HUT) and Unides Design Ay. (Helsinki, Finland). With the latter, groundbreaking hardware for binaural technology has been engineered.

Abstract

The study of human multisensory, especially audio-visual perception, has recently obtained increased attention with the development of virtual reality systems, teleconferencing, computer games and home theatre systems. An immersive sound scape can be created by three-dimensional sound applying head-related transfer functions that “model” human spatial hearing. Already from practice one realizes that the (spatial) hearing sense is truly multi-modal, applying also other senses, such as motion (moving the head), vision, tactile sensing etc. However, due to the technical difficulties and other complexity, serious efforts to investigate the sensory interactions (especially concerning spatial hearing) are still lacking. Therefore, a deep understanding of our second most important sense — hearing — is most incomplete. The author's research focuses on the true spatial hearing including interactions between other sensory modalities, such as audio-visual, audio-motional, audio-visual-motional. This basic research has strong interdisciplinary connections in various fields of science and numerous application areas. Novel research paradigms will be addressed under various objectives.

Keywords: 3-D sound, audio-visual perception, basic research, heterosensory, homosensory, HRTF, multidisciplinary, multi-modal perception, spatial hearing, virtual reality

Desires of author

The author is anticipating extensive discussions on the current status of other scholars' research activities, interests and future plans concerning the various issues on multi-modal perception and applications thereof. Various viewpoints are to be considered, ranging from technical (e.g., audio-visual synchrony), methodological (e.g., perceptual metrics, analysis methods) to psychological (e.g., cognition) aspects.

The author's main interest is in multidisciplinary basic research, focused on a) investigating accurately the human spatial hearing (applying HRTFs) and b) studying sensory interactions between spatial hearing and other sensory modalities, such as audio-visual, audio-motional, audio-tactile, audio-visual-motional and audio-visual-motional-tactile. The ultimate aim is to understand better the enigma of the human being, her/his behaviour and capabilities. To accomplish this, the author is seeking after skilled scholars with overlapping interests for possible collaboration, believing that elaborate methods, strong will and intelligent elucidation will lead to able scientific results.

Introduction

A natural immersive sound scene can be accomplished by three-dimensional sound applying so-called head-related transfer functions (HRTFs) that comprise the homosensory cues of spatial hearing. In reality, spatial hearing applies also other sensory modalities, a fact that is demonstrated by the following examples of cross-modal induction in perception. These matters set the fundamental framework for the author's research and are hence discussed below.

Homosensory cues of spatial hearing

The *primary localization* cues, the *binaural interaural level* and *time differences* (ILDs and ITDs) were formulated to the public already in 1907 by Lord Rayleigh, after collection of his previous hypothesis (1877) and Giovanni Venturi's founding research almost a century earlier. Localization performances a) on the median plane and b) monaurally (see, e.g., Hebrank and Wright 1974ab) prove that ILD and ITD cannot be the only localization cues. In headphone listening sound is often perceived inside the head to the middle (*inside-the-head-locatedness*), which is neither explained by the duplex theory (Yost and Hafter 1987). The necessary *spectral filtering* caused by the body, torso, head and especially the *pinnae* is denoted as the *monaural localization cue*. Obviously, binaural localization yields more precise accuracy than monaural (e.g., Blauert 1997). These cues apply only one sensory modality (hearing) and are thus *homosensory*. They are embodied in the *head-related transfer functions* (HRTFs) that involve measured (or modeled) responses of a sound source in free field to a point in the ear canal (Blauert 1997, Møller 1992). Spatial hearing perception based on HRTFs has been widely investigated in recent years (see, e.g., Wightman et. al 2001 and Møller et. al 2001).

In literature, a number of different spatial hearing models applying monosensory (acoustic) cues have been proposed: *physical*, *psychoacoustical* (*behavioral*) and *functional*. All the models need to generalize over inter-personal differences (and inaccuracies due to various reasons) in human anatomy and perception. Furthermore, strong enough experimental data is required to support the hypothesis.

Schelhammer presented a hypothesis of the sound gathering effect of the pinna already in 1684. Now, after three hundred years of investigation, there still remains research work to solve the puzzle of the homosensory spatial hearing. At least from the neurophysiological point of view, research is still in an infant level. The reason for this is also obvious: the introduction of the modern brain imaging techniques and binaural techniques have finally made possible non-invasive (cortical) measurements, in which human binaural neural processing of natural three-dimensional sounds can be investigated.

