The influence of duration and level on human sound localization

Joyce Vliegen and A. John Van Opstal[a)]
Department of Medical Physics and Biophysics, University of Nijmegen, Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands

(Received 30 March 2003; revised 12 September 2003; accepted 10 January 2004)

The localization of sounds in the vertical plane (elevation) deteriorates for short-duration wideband sounds at moderate to high intensities. The effect is described by a systematic decrease of the elevation gain (slope of stimulus–response relation) at short sound durations. Two hypotheses have been proposed to explain this finding. Either the sound localization system integrates over a time window that is too short to accurately extract the spectral localization cues (*neural integration hypothesis*), or the effect results from cochlear saturation at high intensities (*adaptation hypothesis*). While the neural integration model predicts that elevation gain is independent of sound level, the adaptation hypothesis holds that low elevation gains for short-duration sounds are only obtained at high intensities. Here, these predictions are tested over a larger range of stimulus parameters than has been done so far. Subjects responded with rapid head movements to noise bursts in the two-dimensional frontal space. Stimulus durations ranged from 3 to 100 ms; sound levels from 26 to 73 dB SPL. Results show that the elevation gain decreases for short noise bursts at all sound levels, a finding that supports the integration model. On the other hand, the short-duration gain also decreases at high sound levels, which is in line with the adaptation hypothesis. The finding that elevation gain was a nonmonotonic function of sound level for all sound durations, however, is predicted by neither model. It is concluded that both mechanisms underlie the elevation gain effect and a conceptual model is proposed to reconcile these findings. [DOI: 10.1121/1.1687423]

PACS numbers: 43.66.Qp, 43.66.Ba, 43.66.Mk [AK] Pages: 1705–1713

I. INTRODUCTION

In order to localize a sound, the auditory system relies on binaural and monaural acoustic cues. Binaural cues result from interaural differences in sound level (ILD) and timing (ITD), which relate to sound position in the horizontal plane (azimuth). Monaural cues consist of direction-dependent spectral shape information caused by reflection and diffraction at torso, head, and pinnae (described by head-related transfer functions, or HRTFs). These spectral cues are essential to resolve front–back confusions and to localize sounds in the vertical plane (elevation; see Blauert, 1996, for a review). Although the binaural difference cues are extracted quite reliably under a wide variety of stimulus conditions and spectra, the transformation of the HRTFs into a reliable estimate of sound-source elevation is a challenging problem for several reasons.

First, the spectrum at the eardrum (which will be denoted by the sensory spectrum) is a linear convolution of the (*a priori* unknown) sound-source spectrum with the particular HRTF associated with the unknown sound direction. Thus, in extracting sound-source elevation, the auditory system is faced with an ill-posed problem. One way to deal with this problem would be to incorporate *a priori* assumptions about potential source spectra. For example, if the source spectrum is assumed flat, the sensory spectrum is identical to the HRTF. Yet, subjects are able to localize a variety of broadband sound spectra that are not flat with remarkable accuracy (Oldfield and Parker, 1984; Wightman and Kistler, 1989; Middlebrooks and Green, 1991; Hofman and Van Opstal, 1998). Apparently, the assumptions about potential source spectra are more relaxed.

If the assumption holds that source spectra do not resemble any of the HRTFs, the spectral correlation between the sensory spectrum and each of the HRTFs can be shown to peak exactly at the correct HRTF (Middlebrooks, 1992; see Hofman and Van Opstal, 1998, for details). Such a strategy would allow accurate localization for a large class of nonflat stimulus spectra. However, when amplitude variations within the source spectrum become too large, the localization accuracy of sound elevation deteriorates (Wightman and Kistler, 1989; Hofman and Van Opstal, 2002).

A second problem concerns the presence of considerable spectro-temporal variations in natural sounds. Until recently, localization studies have typically used long-duration stimuli with stationary spectro-temporal properties. Not much is known as to how nonstationary sounds affect sound localization performance.

Hofman and Van Opstal (1998) studied the effects of different spectro-temporal stimulus properties on sound localization performance in the two-dimensional frontal hemifield. The only response variable that depended systematically on the temporal stimulus parameters was the *slope* of the stimulus–response relation for the elevation components (i.e., the elevation gain). In particular, for stimuli with durations shorter than several tens of ms the gain started to decrease with decreasing burst duration. Neither response variability, nor the azimuth responses depended on the stimulus parameters. Based on their results, Hofman and Van Opstal

a)Electronic mail: johnvo@mbfys.kun.nl

(1998) proposed that the sound localization system needs to integrate about 40–80 ms of broadband input to yield a stable estimate of sound-source elevation (the *neural integration hypothesis*).

Recently, an alternative explanation for these data has been put forward (Macpherson and Middlebrooks, 2000). In that proposal, the decrease in gain is due to the so-called "*negative level effect*" reported earlier by Hartmann and Rakerd (1993). In this earlier study, subjects were unable to localize high-level clicks (>86 dB SPL), with errors decreasing for intermediate (74–86 dB) and lower (68–80 dB) sound levels. Hartmann and Rakerd (1993) suggested that this effect was caused by saturation of cochlear excitation patterns. As a consequence, the auditory system would fail to resolve the spectral details of the clicks. For long-duration stimuli, the system would adapt to the high sound level, so that a reliable elevation estimate could be based on later portions of the signal (the *adaptation hypothesis*).

To elaborate on this possibility, Macpherson and Middlebrooks (2000) presented short- (3 ms) and long-duration (100 ms) noise bursts at sensation levels (SL) between 25 and 60 dB. Like Hofman and Van Opstal (1998), they found that elevation gains were lower for short-duration stimuli than for long-duration stimuli, but only at high sensation levels. Moreover, when the short noise bursts were presented within spatially diffuse noise, elevation gain depended on the level of the masker. Elevation gains increased with increasing masker level until a masked sensation level of about 40 dB. These results are at odds with the neural integration hypothesis, which would predict no effect of signal level. However, they are predicted by the adaptation model, as the background noise would activate the putative adaptive mechanism prior to the onset of the 3-ms noise bursts. At higher masker levels, performance decreased, which could be due to a low signal-to-noise ratio.

Macpherson and Middlebrooks (2000) concluded that the results of all three studies can thus be explained by the negative level effect. Note, however, that this mechanism does not specify how and why only the elevation *gain* would be affected by cochlear saturation, and why other parameters, e.g., response variability, or azimuth localization, remain unaffected.

