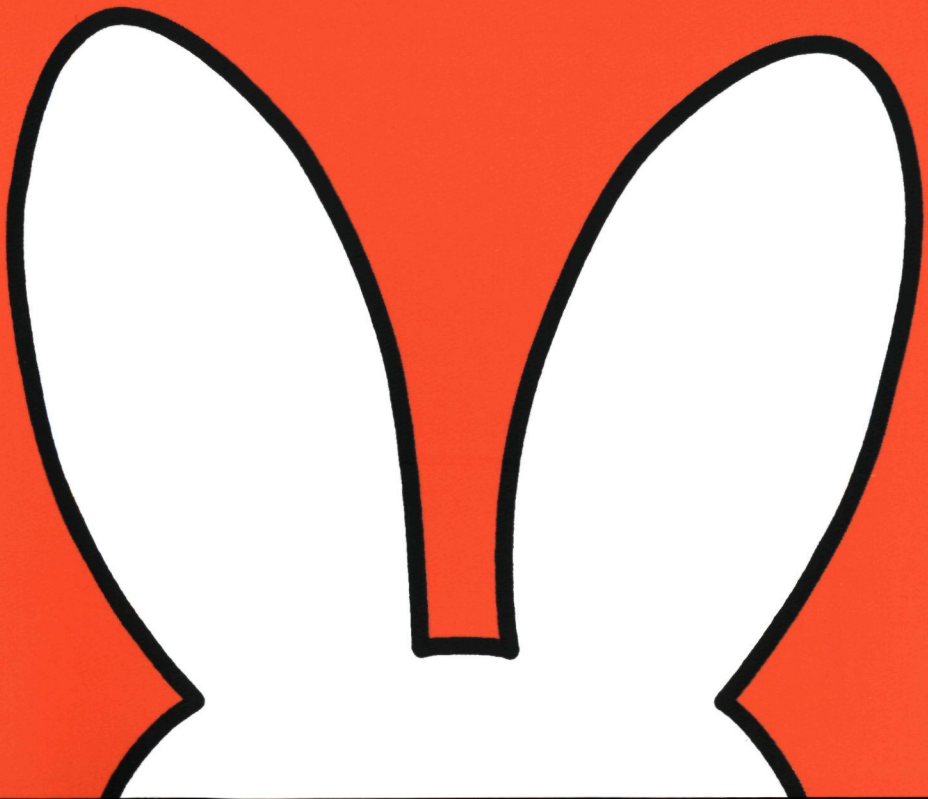
Localization behavior in audiovisual space

Joyce Vliegen



Localization behavior in audiovisual space

Een wetenschappelijke proeve op het gebied van de
Medische Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen,
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op dinsdag 21 november 2006
om 13.30 uur precies

door

Joyce Vliegen

geboren op 16 mei 1974
te Tilburg

Promotores	Prof dr C C A M Gielen Prof dr A J van Opstal
Manuscriptcommissie	Prof dr A R Cools Prof dr A Kohlrausch (Philips Research Laboratories Eindhoven) Dr E Brenner (Vrije Universiteit Amsterdam)

ISBN 90-9021109-8

This research was supported by NWO, Maatschappij- en gedragswetenschappen (MaGW)

Gedrukt door PrintPartners Ipskamp, Enschede

Contents

Chapter 1

Introduction

To safely move around in the world around us, it is important that we know where objects are located. For this, vision is the most important sense organ. We instinctively react to visual cues to avoid obstacles, stop for a red light, look for traffic before crossing a road, etc. But we rely on our ears more than we realize. For instance, we often hear a car coming before we see it. This is partly due to the fact that our field of view is limited, whereas we can hear sounds coming from any direction. In that way sound localization can help us to direct our attention to an event. In this thesis different aspects of sound localization are investigated. Chapters 2 and 3 concern the process of extracting the auditory target coordinates from the sound-source spectrum Chapter 4 studies the eye-head motor responses to the auditory target coordinates. In Chapter 5 a comparison is made to eye-head motor behavior in visual localization. This first chapter gives a short overview of several topics that are important in sound localization: the structure and working of the ear, the different localization cues and the problems the auditory system encounters in extracting these cues from the sound signal, and finally eye-head motor behavior to make an orienting movement toward the perceived target position.

1.1 The ear

Sound enters the ears at the pinnae (see Fig. 1.1), which filter the incoming sound waves in a direction-dependent manner due to their irregular shape. This is important for sound localization in the vertical direction, as will be explained later. Vibrations of the eardrum (tympanic membrane), at the end of the ear canal, set the ossicles of the middle ear in motion (see Fig. 1.1). These three small bones (malleus, incus, stapes) enable an efficient transfer of sound waves from air in the outer and middle ear to fluid in the cochlea in the inner ear. The stapes makes contact with the cochlea at the oval window.

The cochlea is the actual sensory organ of the ear. It consists of three fluid-filled chambers, divided by two membranes the basilar membrane and Reissner's membrane (see Fig. 1.2). The outer two chambers, the scala vestibuli and the scala tympani, connect at the inner tip of the cochlea, the helicotrema.

In response to a pure tone, a traveling wave will form at the basilar membrane, moving from the oval window toward the inner tip of the cochlea. The amplitude of the wave first slowly increases and then decreases abruptly. The location of the amplitude maximum of the envelope of this wave at the basilar membrane depends on the frequency of the tone. The properties of the basilar membrane vary over its length: near the oval window, it is rather narrow and stiff, whereas at the end, near the helicotrema, it is much wider and less stiff. As

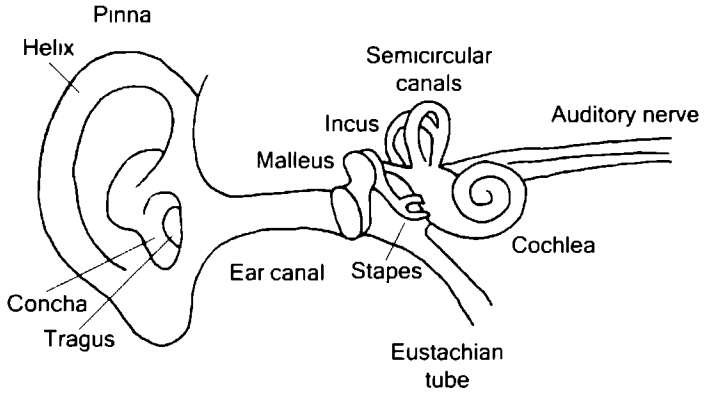


Figure 1.1: *Schematic overview of the ear, with the outer ear (pinna, ear canal), the middle ear with the ossicles (malleus, incus, stapes), and the inner ear (including the cochlea and semicircular canals) The semicircular canals are part of the vestibular system and are important for our sense of balance*

a result, high-frequency sounds produce maximum stimulation near the oval window, whereas low-frequency sounds have a maximum closer to the helicotrema. In this way, the cochlea acts as a frequency analyzer. This frequency-to-place conversion, or tonotopic mapping, is very important for our ability to perceive pitch.

More complex sounds produce a more complex pattern of vibrations at the basilar membrane. A complex tone consisting of two tones with a large enough frequency separation will result in two separate patterns of vibration with two clear maxima that correspond to the frequencies of the two tones. For two tones that are closer in frequency, the two vibration patterns will interact and a complex waveform results. If the tones are close enough in frequency, the two amplitude maxima will merge to one broader peak and the two components will not be resolved by the basilar membrane. The position of the amplitude peak of the envelope of this pattern varies logarithmically with frequency and the relative bandwidths of the excitation patterns are approximately constant, which makes the minimal frequency difference necessary for separation of two tones proportional to their center frequency (CF). Consequently, for a complex harmonic tone consisting of evenly spaced harmonics, the lower harmonics will be resolved by the basilar membrane, but the very high harmonics will not.

Located at the basilar membrane is the organ of Corti (see Fig. 1.2), which mediates the transduction of sound signals to the brain. This organ contains two

groups of hair cells: inner hair cells and outer hair cells. It is currently thought that the outer hair cells play a crucial role in the nonlinear amplification of the basilar membrane vibrations, especially at very low sound levels. Through their rapid shortening and lengthening, in phase with the sound's frequency, they provide positive mechanical feedback to the basilar membrane vibrations, resulting in much better frequency resolution Above the hair cells, touching the stereocilia (the hairs on top of the hair cells), lies the tectorial membrane, which is fixed on one side. If the basilar membrane is set in motion by auditory stimulation, the stereocilia of the inner hair cells will be displaced against the tectorial membrane, which eventually results in the generation of action potentials in the neurons of the auditory nerve.

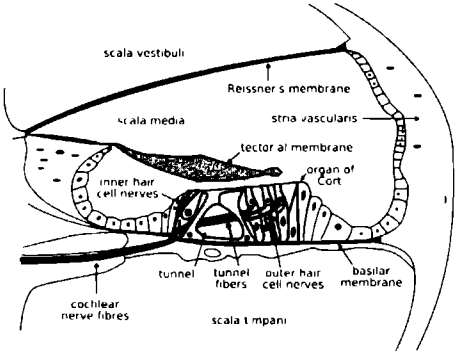


Figure 1.2: *Schematic overview of the cochlea, with the organ of Corti in the scala media The organ of Corti lies on the basilar membrane and contains inner and outer hair cells that make contact with the tectorial membrane through the stereocilia From http //en wikipedia org/wiki/Cochlea*

1.2 Sound localization

For humans, vision and hearing are the two primary senses for getting around in every-day life. In the visual system an event is projected onto the photo receptors of the retina and each point in the visual world corresponds to a point at the retina. Moreover, neighboring spatial positions stimulate neighboring retinal positions. Although our visual acuity in the fovea (the most central and most sensitive part of the retina, with sharpest vision) is very high, it deteriorates rapidly with retinal eccentricity and our visual field is limited. In contrast, we can detect sounds coming from any direction. Thus, sound localization can aid to direct the eyes to an area of interest so that we can further inspect it visually.

Heffner and Heffner (1992) found that mammals with narrow areas of best vision tend to have better sound localization acuity irrespective of head size or life style (diurnal or nocturnal) They concluded that "a primary function of sound localization is to direct the eyes to the source of a sound"

Unlike the visual system, the auditory system does not have a topographic organization, but as described above, is tonotopically organized Auditory localization is based on indirect physical cues that are present in the sound signal due to the interaction of the sound waves and the pinnae, head and body In the horizontal direction (azimuth), we primarily use the fact that our two ears are located on different sides of the head, differences between the two ears in sound level and timing provide information about sound direction in the horizontal plane (binaural cues) For instance, if someone is talking on your right side, the signal in your right ear will have a higher level than in your left ear due to a shadowing effect of the head, and the sound will arrive earlier in your right ear than in your left ear In the vertical direction (elevation), we use cues provided by direction-dependent filtering of the pinnae (monaural cues)

Binaural localization cues

In 1907 Lord Rayleigh put forward the idea that horizontal sound localization is based on two cues that complement each other interaural time differences (ITDs) for low frequencies and interaural level differences (ILDs) for high frequencies (the "duplex theory" of sound localization) Although his research was based on pure tones (produced by tuning forks), almost a century later Macpherson and Middlebrooks (2002) found that also for broadband sounds, the duplex theory provides a useful description of the relative importance of the ITD and ILD cues for the different frequency regions

ITD is the difference in arrival time between the two ears due to the longer pathway for the farther ear and the finite sound velocity (~340 m/s in air) In humans, the maximum pathway difference is about 23 cm, corresponding to an ITD of about 690 μs With ongoing sound signals, there is no interaural difference in onset time, only information on the difference in phase between the two ears (interaural phase difference, or IPD) is available However, for frequencies higher than about 1500 Hz, the wavelength becomes smaller than the difference in path length and the phase difference becomes ambiguous Therefore, ITDs are mainly important for sound localization at low frequencies The ambiguity might be resolved by combining phase information over several frequency bands or by making use of time differences in sound onset or of slow fluctuations in the ongoing envelope of the sound for higher frequencies Despite these limitations, ITDs are considered to be the most important localization cues Wightman and

Kistler (1992) proposed that ITDs may be used primarily to establish the possible sound locations, after which ILDs and spectral cues are used to resolve potential confusions.

Interaural level difference (ILD) is the result of attenuation by the head on the side opposite to the sound source and can amount to as much as 20 dB. ILDs are mainly important for sound localization at frequencies above about 2000 Hz. For lower frequencies the larger wavelength allows the sound to bend around the head with little attenuation.

Monaural localization cues

Although binaural difference cues can differentiate positions in the horizontal direction, for positions on the median plane (the plane trough the center of the head, from front to back) the sound will arrive at both ears simultaneously, and both ITD and ILD will be zero. In fact, on both sides of the head, centered around the interaural axis, a so called "cone of confusions" can be defined. For all positions on this cone the difference in distance between the near and far ear is the same, and thus the ITD and ILD will be the same. Thus, for any azimuth position, the ITD and ILD will be the same whether the sound source is located in the front or in the back, which may result in front-back confusions. To resolve these confusions and to localize sounds in the vertical direction the auditory system relies on monaural spectral cues.

Sound waves from different directions are reflected and diffracted by the pinna in different ways, resulting in direction-dependent patterns of attenuations and amplifications (see Fig. 1.3). These filter functions have become known as head-related transfer functions (HRTFs) HRTFs are mainly important for elevation localization for frequencies above 3 to 4 kHz. Through experience the auditory system is thought to have acquired and stored information about these HRTFs and to be able to use this internal representation of the spectral filters to extract sound-source elevation. Hofman et al. (1998) and Van Wanrooij and Van Opstal (2005) showed that also a new set of HRTFs can be learned. In these studies, subjects had a mold inserted into their ears to distort the pinna cues and within a few weeks they learned to localize with their new ears After a few weeks the molds were taken out and the subjects were tested again with undisturbed ears. Localization accuracy was as high as before the experiment, which shows that the new set of HRTFs did not interfere with the original HRTFs. Moreover, for some time after the molds had been taken out, subjects were still able to localize with the molds.

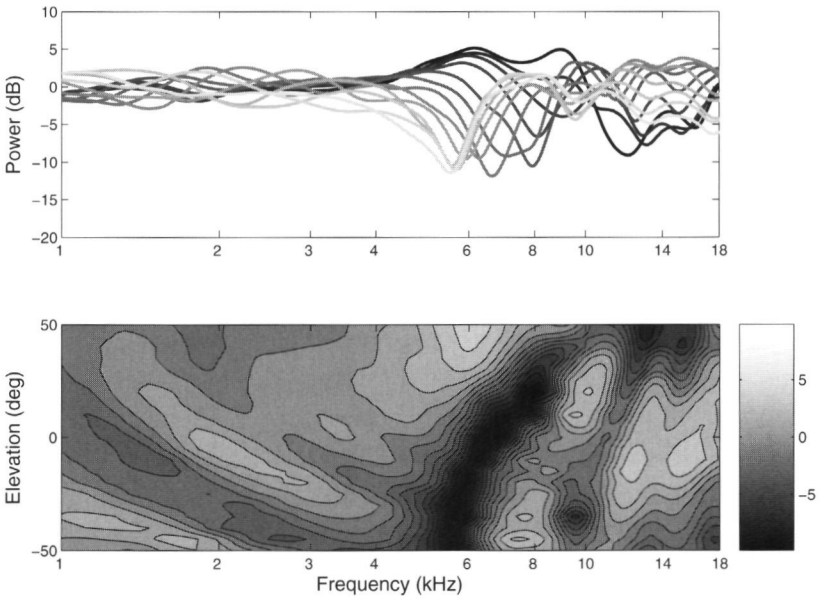


Figure 1.3: *Top: Head-related transfer functions (HRTFs) of listener JO for several elevations, with darker gray values coding for higher elevations. Bottom: Contour plot of HRTFs where gray value codes for amplitude: light gray values indicate peaks in the transfer function whereas dark gray values indicate notches. Note the prominent notch running from about 5.5 kHz at* $\varepsilon = -50^\circ$ *to about 10 kHz for* $\varepsilon = +50^\circ$.

