Perception and Cognition Models for Human Data Interaction
Annotated Bibliography

Sharath Chandra Ramakrishnan
(SharathChandraRam@utdallas.edu)
School of Arts Technology and Emerging Communication
University of Texas at Dallas

Sunday 5th May, 2019

References

[1] G. J. Brown and D. Wang, *Separation of Speech by Computational Auditory Scene Analysis.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 371–402.

> The term auditory scene analysis (ASA) refers to the ability of human listeners to form perceptual representations of the constituent sources in an acoustic mixture, as in the well-known 'cocktail party' effect. Accordingly, computational auditory scene analysis (CASA) is the field of study which attempts to replicate ASA in machines. Some CASA systems are closely modelled on the known stages of auditory processing, whereas others adopt a more functional approach. However, all are broadly based on the principles underlying the perception and organization of sound by human listeners, and in this respect they differ from ICA and other approaches to sound separation. In this chapter, the authors review the principles underlying ASA and show how they can be implemented in

CASA systems. The authors also consider the link between CASA and automatic speech recognition, and draw distinctions between the CASA and ICA approaches.

[2] M. D. Cooke, G. J. Brown, M. Crawford, and P. H. Green, "Computational auditory scene analysis: listening to several things at once." *Endeavour*, vol. 17 4, pp. 186–90, 1993.

The problem of distinguishing particular sounds, such as conversation, against a background of irrelevant noise is a matter of common experience. Psychologists have studied it for some 40 years, but it is only comparatively recently that computer modelling of the phenomenon has been attempted. This article reviews progress made, possible practical applications, and prospects for the future.

[3] G. Eckel, "Immersive audio-augmented environments: the listen project," in *Proceedings Fifth International Conference on Information Visualisation*, 2001, pp. 571–573.

The author reports on the LISTEN project, a research project funded by the European Commission in the context of the Information Society Technology (IST) program. LISTEN, which started in January 2001, will provide users with intuitive access to personalized and situated audio information spaces while they naturally explore everyday environments. A new form of multisensory content is proposed to enhance the sensual impact of a broad spectrum of applications ranging from art installations to entertainment events. This is achieved by augmenting the physical environment through a dynamic soundscape, which users experience over motion-tracked wireless headphones. Immersive audio-augmented environments are created by combining high-definition spatial audio rendering technology with advanced user modeling methods. These allow for adapting the content to the users' individual spatial behavior. The project will produce several prototypes and a virtual reality based authoring tool. Technological innovations will be validated under laboratory conditions, whilst the prototypes will be evaluated in public exhibitions

[4] B. Fröhlich, S. Barrass, B. Zehner, J. Plate, and M. Göbel, "Exploring geo-scientific data in virtual environments," in *Proceedings of the Conference on Visualization '99: Celebrating Ten Years*, ser. VIS '99. Los Alamitos, CA, USA: IEEE Computer Society Press, 1999, pp. 169–173. [Online]. Available: http://dl.acm.org/citation.cfm?id=319351.319371

> This paper describes tools and techniques for the exploration of geo-scientific data from the oil and gas domain in stereoscopic virtual environments. The two main sources of data in the exploration task are seismic volumes and multivariate well logs of physical properties down a bore hole. The authors have developed a props-based interaction device called the cubic mouse to allow more direct and intuitive interaction with a cubic seismic volume. This device effectively places the seismic cube in the user's hand. Geologists who have tried this device have been enthusiastic about the case of use, and were adept only a few moments after picking it up. The authors have also developed a multi-modal visualisation and sonification technique for the dense, multivariate well log data. The visualisation can show two well log variables mapped along the well geometry in a bivariate colour scheme, and another variable on a sliding lens. A sonification probe is attached to the lens so that other variables can be heard. The sonification is based on a Geigercounter metaphor that is widely understood and which makes it easy to explain. The data is sonified at higher or lower resolutions depending on the speed of the lens. Sweeps can be made at slower rates and over smaller intervals to home in on peaks, boundaries or other features in the full resolution data set.

[5] T. Hermann and H. Ritter, "Listen to your Data: Model-Based Sonification for Data Analysis," in *Advances in intelligent computing and multimedia systems*, G. E. Lasker and M. R. Syed, Eds. Int. Inst. for Advanced Studies in System research and cybernetics, 1999, pp. 189–194.

> Sonification is the use of non-speech audio to convey information. The authors in this paper are developing tools for interactive data exploration, which make use of sonification

for data presentation. In this paper, model-based sonification is presented as a concept to design auditory displays. Two designs are described: (1) particle trajectories in a "data potential" is a sonification model to reveal information about the clustering of vectorial data and (2) "data-sonograms" is a sonification for data from a classification problem to reveal information about the mixing of distinct classes.

[6] D. Oldoni, B. D. Coensel, M. Boes, M. Rademaker, B. D. Baets, T. V. Renterghem, and D. Botteldooren, "A computational model of auditory attention for use in soundscape research," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 852–861, 2013. [Online]. Available: https://doi.org/10.1121/1.4807798

In this paper, a model of auditory attention to environmental sound is presented, which balances computational complexity and biological plausibility. Once the model is trained for a particular location, it classifies the sounds that are present in the soundscape and simulates how a typical listener would switch attention over time between different sounds. The model provides an acoustic summary, giving the soundscape designer a quick overview of the typical sounds at a particular location, and allows assessment of the perceptual effect of introducing additional sounds

[7] S. Pauletto and A. Hunt, "Interactive sonification of complex data," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 923 – 933, 2009, special issue on Sonic Interaction Design. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581909000706

In this paper authors present two experiments on implementing interaction in sonification displays: the first focuses on recorded data (interactive navigation) and the second on data gathered in real time (auditory feedback). Complex synthesised data are explored in the first experiment to evaluate how well the known characteristics present in the data are distinguished using different interaction methods, while real medical data (from physiotherapy) are used for the second. The addition of interaction to the exploration of sonified recorded data improves the system usability (efficiency, effectiveness and user satisfaction), and the real-time sonification of complex physiotherapy data can produce sounds with timbral characteristics that audibly change when important characteristics present in the data vary.