Note also that the fixed stimulus level of 70 dB SPL employed by Hofman and Van Opstal (1998) corresponds to the low end of intensities used by Hartmann and Rakerd (1993). Moreover, Frens and Van Opstal (1995) had reported similar gain-duration effects for stimuli of only 60 dB SPL.

The results of Hofman and Van Opstal (1998) and Macpherson and Middlebrooks (2000) are difficult to compare directly because of differences in methodology. First, Hofman and Van Opstal (1998) used a variety of stimulus durations, mixed randomly within a single recording session, whereas Macpherson and Middlebrooks (2000) collected responses to two different stimulus durations (3 and 100 ms) in different blocks of trials. Second, while Hofman and Van Opstal (1998) presented all stimuli at 70 dB SPL, Macpherson and Middlebrooks (2000) employed various intensities, but quantified as sensation levels. These two measures are not readily equated. Third, the pointer used to indicate perceived sound direction differed in the two studies: eye movements (restricted to the 35-deg oculomotor range) by Hofman and Van Opstal (1998) vs head movements over a much larger measurement range by Macpherson and Middlebrooks (2000).

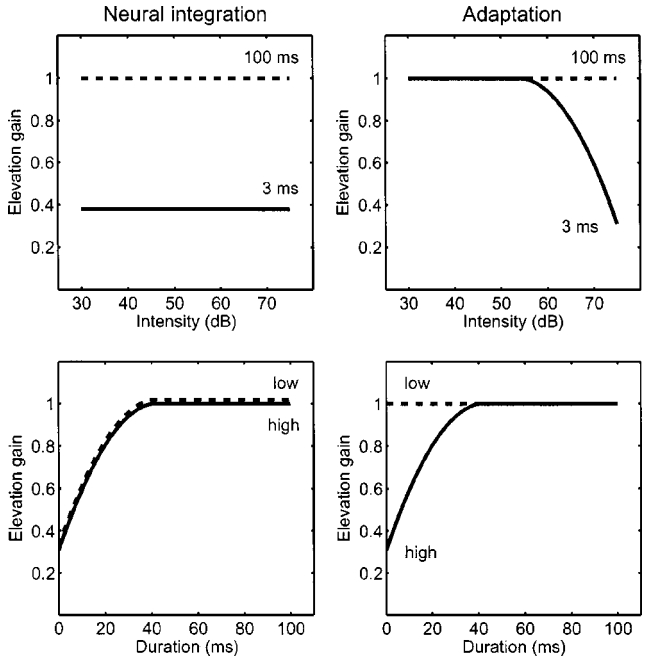

FIG. 1. Predictions of the neural integration model (left) and the cochlear adaptation hypothesis (right). Top row: elevation gain as a function of intensity for two durations. Bottom row: elevation gain as a function of duration for low and high intensities. The adaptation model predicts a decrease of elevation gain for short-duration stimuli at high intensities only, and a stable gain for longer-duration stimuli at all levels. The neural integration model predicts a decrease of the gain with duration at all stimulus levels, while gain is insensitive to stimulus level.

Finally, both studies measured only a small portion of the duration-intensity parameter space, with minor overlap. Therefore, to allow for a better comparison of both data sets, we have included and extended the measurements of both studies by employing a range of noise durations (3–100 ms) and sound levels (26–73 dB SPL). Up to 16 different stimulus conditions were measured within the same recording session, and were randomly interleaved. A summary of the expected results for the two hypotheses is provided in Fig. 1.

II. METHODS

The experiment consisted of three sessions, differing slightly in the parameter values used. In the first session we used durations ranging from 3 to 100 ms. We found that the largest changes in the results occurred for durations between 3 and 30 ms. Therefore, in later sessions we restricted the duration values to this range. These last two sessions consisted of stimuli with the same range in durations, but with different, slightly overlapping, intensity ranges.

A. Subjects

Two female and seven male subjects participated in the experiments. Their age ranged from 22 to 44 years. Two of the subjects (JV and JO) were the authors of this paper. Five other subjects were experienced in sound localization

studies. Subjects FF and JM had no previous localization experience. Before the actual experiment started, these inexperienced subjects were given a short practice session to get familiar with the stimuli and the localization paradigm. All subjects had normal binaural hearing (absolute thresholds within 20 dB HL at frequencies between 250 and 8000 Hz).

Subject JV participated in all three sessions. Subjects JO, FF, and JM participated in the first session only, while the remaining five subjects participated in sessions 2 and 3.

B. Apparatus

Experiments were conducted in a completely dark and sound-attenuated room with dimensions $L \times W \times H = 3.5 \times 2.45 \times 2.45\ m^3$. The room had an ambient background sound level of 20 dBA SPL. Horizontal and vertical head movements were measured with the search-coil technique. Subjects wore a lightweight helmet (about 150 g), consisting of a narrow strap above the ears, which could be adjusted to fit around the subject's head, and a second strap that ran over the head. A small coil was mounted on the latter. Two orthogonal pairs of coils were attached to the room's edges to generate the horizontal (60 kHz) and vertical (80 kHz) magnetic fields. The head-coil signal was amplified and demodulated (Remmel Labs), after which it was low-pass filtered at 150 Hz (Krohn-Hite 4413) and then stored on hard disk at a sampling rate of 500 Hz/channel for subsequent off-line analysis.

Subjects were seated comfortably in the center of the room facing a frontal hemisphere (radius: 1.0 m) that consisted of a thin wooden framework with 12 spokes and five concentric rings. This setup thus defined a polar coordinate system with its origin at the straight-ahead position. Target excentricity, R, is measured as the angle with respect to the straight-ahead position, whereas target direction, ϕ, is measured in relation to the horizontal meridian. For example, $R = 0$ deg corresponds to straight ahead for each ϕ, and $\phi = 0$, 90, 180, and 270 deg (for $R > 0$) corresponds to right, up, left, and down, respectively. On the hemisphere, a total of 58 small broad-range loudspeakers (Monacor MSP-30) were mounted at directions $\phi = 0$, 30, 60, 90,..., 330 deg (corresponding to each of the 12 spokes) and excentricities of $R = 0$, 15, 30, 45, 60, and 75 deg (corresponding to the five rings). At the outer ring ($R = 75$ deg) part of the framework (at downward directions $\phi = 240$, 270, and 300 deg) was removed to allow for space for the subject's legs. A thin glassfiber ended in the center of each speaker, through which a well-defined visual stimulus (0.15-deg diameter, 1.5-Cd/m^2) could be presented that originated from a red and green LED mounted behind the speaker. The peripheral LEDs were used to calibrate the head-coil signals at the start of an experimental session (see below), while the center LED at $(R, \phi) = (0,0)$ deg served as a fixation light at the start of a localization trial. The polar target coordinates (R, ϕ) were transformed into azimuth-elevation angles (α, ε), in the off-line analysis of the data (see Sec. II F and Hofman and Van Opstal, 1998, for details).