An ill-posed problem The sound spectrum arriving at the eardrum (sensory spectrum, $S(f)$) is determined by a convolution of the sound-source signal and the impulse response of the HRTF of the sound-source direction: $S(f) = H(f;\alpha_0;\varepsilon_0) \cdot X(f)$, with α_0 the sound's azimuth angle, ε_0 the elevation angle, f the frequency, H the HRTF, and X the source spectrum. However, both the sound-source spectrum and the sound-source direction are a-priori unknown to the auditory system. This makes the extraction of sound-source elevation from the sensory spectrum an ill-posed problem, as there are infinitely may possible combinations of H and X yielding the same sensory spectrum. Several theories have been put forward to explain how the auditory system may deal with this problem.

Blauert (1969/1970) found that narrowband noises are localized based on their frequency content, rather than on their actual location. He called these frequency bands "directional bands". These frequency bands corresponded largely to important regions of amplification in the HRTF of the perceived direction. Musicant and Butler (1984) elaborated this idea into a model in which sound

localization is based on peaks in the sensory spectrum. They introduced the concept of *covert peak areas*: regions in space from which a narrow band of noise generates a maximum sound pressure level at the ear canal entrance.

Other models are based on a comparison between the sensory spectrum and a database of stored HRTFs. Middlebrooks (1992) proposed that the auditory system assumes natural sounds to have flat, broadband spectra ($X(f) = 1$). In this case, the sensory spectrum is equivalent to the HRTF. Hofman and Van Opstal (1998) suggested that the restrictions on sound spectra that can be localized accurately may be more relaxed. They showed mathematically that as long as the sound-source spectrum does not resemble any of the HRTFs, the similarity between the sensory spectrum and the HRTFs will always be highest for the HRTF of the actual sound direction.

Zakarauskas and Cynader (1993) hypothesized that, as HRTFs have rather steep amplitude spectra, sounds will be localized accurately as long as their spectra are locally flat or have a locally constant slope. Computer simulations indicated that a model based on a flat second derivative yielded the most accurate and robust localization performance. In this thesis, we have tested the predictions of these different models experimentally (Chapter3).

Other localization cues

Under some circumstances, head movements may provide additional localization cues. Front-back confusions may be resolved by comparing the perceived change in azimuth direction with head movement. Alternatively, head movement can be used in an active search to find the direction of maximal intensity, or the direction at which the input at the two ears coincides. Perrett and Nobel (1997b) found that head rotation can help to resolve front-back confusions for low-pass filtered, high-pass filtered and broadband noise. More importantly, head rotation was found to restore elevation localization for low-pass filtered noise (for which spectral cues are insignificant) and in conditions in which pinna cues are distorted. For high-pass filtered sounds, no effect of head rotation was found. They posed that with head rotation, interaural differences change differently for positions in the front and in the back. Moreover, the rate of change of the ITDs depends on the elevation of the sound: for positions on the horizon the rate of change is maximal, whereas for positions directly overhead or below the rate is zero. However, in their experiments the stimuli were 3 s long For shorter sounds, head movements do not seem to be an efficient cue. Goossens and Van Opstal (1999) showed that the elevation localization of 800-ms pure tones did not improve when the subjects were free to move their head. Note however, that the task was to make a rapid orienting movement and subjects

were not allowed to make slow searching movements to localize the sound In a review, Middlebrooks and Green (1991) concluded that head movement can only improve sound localization if the sound duration is long enough to allow the listener to tune in on the sound

1.3 Reference frames

In localizing sounds, a natural response is to make an orienting eye-head movement toward the sound source This may seem a relatively easy task, but the localization system encounters various problems in the programming of this response Firstly, as the ears are fixed on the head, sound-localization cues are head centered Therefore, an auditory target will initially be coded in a head-centered, or craniocentric, reference frame Conversely, although the eyes are in a fixed position in the head, they can rotate and assume different orientations (note that in this thesis position is often used as synonymous to orientation) Under natural circumstances, the eyes typically are in an eccentric orientation in the head, resulting in a misalignment of the head-motor vector and the eye-motor vector to reach the target For the eye-motor system to be able to guide the eyes to the direction of a sound source, the head-centered target coordinates have to be transformed into eye-centered motor commands For this transformation a signal about eye position in the head is needed (see Fig 1 4A) For visual localization the situation seems more straightforward Visual cues define an eye-centered, or oculocentric, reference frame, therefore no coordinate transformation is necessary for the motor commands to the eyes (see Fig 1 4B) However, usually the head is also involved in the orienting response As the eyes have a limited motor range ($\pm$ 45°), the head moves together with the eyes to bring them on target This combined eye-head movement is called a gaze shift Gaze is defined as the eye orientation in space and is well approximated by the head orientation in space plus the eye orientation in the head To make a head movement to a visual target, an eye-position signal is needed to compute the head-motor command (see Fig 1 4B)

Another challenging problem for sensorimotor behavior is that under most natural circumstances, the eyes and head are moving continuously, so that also during the presentation of the sound signal, the head and eyes may be moving, both relative to space and relative to each other Before an orienting response to the perceived target can be made, eye and head position will then have changed and the localization system will have to compensate for these movements to accurately foveate the target In typical localization experiments however, both the eyes and head start in the straight-ahead direction and a single target is

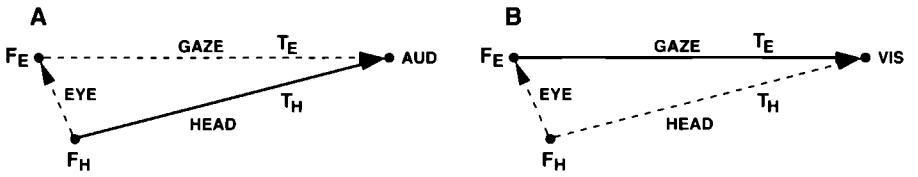


Figure 1.4: *Localization of a single auditory* **(A)** *and visual* **(B)** *target (AUD and VIS, respectively) For auditory targets information about eye position in the head (EYE) is needed to transform the initial craniocentric target coordinates (T_H) into an oculomotor error (T_I) to enable an accurate gaze shift (GAZE) to the target For visual targets the initial retinal target coordinates have to be transformed into a craniocentric motor error for an accurate head movement (HEAD) to the target F_I $_H$ indicates the fixation position of the eyes and head, respectively The continuous lines represent the initial auditory and visual target vector (T_H and I_E, respectively) The dashed line indicates the transformed auditory and visual target vector (I_E and T_H, respectively)*

presented. In this situation the eye-centered, head-centered, and body-centered reference frames all coincide and the craniocentric (auditory) or retinal (visual) target location has a one-to-one correspondence to the motor commands to the eyes and head for accurate localization of the target.

Goossens and Van Opstal (1999) investigated a more complex situation by means of a double-step experiment in which a visual and an auditory target were presented shortly after each other, before initiation of any eye-head movement. Subjects were asked to make eye-head localization movements to both targets in their order of presentation, resulting in a gaze shift to the auditory target after an intervening saccade to the visual target. Thus, for accurate localization of the second, auditory, target the initial craniocentric coordinates of this target had to be updated by the intervening eye-head movement to the first, visual, target (see Fig. 1.5). They found that their subjects made spatially accurate localization responses in which the preceding eye-head movements were fully accounted for.

One way in which the gaze-motor system may update the initial craniocentric target position to obtain the motor commands for accurate localization is by efference information on the intervening eye-head movement. Recordings of neurons in cortical areas of the monkey brain have suggested that this target update may already have been performed *before* the first movement is actually generated (Duhamel et al , 1992; Colby et al., 1995; Walker et al., 1995; Umeno and Goldberg, 1997). This corresponds to a *preprogrammed strategy*, based on feedforward information, as opposed to an updating mechanism that is based on feedback about the actual movement (either during or after the motor act). This preprogrammed update strategy is called "predictive remap-

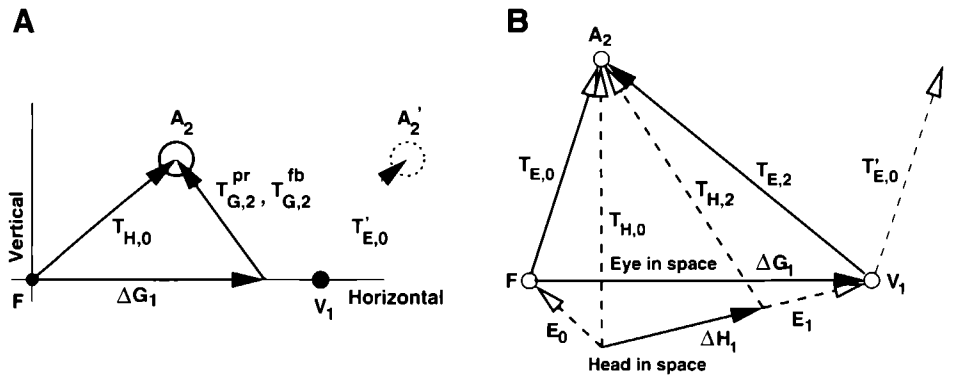


Figure 1.5: *In the static double-step condition, both targets (V_1 and A_2) are presented before initiation of eye-head movement (during fixation, F) At target presentation, the craniocentric target position is $T_{H\,0}$ and eye-in-head position is E_0 After the gaze shift to the first target (ΔG_1), the eye and head motor error for the second target are $T_{E\,2}$ and $T_{H\,2}$, respectively, and eye-in-head is E_1 Vector $T'_{E\,0}$ indicates the hypothetical localization movement to the second target if the localization system would not compensate for the first gaze movement* **(A)** *shows the predicted gaze shift to the second target according to the predictive remapping model ($T^{pr}_{G\,2}$) and according to the dynamic feedback model ($T^{fb}_{G\,2}$) In this condition, both models predict an accurate gaze response to the second target* **(B)** *shows the vectors that play a role in accurate eye-head motor behavior in an auditory-visual double step*

ping" Formally, one could write the coordinate transformation according to this scheme as: $T^{pr}_{G\,2} = T_{H\,0} - \Delta G^{pre}$, with $T^{pr}_{G\,2}$ the motor error to reach the target, $T_{H\,0}$ the initial craniocentric location of the target, and ΔG^{pre} the displacement of the eyes and head to the first target as predicted by the gaze-control system. Alternatively, the system could rely on *feedback* information on the actual movement. According to this strategy, the target is updated by the actual movement vector, ΔG^{act} as $T^{fb}_{G\,2} = T_{H\,0} - \Delta G^{act}$

Under most circumstances the preprogrammed eye-head movement (ΔG^{pre}) and the actual movement (ΔG^{act}) will be the same, and both schemes will predict the same target location. Hence, both these models can account for the accurate localization responses in the static double-step condition, in which both targets are presented *before* initiation of eye-head movement (see Fig. 1.5).

Dissociating feedforward from feedback programming As described in the previous section, the static double-step experiment does not allow for a dissociation of the feedforward and feedback programming schemes. However, with a slightly different version of this paradigm, the two schemes make different predictions. In the so-called dynamic double-step paradigm (Fig. 1.6), the second target is presented in midflight of the gaze shift to the first target. In this situ-

ation, the gaze control system is denied any prior information about the timing and location of the second target (and thus about the gaze shift after target presentation), and is therefore unable to make an accurate prediction of the updated target location. Indeed, according to this scheme, the update would be based on $T_{G,2}^{pr} = T_H^* - \Delta G_1$, with T_H^* the craniocentric error at the time of target presentation, and ΔG_1 the first gaze shift The system would make an error in the opposite direction of the first gaze shift (overcompensation) corresponding to the eye-head movement before target presentation.

In contrast, the feedback model would relie on the actual motor information, and would incorporate the partial gaze shift after target presentation, ΔG^* to update the target: $T_{G,2}^{fb} = T_H^* - \Delta G^*$. Figure 1.6 shows the full motor scheme for eye and head in space in the dynamic double step, and the required transformations that are needed to let both eye (ΔG_2) and head (ΔH_2) orient toward the second target. Our results, which are described in Chapters 4 (visual-auditory double step) and 5 (visual-visual double step) show that localization responses were still accurate when the second target was presented during the intervening eye-head movement to the first target, for both auditory and visual second targets.

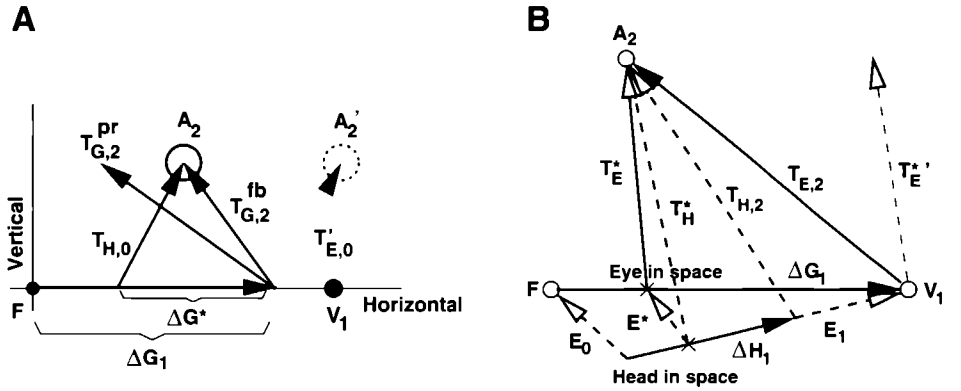


Figure 1.6: *In the double-step condition the second target (A_2) is presented during the eye-head movement toward the first target (ΔG_1, target presentation indicated by $\times$) At target presentation, the craniocentric target position is T_H^* and eye-in-head position is E^* Vector $T_E^{*\prime}$ indicates the hypothetical localization movement to the second target if the localization system would not compensate for the first gaze movement* **(A)** *shows the predicted gaze shift to the second target according to the predictive remapping model ($T_{G,2}^{pr}$) and according to the dynamic feedback model ($T_{G,2}^{fb}$) In this condition, according to the predictive remapping model the predicted gaze shift to the second target will have a localization error corresponding to the gaze movement before target presentation The dynamic feedback model still predicts an accurate gaze shift to the second target* **(B)** *shows the vectors that play a role in accurate eye-head motor behavior in an auditory-visual double step Same format as Figure 1 5*

Displacement vs. position feedback An important issue in gaze control studies concerns the nature of the feedback signals involved in target updating. Most research in this field has been done on oculomotor behavior to visual targets in conditions where the head was fixed (so that gaze displacement equals eye displacement). However, the models developed to explain the results of these studies can also be generalized for eye-head motor behavior in head-free conditions for both visual and auditory targets. Here I will describe the models as they were originally developed, for oculomotor responses to visual targets with a fixed head position. One of the main questions of this field of research is whether the targets remain in eye-centered coordinates (e.g. Jürgens et al. 1981), which are updated by *eye-displacement* information, or whether they are first transformed into a supraretinal (e.g. head-centered) reference frame (Robinson, 1975; Van Gisbergen et al. 1981), which is updated by continuous feedback about *eye position*

So far, neurophysiological research has favored the displacement scheme as it contains signals that are found throughout the visuomotor system: retinal error, eye displacement, and eye-motor error. The latter, T_E, is obtained by subtracting eye displacement, ΔE, from the initial retinal error, $T_{E\,0}$: $T_E = T_{E\,0} - \Delta E$ (see Fig. 1.7A). The eye displacement signal ΔE is thought to be generated by a resettable integrator that integrates eye velocity during the saccade and is reset (presumably with a time constant of several tens of ms) to zero after each saccade.

