Interval scaling of virtual sound sources when listening with one ear

Daniel E. Shub[1] and Virginia M. Richards[2]

1 *National Biomedical Research Unit in Hearing, School of Psychology, University of Nottingham, United Kingdom*

2 *Department of Psychology, University of Pennsylvania, Philadelphia, United States of America*

When listening monaurally to sounds of fixed level, listeners can discount the "un-natural" infinite interaural level difference (ILD) and use loudness cues to determine the azimuthal location of a sound source. Here, we investigated the ability to use a spectral shape cue to indicate the azimuthal location of a virtual sound source. Subjects positioned a visual pointer to indicate the relative similarity between the target stimulus (a random-level multi-tone stimulus with a parametrically varied bandwidth and spectral density) and a perceptual anchor. For all of the bandwidths and spectral densities tested, the correlation between the responses of the subjects and the source's azimuth exceeded 0.22, with a maximum of 0.81. In contrast, the correlation between the subjects' responses and the stimulus levels were near zero. A linear decision model that utilized the spectral shape of the perceptual anchor as a template accurately predicted the dependence of subjects' responses on the source location. Overall, the psychophysical results coupled with the model predictions suggest that subjects can ignore uninformative ILD information and use spectral shape information to determine the azimuthal location of a sound source when the sources' level is random.

INTRODUCTION

For a monaurally presented stimulus the effective interaural level difference (ILD) is consistent with a source that is located "at the ear", independent of the actual location of the sound source. When monaural listeners are asked to describe the perceived location of a sound, there is a strong bias towards the hearing side and there is little dependence of the perceived location on the actual source location (e.g., Slattery and Middlebrooks, 1994; van Wanrooij and van Opstal, 2007; van Wanrooij and van Opstal, 2004; Wightman and Kistler, 1997). Blauert (1982), however, noted that, in regards to binaural localization, "an estimation of the direction of the sound source is not equivalent to a description of the position of the auditory event". This distinction is particularly relevant for monaural localization because of the uninformative ILD. Monaural listeners may be able to estimate accurately the location of a sound source by incorporating information from the overall level (loudness) and spectral shape (timbre) and ignoring the ILD and the perceived location. The current study uses an interval-scaling paradigm to investigate the ability of monaural listeners to estimate the azimuth of a sound source.

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.

There have been a few studies of the ability of monaural listeners to identify the azimuth of a sound source when there are a small number of potential locations. Performance in monaural azimuth discrimination tasks (i.e., two potential locations) with fixed level sources (Häusler *et al*., 1983) and random level sources (Shub *et al*., 2008) demonstrate that the frequency and location dependent nature of the head shadow (e.g., Shaw, 1974; Shaw and Vaillancourt, 1985) provides information about the source location. Shub *et al*. (2008) postulated that a model of spectral shape analysis based on weighted linear combinations of the observed levels within each frequency channel (e.g., Durlach *et al*., 1986; Berg, 2004) could be used to predict discrimination performance. This type of modeling approach could likely be extended to predict performance in monaural azimuth identification tasks (Fisher and Freedman, 1968; Freedman and Fisher, 1968; Shub *et al*., 2008), where the number of potential locations does not exceed the limits of working memory (Miller, 1956).

In "true localization" tasks, the number of potential locations far exceeds the limits of working memory and the responses are not limited to a discrete set of locations. The estimated location in these types of tasks is still potentially derived from a weighted linear combination of the observed levels within each frequency channel, but it is not obvious what weighting pattern should be used. One solution to this problem is to ask the subjects to make a comparison to a perceptual cue/anchor (i.e., Braida *et al*., 1984) and then assume that this cue/anchor provides the weighting template. This weighted linear combination approach assumes that estimating the azimuth of a source is based on spectral shape processing mechanisms as opposed to the perceived location which likely depends on specialized "localization" mechanisms.