[8] J. R. Pomerantz and M. C. Portillo, "Emergent features, gestalts, and feature integration theory," in *From Perception To Consciousness: Searching with Anne Treisman*, Jan 2014.

The question that lingered on with Treisman was of did all perceptual wholes go through the 2-stage process of pre-attentive processing followed by attention selection? In a 1977 study, Pomerantz and colleagues devised an experiment in which they found that when certain elements are added to modify the form of a stimulus, new features emerge from what was previously thought to to be basic and primal individual features.

[9] B. Poppinga, C. Magnusson, M. Pielot, and K. Rassmus-Gröhn, "Touchover map: Audio-tactile exploration of interactive maps," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '11. New York, NY, USA: ACM, 2011, pp. 545–550. [Online]. Available: http://doi.acm.org/10.1145/2037373.2037458

This article reports on a preliminary study, which investigates if vibration and speech feedback can be used in order to make a digital map on a touch screen device more accessible. The researcher test if vibration feedback combined with speech, triggered as the finger moves over relevant map objects, works to make sense of the map content. The study results indicate that it is indeed possible to get a basic overview of the map layout even if a person does not have access to the visual presentation. In the conclusions the interaction problems are indentified and suggestions for future improvements are given

[10] G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. van Dantzich, "Data mountain: Using spatial memory for document management," in *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '98. New York, NY, USA: ACM, 1998, pp. 153–162. [Online]. Available: http://doi.acm.org/10.1145/288392.288596

[11] N. Saint-arnaud and K. Popat, "Analysis and synthesis of sound textures," in *in Readings in Computational Auditory Scene Analysis*, 1995, pp. 125–131.

> The sound of rain or of a large crowd are examples of sound textures. based probability model is used to characterize the high level of sound textures. The model is then used to resynthesize textures that are perceptually similar to originals (training data). They authors present a method for resynthesis of sound textures, like the sound of rain, large crowds, fish tank bubbles, photocopiers and myriad others. Defining sound texture is no easy task. Most people will agree that the noise of a fan is a likely sound texture. Some other people would say that a fan is too bland, that it is only a noise. The sound of rain, or of a crowd are perhaps better textures. But few will say that one voice makes a texture.

[12] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc*, vol. 47, no. 9, pp. 675–705, 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12095

> The theory and techniques for virtual acoustic modeling and rendering are discussed. The creation of natural sounding audiovisual environments can be divided into three main tasks: sound source, room acoustics, and listener modeling. These topics are discussed in the context of both non-real-time and real-time virtual acoustic environments. Implementation strategies are considered, and a modular and expandable simulation software is described.

[13] S. Smith, R. D. Bergeron, and G. G. Grinstein, "Stereophonic and surface sound generation for exploratory data analysis," in *Proceedings*

of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '90. New York, NY, USA: ACM, 1990, pp. 125–132. [Online]. Available: http://doi.acm.org/10.1145/97243.97264

The analysis and interpretation of very high dimensional data require the development and use of data presentation techniques that harness human perceptual powers. The University of Lowell's Exploratory Visualization project (Exvis) aims at designing, implementing, and evaluating perceptually-based tools for data presentation using both visual and auditory domains. This paper describes several auditory data presentation techniques, including the generation of stereophonic sound with apparent depth and sound that appears to emanate from a two-dimensional area. Both approaches can produce sound with auditory texture.

[14] J. Thompson, J. Kuchera-Morin, M. Novak, D. Overholt, L. Putnam, G. Wakefield, and W. Smith, "The allobrain: An interactive, stereographic, 3d audio, immersive virtual world," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 934 – 946, 2009, special issue on Sonic Interaction Design. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581909000688

This paper describes the creation of the Allobrain project, an interactive, stereographic, 3D audio, immersive virtual world constructed from fMRI brain data and installed in the Allosphere, one of the largest virtual reality spaces in existence. This paper portrays the role the Allobrain project played as an artwork driving the technological infrastructure of the Allosphere. The construction of the Cosm toolkit software for prototyping the Allobrain and other interactive, stereographic, 3D audio, immersive virtual worlds in the Allosphere is described in detail. Aesthetic considerations of the Allobrain project are discussed in relation to world-making as a means to understand and explore large data sets.

[15] H. Zhao, C. Plaisant, and B. Shneiderman, ""i hear the pattern": Interactive sonification of geographical data patterns," in *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI

EA '05. New York, NY, USA: ACM, 2005, pp. 1905–1908. [Online]. Available: http://doi.acm.org/10.1145/1056808.1057052

Interactive sonification (non-speech sound) is a novel strategy to present the geographical distribution patterns of statistical data to vision impaired users. The authors discuss the design space with dimensions of interaction actions, data representation forms, input devices, navigation structures, and sound feedback encoding. Two interfaces were designed, one using a keyboard and another using a smooth surface touch tablet. A study with three blind users shows that they are able to perceive patterns of 5-category values on both familiar and unknown maps, and learn new map geography, in both interfaces.