Instead, the supraretinal model relies on feedback from eye position, E, rather than eye displacement. First, the target is transformed into a head-centered reference frame: $T_H = T_{E\,0} + E_0$ with E_0 the eye position at target onset. After the saccade is finished, the updated target location is found by subtracting the current eye position, E_1, from T_H : $T_E = T_H - E_1$ (see Fig. 1.7B).

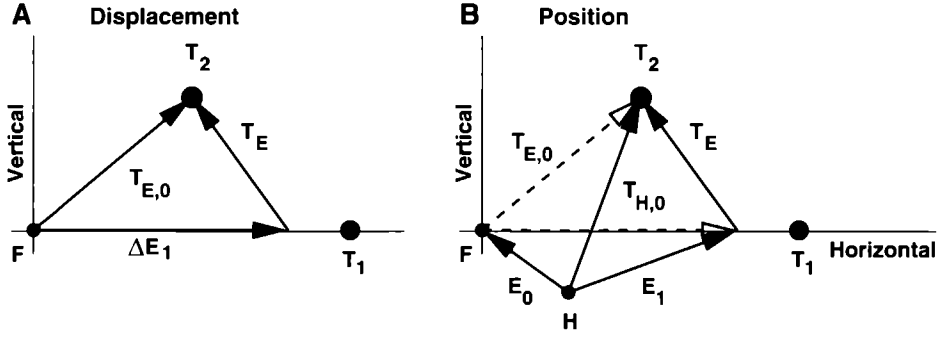


Figure 1.7: *Updating of the target position according to a displacement scheme* **(A)** *and a position scheme* **(B)**

	displacement		position
	visual	auditory	visual and auditory
gaze	$\Delta G_2 = T_L - \Delta G_1$	$\Delta G_2 = T_H - \Delta H_1 - E_1$	$\Delta G_2 = T_S - H_1 - E_1$
head	$\Delta H_2 = T_L - \Delta G_1 + E_1$	$\Delta H_2 = T_H - \Delta H_1$	$\Delta H_2 = T_S - H_1$

Table 1 1 *Coordinate transformations according to the displacement scheme and the position scheme for gaze and head, for both visual and auditory targets*

Note that these schemes can be readily extended to include head and body movements, in which case the target in the position model is transformed into a body-centered, or even a world reference frame (spatial coordinates T_S, see Tab 1 1) Note that the transformations for visual and auditory targets are the same For the displacement model, different transformations are needed for auditory and visual targets (see Tab 1 1)

Again, most behavioral visuomotor studies (typically head-fixed) do not allow to dissociate these possibilities However, when considering orienting in a multimodal environment (auditory and visual), with the eyes and head, and under static and dynamic conditions, our experiments (described in Chapters 4 and 5) provide a number of arguments as to why a position-updating scheme is more parsimonious than a displacement scheme

1 As illustrated in Figure 1 4, even the displacement model cannot purely rely on displacement signals to explain why both eye and head can make goal-directed responses to a visual and an auditory target, as eye-position information is always needed

2 In dynamic double-step condition (Fig 1 6) the partial displacements of the eyes (ΔG^*) and head (ΔH^*) after target presentation are needed to update the target location For the displacement scheme, which is based on the operation of resettable integrators, this would require the immediate stopping and restarting of the integrators in midflight of the gaze shift, as any delay would introduce systematic errors Recent studies suggest that this resetting takes several tens of ms (Nichols and Sparks, 1995) However, no such errors were observed in our data

3 A target representation in body (or world) coordinates greatly simplifies the neural computations to update the motor errors for eyes and head to targets, as the underlying transformations for auditory vs visual, and for static vs dynamic will be identical (see Tab 1 1) For displacement updating however, these transformations will strongly depend on target

modality and on stimulus timing In our experiments we observed no systematic differences between the different conditions, suggesting that the underlying transformations were indeed quite similar

1.4 Research topics of this thesis

In Chapter 2 the influence of sound duration and intensity on elevation localization is investigated In previous research it was found that elevation gain decreases for short noise bursts at high sound levels This could be due to either cochlear saturation at high sound levels, or to the fact that the signal is too short to allow reliable extraction of spectral localization cues, impeding a stable estimation of sound-source elevation Our experiments try to reconcile these two models

Chapter 3 examines how the auditory system is able to extract sound-source elevation from the sensory spectrum, when both the sound-source spectrum and the sound-source direction are unknown Different spectral localization models were tested by measuring localization responses to randomly shaped rippled amplitude spectra that were presented from a fixed speaker position

In Chapter 4, a visual-auditory double-step localization paradigm was used to investigate eye-head motor behavior in situations where the eyes and head are not initially aligned and make an intervening gaze shift before making a localization movement to the auditory target (static condition) and where the eyes and head are moving during presentation of the target (dynamic condition) In these situations the initial craniocentric target representation does not correspond to the eye-motor error and head-motor error to reach the target and the intervening gaze shift has to be taken into account for accurate localization of the target

In Chapter 5 these experiments were repeated with visual-visual static and dynamic double-step conditions to be able to compare auditory and visual localization behavior and to put the results of Chapter 4 in a coherent framework

Chapter 2

The influence of duration and level on human sound localization

The localization of sounds in the vertical plane (elevation) deteriorates for short-duration wide-band 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 we test these predictions 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 Our finding that elevation gain was a nonmonotonic function of sound level for all sound durations, however, is predicted by neither model. We conclude that both mechanisms underlie the elevation gain effect and propose a conceptual model to reconcile these findings.

Adapted from Vliegen J and Van Opstal AJ (2004) *The influence of duration and level on human sound localization* **J Acoust Soc** 115 1705-1713

2.1 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, 1989a; 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 non-flat stimulus spectra However, when amplitude variations within the source spectrum become too large, the localization accuracy of sound elevation deteriorates (Wightman and Kistler, 1989a; 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 non-stationary 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 (1998) proposed that the sound localization system needs to integrate about 40 to 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 to 86 dB) and lower (68 to 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 dB 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 like 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 level 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° oculomotor range) by Hofman and Van Opstal (1998), vs head movements over a much larger measurement range in Macpherson and Middlebrooks (2000)

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 to 100 ms) and sound levels (26 to 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 Figure 2 1

2.2 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

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

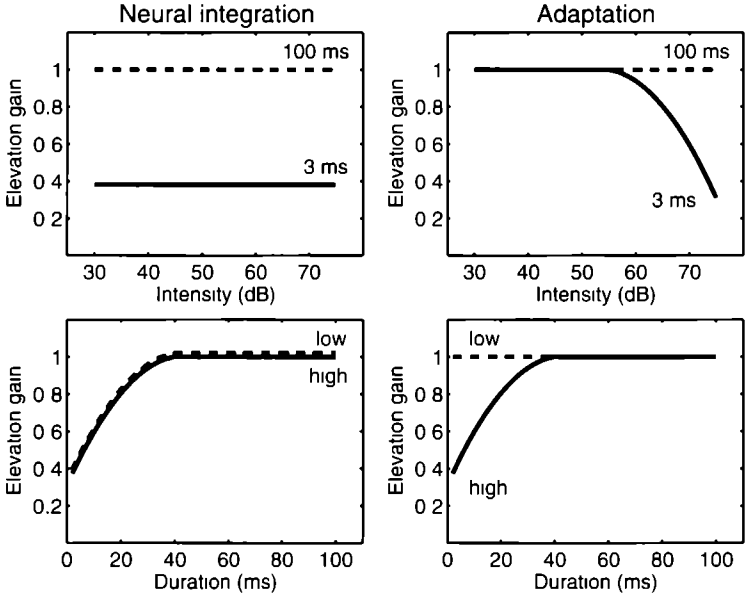


Figure 2 1: *Predictions of the neural integration model (left) and the cochlear adaptation hypothesis (right) Top row Elevation gain as a function of intensity for two durations Bottom row Elevation gain as a function of duration for low and high intensities The adaptation model predicts a decrease 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*

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 two and three

Apparatus

Experiments were conducted in a completely dark and sound-attenuated room with dimensions LxWxH = 3.5 x 2.45 x 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 eccentricity, 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, $R = 0^\circ$ corresponds to straight ahead for each ϕ, and $\phi = 0$, 90, 180 and 270° (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° (corresponding to each of the 12 spokes) and eccentricities of $R = 0$, 15, 30, 45, 60 and 75° (corresponding to the five rings) At the outer ring ($R = 75^\circ$) part of the framework (at downward directions $\phi = 240$, 270 and 300°) was removed to allow for space for the subject's legs A thin glass fiber ended in the center of each speaker, through which a well-defined visual stimulus (0 15° 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 $[R\ \phi] = (0,0)^\circ$ served as a fixation light at the start of a localization trial The polar target coordinates (R, ϕ) were transformed into azimuth - elevation angles $(\alpha\ \varepsilon)$, in the offline analysis of the data (see Data Analysis 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°, 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)

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

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

Table 2.1: *Detection thresholds in dB SPL for all subjects for the four stimulus durations employed in the second and third session*

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 offset 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)

Sensation levels

For the six subjects that participated in sessions two and three, free-field detection thresholds for broad-band 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 2.1 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.

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 eccentric (exceeding 20°), resulting in potentially large (and variable) undershoots of the measured head position if subjects use both eyes and head to point to the

target (Goossens and Van Opstal, 1997b) 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 toward 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 to 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

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, 1997b) 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:

$$\begin{aligned} \alpha_R &= G_\alpha \cdot \alpha_T + b_\alpha \\ \varepsilon_R &= G_\varepsilon \cdot \varepsilon_T + b_\varepsilon \end{aligned} \tag{2.1}$$

where $(\alpha_R, \varepsilon_R)$ are the head-movement response components, $(\alpha_I, \varepsilon_I)$ 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 Results)

2.3 Results

Typical localization results of the first experimental session are presented in Figure 2 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

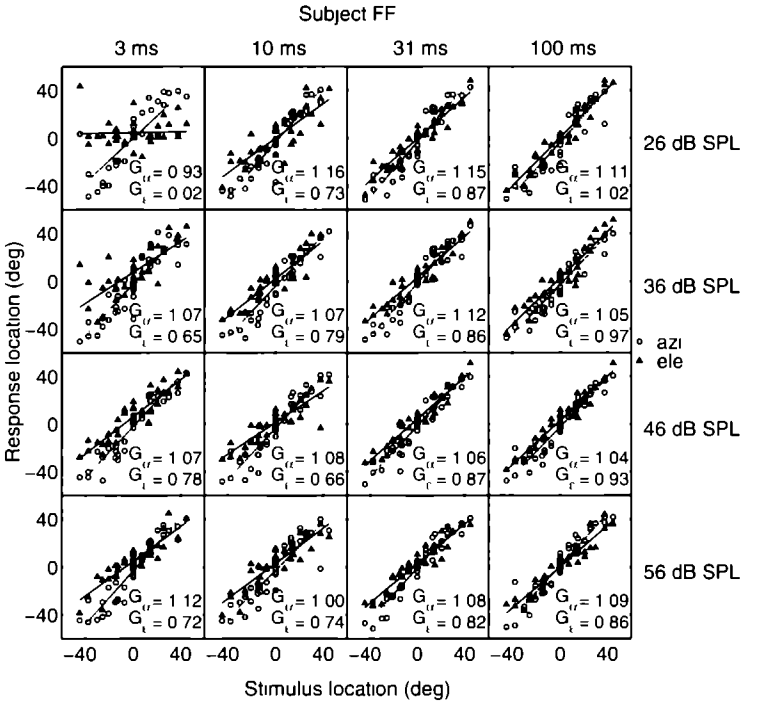


Figure 2 2 *Stimulus response relations for azimuth (○) and elevation (▲) response components of subject FF for four different stimulus durations (columns) and four different stimulus intensities (rows) Data taken from session one Best-fit regression lines (dotted azimuth, solid elevation) are also shown,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

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 experiment, 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 Figure 2.3 the gains for the azimuth response components of sessions two and three are plotted as a function of stimulus duration for all intensities

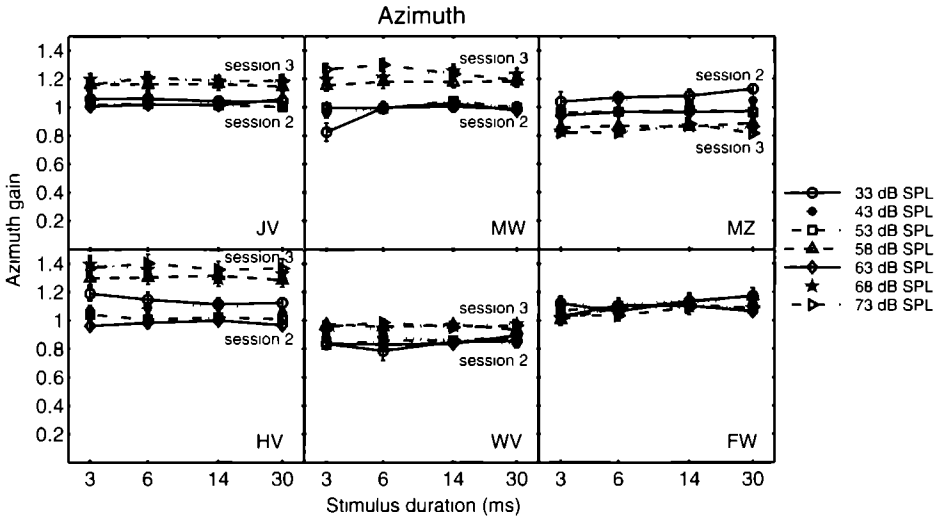


Figure 2.3: *Azimuth gains as a function of stimulus duration for all six subjects of sessions two and three The different line styles and symbols in each panel correspond to the different stimulus levels. Note the absence of any consistent trend and apparent separation of the obtained gain values for the two sessions in all but one subjects.*

and all six subjects. For most subjects, 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, each one corresponding to a different recording session (two or three). 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 Figure 2.4 in the same format as Figure 2.3. 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 to 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 Figure 2.4, elevation gain also appeared to vary with

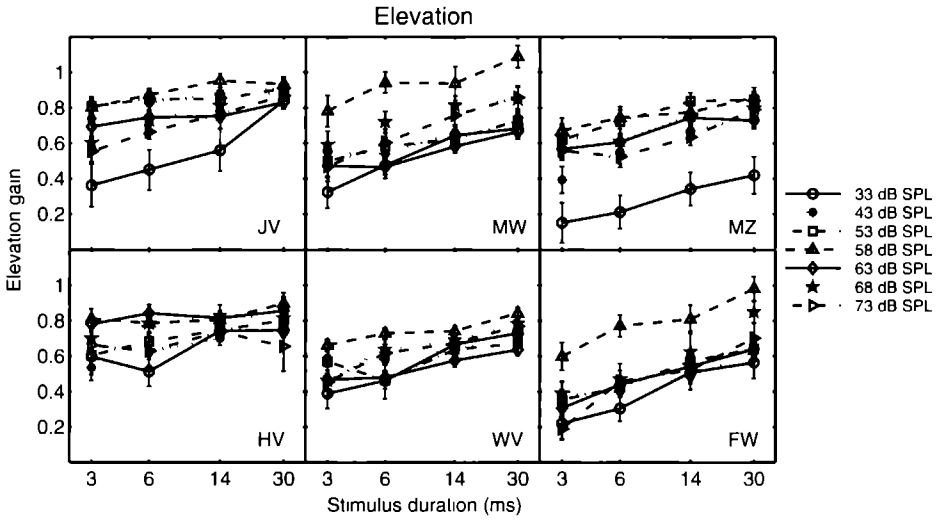


Figure 2.4: *Elevation gains as a function of stimulus duration for all six subjects of sessions two and three. Same format as Figure 2.3. Note the clear effect of stimulus duration on elevation gain for all subjects and at all stimulus levels*

stimulus intensity. This feature is better illustrated in Figure 2.5, 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 two and three. 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 effect, was not predicted by either the neural integration hypothesis or the adaptation hypothesis.