In the first experimental session, only the speakers at the first three rings of the hemisphere were used ($R = 0$, 15, 30, 45 deg; $N = 37$ locations). For the second and third session the speakers of all five rings were used, except for the central speaker at straight ahead; $N = 57$.

The height of the chair was adjusted to align the center of the subject's head with the center of the hemisphere. Walls, ceiling, and floor, as well as the spokes and rings of the hemifield, were covered with black sound-absorbing foam that eliminated acoustic reflections down to 500 Hz (Schulpen Schuim, The Netherlands).

TABLE I. Detection thresholds in dB SPL for all subjects for the four stimulus durations employed in the second and third session.

Duration	JV	MW	HV	MZ	WV	FW
3 ms	23	30	31	27	25	31
6 ms	21	25	25	19	18	19
14 ms	20	21	13	17	18	21
30 ms	20	21	13	12	17	20

C. Stimuli

Acoustic stimuli were generated digitally with a Tucker-Davis System II, using a TDT DA1 16-bit digital-to-analog converter (50-kHz sampling rate). Stimuli were then passed to a TDT PA4 programmable attenuator, which controlled the sound level. All stimuli consisted of independently generated Gaussian white noise with 0.5-ms sine-squared on- and off-set ramps.

In the first session, durations of 3, 10, 31, and 100 ms were used, with intensities of 26, 36, 46, and 56 dB SPL (a total of 592 trials per run and two or three runs per subject). In the second and third session, durations of 3, 6, 14, and 30 ms were used. Sound levels were at 33, 43, 53, and 63 dB SPL for the second session (one run of 912 trials) and at 58, 68, and 73 dB SPL for the third session (one run of 684 trials).

D. Sensation levels

For the six subjects that participated in sessions 2 and 3, free-field detection thresholds for broadband noise bursts of 3, 6, 14, and 30 ms were determined. Sounds were presented from the center speaker in the sound-attenuated room. Listeners performed a two-interval, two-alternative, forced-choice task where sound level was controlled by a three-down, one-up adaptive tracking procedure (Levitt, 1971). For all subjects, thresholds decreased with increasing noise duration. Table I summarizes the results of these measurements for all subjects. From these data sensation levels (SL) were computed by subtracting the thresholds from the SPL values of the stimuli as recorded at the level of the subject's head.

E. Recording paradigm

All measurements were performed in darkness. When making a head saccade in darkness, the eyes will typically not remain centered in the head. Especially for peripheral target locations, the position of the eyes in the head will be quite excentric (exceeding 20 deg), resulting in potentially large (and variable) undershoots of the measured head posi-

tion if subjects use both eyes and head to point to the target (Goossens and Van Opstal, 1997). To circumvent this potential problem, a thin aluminum rod with a dim red LED (0.15 Cd/m^2) attached to its end protruded from the helmet's left side. The rod was adjusted such that the LED was positioned in front of the subject's eyes at a distance of about 40 cm. At the start of a trial, the subject had to align this rod LED with the central LED of the hemisphere, while keeping his head in a comfortable straight-ahead position. The rod LED thus served as a head-fixed pointer during the experiments. Pointing with the LED to the perceived location of the target ensured that the eyes remained at a fixed, central position in the head while pointing.

Each recording session started with a calibration run in which the subject had to align the rod LED with each of the LEDs on the hemisphere. After calibration, head position was known with an absolute accuracy of 3% or better over the entire measurement range.

In subsequent blocks, the sound stimuli were presented. Each trial started by presenting the central fixation LED. After a randomly selected fixation period of 1.5 to 2.0 s, the fixation LED was switched off and 400 ms later the sound stimulus was presented at a peripheral location. The subject's task was to point the rod LED as quickly and as accurately as possible towards the perceived sound location. No feedback was given about performance. As stimuli were always extinguished well before the initiation of the head movement (typical reaction times about 200–300 ms), all experiments were conducted under fully open-loop conditions.

For all experiments, the order of stimulus conditions and positions was randomized throughout a session.

F. Data analysis and statistics

The coordinates of the target locations and head-movement responses are described in a double-pole coordinate system, in which the origin coincides with the center of the head. The horizontal component, azimuth α, is defined as the direction relative to the vertical median plane, whereas the vertical component, elevation ϵ, is defined as the direction relative to the horizontal plane through the ears (Knudsen and Konishi, 1979).

From the calibration run, the raw head position signals and the corresponding LED coordinates were used to train two three-layer backpropagation neural networks that mapped the raw data signals to the calibrated head position signals (azimuth and elevation angles, respectively). This was done to account for minor cross talk between horizontal and vertical channels and minor inhomogeneities in the magnetic fields (Goossens and Van Opstal, 1997). Goal-directed head movements were identified in the calibrated response data. The endpoint of the first head movement after stimulus onset, where response azimuth and elevation were stable, was defined as the response position.

Head saccades with a reaction time *re*: stimulus onset of less than 80 ms or above 800 ms were discarded from further analysis. Earlier responses are assumed to be predictive and are usually very inaccurate. Later responses are considered to be caused by inattention of the subject. Typically, less than 2% of the responses had to be discarded on the basis of these criteria.

For each stimulus condition (fixed stimulus duration and sound level), a linear regression line was fitted through the stimulus–response relations for azimuth (α) and elevation (ε) components, respectively, by applying the least-squares error criterion

$$\alpha_R = G_\alpha \cdot \alpha_T + b_\alpha ,$$
$$\varepsilon_R = G_\varepsilon \cdot \varepsilon_T + b_\varepsilon , \tag{1}$$

where (α_R, ϵ_R) are the head-movement response components, $(\alpha_T, \varepsilon_T)$ are the target coordinates; $(G_\alpha, G_\varepsilon)$ are the slopes of the regression lines (here called the response gain), and $(b_\alpha, b_\varepsilon)$ (in deg) are the offsets (response bias). The bootstrap method was used to estimate the standard deviations of the slopes, offsets, and Pearson's linear correlation coefficients (Press *et al.*, 1992).