The relative importances of the estimated and perceived locations on the benefits of spatial hearing are unknown. Resolving conflicts between the perceived and estimated locations may increase the cognitive load resulting in slower orientations to sounds of interest. Being able to estimate the location of a sound of interest, however, may be sufficient for maximizing the signal-to-noise ratio. It is unclear if spatial release from masking depends on the perceived position of an auditory event or the estimated location of the source (Freyman *et al*., 1999). Understanding the distinction made by Blauert (1982) is of critical importance for characterizing the advantages of bilateral auditory assistive devices (e.g., hearing aids and cochlear implants) over unilateral assistive devices.

METHODS

The current study uses an interval-scaling paradigm and a model of spectral shape analysis to investigate the ability of monaural listeners to estimate the azimuth of a virtual sound source. The virtual sources were normalized in energy to reduce the dependence of the loudness on location. A three-interval paradigm, with the second interval acting as the target and the first and third intervals acting as anchors, was used. The first and third intervals were always presented from the extreme left and right, respectively. The azimuth and overall level of the target was randomly chosen on each trial. Normal-hearing subjects, listening monaurally, responded with a

continuous visual pointer (a graphical slider presented on a computer monitor) to indicate the relative similarity between the target and the anchors. The stimulus was multi-tone (sum of sinusoids) with a parametrically varied lowest frequency, highest frequency, and component spacing. The measured responses were ultimately compared to the predictions of a model of spectral shape analysis.

Subjects

Three subjects (S1, S2, and S3) participated in the experiment. The subjects were between 19 and 33 years old and had pure tone thresholds less than or equal to 20 dB HL at frequencies of 250, 500, 1000, 2000, 4000, 6000, and 8000 Hz in both ears. Subject S1 had extensive prior experience in psychoacoustic experiments. The subjects received an hourly wage for their participation. The subjects were given at least 10 h to familiarize themselves with the task.

Stimuli

The stimulus was a multi-tone complex with logarithmically spaced components. The durations of the stimuli were 250 ms including 5 ms onset and offset cosine-squared ramps. The spacing (1, 2, 4, or 8 components per octave), the lowest frequency (250, 500, or 1000 Hz) and the highest frequency (4, 8, or 16 kHz) were parametrically varied. Prior to spatial processing, each component had a level of 53.5 dB SPL independent of the component spacing, lowest frequency component, and high frequency component.

The spatial processing utilized non-individualized head-related transfer functions (HRTFs) measured by Algazi *et al*. (2001) on the Knowles Electronics Manikin for Acoustic Research (KEMAR) to simulate 37 different spatial locations (5° separation spanning the frontal hemi-field). For each location, the HRTF was scaled to have unit energy to reduce the dependence of the loudness on location. The spatial processing consisted of scaling the amplitude of each component in the multi-tone complex by the square root of the energy within a 1/3 octave band centered at the component frequency of the normalized HRTF. The components were then added together in sine phase on the first and third intervals and were added together with random phases on the second interval. Finally, the overall level of the second interval was randomly adjusted (20 dB wide uniform distribution center on the nominal level) on every trial. After the spatial processing (and the overall level randomization) the highest possible overall level was exactly 90 dB SPL and the lowest possible level of any component was 32 dB SPL. The stimuli were generated with Tucker-Davis Technology System 3 hardware (RP2.1 running at 48828.125 Hz) and were presented at the left ear of the subjects over Sennheiser HD 410 SL headphones.

Paradigm

An interval-scaling paradigm with three intervals and correct answer feedback was used. The first and last intervals served as cues. The first interval was always presented from the leftmost virtual location; the third interval was always presented from the rightmost virtual location. The second interval was randomly presented from one of 37 different locations on the frontal hemi-field. There was 200 ms between intervals. The subjects responded with a continuous visual pointer presented on a computer monitor to indicate the relative similarity between the target and the perceptual anchors.

Data collection was blocked such that on a given day subjects completed two stimulus conditions. On every day, the subjects completed 600 trials (10 blocks x 60 trials) with the first stimulus condition before any of the 600 trials for the second stimulus condition were run. The order of the conditions was random.