It should be noted that, as in Figure 2.3, 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 Figure 2.6. 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.

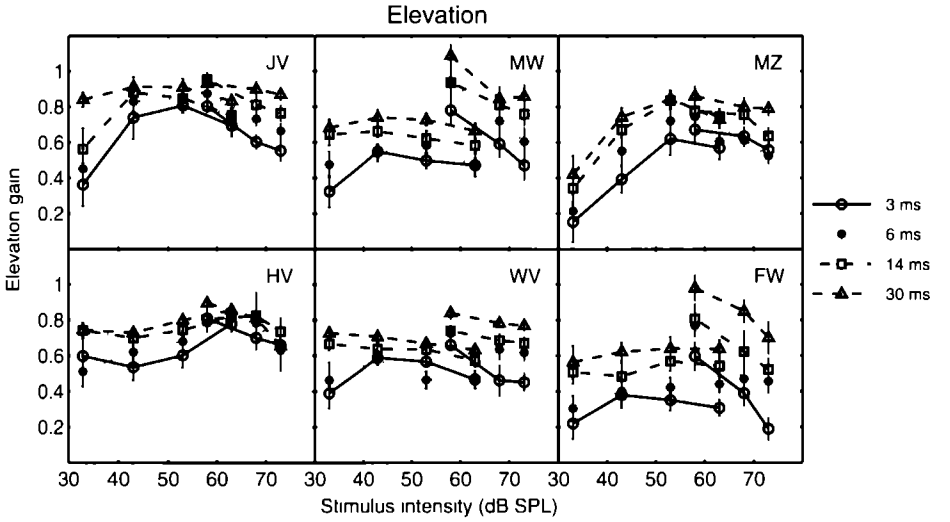


Figure 2.5: *Elevation gains as a function of intensity (in dB SPL) for all six subjects of sessions two and three. The different line styles and symbols correspond to the different stimulus durations. Note the consistent nonmonotonic changes of elevation gain with stimulus level. Note also that both a positive and a negative level effect were observed for all stimulus durations.*

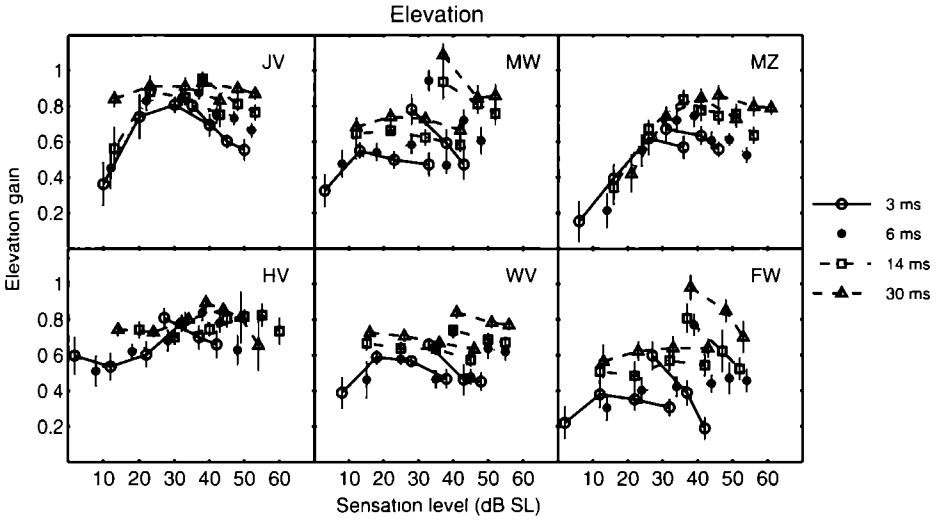


Figure 2.6: *Elevation gains as a function of sensation level for all six subjects of sessions two and three. Same format as Figure 2.5.*

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 Eqn. 2.1 was taken to be a function of both intensity and duration, yielding $G_\epsilon(D,I)$. The shape of this function was estimated on the basis of the results shown in Figures 2.2, 2 4, 2.5 and 2.6 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 and 3195 data points, see 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 Tab. 2.3 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 and 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 (Eqn. 2.1) on the data from session three only (for which L = 58, 68, 73 dB SPL):

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

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_I and β_D are the (dimensionless) partial regression coefficients. The resulting regression parameters for each subject are listed in Table 2.2 (left portion) Note that all coefficients for stimulus level are indeed negative, while for sound duration they are positive. More importantly, the absolute values of

Subject	Negative level			Positive level		
	β_L	β_D	R^2	β_I	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	

Table 2.2: *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) (Eqn 2 2)*

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 (Eqn. 2.2) 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 2.2 (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, Figure 2.7 shows the response variability (defined as the mean squared error around the regression line) of the data from sessions two and three 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°).

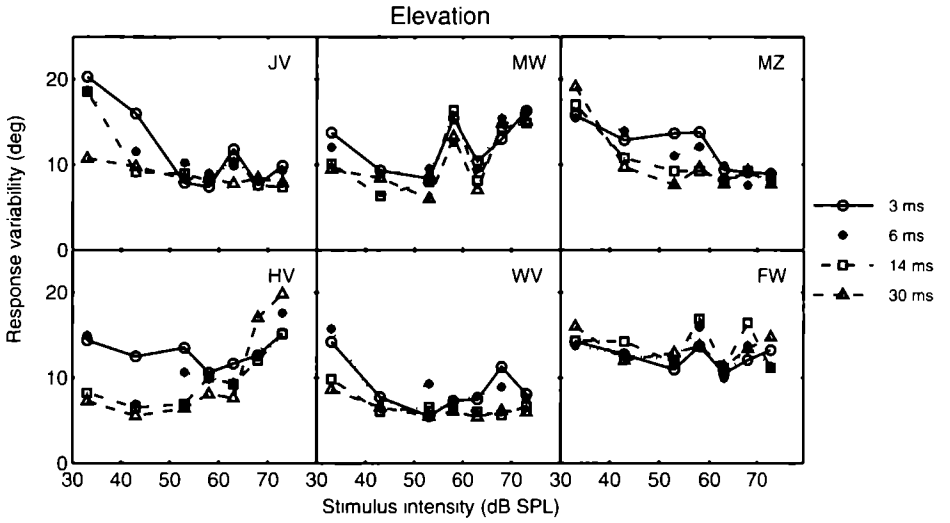


Figure 2.7: *Variability of elevation responses as a function of stimulus intensity. The different line styles and symbols correspond to the different stimulus durations. For most stimuli the variability is comparable; in contrast to the effects on response gain values, the variability does not change with recording session.*

2.4 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. 2.2, 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. 2.2, 2.4 to 2.6). Neither response bias (not shown), nor response variability (Fig. 2.7) was systematically related to the stimulus parameters. Our results are summarized in Fig. 2.8, which plots, in the format of Fig. 2.1, the prediction of Eqn. 2.4 (see appendix) applied to the pooled elevation gain data of subject JV. A comparison of Figs. 2.8 and 2.1 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

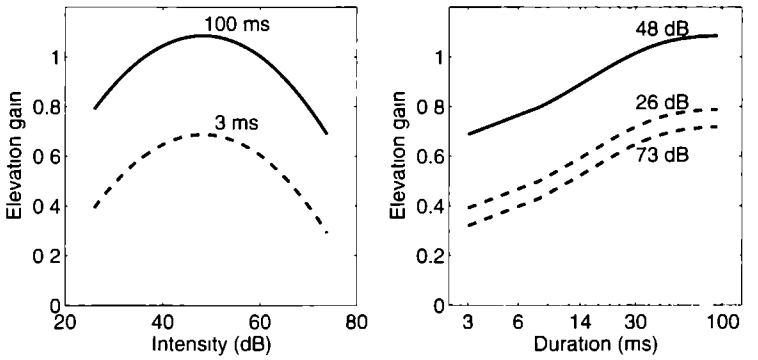


Figure 2.8: *Schematic summary of the results, presented in the same format as Fig 2 1 Curves are based on the parameters of a nonlinear regression (Eqn 2 4) on the elevation gain data of all three sessions for subject JV Note the logarithmic x-axis*

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.

Elevation gains varied in a nonmonotonic way with sound intensity (Figs. 2.5, 2.6). 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 to 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 (Tab. 2.2). 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. 2.3 to 2.6 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 to 63 dB vs 58 to 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 inter-burst 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 relie 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 inter-session 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. 2.9). 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 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 due to the higher SNR. This effect would account for the positive level effect observed in our data. In the

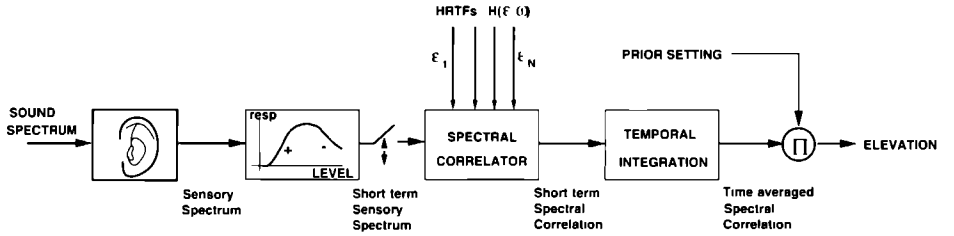


Figure 2 9 *Extension of the conceptual model of Hofman and Van Opstal (1998) in which the output of the short-term integration stage (which embodies a multiple look on the sensory spectrum over short (< 5 ms) time windows) depends on sound level The latter may be due to cochlear nonlinearities and/or 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*

same vein, also 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 - section Maatschappij- en gedragswetenschappen, MaGW, project nr 410-20-301, 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

	Elevation gain				Bias			
Subject	$\beta_1(\cdot 10^{-2})$	$\beta_2(\cdot 10^{-4})$	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

Table 2.3: *Partial regression coefficients for the multiple nonlinear regression on the data for all subjects and all stimulus conditions (Eqn 2 4)*

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:

$$\varepsilon_R = G_\varepsilon(L,D)\cdot\varepsilon_I + b_\varepsilon \tag{2.3}$$

Based on the stimulus-specific linear regressions, plotted in Fig. 2.2, and the resulting gains, plotted in Figs. 2.4 to 2.6, the following function was chosen to capture the observed effects.

$$\begin{aligned} G_\varepsilon(L,D) &= \beta_1\cdot L + \beta_2\cdot L^2 + \beta_3\cdot(1-e^{-\beta_4 D}) \\ b_\varepsilon &= \beta_5 \end{aligned} \tag{2.4}$$

with L stimulus level (in dB SPL), D duration (in ms). Fit parameters β_1 to β_5 were obtained by minimizing the mean-squared error between model and data. The resulting regression parameters for each subject are listed in Table 2.3

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 and 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 and 30 ms.

Chapter 3

Reconstructing spectral cues for sound localization from responses to rippled noise stimuli

Human sound localization in the vertical plane (elevation) relies on an analysis of the complex spectral shape cues provided by the pinnae. However, because the actual free-field stimulus spectrum is a-priori unknown to the auditory system, the problem of extracting the elevation angle from the sensory spectrum is ill-posed. In this study we tested different spectral localization models by eliciting head movements toward broadband noise stimuli with randomly shaped rippled amplitude spectra emanating from a speaker at a fixed location, while varying the ripple bandwidth between 1.5 and 5.0 cycles/octave. Six listeners participated in the experiments. From the distributions of localization responses toward the individual stimuli, we estimated the listeners' spectral-shape cues underlying their elevation percepts, by applying a statistical analysis based on maximum-likelihood estimation.

The resulting reconstructed spectral cues appeared to be invariant to the considerable variation in ripple bandwidths, and for each listener had a remarkable resemblance to their idiosyncratic head-related transfer functions (HRTFs). These results are not in line with models that rely on the detection of a single peak or notch in the amplitude spectrum, nor with a local analysis of first- and second-order spectral derivatives. Instead, our data support a model in which the auditory system performs a cross-correlation analysis between the sensory input at the eardrum and stored representations of HRTFs, to determine the perceived elevation angle.

Submitted for publication

3.1 Introduction

Human directional hearing relies on the processing of acoustic cues that originate from the interaction of sound waves with the head and pinnae. Sound localization in the horizontal plane (*azimuth*) utilizes binaural differences in sound arrival time and phase for frequencies up to about 1.5 kHz, and in sound level for frequencies exceeding about 3 kHz.

For localization in the vertical plane (*elevation*), the auditory system employs the fact that sound waves (above 3 to 4 kHz) arriving at the ears are reflected and refracted within the asymmetrical pinna aperture before reaching the eardrum, which results in an elevation-dependent pattern of amplifications and attenuations of the amplitude spectrum in the ear canal. These patterns are known as head-related transfer functions, or HRTFs (e.g. Wightman and Kistler, 1989a,b), and the auditory system is able to extract the sound-source elevation angle from these spectral shape cues (see Blauert, 1997, for a review). It is generally assumed that the auditory system has acquired, and stored, knowledge about the HRTFs through learning and interacting with the acoustic environment. Indeed, recent studies by Hofman et al. (1998) and by Van Wanrooij and Van Opstal (2005), in which the pinna geometry was altered by inserting a small mold in the concha, showed that the human auditory system can learn new sets of HRTFs within one to a few weeks. Although the driving force for this learning is yet to be established, it is likely to be guided by feedback from the environment, e.g. by combining information about self motion (head and body movements) and information from the visual system with the acoustic input and the associated sound-localization errors (e.g. Zwiers et al., 2003).

In this paper, we study the mechanisms that may underly the neural mapping from the spectral shape cues into an estimate of sound source elevation.