To quantify the effects of stimulus duration and sound level on the stimulus–response relation, we also performed a nonlinear regression on the entire data set (all stimulus conditions and recording sessions pooled; elevation data only). In this regression, the elevation gain, G_ε was a (nonlinear) function of duration and sound level (five free parameters; see the Appendix for details).

Finally, to enable a quantitative comparison of the relative contributions of stimulus duration and stimulus level on the response elevations across the different stimulus conditions, we also performed two normalized multiple-linear regressions on two relevant cross sections through the data (see Sec. III).

III. RESULTS

Typical localization results of the first experimental session are presented in Fig. 2, which shows the endpoints of the azimuth and elevation components of the head-movement responses of subject FF together with the fitted linear regression lines. We found for all subjects that sound-source azimuth (○) was localized accurately with performance remaining rather stable for all test conditions. In contrast, the elevation response components (▲) depended strongly on the different stimulus parameters. Correlation coefficients for the stimulus–response relations were typically high. Both for the azimuth and elevation response components they were found to be close to 1.0, except for the shortest stimuli at the lowest sound level (26 dB), where correlations dropped to around zero for two subjects for both azimuth and elevation. These stimuli were probably close to, or even below, the detection threshold for these subjects. Azimuth gains were stable for all conditions, except for the 3-ms condition at the lowest intensity, where gains were considerably lower for those same two subjects. For the other two subjects in this stimulus condition azimuth gains decreased only slightly. For the elevation responses, gains appeared to increase with increasing duration for all stimulus levels. For stimulus durations between 30 and 100 ms, the response gain leveled off. For fixed durations the slope of the elevation regression line also varied with stimulus level.

In Fig. 3 the gains for the azimuth response components of sessions 2 and 3 are plotted as a function of stimulus duration for all intensities and all six subjects. For most sub-

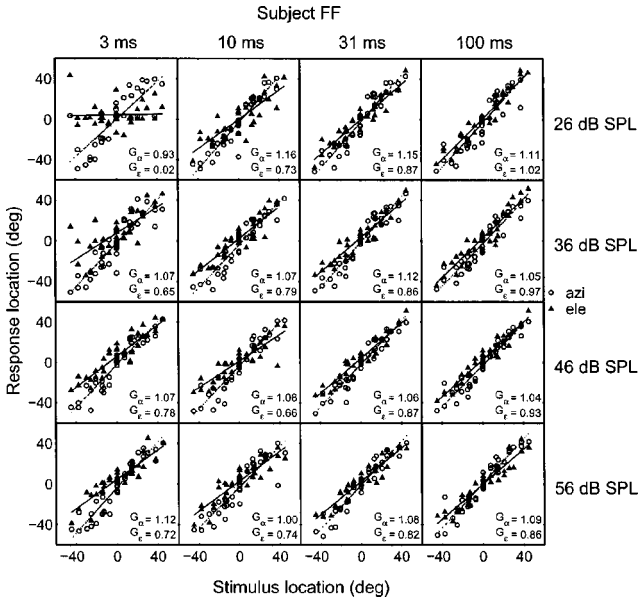

FIG. 2. Stimulus response relations for azimuth (○) and elevation (▲) response components of subject FF for four different stimulus durations (columns) and four different stimulus intensities (rows). Data taken from session 1. Best-fit regression lines (dotted: azimuth, solid: elevation) are also shown, together with the values of the azimuth and elevation gains.

jects, gains were around 1.0 or slightly higher, except for subject HV, whose gains were around 1.4 in the third session. Gains remained stable across the different stimulus conditions. Note also that the gain values could vary considerably between sessions. This is apparent for most of the subjects, for whom the data appear to split into two separate clusters,

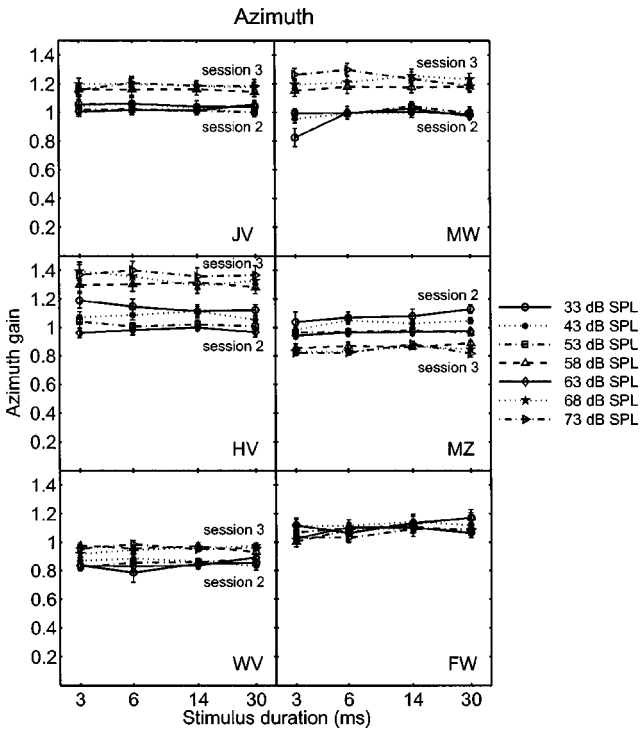

FIG. 3. Azimuth gains as a function of stimulus duration for all six subjects of sessions 2 and 3. The different line styles and symbols in each panel correspond to the different stimulus levels. Note the absence of any consistent trend and apparent separation of the obtained gain values for the two sessions in all but one subjects.

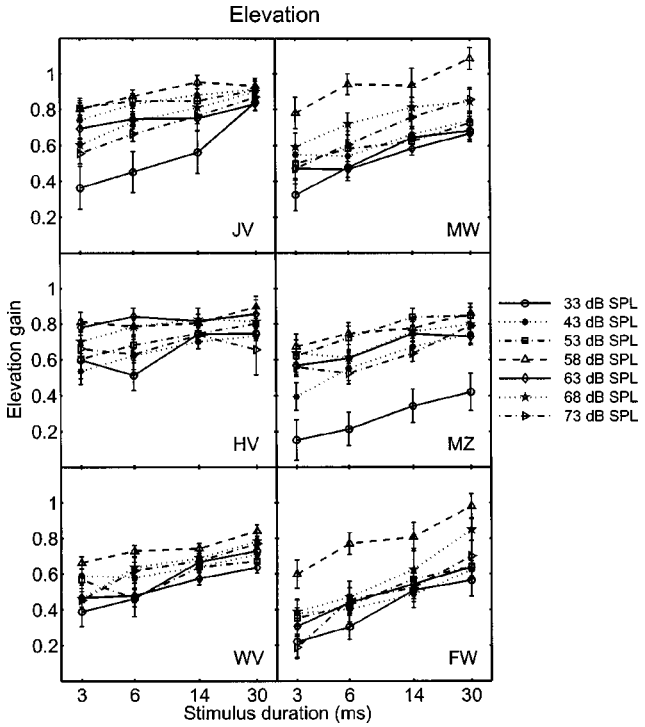

FIG. 4. Elevation gains as a function of stimulus duration for all six subjects of sessions 2 and 3. Same format as Fig. 3. Note the clear effect of stimulus duration on elevation gain for all subjects and at all stimulus levels.

each one corresponding to a different recording session (2 or 3). It might be due to simple day-to-day variation or to the different intensity ranges used in the two sessions.