Prior to beginning testing, the subjects were instructed that on each trial they would hear three sounds and needed to give a response via a visual pointer (a graphical slider presented on a computer monitor) to indicate the relative similarity between the target (the second sound) and the anchors (the first and third sounds). Subjects were told that the loudness of the target would be random and that they should ignore that aspect of the sound. They were also told that the sounds were generated by adjusting a control slider that was identical to the response slider. Further, they were told that the sound in the first interval was made by adjusting the control slider all the way to the left and that the sound in the third interval was made by adjusting the control slider all the way to the right. They were instructed to adjust the response slider to indicate where the control slider must have been to make the sound in the second interval. Finally, they were told that after they made their response, the correct location on the slider would be shown to them and that they should use this feedback to improve. Additionally, after every 60 trials a summary scatter plot presented their response as a function of the correct response. Importantly, spatial location was never mentioned to the subjects and it is likely that the subjects were unaware that the slider corresponded to spatial location.

Model

The modeling approach applied here assumes that the location of the continuous visual pointer is a result of a spectral shape analysis. The ability to discriminate changes in spectral shape has been successfully modeled using a weighted linear combination of the observed levels within each frequency channel (Durlach *et al*., 1986; Berg, 2004). A weighted linear combination model of spectral shape discrimination can be extended to interval-scaling by defining the pointer location on each trial as being equal to $a(\vec{w} - \bar{w})^T \vec{x} + b$, where $\vec{x}$ is a vector of the levels of the components of the target on that trial, and $\vec{w}$ is the difference between the levels of the components of the two perceptual cues, $\bar{w}$ is the mean of $\vec{w}$, and a and b are the free parameters of the model. For each subject and stimulus condition, the parameters a and b are chosen to maximize the percentage of the variance, in the mean as a function of location

of the measured responses, accounted for by the model. The model predictions are completely defined, apart from a linear transformation (defined by the parameters *a* and *b*), by the spectral shape of the stimulus. For simplicity in presenting the data, the linear transformation is applied to the responses of the subjects, as opposed to the model predictions, so that only a single prediction is made for each stimulus condition.

RESULTS

For the 36 different stimulus conditions tested, the responses of the subjects showed a systematic dependence on the virtual location of the target. Figure 1 shows the responses as a function of the virtual location of the target for a representative subject (S1) in two different stimulus conditions. The effect of the minimum frequency, maximum frequency, and number of components per octave in the stimulus substantially influenced the dependence of the response on the target location. In some conditions, there was a nearly linear relationship between the median response pointer positions and the target location (cf. the left panel of Fig. 1). In other conditions, there was a more step like relationship (cf. the right panel of Fig. 1). The inter-quartile range of the response pointer positions, for a given target location, were also variable across conditions. No single metric fully characterizes the dependence of the response pointer positions on the target location.

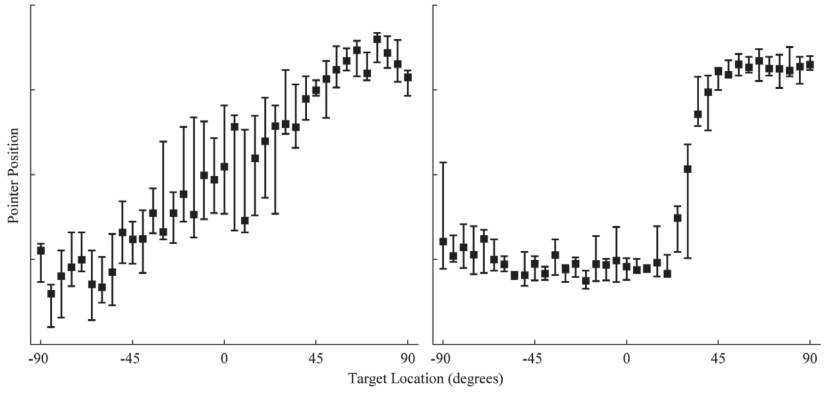

Fig. 1: The left panel shows the median responses (and inter-quartile ranges) as a function of the target location for subject S1 with a stimulus that has 8 components per octave between 500 and 8000 Hz. The right panel is the same except for a stimulus that has 4 components per octave between 500 and 4000 Hz.