The auditory system faces a fundamental problem in determining the elevation angle of the sound source, ε_s, as the acoustic pressure at the eardrum, $s(t;\varepsilon_s)$ (here denoted as the *sensory signal*), results from a convolution of the sound-source pressure in the free field, $x(t)$, the direction-dependent acoustic filter of the head and pinna, $h_{\text{pinna}}(t;\varepsilon_s)$, and the (direction-independent) filtering of the ear canal, $h_{\text{canal}}(t)$:

$$s(t;\varepsilon_s) = h_{\text{canal}}(t) \star h_{\text{pinna}}(t;\varepsilon_s) \star x(t) \tag{3.1}$$

where $\star$ indicates convolution. Fourier transformation of Eqn. 3.1, followed by taking the logarithm of the amplitude spectrum and frequency results in a spectral representation of the sensory signal, as it is thought to be represented in the auditory system:

$$\log S(\omega;\varepsilon_s) = \log X(\omega) + \log H(\omega;\varepsilon_s) \tag{3.2}$$

with ω the frequency in octaves, $S(\omega;\varepsilon_s)$ the sensory spectrum, $X(\omega)$ the sound-source spectrum, and $H(\omega;\varepsilon_s)$ the combined transfer characteristic of head, pinna, and ear canal (the HRTF).

Both the sound-source spectrum and the HRTF associated with the source direction are a-priori unknown to the auditory system, which renders the estimation of sound-source elevation on the basis of spectral filtering an ill-posed problem. In order to deal with this problem, the auditory system is thought to make certain assumptions about the source spectrum. Different mechanisms have been proposed in the literature to explain how the elevation angle may be extracted from the sensory input.

Models for elevation localization

Essentially, two types of models have been put forward. In the first type, the auditory system searches for a particular feature in the sensory spectrum (e.g. a spectral peak or a notch), which is compared to stored knowledge about the HRTFs. The localization percept is then determined by the HRTF containing that particular feature in its amplitude spectrum. In the second type, the entire spectrum is analyzed and compared to the spectral shapes of the stored HRTFs.

The CPA model Blauert (1969/1970) found that narrow-band noises are localized on the basis of their frequency content rather than their actual location and that the frequency band corresponds to an important region of amplification in the HRTF that is associated with the perceived location. He called these frequency bands "directional bands". Musicant and Butler (1984) extended these experiments and found similar results. This led Butler and colleagues to propose that the peaks in the sensory spectrum act as a natural cue for sound localization. They introduced the concept of the *covert peak area* (CPA; Butler, 1987; Butler and Musicant, 1993; Musicant and Butler, 1984; and Rogers and Butler, 1992), which is defined as the region in space from which a narrow band of noise generates a maximum sound pressure level at the ear canal entrance. Rogers and Butler (1992) bandpass-filtered noise stimuli to contain only frequencies associated with a particular CPA for "down" or "up" locations in the vertical plane for a specific listener. Monaural elevation judgments were in general agreement with the CPA theory. Moreover, Butler and Musicant (1993) found that for broadband noise stimuli in which selected frequency segments were attenuated,

binaural localization judgments were displaced away from the CPAs associated with the attenuated frequency regions. These findings demonstrate that energy peaks in the sound spectrum influence sound localization.

Cross-correlation models A model of the second type was first formulated by Middlebrooks (1992), who proposed that to solve the ill-posed problem, the auditory system assumes that spectra of natural sounds are broadband and flat ($X(\omega) = \text{constant}$). In this case, the spectrum at the eardrum is entirely dominated by the HRTF associated with the sound-source direction In his model, the auditory system performs a cross-correlation between the sensory spectrum and a library of stored broadband HRTFs.

However, the assumption of a flat source spectrum may be too strict for adequate sound-localization performance. Indeed, sound-localization studies indicate that there is a considerable tolerance as to the shape of the amplitude spectrum For example, in a study with virtual stimuli filtered with HRTFs, Kulkarni and Colburn (1998) found that spectral details may not be very important in sound localization; considerable smoothing of the HRTFs was allowed without affecting the perceived elevation.

In their formulation of the cross-correlation model, Hofman and Van Opstal (1998) showed mathematically that as long as source spectra do not resemble any of the HRTFs (i.e. the correlation between the source spectrum and the HRTFs is low), the cross-correlation between the sensory spectrum and the set of HRTFs will be guaranteed to peak at the HRTF of the actual sound direction Hence, if localization were based on determining the HRTF of maximum cross-correlation, it will be accurate for a broad class of spectral shapes. Mislocalization will occur only if the source spectrum does correlate well with one or several of the HRTFs.

Local first and second derivative models An alternative model that analyzes the entire spectral shape of the sensory spectrum, and that is not restricted to flat source spectra, was proposed by Zakarauskas and Cynader (1993). They hypothesized that as the amplitude spectra imposed by the HRTF are rather steep, the auditory system should have no problem in localizing sounds for which either the source spectrum is locally flat, or the slope of the spectrum is locally constant. They developed two computational models of spectral cue localization that were based on the first and second derivatives of the sensory spectrum. Their computer simulations indicated that a system assuming a source spectrum with a flat second derivative would yield more accurate and robust localization performance than the model based on a locally flat first derivative.

Testing the different models

Recently, Hofman and Van Opstal (2002) performed an experiment in which they exploited the prediction from the cross-correlation model, that if the sound-source spectrum resembled any of the stored HRTFs, the perceived elevation would be mislocalized in the direction of that HRTF They presented listeners with a large set of broadband sounds that had randomly shaped amplitude spectra, emanating from a fixed speaker at the straight-ahead location The elevation distributions of eye-movement localization responses for the entire set of rippled-noise stimuli were used to reconstruct the potential spectral features underlying the localization of sound-source elevation

The reconstruction was based on a linear weighting of the rippled spectra, in which the maximum likelihood estimate of each stimulus served as its weighting factor Interestingly, the resulting reconstructed spectral shapes appeared to resemble the listeners' actual HRTFs, and thus seemed to support the hypothesis that perceived sound-source elevation may be determined by the entire spectral shape of the HRTF, rather than by a single prominent spectral feature However, the experiment was not specifically designed to dissociate the different models described above

In the current study, we have extended this paradigm, with the aim to test the predictions of the different models outlined above In particular, we have applied different sets of stimuli, in which the shapes of the amplitude spectra were determined by different ripple bandwidths If a particular spectral feature (like a peak, or a notch) determined the elevation percept, the reconstruction should yield this particular feature, irrespective of the ripple bandwidth On the other hand, as the first- and second-order local spectral derivatives of the rippled stimuli systematically varied with ripple bandwidth, the model of Zakarauskas and Cynader (1993) predicts that the spatial range of stimulus mislocalizations and ripple bandwidth will co-vary the faster the amplitude ripples, the higher the local spectral first- and second-order derivatives, resulting in larger mislocalizations Finally, if the reconstructions for the different stimulus sets prove to be similar to the spectral shapes of the listener's HRTFs, and invariant to variations in the ripple bandwidth, this would support the spectral cross-correlation model

Our results show that the psychophysical reconstructions of spectral features yield similar results for the different stimulus sets, and that the distributions of stimulus mislocalizations do not depend in a systematic way on ripple bandwidth We therefore propose that the auditory system performs a cross-correlation analysis on the actual sensory spectrum and the set of stored HRTF representations, and does not seem to rely on either a single peak or notch in the spectrum, or on a criterion based on local first or second spectral derivatives Because the spatial

range of localization responses was strongly influenced by the speaker's location, our results also indicate that the perceived elevation is not simply determined by the site of maximum cross-correlation.

3.2 Methods

Generating broad-band rippled noise stimuli

Details on the generation of the random-spectral shape stimuli have been provided in Hofman and Van Opstal (2002). Briefly, a set of 175 broad-band stimuli was derived from a long array of Gaussian-distributed amplitudes (the "root sequence") that was low-pass filtered at a given bandwidth, here termed the *ripple bandwidth*. A particular stimulus was created from this filtered sequence by selecting a window of 100 samples, representing a random-shaped amplitude spectrum that extended over 3 octaves, from 2.5 to 20.0 kHz. The windowed stimulus sequence, which served as a filter to create the actual sound, was smoothed by sine-squared on- and offset ramps of 0.5 octave width. The filter was subsequently extended to lower frequencies with a flat 1.0 to 2.5 kHz band and was applied to Gaussian white noise (1.0 to 20.0 kHz) to generate the actual spectrally rippled sound-pressure wave. All stimuli were generated by Matlab software (The Mathworks, Natick, MA, version 6.5).

To create the next stimulus, the window was shifted across the root sequence by 1/6 octave Subsequent stimuli in the set thus had similar shapes, shifted by 1/6 octave. In the study of Hofman and Van Opstal (2002), the low-pass filter applied to the root sequence had a steep cut-off at 3.0 cycles/octave (c/o). Figure 3.1 illustrates three typical examples of subsequent random rippled spectra, filtered at 3.0 c/o

In the present experiments, three different sets of 175 broad-band (1.0 to 20 kHz) stimuli were generated, with ripple bandwidths set at 1.5 c/o, 3.0 c/o, and 5.0 c/o. The stimuli were presented at an intensity of 60 dBA SPL, measured at the site of the listener's head and had a duration of 250 ms with 5-ms sine-squared on- and offset ramps.

Examples of representative amplitude spectra from each of the three stimulus sets are shown in Figure 3.2, together with a typical DTF (taken from listener JG at 0° elevation), for comparison. Note that the spectral width of the amplitude variations in the DTF fall between those of the 1.5 c/o and 3 0 c/o stimuli, but that the variations in the 5 c/o stimulus are clearly much faster.

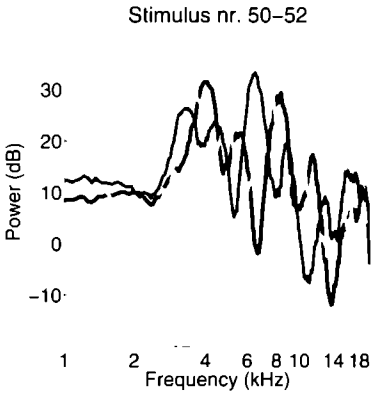


Figure 3.1: *Three examples of subsequent rippled spectra $X_k(\omega)$, taken from the 3 c/o stimulus set, for k = 50, 51, and 52 (in black, light gray, and dark gray respectively). All three spectra are flat from 1 to 2.5 kHz, and ripples (peak-peak amplitudes up to about 25 dB) run between 3.5 to 20 kHz. Note the relative shift of the peaks and notches between the subsequent stimulus ranks.*

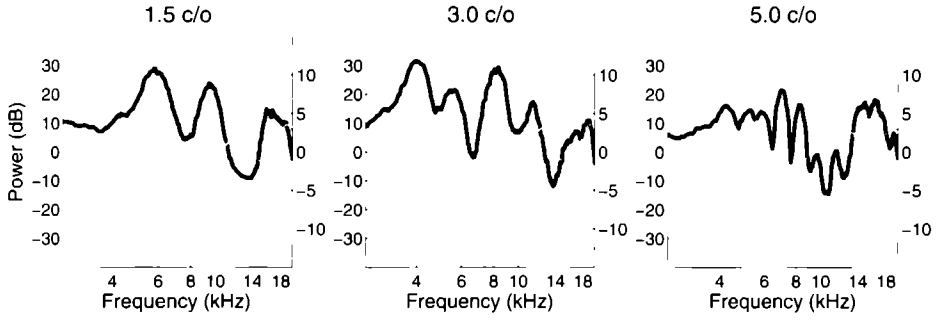


Figure 3.2: *Three representative examples of rippled spectra with a bandwidth of 1.5, 3.0, and 5.0 c/o (black lines), together with the DTF of 0 elevation of listener JG (gray dashed line) drawn in the same panels, for comparison. The scale of the DTF is plotted on the right-hand side. The three spectra differ markedly in their amount of spectral variation. Note that the width of the amplitude variations of the DTF appears to fall between those of the 1.5 c/o and the 3.0 c/o stimuli, but that the ripples in the 5 c/o spectrum are much faster.*

Listeners

Six listeners took part in the experiments: four males and two females. Their ages ranged from 22 to 46 years. Three of the listeners (JO, JV, and TE) were the authors. The other listeners (JG, MW, and RK) were informed about the actual speaker position, but were kept naive about the purpose of the experiments. All listeners had normal hearing, and had experience with the type of sound-

localization experiments carried out in the laboratory.

Experimental setup

Experiments were conducted in a dark and sound-attenuated room with dimensions 3 x 3 x 3 m^3. The walls, floor, ceiling, and large objects were covered with acoustic foam that effectively absorbed reflections above 500 Hz. The ambient background level in the room was 30 dBA SPL. Listeners were seated comfortably on a chair in the center of the room, facing an acoustically transparent thin-wire frontal hemisphere with a radius of 0.85 m, the center of which was aligned with the center of the listener's head. On this hemisphere 85 red/green light-emitting diodes (LEDs) were mounted at seven eccentricities: $R = [0, 2, 5, 9, 14, 20, 27, 35]^\circ$, relative to straight-ahead ($[R.\Phi] = [0, 0]^\circ$ and at twelve directions, given by $\Phi = [0, 30, ..., 330]^\circ$, where $\Phi = 0^\circ$ is rightward and $\Phi = 90^\circ$ is upward. These LEDs were used for calibration of the head-coil measurements (see below) and for providing a fixation light at the start of each localization trial.

Sound stimuli were delivered through a broad-range lightweight speaker (Philips AD-44725) mounted on a two-link robot. The robot consisted of a base with two nested L-shaped arms, each driven by a stepping motor (Berger-Lahr VRDM5). To hide the speaker from view, the wire hemisphere was covered with thin black silk.

Two PCs controlled the experiment. One PC-486 was equipped with the hardware for data acquisition (Metrabyte DAS16), stimulus timing (Data Translation DT2817), and digital control of the LEDs (Philips I2C). The other PC-486 generated the acoustic stimuli upon receiving a trigger from the DT2817. The output of this PC was passed through a DA-converter (Data Translation DT2821) at a sampling rate of 50 kHz, fed into a bandpass filter (Krohn-Hite 3343) with a flat passband between 0.2 kHz and 20 kHz, amplified (Luxman A-331), and passed to the speaker. An equalizer (Behringer Ultra-Curve) flattened the speaker transfer characteristic within 5 dB in the passband.