The data for the elevation gains obtained from these same sessions are shown in Fig. 4 in the same format as Fig. 3. Although the absolute gain values differed between subjects, qualitatively similar patterns emerged for all subjects in both recording sessions. Elevation gain covaried with sound duration for all stimulus intensities, although the effect was most prominent at low and high levels. Gains were lowest for the 3-ms bursts at 33 dB SPL, where elevation gains were typically around 0.2–0.4. The fact that elevation gain increased with increasing sound duration for all stimulus levels, and not just for the highest stimulus levels, provides support for the neural integration hypothesis and is inconsistent with the adaptation hypothesis.

As can be noted in Fig. 4, elevation gain also appeared to vary with stimulus intensity. This feature is better illustrated in Fig. 5, which shows elevation gain as a function of absolute sound level (in dB SPL) for all stimulus durations and all subjects who participated in sessions 2 and 3. The gains were lowest for the lowest sound intensities, and especially for the shortest noise bursts. For intermediate sound levels, gains increased to a maximum value, to decrease again for higher sound levels. This latter phenomenon is reminiscent of the negative level effect reported by Hartmann and Rakerd (1993) and Macpherson and Middlebrooks (2000). It can be seen, however, that gains varied with intensity for all stimulus durations, not only the shortest ones, although the changes tended to be smaller for longer stimulus durations. The fact that elevation gains increased with increasing sound level for low intensities, a positive level

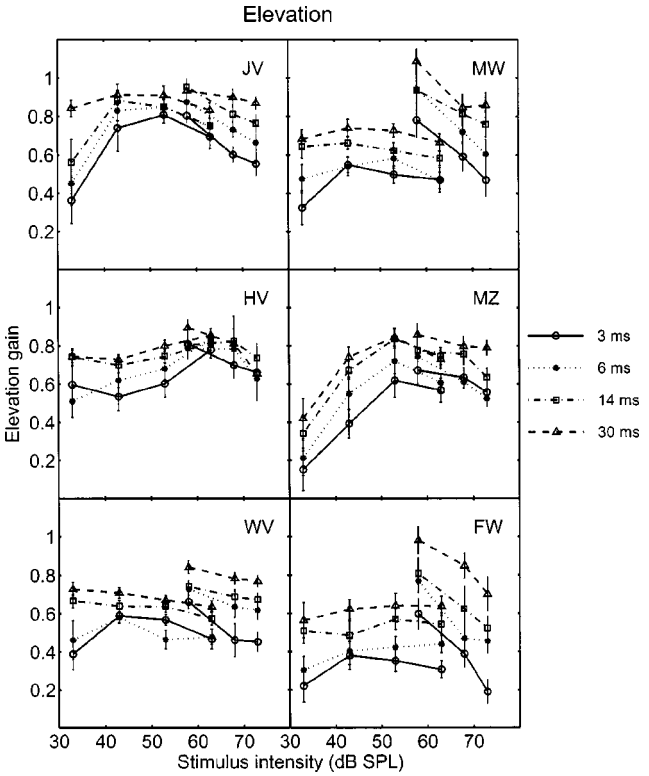

FIG. 5. Elevation gains as a function of intensity (in dB SPL) for all six subjects of sessions 2 and 3. The different line styles and symbols correspond to the different stimulus durations. Note the consistent nonmonotonic changes of elevation gain with stimulus level. Note also that both a positive and a negative level effect were observed for all stimulus durations.

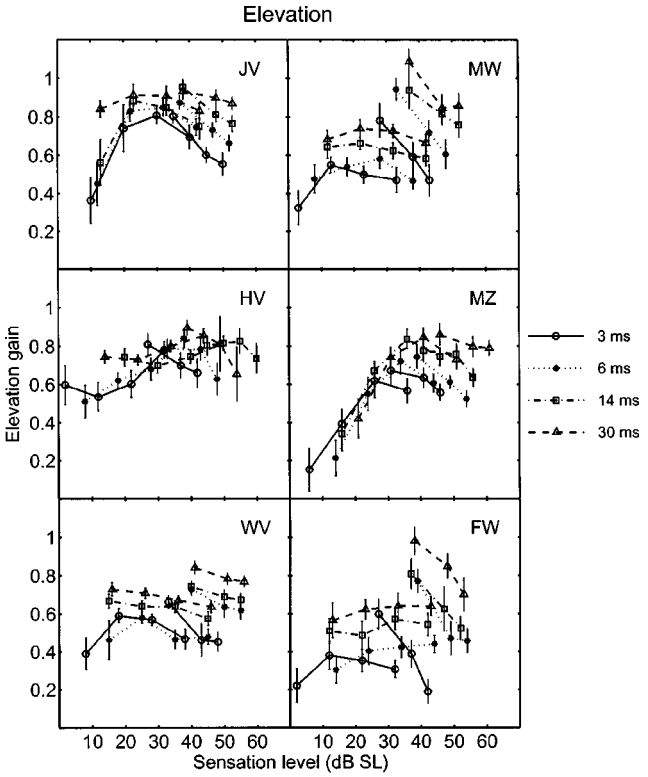

FIG. 6. Elevation gains as a function of sensation level for all six subjects of sessions 2 and 3. Same format as Fig. 5.

effect, was not predicted by either the neural integration hypothesis or the adaptation hypothesis.

It should be noted that, as in Fig. 3, three of the subjects (MW, WV, FW) showed different gain values for similar stimulus conditions in the two sessions, with higher gains in session 3 than in session 2.