Multi-modal interactions in spatial hearing

If the location of the auditory event does not unequivocally correlate with the sound signals at the eardrums, also further supplemental sensory information is needed (Blauert 1997). In practice this means that at least in ambiguous cases humans incorporate inter-sensory information for determining the locations and distances of sound sources. The *bone-conduction* theories are homosensory, other theories of spatial hearing are heterosensory. The latter involve interaction between audition and other modalities. They are called *motional*, *visual*, *vestibular* and *tactile* theories.

The so-called *motional* (or *motoric*) theories describe relationships between the position of the auditory event and the variations to the ear input signals during head movements (Blauert 1997). They also characterize changes in other attributes, such as loudness and tone color, which the subject can utilize in sound localization. Humans (and animals) tend to move naturally their heads; often only a slight head movement will remarkably improve the localization accuracy.

It is also evident that *vision* has a powerful effect on spatial hearing, usually seeing the sound source improves the localization. Auditory localization is rather poor compared to vision under everyday conditions, typical errors of localization are 4-10° for the horizontal plane and much worse for elevations. In the sharpest detection area, i.e., the forward horizontal plane, a minimum angular separation between two (pure-tone) sound sources as small as 1° can be detected (Blauert 1997). According to Begault (1999), in multi-modal interaction experiments (e.g., audio-visual) the absolute accuracy of a particular modality should be regarded secondary. Instead, the evaluations have to focus on the overall "quality" of the perception, where positions produced by two modalities are judged relative to one another.

Vestibular theories present the organ of balance would have some direct effect on the spatial hearing, besides the obvious indirect influence in strong accelerations etc. These hypotheses, as well as *tactile* and *vibro-tactile* (e.g., “feeling the (live) music”) *theories*, present that interaction between modalities can make a particular component of the modality in question either more or less noticeable. Formerly, these approaches have not been considered meaningful in regard to spatial hearing. However, this attitude has changed as the research on multimodal interaction and perception has gained more interest. For example, force feedback is regarded an important technology development for virtual reality applications (Begault 1999).

Cross-modal induction in human perception

In certain conditions (temporal or spatial asynchrony or contradictory stimuli etc.) cross-modal induction is produced in human perception. This non-typical operation, e.g., modal override or fusion in sensory processing, presents useful starting points for the research of human perception and, e.g., cognitive systems. In the following, such cases of audio-visual perception are presented. These topics are the basic problem fields of *auditory scene analysis (ASA)*, which is a method to understand brain and auditory system processing of complex sound environments.

In some cases vision overrides the auditory cues in (spatial) hearing; e.g., when watching television the sound seems to come from the screen, though it actually comes from the loudspeakers nearby. This so-called *ventriloquism effect* is defined as the spatially biased perception of the auditory stimulus from the same point as the visual stimulus (Shinn-Cunningham *et al.* 1997). Vision does not only reinforce the spatial auditory perception, especially in equivocal directions, but it also gives a great improvement in distance judgments of sound sources. However, vision can also diminish an auditory event. This can happen, e.g., in a concert hall that gives to the listener an inconsistent image between the auditory and visual space. Once the person closes her/his eyes, the music seems to sound better, because vision gives no disturbing information. Also, steering attention to one modality may improve one’s performance, e.g., one can concentrate better to music with eyes closed. Furthermore, Paulsen and Ewertsen (1966) report a so-called *audio-visual reflex* as an involuntary turning of the eyeballs to front the direction of the sound source. This reflex requires at least some level of awareness of the direction of the sound source.

Many audio-visual psychophysical studies examine the influence of visual stimulation to auditory detection and vice versa. Welch and Warren (1986) present in their review that it is easier to detect auditory signal with the presence of a visual stimulus. In overall, the sensitivity to audio-visual events is increased under multi-modal conditions. The effect of *perceptual defense* states that presenting emotionally reserved auditory stimuli (i.e., words) raises the threshold of visual perception, and vice versa (Hardy and Legge 1968). These results have later been found controversial because of strong differences between subjects. In any case, the theory has proven to be beneficial in revealing pathological cases, such as various syndromes.

Audio-visual interaction on speech intelligibility is well-known from the *McGurk effect*, where conflicting audio-visual cues affect the intelligibility, and can create a fusion response that differs from both the auditory and visual stimulus (McGurk and McDonald 1976).