Measurement of the perceived sound direction

Listeners were asked to respond by turning their head to the perceived direction of the sound source. The 2D orientation of the listener's head was measured with the magnetic search-coil induction technique (Collewijn et al., 1975). Two orthogonal 3 x 3 m^2 sets of coils, attached to the walls, floor and ceiling of the room, generated a horizontal (30 kHz) and vertical (40 kHz) oscillating magnetic field. Listeners wore a lightweight helmet, consisting of a narrow strap above the

ears, that could be adjusted to fit around the head, and a second strap that ran over the head A small coil was mounted on the latter

A pliable aluminum rod with a dim red LED at its end was attached to the helmet at a distance of about 0 40 m in front of the listener's eyes, such that it was approximately aligned with the center LED of the hemisphere with the head in a comfortable, straight-ahead orientation Listeners were instructed to use this rod-LED as a pointer to indicate the perceived sound direction In this way, it was guaranteed that listeners always pointed their head with the eyes in a fixed, roughly straight-ahead, orientation in the head A firm neck rest allowed for a reproducible and stable initial orientation of the listener's head at the start of each trial throughout the session

Experimental paradigm

Each experimental session started with a calibration run in which the listener had to align the rod LED with 36 peripheral LEDs on the hemisphere, presented in random order Subsequently, the experiments with the rippled noise stimuli were performed Stimuli with different ripple bandwidths were presented in separate recording sessions All stimuli were presented with the speaker at the straight-ahead position ($[\alpha.\varepsilon] = [0, 0]^\circ$)

An experimental session consisted of two runs in which each of the 175 stimulus spectra was presented once in randomized order Each stimulus spectrum was thus presented twice in one session A trial always started with an initial fixation light at $\alpha = -14^\circ$ or $+14^\circ$ azimuth (randomly selected), and at 0° elevation These fixation locations were chosen rather than straight ahead, like in the study of Hofman and Van Opstal (2002), because in that study the data showed a suppression around 0° elevation, which may have been due to a tendency of listeners to make a response movement, even when they did not have a clear spatial percept By having the initial fixation point away from the midline, we ensured that listeners always generated a goal-directed head movement, even if the sound were to be perceived at zero elevation After a randomized period between 0 9 and 1 1 seconds, the fixation LED was switched off and the sound was presented Head position was measured for 3 0 sec after onset of the fixation light Listeners were instructed to reorient their head as fast and as accurately as possible in the apparent sound direction Although listeners were aware of the fixed speaker at the straight-ahead location, they were encouraged to respond to the perceived apparent sound direction, rather than to the (remembered) actual speaker location

All listeners participated in at least one session with each of the three different ripple bandwidths Listener MW performed in two sessions with the 3 0

c/o-bandwidth stimuli. Listener TE participated in three sessions with the 3.0 c/o-bandwidth stimuli, and in two sessions for both the stimuli with a 1.5 and a 5.0 c/o bandwidth.

Listener JO participated in four additional sessions in which the speaker was positioned at $[\alpha, \varepsilon] = [0, 80]^\circ$, i.e. above the listener. For these experiments, only the stimuli with the ripple bandwidth of 3.0 c/o were used. In this condition, the listener perceived some of the stimuli in the rear hemisphere. In that case, he was instructed to point in the direction of the perceived sound location, mirrored with respect to the frontal plane (i.e. in the frontal hemifield). To be able to identify off-line in which trials the listener perceived stimuli at the rear, he was instructed to push a button as soon as the pointing was completed.

Data analysis

The coordinates of target locations and the endpoints of the localization responses were all described in a double-pole azimuth-elevation coordinate system, in which the origin coincides with the center of the head (Knudsen and Konishi, 1979). The azimuth angle, α, is defined as the angle within the horizontal plane with the vertical midsaggital plane, whereas the elevation angle, ε, is defined as the direction within a vertical plane with the horizontal plane through the listener's ears. The relation between the $[\alpha, \varepsilon]$ coordinates and the polar $[R, \Phi]$ coordinates defined by the LED hemisphere (see above) is given in Hofman and Van Opstal (1998).

The raw head position signals and the corresponding LED coordinates from the calibration run 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). The networks corrected for small inhomogeneities of the magnetic fields and a slight crosstalk between the horizontal and vertical channels that resulted from small deviations from perfect orthogonality.

A custom-made PC program (Hofman and Van Opstal, 1998) was used to identify saccades in the calibrated head-position signals on the basis of preset velocity criteria for saccade onset and offset, respectively. The endpoint of the first saccade after stimulus onset was defined as the response position. All saccades were visually checked and corrected if necessary. Saccades with latencies shorter than 80 ms or longer than 800 ms were discarded from further analysis. Earlier responses are usually predictive, whereas later responses are considered to be caused by inattention. In the case of rear responses, in the experiment with the overhead speaker position, the search coil signal was immediately reset to zero at the moment the listener pushed the button. The occurrence of such

resets was labeled to indicate a percept in the rear hemifield. For these trials, the perceived elevation in the rear hemisphere, ε_r, was calculated by adding 180° to the measured frontal elevation, ε_f (see above).

Directional transfer functions (DTFs) Head-related transfer functions (HRTFs) were measured for all listeners for 25 different elevations ($\varepsilon = -60°$, $-55°$, , 55°, 60°) and at a fixed azimuth, $\alpha = 0°$. A periodic flat-spectrum Schroeder-phase signal (FM-sweep-like signal, Schroeder, 1970) was used as a stimulus. It consisted of 20 periods of 20.5 ms, which added up to a total stimulus duration of 410 ms. The spectrum was flat within 0.2 to 20 kHz, and the sound level at the listener's head was 65 to 70 dB SPL.

Pressure waveforms near the entrance of the ear canal were measured with a miniature microphone (Knowles EA1842) with a thin probe tube (1.5 mm diameter) attached to it. The tube was kept in place by a thin small ring at the end of a custom-made thin metal rod that was positioned to the side of the head with a head band. The listener was seated in a chair in the center of the experiment room.

The microphone signal was amplified by a measurement amplifier (Brüel & Kjær 2610), subsequently fed into a bandpass filter (Krohn Hite 3343, passband 0.2 to 20 kHz), and finally sampled at 50 kHz by a data acquisition board (Data Translation DT3818). From period 2 to 19 of the sampled (periodic) waveforms, the average signal per period was computed (containing 1024 samples), and transformed into 512 spectral bins (resolution 48.8 Hz) by means of the Fast Fourier Transform. The directional transfer functions (DTFs) were then computed by subtracting the mean amplitude spectrum (in decibels) computed over the entire set of HRTFs.

Reconstruction of elevation-related spectral shapes The data from the localization experiments were used to reconstruct elevation-specific spectral shapes, by applying the method described in Hofman and Van Opstal (2002). In short, smooth response distributions to a given rippled stimulus, $X_k(\omega)$, here denoted by $p(\varepsilon|X_k(\omega))$, with $k \in 1 \cdots 175$, were constructed by replacing each elevation response to that stimulus, ε_{nk}, by a normalized Gaussian, $G[(\varepsilon - \varepsilon_{nk}); \sigma_\varepsilon]$, centered at the response elevation, with a width in the elevation direction of $\sigma_\varepsilon = 2°$, and by summing all Gaussians (typically, $n \in [1.2]$, see e.g. Fig. 3.6). Thus, $p(\varepsilon|X_k(\omega))$ can be interpreted as the probability of a response to elevation ε, when stimulus $X_k(\omega)$ is presented. When the resulting probability distributions are (close to) unimodal, there is a unique stimulus-response relationship (see e.g Figs 3.5 and 3.6).

In this study, we aimed to estimate the spectral features that underly the percept of elevation angle ε, which we here denote by $P(\omega;\varepsilon)$ As a first step toward this estimate, we need to determine the probability that stimulus $X_k(\omega)$ was presented, given a particular response elevation, ε. This conditional probability, described by $p(X_k(\omega)|\varepsilon)$, can be computed from Bayes' rule, according to:

$$p(X_k(\omega)|\varepsilon) = \frac{p(\varepsilon|X_k(\omega)) \cdot p(X_k(\omega))}{p(\varepsilon)} \tag{3.3}$$

in which $X_k(\omega)$ indicates the log power spectrum of the presented stimulus, and $p(X_k(\omega))$ is the (expected) unconditional probability of spectral shape $X_k(\omega)$, known as the *prior* distribution. The normalization factor, $p(\varepsilon)$, represents the total distribution of elevation responses irrespective of the stimulus spectrum, and equals the normalized distribution of all elevation responses (see e.g. Fig. 3.3B). Note, that in Eqn. 3.3 both $p(\varepsilon)$ and the conditional probability $p(\varepsilon|X_k(\omega))$ can be extracted from the experimental data. However, the prior, $p(X_k(\omega))$, is in principle unknown, as it may be partly determined by (idiosyncratic) expectations about stimulus spectra, by previous experience, or by other covert factors that could not be controlled in the experiment. To circumvent the problem of having to estimate the prior distribution, we therefore took a more pragmatic approach by determining the unconditional probability of stimulus $X_k(\omega)$ from the actual distribution of applied stimulus spectra. This distribution was taken to be *uniform* in our experiments, as the occurrence of each particular stimulus was equally likely, and completely randomized. Thus, $p(X_k) \equiv 1/175$, for all k, and as a consequence, Bayes' rule becomes a maximum likelihood estimate (MLE):

$$p(X_k(\omega)|\varepsilon) = \frac{p(\varepsilon|X_k(\omega))}{175 \cdot p(\varepsilon)} = \frac{\sum_{n=1}^{N_k} G[(\varepsilon - \varepsilon_{nk});\sigma_\varepsilon]}{175 \cdot p(\varepsilon)} \tag{3.4}$$

with N_k the number of repetitions of the stimulus (typically, $N_k = 2$).
If the MLE for $p(X_k(\omega)|\varepsilon)$ is large, more responses to elevation ε are made for stimulus spectrum $X_k(\omega)$ than for other spectra, and thus the stimulus will contain spectral features that contribute significantly to the perceived elevation angle ε.

As will become apparent from the data, a given elevation is typically perceived for a number of rippled spectral shapes. This suggests that the different spectra may either contain a common spectral feature that is responsible for the percept, or a number of different spectral features that all contribute to the same perceived elevation angle. Thus, to estimate the actual spectral features

underlying the perceived elevation angle, $P(\omega;\varepsilon_p)$, all rippled spectra that gave rise to that percept should somehow be incorporated. If, for example, ε_p would be determined by a single spectral feature, say, a peak or a notch (see Introduction), it should emerge as the common feature in all contributing rippled spectra. Following the procedure of Hofman and Van Opstal (2002), we adopted a linear weighting scheme to estimate $P(\omega;\varepsilon_p)$.

From the response-dependent MLEs, $p(X_k(\omega)|\varepsilon)$, extracted from the data for each rippled spectrum through Eqn. 3.4, we estimated the elevation-dependent spectral features giving rise to a perceived elevation angle, ε_p, by taking a weighted sum of the rippled stimulus spectra $X_k(\omega)$ that contributed to this perceived elevation, and let the MLEs act as linear weights (Hofman and Van Opstal, 2002):

$$P(\omega;\varepsilon_p) = \sum_{k=1}^{N_{\varepsilon_p}} p(X_k(\omega)|\varepsilon_p) \cdot X_k(\omega) \tag{3.5}$$

with N_{ε_p} the number of stimuli contributing to percept ε_p.

Note, that in the actual experiments, the sensory spectrum was not equal to the free-field rippled spectrum $X_k(\omega)$, because it is determined by the multiplication between the free-field stimulus spectrum and the HRTF associated with the straight-ahead speaker location at $\varepsilon_0 = 0°$, $H(\omega;\varepsilon_0)$. Thus, the sensory spectrum at the eardrum (in dB) for stimulus k is:

$$\log S_k(\omega;\varepsilon_0) = \log X_k(\omega) + \log H(\omega;\varepsilon_0) \sim \log X_k(\omega) + \log[\gamma \cdot \mathrm{DTF}(\omega;\varepsilon_0)] \tag{3.6}$$

In the study of Hofman and Van Opstal (2002), this aspect was not incorporated in the analysis. In the present paper, we have accounted for this difference to better estimate the perceptual spectral features. The HRTF and the DTF are related via the direction-independent transfer characteristic of the ear canal. Especially for higher frequencies the exact position of the recording microphone may become critical, and influence the intensities of peaks and notches in an unknown way. In our reconstructions, we therefore simplified the relation between HRTF and DTF by introducing a linear scaling parameter, γ.
Thus, Eqn 3.5 was modified to:

$$P_0(\omega;\varepsilon_p) = \sum_{k=1}^{N_{\varepsilon_p}} p(S_k(\omega;\varepsilon_0)|\varepsilon_p) \cdot S_k(\omega;\varepsilon_0) \tag{3.7}$$

in which the subscript 0 in $P_0(\omega;\varepsilon_p)$ indicates that the free-field stimuli have been converted into estimated sensory spectra.

In the reconstruction algorithm, we implicitly assumed that the localization response depends only on the sensory spectrum, and is not influenced by other factors such as head orientation, or expectancy Moreover, we assumed, for simplicity, that in case a stimulus contained spectral shape features relating to different elevations, the listener's response was not determined by averaging over the different elevation angles

Spectral correlation

Quantitative comparisons between the profiles of two magnitude spectra, $A(\omega)$ and $B(\omega)$, were performed by computing the spectral correlation coefficient, C_{AB} (Hofman and Van Opstal, 1998)

$$C_{AB} \equiv \left\langle \left(\frac{A(\omega) - \bar{A}}{\sigma_A} \right) \left(\frac{B(\omega) - \bar{B}}{\sigma_B} \right) \right\rangle \tag{3 8}$$

with the spectral mean of $\bar{X}$ defined as

$$\bar{X} = \langle X(\omega) \rangle \equiv \frac{1}{\omega_2 - \omega_1} \int_{\omega_1}^{\omega_2} d\omega X(\omega) \tag{3 9}$$

and the variance σ_X^2

$$\sigma_X^2 = \langle (X(\omega) - \bar{X})^2 \rangle \tag{3 10}$$

Here, we took $\omega_1 = 4$ kHz and $\omega_2 = 14$ kHz The amplitude spectrum function, $A(\omega)$ or $B(\omega)$, was specified in decibels, and frequency, ω, was given in octaves One can interpret C_{AB} as a similarity index that lies in the range $[-1,+1]$ Maximum similarity corresponds to $C_{AB} = +1$, and occurs when $A(\omega)$ can be expressed as $A(\omega) = p\ B(\omega) + q$ (with p and q constants) No similarity occurs when the C_{AB} index is close to zero or becomes negative

3.3 Results

Although the speaker was always positioned at $[\alpha\ \varepsilon] = [0, 0]$, and listeners were aware of this fact, head movement responses were distributed over a considerable range of elevations for most listeners Note that prior knowledge about the speaker position is in fact a disadvantage for the listener, as the stimuli may appear to come from any direction Listeners were encouraged to indicate the perceived direction of the sound and ignore the (remembered) speaker position However, listeners did not make random localization movements away from the center, as the responses to the different presentations of each stimulus - either

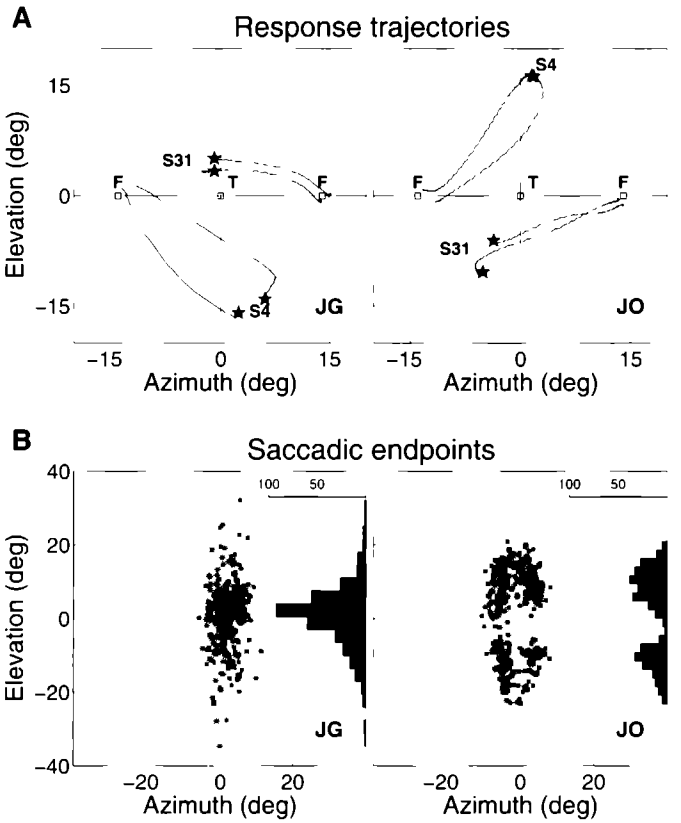


Figure 3.3: **(A)** *Response trajectories of head movements of listeners JG (left) and JO (right) to two stimuli (nr. S4 and nr. S31). F indicates the two fixation LEDs and T the speaker position Note that listeners perceived the stimuli at different elevations, and that their responses were consistent* **(B)** *Response scatter to all stimuli for listeners JG and JO, together with the elevation response distributions. Data for the spectra with a ripple bandwidth of 3 c/o.*

in two runs within one session, or in separate sessions performed on different days - were very consistent.