The correlation between the target location and each response penalizes both deviations from a linear fit of the mean of the responses, conditioned on the target location, and the spread around the mean (e.g., the standard deviation) of the responses. Figure 2 shows the correlation between the response pointer position on every trial and the target location for the three subjects in the 36 different stimulus conditions. The correlations range between 0.22 and 0.81 and are statistically different from zero for every subject and condition tested. Although the correlation metric obscures some of the details of the response patterns, it does allow the general conclusion that the responses of the subjects depend on the target location.

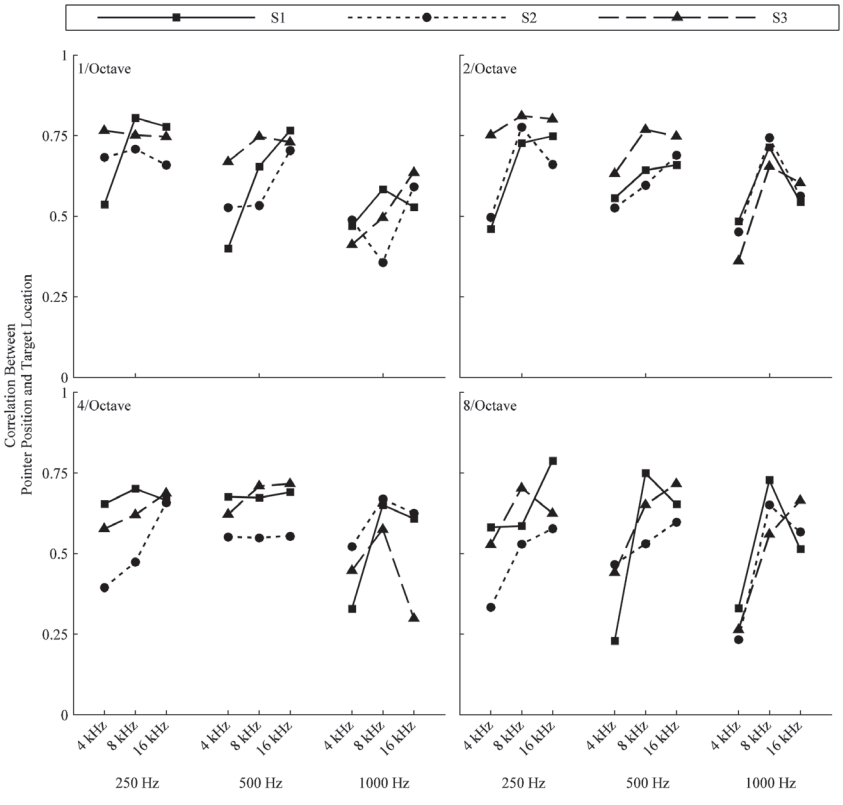

Fig. 2: The correlation between the subject's responses and the target location for each of the 36 different stimulus conditions is shown. Each panel contains the correlation for stimuli with a given number of components per octave.

The model of spectral shape processing provides an alternative means of characterizing performance. The left and right panels of Fig. 3 show the mean response, after a linear transformation, as a function of the target location for the same two conditions as in Fig 1. In addition, the model predictions are also plotted as solid lines. Even though the two different stimulus conditions lead to substantially different dependencies of the responses on the target location, there is good agreement amongst the subjects. Further, the model is able to predict how the responses depend on the target location.

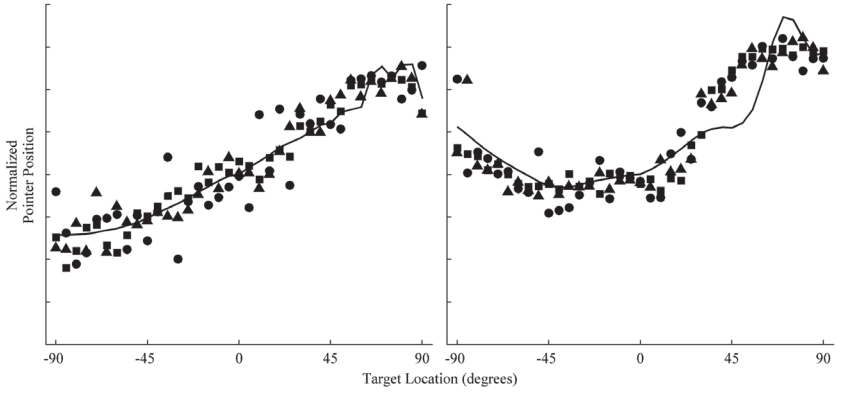

Fig. 3: The left panel shows the mean responses as a function of the target location for all three subjects (different symbols) with a stimulus that has 8 components per octave between 500 and 8000 Hz. The solid line is the prediction of the model based on the spectral shape. The right panel is the same except for a stimulus that has 4 components per octave between 500 and 4000 Hz.