The *cocktail-party effect* is known as the ability to focus listening attention to on a single talker amidst a cacophony of conversations and background noise (Cherry 1953). The ability still exists when listening to high quality binaural recordings. Regardless of the wide-scale research, the underlying explanation for this effect is still not clear. Apparently, it is linked to human speech production system, auditory system and/or high-level perceptual and language processing systems.

Research paradigms

The basis for the author’s research is the employment of the HRTF data that he has measured since the year 1997. He has set the following research paradigms that have not been solved in full by other scholars:

- What are the features that constitute the idiosyncratic features in individual HRTFs?
 - What are the roles of clothes, hair, hairstyle and headgear?
 - What is the effect of anatomical attributes, such as cranial and pinnae size?

- To what extent are these idiosyncratic features of HRTFs necessary for reproducing perceptually accurate 3-D sound?
- What are the errors and consequences in using non-individual HRTFs for perceptual experiments, compared to the use of individual HRTFs?
- Is it possible to find a (more) generic HRTF model (based on the previous)?
 - Would such a model be good enough for scientific experiments, i.e., What are the errors and consequences in using such a generic HRTF model for perceptual experiments, compared to the use of individual HRTFs?
- What is the accuracy of auditory perception considering the whole 3-D space?
 - How does the perception change in the presence of distracters in other sensory modalities?
 - How does the perception change in the presence of supporting information from other modalities?
- What are the neurophysiological (cortical) findings related to the paradigms above, and what do they reveal about human multimodal information processing?

These paradigms will be addressed in the research objectives discussed below. The objectives are highly inter-related, and the weight is put to basic research — to understand how the human spatial hearing really works in its all complexity.

Objective I: Analysis of HRTF quality

The investigation of the *fundamental HRTF data quality* (Riederer 1998a, Riederer and Karjalainen 1998, Riederer 2000) is vital, because only this way the basis for the whole spatial hearing research can be confirmed. Repeatability investigations demonstrate the (high) *measurement system quality* (Riederer 1998b, Riederer 2000). Non-quantitative characteristics of HRTFs, based on the individual anatomy, are investigated by *structural HRTF analysis*.

Objective II: Quantitative analysis of HRTFs

Quantitative matters are strongly interlinked to Objective I, allowing direct utilization of methods and results between Objective I & II. The latter aims to the most enchanting research result on spatial hearing ever: a *generic HRTF model*. Such a model would give a *deeper universal understanding of spatial hearing*: to comprehend in detail (e.g., as a function of azimuth and elevation angle, *person-independently*) how the basic binaural cues are constituted. It is most obvious that there would not exist a *single pair of ears* (“golden HRTFs”) but perhaps independent groups with common (idiosyncratic) features (e.g., “big heads/ears”, “small heads/ears”). The Objective II concentrates on various issues around *HRTF classification* (Riederer 2000) and *distance-dependence HRTF analysis* (Riederer 1998a, Huopaniemi and Riederer 1998).

Objective III: Conversion methods

Objective III concentrates on digital signal processing issues, in order to make possible the empirical verification to the results of Objectives I and II, when applying binaural recordings and multichannel recordings (e.g., Dolby Digital program material) as test stimuli. The focus is to produce natural three-dimensional sound listening experiences by headphones for perceptual studies discussed in Objectives IV & V. This necessitates the implementation of *equalization* (free-field, diffuse field) and *conversion methods*, e.g., multichannel–binaural and individual binaural–generic binaural.

Objective IV: Psychophysiological studies

Careful psychophysiological experiments have been planned in order to address the research paradigms stated earlier. Differing from all the published research, a large amount of directions will be covered. This will reveal the true capability of human spatial hearing *covering the whole 3-D auditory space*. In order to make this possible, novel methods have been devised. Basically, virtual sources (sounds created via HRTFs, reproduced by headphones) are compared to the reference sources (loudspeakers at fixed positions). The subject is sitting on a rotating turntable; thus a great number of sound incidents are efficiently investigated. *Auditory, visual and motional, uni-, bi-, and tri-modal interactions* are to be investigated.

Objective V: Neurophysiological studies

Collaborative research (since 1997) continues with the Low Temperature Laboratory (LTL), Brain Research Unit. Individual and non-individual HRTFs will be applied for the investigation of the binaural neural processing with the 306-channel MEG instrument, utilizing the custom-made high quality tube headphones. The special interest is in comparing the metrics in the cortical representations of the three-dimensional auditory space in front-back, left-right and horizontal planes. Also multimodal experiments applying both EEG and MEG instruments will be performed.