As an example, Figure 3.3A shows typical responses to both presentations of two different rippled stimuli, each with a ripple bandwidth of 3 c/o (stimulus numbers S4 and S31) for two listeners, JG and JO. Note that the listeners responded to different elevations for both stimuli; JG localized S31 around 5 degrees upward from the center location, and S4 around 15 degrees downward, whereas JO localized S31 around 7 degrees downward, and S4 around 15 degrees upward.

Bandwidth	1 5 c/o	3 0 c/o	5 0 c/o	upward
JG	−0.07 (8 24)	1.16 (7 77)	1.13 (6 07)	
JO	8.82 (14 37)	5.09 (11 59)	8.97 (12 27)	27 22 (24 09)
JV	0.75 (4 22)	0 59 (3 56)	3.95 (1 97)	
MW	21.10 (5 24)	15.05 (5 48)	20.08 (8 30)	
RK	1.82 (4 73)	2.77 (3 35)	2.54 (4 10)	
TE	6.10 (10 40)	13.23 (10 86)	5.58 (8 01)	

Table 3.1: *Medians and standard deviations (between brackets) of the response distributions for the three stimulus sets with different ripple bandwidths for all listeners*

Note also the reproducibility of the listeners' responses, as the two head movement trajectories to each stimulus end at the same locations. Pearson's correlation between the elevation components of the first and second response to each stimulus of the entire data set is high for both listeners: $r = 0.82$ for JG, and $r = 0.86$ for JO (see below, Fig. 3.5 and Tab. 3.2). However, the correlation between the responses of JG and JO was much lower ($r = 0.35$), which indicates that they typically perceived the same stimulus at quite different locations.

Figure 3.3B shows the distributions of all responses for the 3 c/o stimuli of both listeners These distributions are proportional to $p(\varepsilon)$ in Eqn. 3.3 (Methods). Both had a response range that extended from about -25° to $+25^\circ$ in elevation. In azimuth, the response endpoints remained close to the midline at $\alpha = 0^\circ$. Whereas the majority of the responses of listener JG clustered around the real speaker location at 0° elevation, the responses of listener JO were mainly directed away from this location.

Table 3.1 provides the medians and standard deviations for the elevation responses of all listeners and all three stimulus sets. For listeners JG, JV, and RK the responses were distributed around 0° elevation, whereas the responses of listeners MW and TE were skewed toward upward elevation angles. More importantly, however, is the observation that there was no systematic difference between the response distributions for the three stimulus sets: the median values, standard deviations, and response ranges were independent of the ripple bandwidth for each of the listeners. This important point is illustrated in Figure 3.4, which shows the response range as a function of the ripple bandwidth. No systematic relation emerged, indicating that the response distributions were insensitive to the variations in the amplitude spectra.

Not only did listeners respond over a considerable elevation range, their responses to the different presentations of each stimulus were also quite consistent. Figure 3.5 shows for three of the listeners the two response elevations for

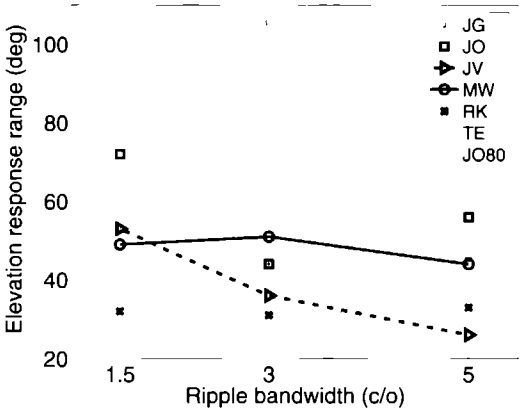


Figure 3.4: *Elevation response range as a function of ripple bandwidth for all listeners. The star at 3 c/o is the response range of listener JO with the speaker at [α.ε] = [0, 80] . Note the absence of a systematic relation between stimulus bandwidth and response range.*

each stimulus plotted against each other. Listeners JG and JO responded over a larger elevation range than listener JV and their correlations were higher ($r =$ 0.83 and 0.86, vs. 0.48, respectively). The correlations between the responses of the two runs are given in Table 3.2 for all listeners and all three stimulus sets. Correlations were usually high and always significant. Listeners JG and JO, who were the most experienced in these sound localization experiments, yielded correlations between 0.82 and 0.92, but also listener RK, who was rather inexperienced in these experiments, showed high correlations. For listeners MW and TE, who performed in more than one session, we also calculated the correlation of responses between sessions. For listener MW the correlation between the two sessions with 3 c/o stimuli was 0.61. For listener TE the correlations between the different sessions of a given ripple bandwidth varied between 0.50 and 0.61. These values are in the same range as the correlations between the two runs within one session, which shows that listeners consistently assigned a particular elevation to a given spectral shape.

An important requirement for the reconstruction procedure is that the conditional response probabilities, $p(\varepsilon|X_i(\omega))$, are described by (near-)unimodal distributions, indicating that a given stimulus yielded a unique elevation percept. As described in the Methods, each elevation endpoint was replaced by a normalized Gaussian (width 2°). Response Gaussians for the same stimulus were subsequently added to estimate the elevation response distribution for a given stimulus. To illustrate this procedure, Figure 3.6 shows fifteen consecutive stimulus spectra and their corresponding response distributions for listener JG

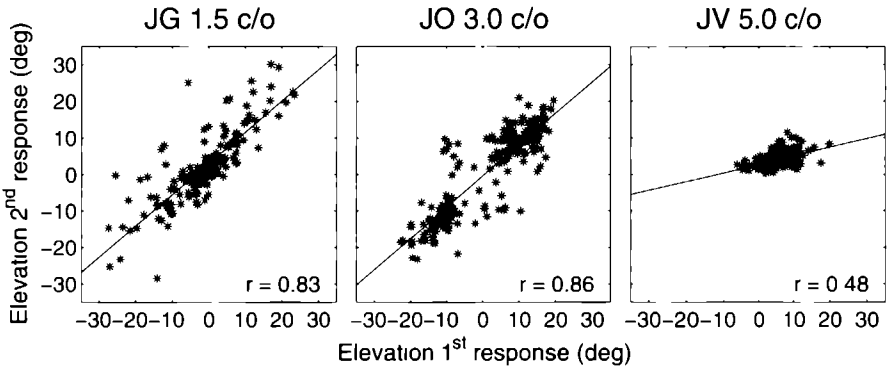


Figure 3.5: *Elevation responses of the second run vs elevation responses of the first run together with a regression line and the correlation between the two runs (indicated in the lower right corner) Data of listener JG for the 1 5 c/o stimuli, listener JO for the 3 0 c/o stimuli and listener JV for the 5 0 c/o stimuli Both JG and JO responded over a considerable range and show a high correlation between the two presentations of one stimulus Listener JV responded over a much more restricted range and the correlation between the two runs was lower, although still highly significant*

Bandwidth	1 5 c/o	3 0 c/o	5 0 c/o	upward
JG	0 83	0 82	0 82	
JO	0 87	0 86	0 92	0 89, 0 90, 0 87, 0 84
JV	0 57	0 64	0 48	
MW	0 73	0 72, 0 62	0 53	
RK	0 83	0 74	0 68	
TE	0 63, 0 73	0 61, 0 54, 0 55	0 70, 0 58	

Table 3.2: *Correlations between the elevation components of the two responses toward each stimulus for all three stimulus sets and all listeners*

for the 3 c/o stimuli. In line with the high correlation observed in Figure 3.5, the majority of response distributions were indeed unimodal. As explained in the Methods, in stimulus spectra with successive indices, k, the window defining the filter shape to create the stimulus was shifted by 1/6 octave across the rippled root sequence. As a result, spectral features shift to lower frequencies by 1/6 octave for consecutive rippled spectra This aspect of the stimuli is apparent in the left panel of Figure 3.6. For example, the peak at 14 kHz in stimulus 127 is located at 7 kHz in nr. 133, and at 3.5 kHz in nr. 139. Note, that as the spectral features move from higher to lower frequencies, the response distributions tend to shift in elevation in a systematic way.

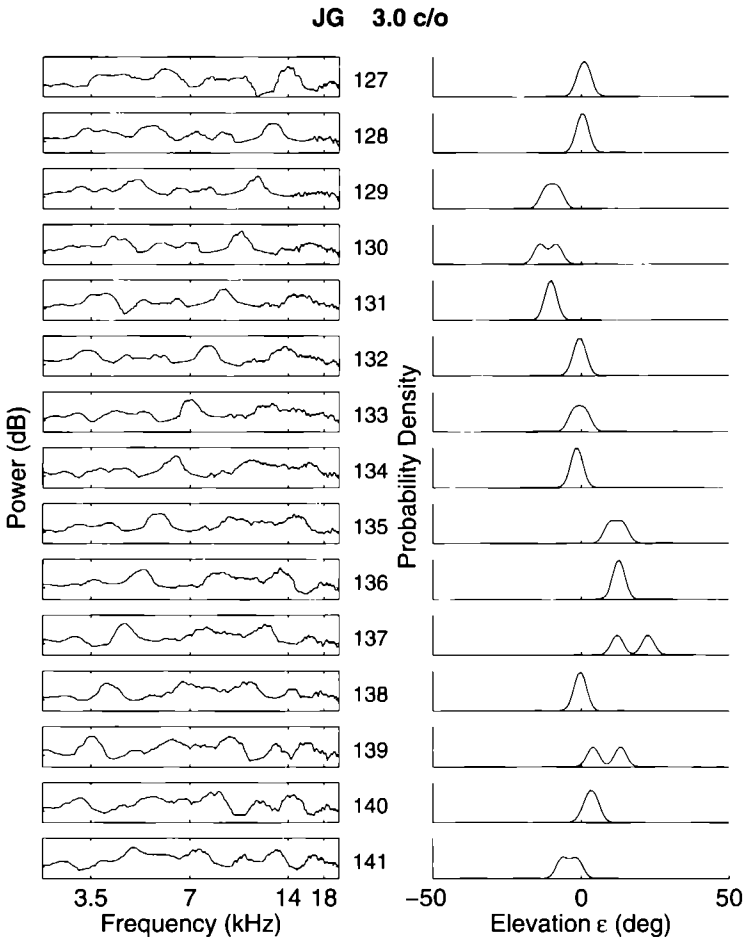


Figure 3.6: *Subsequent stimulus spectra (left, stimulus rank indicated on the right) with their conditional response probability density distributions (right) for listener JG for the 3 c/o stimuli. Stimuli were presented in random order in the experiment, but are shown here in ranked order for illustrative purposes. Note that certain stimulus sequences, e.g. 127 to 129, 130 to 133, 134 to 138, and 139 to 141 yielded a systematic shift in the perceived elevation. Note also that a given elevation, e.g. 5° upward, is perceived by different stimulus spectra (here: 127, 131, 134, and 139).*

Conditional response distributions as are shown in Figure 3.6, together with the normalized unconditional distributions of elevation responses as shown in the histograms of Figure 3.3B, were then used to compute the MLE for each stimulus, given the response elevation, $p(X_k(\omega)|\varepsilon)$, by applying Eqn. 3.4. These MLE estimates were subsequently used as linear weights for each of the corre-

sponding stimulus spectra to construct the elevation-dependent spectral shapes underlying the percept of a given elevation angle, $P(\omega;\varepsilon_p)$, with Eqn. 3.5. In Figure 3.7 we show the reconstructed perceptual spectral features for listener JG over the 3 to 18 kHz band for sounds with a ripple bandwidth of 3 c/o, in the same format as in the study of Hofman and Van Opstal (2002). The abscissa represents the listener's perceived elevation, ε_p, while the amplitude of the spectra (in dB) is encoded in gray scale: light shades correspond to a peak in the spectrum, while dark shades indicate a spectral notch. Note that the reconstructed perceptual spectral features have a rich structure. Rather than a single peak or notch, a complex combination of peaks and notches appears to underly the listener's perceived elevation direction. Thus, this result does not support models that explain the extraction of elevation on the basis of a CPA, or a single notch (see Introduction).