Figure 4 shows the percentage of the variability in the mean responses, as a function of location, accounted for by the model as a function of the predictable variance. The predictable variance is defined as $100(2\rho/(1+\rho))$, where ρ is the correlation between the mean responses, as a function of location, measured during the first and second halves of data collection (Ahumada and Lovell, 1971). The percentage of the variance accounted for by the model depends on the predictable variance. In most cases, the model is predicting slightly less than the total amount of predictable variance. In a few cases, the model is slightly over fitting the data (values above the major diagonal) and in some cases, the model fails to predict all of the predictable variance (values below the major diagonal). Overall, a model that assumes that the subjects position the response slider based on a linear comparison between the spectral shape of the stimulus in the second interval and the difference between the spectral shapes of the perceptual anchors (i.e., the first and third intervals), accurately predicts the response properties over a large range of stimulus bandwidths and spectral densities.

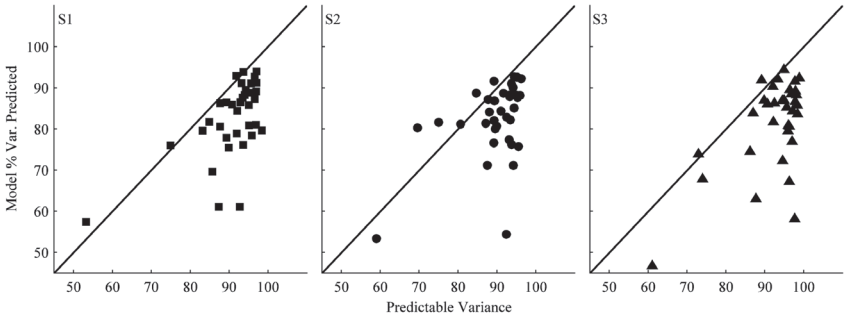

Fig. 4: Each panel shows the percentage of the variability in the mean responses, as a function of location, accounted for by the model as a function of the predictable variance for the 36 different stimulus conditions for a single subject.

DISCUSSION

In the current study, the correlation between the response pointer and the azimuth of the virtual target ranged between 0.22 and 0.81. At first glance, these correlations do not seem substantially different from the correlations between the perceived location and actual location reported by van Wanrooij and van Opstal (2004), which ranged between 0.20 and 0.55. There are two problems with this comparison. First, the stimuli used by van Wanrooij and van Opstal (2004) had energy between 1 and 20 kHz, but in the current study many of the lowest correlations were measured in conditions in which the stimulus had only middle frequencies (i.e., between 1 and 4 kHz). Second, van Wanrooij and van Opstal (2004) did not normalize the levels of their sources (i.e., the overall level of the sound was correlated with the actual location) and they found that the perceived location depended more (and in 1/3 of their subjects entirely) on the overall level. It therefore seems that, as postulated by Blauert (1982), estimating the source location is different than reporting the perceived location.

It is unclear if being able to estimate the location of a source with an unknown level is beneficial to individuals who are effectively monaural (e.g., users of unilateral cochlear implants). Monaural performance is substantially worse than binaural performance; for example, van Wanrooij and van Opstal (2004) measured correlations between 0.95 and 0.99 for binaural localization. Drennan *et al*. (2007), however, found degrading binaural cues to produce a correlation of 0.70 (similar to the performance of the monaural subjects in the current study for a range of stimulus bandwidths) had almost no effect on the amount of spatial release from masking. Although the monaural performance measured here is worse than binaural performance, potentially monaural listeners have sufficient information to navigate an acoustic world, maximize the signal-to-noise ratio, and achieve the maximal amount of spatial release from masking. More research needs to be conducted before we can accurately characterize the benefits of bilateral and binaural hearing.