For a better comparison with the data of Macpherson and Middlebrooks (2000), elevation gains are plotted as a function of sensation level in Fig. 6. Elevation gains increased strongly at the lower sensation levels; above about 45 dB SL the gains decreased. This trend was obtained for all stimulus durations.

In order to describe the effects of stimulus duration and intensity for the entire data set, we performed a nonlinear regression on all elevation responses of a given subject, pooled across recording sessions and stimulus parameters. To that end, the gain in the regression model of Eq. (1) was taken to be a function of both intensity and duration, yielding $G_\varepsilon(D,I)$. The shape of this function was estimated on the basis of the results shown in Figs. 2, 4, 5, and 6. Thus, the intensity dependence of the elevation response gain was described by a simple parabolic function, to incorporate both the positive and negative level effects. The effect of stimulus duration was described by a saturating exponential, which levels off for long durations. The response bias had a fixed value. The regression model had five free parameters, $\beta_1-\beta_5$, which were found by minimizing the mse (fitting between 1005–3195 data points; see the Appendix for details). The model yielded a good description of the data, with consistent parameter values for the different subjects and recording sessions, and high R^2-values (see Table III for results). On the basis of these results we estimated the stimulus intensity for which the elevation response gains reached a maximum at $-\beta_1/(2\beta_2)$. Values were typically between 50–70 dB SPL, with a median of 62 dB SPL.

According to the adaptation hypothesis, the negative level effect is obtained for short-duration stimuli only. Our data, however, suggest that a negative level effect occurs at all stimulus durations. Although this observation is supported by the nonlinear regression model, it is not possible to quantitatively compare the strength with which each stimulus parameter influences the elevation responses because the different variables are expressed in different units. A simpler way to quantify these effects would therefore be to convert to dimensionless variables (i.e., normalization).

To restrict the analysis to the negative level effect only, it is necessary to incorporate only that section of the data where it occurs (for the highest stimulus levels). To that end, we performed a multiple linear regression on the normalized elevation gains ($N=12$) obtained by linear regression [Eq. (1)] on the data from session 3 only (for which $L=58$, 68, 73 dB SPL)

$$\hat{G}_\epsilon=\beta_L\cdot\hat{L}+\beta_D\cdot\hat{D},\ \text{ with }\ \hat{X}\equiv\frac{X-\mu_X}{\sigma_X}, \tag{2}$$

with μ_X and σ_X the mean and variance of the respective variable (L is stimulus level in dB SPL, D is duration in ms, and G_ε is the measured elevation gain). In this regression, β_L and β_D are the (dimensionless) partial regression coefficients. The resulting regression parameters for each subject are listed in Table II (left portion). Note that all coefficients

TABLE II. Normalized partial regression coefficients for stimulus duration and intensity for the negative level effect (data of the third session: 58, 68, 73 dB SPL), and for the positive level effect (data of the second session: 33, 43, 53 dB SPL) [Eq. (2)].

	Negative level			Positive level		
Subject	β_L	β_D	R^2	β_L	β_D	R^2
JV	−0.61	0.68	0.79	0.69	0.50	0.66
MW	−0.68	0.65	0.85	0.27	0.83	0.70
HV	−0.75	0.32	0.59	0.27	0.80	0.64
MZ	−0.54	0.78	0.86	0.86	0.43	0.91
WV	−0.44	0.78	0.76	0.10	0.78	0.54
FW	−0.62	0.73	0.90	0.32	0.88	0.84
Mean	−0.61	0.66	⋯	0.42	0.70	⋯

for stimulus level are indeed negative, while for sound duration they are positive. More importantly, the absolute values of the two parameters are roughly equal, indicating that at high stimulus levels both stimulus factors influence the elevation gain to a comparable degree.

A positive level effect was obtained for lower stimulus levels and for all stimulus durations. To quantify this effect we performed a multiple linear regression [Eq. (2)] on the normalized elevation gains ($N=12$) for the lower stimulus levels ($L=33$, 43, 53 dB SPL). The resulting regression parameters are listed in Table II (right portion). The coefficients for stimulus level are all positive and their absolute values are slightly smaller than for the negative level effect. For sound duration, the values are roughly equal to the duration values for the negative level effect.

If this positive level effect were entirely due to a poor signal-to-noise ratio (SNR), the response variability would be expected to systematically vary with stimulus duration and sound level in a similar way as the response gain. To test for this, Fig. 7 shows the response variability (defined as the mean-squared error around the regression line) of the data from sessions 2 and 3 as a function of stimulus intensity for the different stimulus durations. Note that only for the lowest stimulus intensities and shortest durations was the response variability higher than for the other conditions for most subjects. Only for subjects MW and HV did the variability increase for high intensities, but this was true for all durations. Interestingly, the variability obtained for the high-intensity, short-duration stimuli was indistinguishable from the other stimulus conditions. For the majority of stimulus levels, the variability is quite comparable (around 10 deg).

IV. DISCUSSION

By systematically varying both sound duration and sound level within the same experimental session, the current experiments confirm and extend recent reports by Hofman and Van Opstal (1998) and Macpherson and Middlebrooks (2000), and provide more insight into the combined effects of these stimulus parameters on human sound localization.

The results show that the azimuth response components remained virtually unaffected for all stimulus conditions (Figs. 2, 3) except for stimuli with an intensity around the detection threshold. However, the response *elevation* gain was strongly affected by both stimulus parameters (Figs. 2, 4–6). Neither response bias (not shown), nor response variability (Fig. 7) was systematically related to the stimulus parameters. Our results are summarized in Fig. 8, which plots, in the format of Fig. 1, the prediction of Eq. (A2) (see the Appendix) applied to the pooled elevation gain data of subject JV. A comparison of Figs. 8 and 1 indicates that neither the neural integration model, nor the adaptation model explains the data well.

For all subjects, elevation gains increased with increasing sound duration, until a plateau was reached for durations above 30 ms. Although the effect was most conspicuous at the lowest and highest sound levels, it was apparent for all stimulus intensities tested. These results, especially for the higher intensities, are in good agreement with the results reported by Hofman and Van Opstal (1998), who tested their subjects at 70 dB SPL.