Postscriptum

The theoretical and empirical results of the HRTF investigation by the author will be utilized in the audio-visual speech perception research at the Cognitive Science and Technology group, Laboratory of Computational Engineering (HUT). The author is also at the Finnish Graduate School of Electronics, Telecommunication and Automatization, and its financial support is greatly acknowledged.

The author's general research topics of the author involve experimental and theoretical analysis of hearing, vision, attention, motor control and haptic perception. The behavioral studies will touch many of the matters described in Introduction, including evaluation of reaction time, (a)synchrony, modal override and fusion, errors in perception etc. Also fundamental and applied exploration on virtual reality and virtual environments could be considered. Ultimately, a more comprehensive model of the human spatial hearing would be postulated.

References

- Begault D. R., 1999. Auditory and non-auditory factors that potentially influence virtual acoustic imagery. *Proceedings of 16th Audio Engineering Society conference*, Rovaniemi, April 10-12, pp. 1-14.
- Blauert J., 1997. *Spatial Hearing, The Psychophysics of Human Sound Localization*. Revised Edition, MIT Press, Cambridge, Massachusetts, USA, 494 p.
- Cherry E. C., 1953. Some experiments on the recognition of speech with one and with two ears. *J. Acoust. Soc. Amer.*, vol. 25, pp. 975-979.
- Hardy G. R. and Legge D., 1968. Cross-modal induction of changes in sensory thresholds. *Quarterly Journal of Experimental Psychology*, vol. 20, pp. 20-29.
- Hebrank, J., and Wright, D. 1974a. Are two ears necessary for localization of sound sources on the median plane? *J. Acoust. Soc. Am.*, vol. 56, no. 3, pp. 935-938.
- Hebrank, J., and Wright, D. 1974b. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.*, vol. 56, no. 6, pp. 1829-1834.
- Huopaniemi J. and Riederer K. A. J., 1998. Measuring and modeling the effect of source distance in head-related transfer functions. In *ICA/ASA 1998 conference*, Seattle, USA, 20-26.6.1998.
- McGurk H. and McDonald J., 1976. Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Møller H. et. al, 2001. Department of Acoustics, Aalborg University, Denmark. Publications list at <http://acoustics.auc.dk/publications/pubframe.html>
- Møller, H. 1992. Fundamentals of binaural technology. *Applied Acoustics*, vol. 36, pp. 171-218.
- Paulsen J. and Ewertsen H. W., 1966. Audio-visual reflex. *Acta Oto-laryngol. Suppl.*, vol. 224, pp. 217-221.
- Rayleigh Lord (Strutt J. W.) (Ed.), 1907. On our perception of sound direction. *Philosophical magazine*, vol. 13, pp. 214-232. Cited in Blauert (1997).
- Riederer K. A. J and Karjalainen M., 1998. DSP aspects of head-related transfer function measurements. In *IEEE Nordic Signal Processing Symposium, NORSIG'98*, Vigso Holiday Resort, Denmark, 8-11.6.1998.
- Riederer K. A. J., 1998a. *Head-related transfer function measurements*. Master's Thesis. Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, 134 p.
- Riederer K. A. J., 1998b. Repeatability analysis of head-related transfer function measurements, *105th Audio Engineering Society Convention*, San Francisco, Sept. 26-29. Preprint no. 4846, 62 p.
- Riederer K. A. J., 2000. Computational Quality Assessment of HRTFs In *European Signal Processing Conference EUSIPCO (X)*, 5-8.9.2000, Tampere, Finland.
- Schelhammer G. C., 1684. De auditu, liber unus. Lugduni Batavorum. Cited in Békésy G. von, 1960. *Experiments in hearing*. McGraw-Hill, New York, USA. Cited in Blauert (1997).
- Shinn-Cunningham B., Lehnert H., Kramer G., Wenzel E. and Durlach N., 1994. Auditory displays, pp. 611-663. In Gilkey R. and Anderson T. (Eds.) (1997), *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum Associates, New Jersey, USA, 795 p.
- Welch R. B., and Warren D. H., 1986. Intersensory Interactions. In *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Ch. 25. New York, Wiley.
- Wightman F. L. et. al, 2001. Hearing Development Research Laboratory, Waisman Center, University of Wisconsin. List of spatial hearing publications at <http://www.waisman.wisc.edu/hdrl/index.html>
- Yost W. A and Hafter E. R., 1987. Lateralization. In Yost W. A. and Gourevitch G. (Eds.) *Directional Hearing*. Springer-Verlag, New York, pp. 49-84.