The model of spectral shape processing accurately predicted the mean location of the response pointer. For some stimuli, the mean response pointer location, as a function of the target location, was not well fit by a straight line, but was well predicted by the model (cf. the right panel of Fig. 3). For conditions that lead to this type of performance, it may be possible to increase the correlation between the response pointer and the target location by providing subjects with a different set of perceptual anchors (i.e., a set that are tailored to yield a linear relationship). It also may be possible to teach unilateral cochlear implantees how to estimate the location of a sound by deriving perceptual anchors that take into account the cochlear implant processing.

CONCLUSIONS

Monaurally listening subjects used a visual pointer to indicate the relative similarity between a target and two perceptual anchors. The target varied in both overall level and azimuth. There was a high correlation between the pointer position and target azimuth (between 0.22 and 0.81), but there was a complicated dependence of the pointer position on the target azimuth that was not captured by a line. This dependence was influenced by the specifics of the multi-tone stimulus (lowest frequency component, the highest frequency component, and the component spacing). For each subject and stimulus conditions tested, the dependence of the pointer position on the target azimuth was predicted by a model of spectral-shape processing that had only two free parameters. These results suggest that subjects with monaural hearing can be trained to use the spectral shape to estimate the location of a sound with an unknown level.

ACKNOWLEDGEMENTS

This work was supported by NIH NIDCD R01 DC02012 and F32 DC 009384. The data were collected at the University of Pennsylvania. We would like to thank Dr. Barrie Edmonds for helpful comments on a previous version of this manuscript.

REFERENCES

Ahumada, A. Jr., and Lovell, J. (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751-1756.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics (New Paltz NY), pp. 99-102.

Berg, B. G. (**2004**). "A molecular description of profile analysis: decision weights and internal noise," J. Acoust. Soc. Am. **115**, 822-829.

Blauert, J. (**1982**). "Binaural localization," Scand. Audiol. Suppl., **15**, 7-26.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722-731.

Drennan, W. R., Ho Won, J., Dasika, V. K., and Rubinstein, J. T. (**2007**). "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise," J. Assoc. Res. Otolaryngol. **8**, 373-383.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. **80**, 63-72.

Freedman, S. J., and Fisher, H. G. (**1968**). "The role of the pinna in auditory localization," in *The Neuropsychology of Spatially Oriented Behavior*, edited by S. J. Freedman (Dorsey Press, Homewood, Illinois), pp. 135-152.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578-3588.

Fisher, H. G., and Freedman, S. J. (**1968**). "Localization of sound during simulated unilateral conductive hearing loss," Acta Otolaryngol. **66**, 213-220.

Häusler, R., Colburn, S., and Marr, E. (**1983**). "Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests," Acta Otolaryngol. Suppl. (Stockh.) **400**, 1-62.

Miller, G. A. (**1956**). "The magical number seven plus or minus two: some limits on our capacity for processing information," Psychol. Rev. **63**, 81-97.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848-1861.

Shaw, E. A. G., and Vaillancourt, M. M. (**1985**). "Transformation of sound-pressure level from the free field to the eardrum presented in numerical form," J. Acoust. Soc. Am. **78**, 1120-1123.

Shub, D. E., Carr, S. P., Kong, Y., and Colburn, H. S. (**2008**). "Discrimination and identification of azimuth using spectral shape," J. Acoust. Soc. Am. **124**, 3132-3141.

Slattery, III, W. H., and Middlebrooks, J. C. (**1994**). "Monaural sound localization: acute versus chronic unilateral impairment," Hear. Res. **75**, 38-46.

van Wanrooij, M. M., and van Opstal, A. J. (**2007**). "Sound localization under perturbed binaural hearing," J. Neurophysiol. **97**, 715-726.

van Wanrooij, M. M., and van Opstal, A. J. (**2004**). "Contribution of head shadow and pinna cues to chronic monaural sound localization," J. Neurosci. **24**, 4163-4171.

Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050-1063.