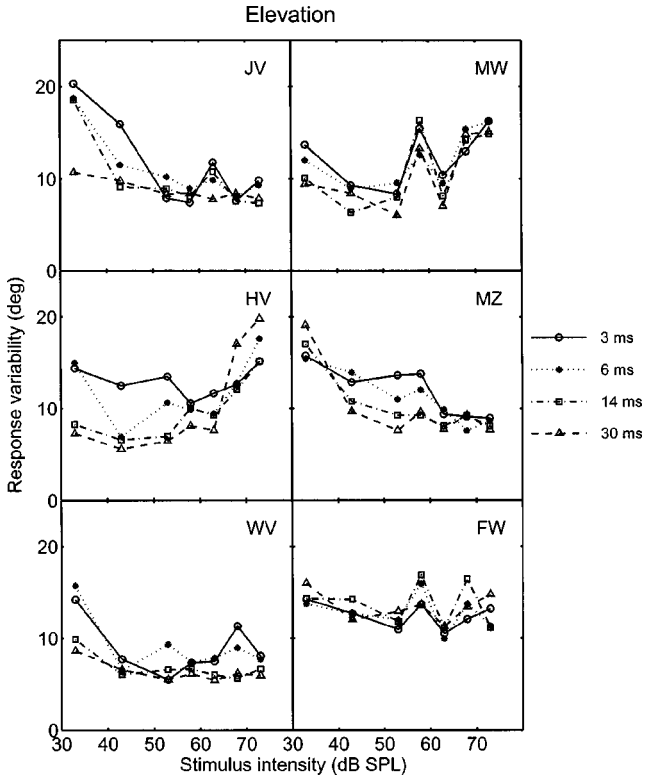

FIG. 7. Variability of elevation responses as a function of stimulus intensity. The different line styles and symbols correspond to the different stimulus durations. For most stimuli the variability is comparable; in contrast to the effects on response gain values, the variability does not change with recording session.

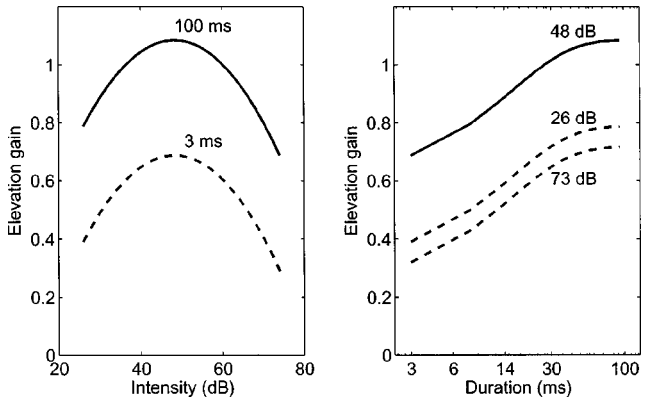

FIG. 8. Schematic summary of the results, presented in the same format as Fig. 2. Curves are based on the parameters of a nonlinear regression [Eq. (A2)] on the elevation gain data of all three sessions for subject JV. Note the logarithmic *x* axis.

Elevation gains varied in a nonmonotonic way with sound intensity (Figs. 5, 6). At low sound levels gains were low; they increased for intermediate sound levels (positive level effect), and decreased again for stimulus levels above about 55–65 dB SPL (negative level effect). When elevation gains are plotted as a function of sensation level, our results are in good agreement with the findings of Macpherson and Middlebrooks (2000).

In contrast to Macpherson and Middlebrooks (2000), however, our results indicate that both stimulus parameters affect the localization of sound-source elevation to a comparable degree (Table II). A possible reason for this difference might be that in the present study all stimulus conditions were randomly interleaved instead of presented in separate blocks of trials with fixed duration. As is illustrated, e.g., in Figs. 3 to 6 there can be considerable day-to-day variation in the absolute values of the obtained gains. Such a variability might potentially mask the effects.

This variability in our results between sessions could be due to simple day-to-day variation, or it could be the result of the differences in the intensity range used (33–63 vs 58–73 dB SPL).

Taken together, our results extend the findings of Hofman and Van Opstal (1998) and Macpherson and Middlebrooks (2000) and provide a more complete picture of the effect of sound duration and intensity on localization behavior. The data indicate that the negative level effect is not sufficient to account for the gain–duration relation which was found to persist for lower stimulus levels too.

We therefore propose that the gain–duration effect is indicative of a neural integration mechanism that accumulates evidence in order to "construct" its best estimate of sound-source elevation. As noted by Macpherson and Middlebrooks (2000), the negative level effect clearly does not fit into such a scheme, but rather provides support for the adaptation model. Note, however, that the consistent effects on elevation gain of other temporal stimulus parameters like sweep duration or interburst interval for long-duration (500 ms) stimuli at 70 dB SPL (Hofman and Van Opstal, 1998) are not readily explained by saturation of cochlear excitation patterns.

The conceptual neural-integration model put forward by Hofman and Van Opstal (1998) provides an explanation for the consistent finding that elevation *gain* is affected by the temporal stimulus parameters. In short, it proposes that the gain reflects the confidence level about the system's final estimate of sound-source elevation. This confidence is obtained by the internal correlation of the sensory spectrum [repeatedly sampled over short (<5 ms) time windows] with learned and stored representations of the subject's spectral cues, and subsequently averaged over a longer time window (several tens of ms). Clearly, this model should be extended to accommodate the level-dependent effects described in the present study.

In the absence of any certainty about stimulus location (e.g., due to low SNR), the default estimate might primarily rely on nonacoustic factors like prior knowledge about potential source locations. For example, in the current experiment this would be on average the straight-ahead location within the frontal hemifield. These factors may thus set the default gain of the internal estimate to zero, as well as an initial response bias (an average expected location). The actual response of the subject would thus be determined by a relative weighting of the prior expectation and the accumulated acoustic evidence for the veridical sound elevation. Idiosyncratic day-to-day variation of the weighting factor could underlie the intersession variability in observed gains.

It is straightforward to appreciate how the dynamic correlation model of Hofman and Van Opstal (1998) could be extended to incorporate the nonlinear influence of stimulus level (Fig. 9). At low stimulus levels and short durations the accumulated evidence remains low; hence, the response gain will be low too. Note that the observed gains were not zero

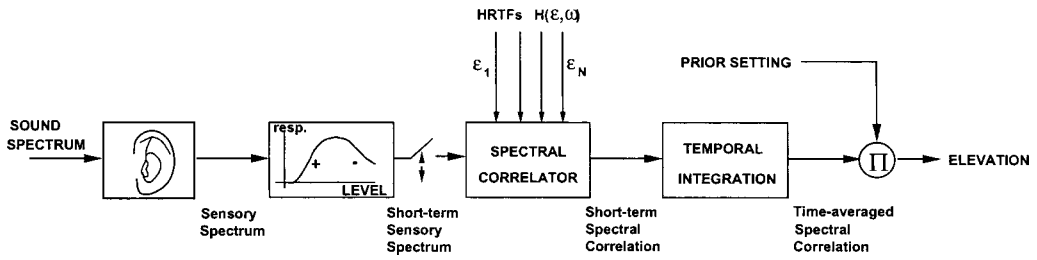

FIG. 9. Extension of the conceptual model of Hofman and Van Opstal (1998) in which the output of the short-term integration stage (which embodies a "multiple look" on the sensory spectrum over short (<5 ms) time windows) depends on sound level. The latter may be due to cochlear nonlinearities and/or neural tuning properties. Following the spectral correlation stage (comparison of the short-term sensory spectrum with stored HRTFs), a dynamic estimate of elevation is generated by averaging over a longer time window of several tens of ms. The output of this final stage is weighted against a preset default estimate, that may be based on prior expectation.

TABLE III. Partial regression coefficients for the multiple nonlinear regression on the data for all subjects and all stimulus conditions [Eq. (A2)].

Subject	Elevation gain				Bias			
	$\beta_1(\cdot 10^{-2})$	$\beta_2(\cdot 10^{-4})$	β_3	$\beta_4(\cdot 10^{-2})$	β_5	R^2	N	Session
JV	2.5	−2.3	0.46	3.3	−3.2	0.75	3195	1,2,3
FF	1.8	−1.8	0.52	8.6	1.8	0.79	1727	1
JM	1.0	−0.73	0.58	5.9	4.1	0.66	1616	1
JO	0.082	1.1	0.44	7.5	7.0	0.58	1005	1
MW	1.5	−0.99	0.35	7.1	9.7	0.71	1584	2,3
HV	2.1	−1.7	0.18	9.8	2.0	0.77	1585	2,3
MZ	1.3	−0.65	0.34	3.3	8.4	0.72	1562	2,3
WV	1.5	−1.2	0.31	8.8	8.6	0.82	1577	2,3
FW	0.72	−0.28	0.46	5.7	11.7	0.50	1592	2,3

for this condition, and that responses appeared to correlate well with the actual stimulus locations. Increasing the stimulus level will in turn improve the correlation, since the signal exceeds the critical SNR sooner. This effect would account for the positive level effect observed in our data. In the same vein, longer stimulus durations accumulate more and more evidence about the veridical sound elevation.

The nonlinear effect of stimulus level (positive and negative gain changes) reported in this paper could in principle be attributed to cochlear mechanisms (e.g., nonlinear amplification at low levels, and compression, or even clipping, at high levels). Alternatively, it might be due to central neural processing mechanisms (like neural saturation, or neural tuning to a specific optimal sound level; e.g., Ryan and Miller, 1978), or to both mechanisms. On the basis of the current experiments it is not possible to dissociate these possibilities.

ACKNOWLEDGMENTS

This research was supported by the Netherlands Organization for Scientific Research (NWO—MaGW, Project No. 410.203.01, JV) and the University of Nijmegen (AJVO). The authors would like to thank Dr. Armin Kohlrausch, Dr. Leslie Bernstein, Dr. Wes Grantham, and one anonymous reviewer for their valuable comments on an earlier version of this paper. We would like to acknowledge the valuable technical assistance of T. van Dreumel and H. Kleijnen for building the LED-speaker hemisphere. We thank Floor Franssen and Joost Maier for their help in carrying out the first series of experiments. We are also grateful to Dr. H. Versnel for setting up the data acquisition software. Finally, we thank our subjects for their time and effort.

APPENDIX:

In the nonlinear regression model of the elevation responses, the gain, G_ε was taken as a function of stimulus duration, D, and sound level, L

$$\varepsilon_R = G_\varepsilon(L,D)\cdot\varepsilon_T + b_\varepsilon . \tag{A1}$$

Based on the stimulus-specific linear regressions, plotted in Fig. 2, and the resulting gains, plotted in Figs. 4–6, the following function was chosen to capture the observed effects:

$$G_\varepsilon(L,D) = \beta_1\cdot L + \beta_2\cdot L^2 + \beta_3\cdot(1-e^{-\beta_4\cdot D}),$$

$$b_\varepsilon = \beta_5, \tag{A2}$$

with L stimulus level (in dB SPL), D duration (in ms). Fit parameters $\beta_1-\beta_5$ were obtained by minimizing the mean-squared error between model and data. The resulting regression parameters for each subject are listed in Table III.

Note that $-\beta_1/(2\beta_2)$ provides an estimate of the stimulus level that yields the highest elevation gain. For the subjects in this study, this optimal sound level was typically between 50–70 dB SPL. The value of $1/\beta_4$ determines at which stimulus duration the elevation gain is estimated to reach 63% of its maximum value: this yields values between 10–30 ms.

Blauert, J. (**1996**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Frens, M. A., and Van Opstal, A. J. (**1995**). ''A quantitative study of auditory-evoked saccadic eye movements in two dimensions,'' Exp. Brain Res. **107**, 103–117.

Goossens, H., and Van Opstal, A. (**1997**). ''Human eye–head coordination in two dimensions under different sensorimotor conditions,'' Exp. Brain Res. **114**, 542–560.

Hartmann, W. M., and Rakerd, B. (**1993**). ''Auditory spectral discrimination and the localization of clicks in the sagittal plane,'' J. Acoust. Soc. Am. **94**, 2083–2092.

Hofman, P. M., and Van Opstal, A. J. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.

Hofman, P. M., and Van Opstal, A. J. (**2002**). ''Bayesian reconstruction of sound localization cues from responses to random spectra,'' Biol. Cybern. **86**, 305–316.

Knudsen, E. I., and Konishi, M. (**1979**). ''Mechanisms of sound localization in the barn owl (*Tyto alba*),'' J. Comp. Physiol. **133**, 13–21.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Macpherson, E. A., and Middlebrooks, J. C. (**2000**). ''Localization of brief sounds: Effects of level and background noise,'' J. Acoust. Soc. Am. **108**, 1834–1849.

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C* (Cambridge University Press, Cambridge, MA).

Ryan, A., and Miller, J. (**1978**). ''Single unit responses in the inferior colliculus of the awake and performing rhesus monkey,'' Exp. Brain Res. **32**, 389–407.

Wightman, F. L., and Kistler, D. J. (**1989**). ''Headphone simulation of free field listening. I Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.