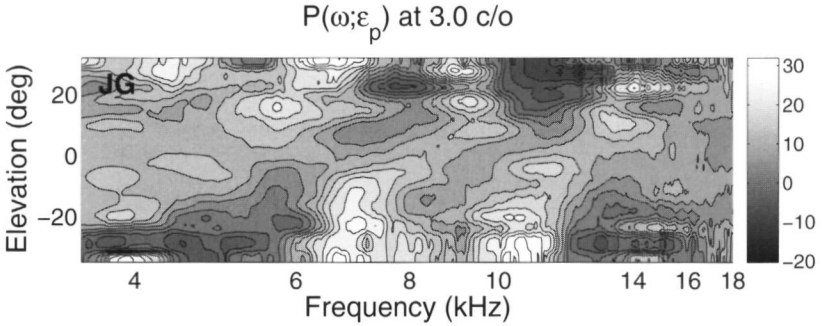


Figure 3.7: *Reconstructed spectral features for the 3 c/o stimuli for listener JG, based on the free-field stimulus spectra (Eqn. 3.5). Abscissa indicates perceived elevation angle (in deg), ordinate frequency (log scale). Amplitude (in dB) is encoded in grayscale, where light shades indicate peaks, and dark shades indicate notches in the spectrum.*

The reconstructed perceptual spectral shapes in Figure 3.7 are based on the free-field stimulus spectra, $X_k(\omega)$, rather than on the sensory spectra, $S_k(\omega)$, delivered to the eardrum. As outlined in the Introduction and the Methods, the latter are more relevant to the auditory system. To estimate the sensory spectrum, the contribution of the DTF for the straight-ahead speaker location was added to the free-field spectrum (Eqn. 3.6; here $\gamma = 1$), and the corrected perceptual features were then reconstructed by applying Eqn. 3.7. The results are shown in Figure 3.8 for the three ripple bandwidths for listener JG, together with the listeners' DTFs for comparison (bottom panel). Note the changes in the perceptual spectral shapes, when compared to the uncorrected data of Figure 3.7. More importantly, however, the three reconstructed spectra resulted to be remarkably similar, not only for the different ripple bandwidths, but also to the

listener's DTFs Prominent features in the DTFs, like the notch running from about 5 kHz for downward elevations, to about 8 kHz for upward elevation angles, are clearly visible in the three reconstructed perceptual spectral shapes But also the diagonal peak from about 10 kHz to 15 kHz, the secondary notch around 10 kHz, the peak around 5 kHz for upward elevations, and the peak near 7 kHz for downward locations were found in the different reconstructions This result is quite remarkable, considering that the entire reconstruction is based on only two (open-loop) head-movement responses per stimulus, that the stimulus sets themselves were highly dissimilar (e g Fig 3 2), and the underlying model is extremely simple (linear weighting of stimulus spectra)

To quantify the similarity between the different reconstructed patterns, Figure 3 9A shows the spectral correlations between the three sets of reconstructed spectra of listener JG (see Methods) Grayscale values indicate correlations between 0 5 to 1, with darker gray shades corresponding to higher correlations Figure 3 9B shows the correlations between the three sets of reconstructed spectral shapes and the DTFs of listener JG, with grayscale values indicating correlations between 0 2 and 1 To construct these latter plots, the reconstructed spectra had to be resampled to match the frequency binning of the DTFs to enable the comparison For both comparisons the correlations are highest around the diagonal and decrease for elevations away from the diagonal, although the results for the 5 c/o stimuli were more variable

Note that despite the remarkable similarities, an important difference between the reconstructed spectral shapes and the measured DTFs was found in the range of the elevation angles (roughly, between -20 and $+20^\circ$ for the perceived elevations, vs -60 to $+60^\circ$ for the DTFs, see Methods) Clearly, the prominent spectral shape of the DTF corresponding to the straight-ahead speaker location was unavoidably added to the random spectral shapes of the free-field stimuli, and therefore the listener's elevation percept was likely to be influenced by the presence of this DTF in the sensory spectrum As explained in the Methods, the reconstruction assumes that the listener's percept was determined by a unique elevation angle, and was not designed to cope with the possibility that the system may actually average across different potential elevation angles to determine the perceived elevation To illustrate the influence of the speaker-induced DTF on the sensory spectrum, we conducted a series of four experiments with the 3 0 c/o stimuli with listener JO, in which we positioned the speaker at $[\alpha\ \varepsilon] = [0, 80]^\circ$ The DTF corresponding to the upward direction is considerably flatter than for the frontal direction (the spatial resolution in upper space is worse too), and therefore will "pollute" the sensory spectrum to a lesser extent

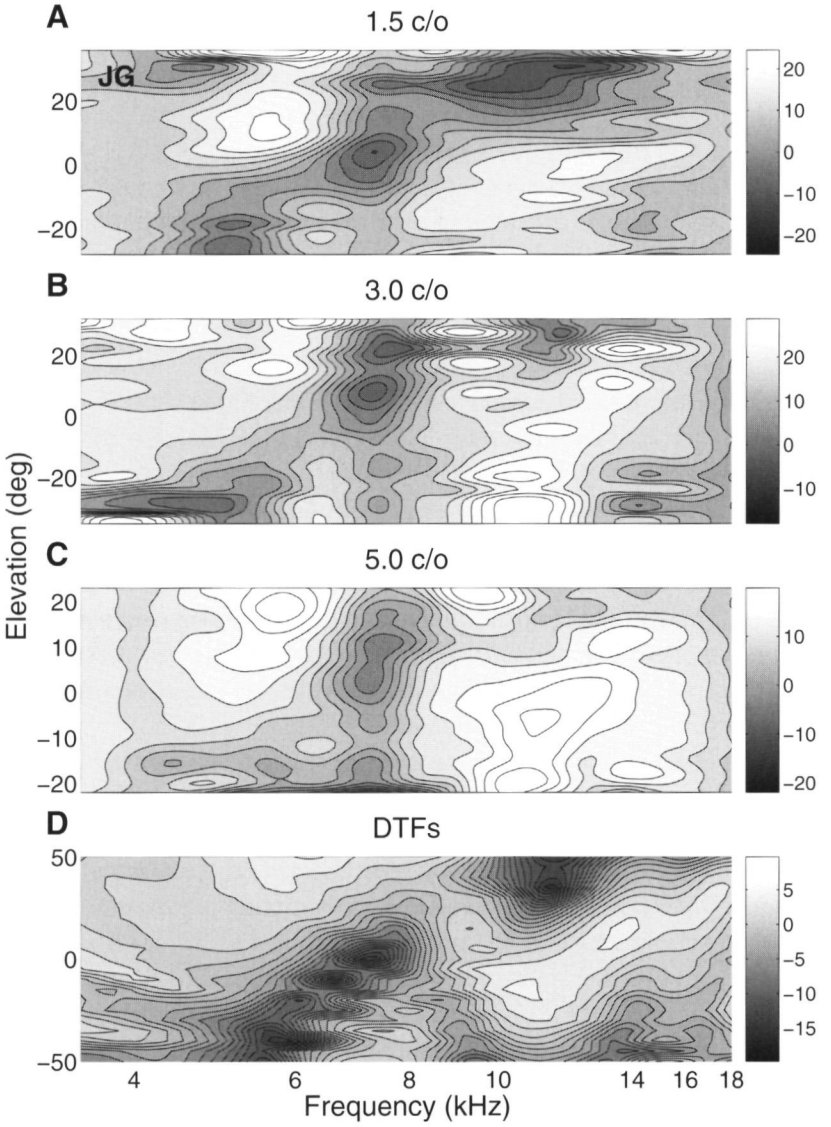


Figure 3.8: *Reconstructed spectral features for the three different amplitude-spectral bandwidths, based on the sensory spectra, together with the DTF for listener JG. Same format as Fig. 3.7. Note that the three reconstructed spectra appear to be very similar, despite the considerable differences in the underlying stimulus spectra (e.g. Fig. 3.2). Note also the remarkable similarities in the reconstructed patterns with the DTFs of this listener. For example, the diagonal notch from about 5 to 8 kHz, and the peak running from around 10 to 15 kHz, can be seen both in the reconstructed spectra and in the DTFs.*

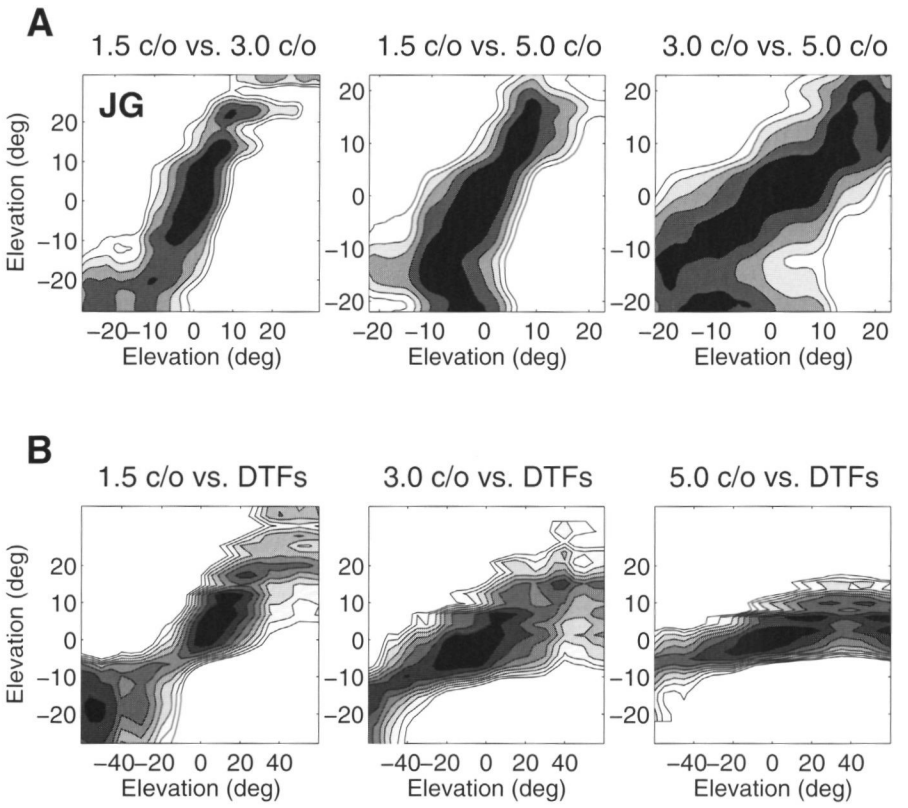


Figure 3.9: **(A)** *A measure of the similarity between the different reconstructed spectral shapes of listener JG is given by the correlation matrix, $C(\varepsilon_1, \varepsilon_2)$. Grayscale values indicate correlation values between 0.5 and 1 with darker grayscale values for higher correlations.* **(B)** *The correlation matrix for the three reconstructed spectra and the DTFs of listener JG. Grayscale values indicate correlation values between 0.2 and 1. The correlation matrix is high only for locations on or near the main diagonal.*

Indeed, in these experiments the listener's elevation responses covered a much larger range. Figure 3.10A shows the responses of the second run versus the responses of the first run for all of these sessions. Although most responses were in the frontal hemisphere ($\varepsilon_p \in [-90, +90]^\circ$), front-back confusions and reversals now also occurred. Data points with $\varepsilon_{1,2} > 90^\circ$ indicate stimuli for which the sound was consistently perceived at a rear location. A front-back confusion occurred when data points in Figure 3.10A ended in either the upper left, or the lower right corner of this plot. Front-back confusions were not random; they tended to cluster around the diagonal with a slope of minus one, indicating that the listener had a clear elevation percept, but only front vs. back was

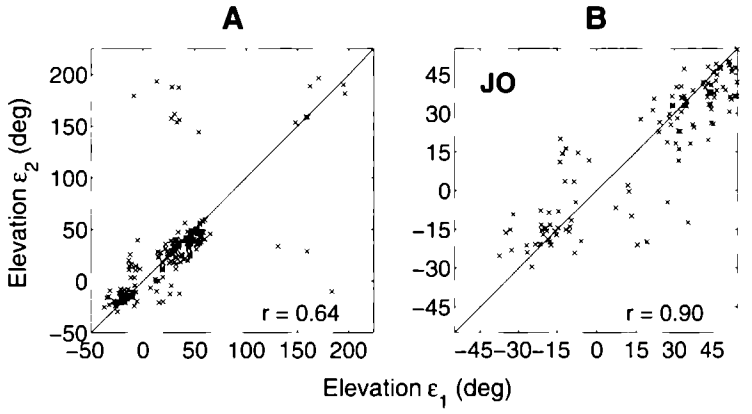


Figure 3.10: *Elevation responses of the second run vs. elevation responses of the first run for the experiment with the speaker at (0,80)° for listener JO (pooled data for all four sessions). Correlations between the two runs are indicated in the lower right corner.* **(A)** *shows the data including the responses to the rear hemisphere ($\varepsilon_p > 90°$. The correlation is low, mainly mainly because of front-back confusions. In* **(B)**, *all responses to the rear and all front-back confusions were removed, leading to a substantially higher response correlation over a larger elevation range than for the straight-ahead speaker (cf. Fig. 3.5B).*

ambiguous. In Figure 3.10B we show the correlation plot when all rear and confused responses were removed. The correlations of the data (here: $r = 0.91$) were then very similar as for the experiment with the speaker at $[\alpha.\varepsilon] = [0, 0]°$ ($r = 0.86$; see Tab. 3.2), but distributed over a much larger elevation range.

From these frontal responses we again reconstructed the perceptual spectral shape functions. The results are shown in Figure 3.11, which compares the spectra for the two speaker positions (straight ahead, Fig. 3.11A, corrected for the sensory spectrum, here $\gamma = 0.5$, see Eqn. 3.6; upward: Fig. 3.11B, uncorrected) with the listener's DTFs (Fig. 3.11C). The same spectral features can be seen in the two reconstructed spectra and in the listener's DTFs, like a notch running from 5 kHz for low elevations to about 8 kHz for high elevations, and a peak from 11 kHz to 14 kHz. However, the perceived elevation range increased from $[-20,+20]°$ to about $[-30,+60]°$ by moving the speaker to $[\alpha.\varepsilon] = [0, 80]°$.

3.4 Discussion

In this paper, we have extended the method of Hofman and Van Opstal (2002) to estimate the spectral features that underly the percept of sound-source elevation in a number of ways. First, by applying the algorithm for different ripple

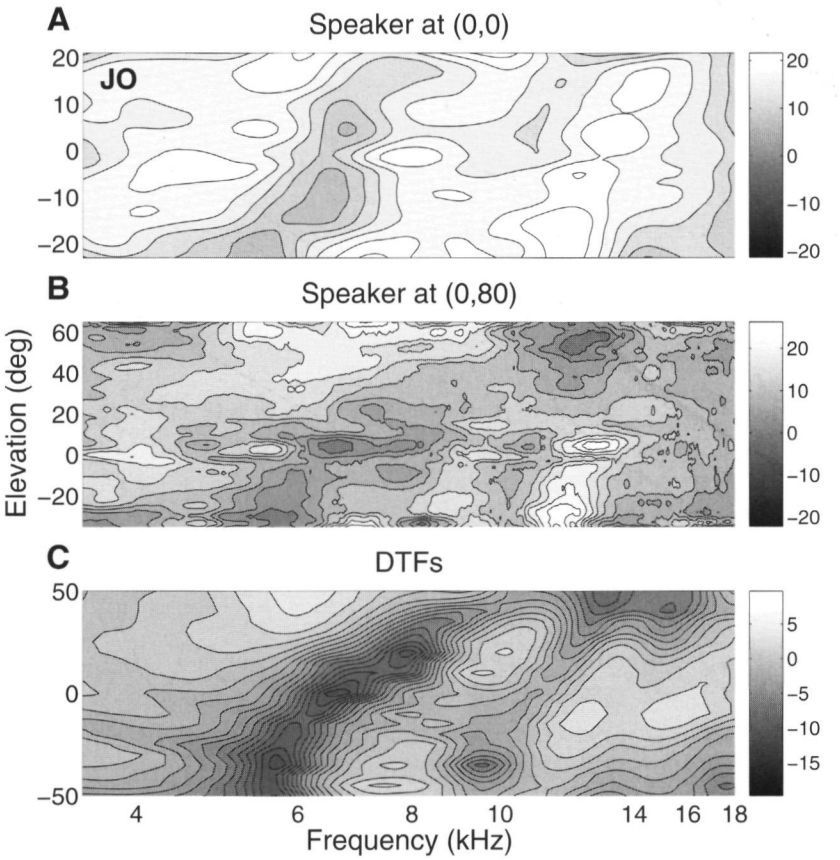


Figure 3.11: *Reconstruction of the spectral features for listener JO, based on the elevation responses to the 3 c/o stimuli from the frontal speaker location* **(A)**; *corrected for the straight-ahead DTF, weight 0.5) and from the overhead speaker location* **(B**; *only responses into the frontal hemifield; not corrected for the speaker's DTF).* **(C)** *shows the DTFs of listener JO. Again, there is a high similarity between the two reconstructed spectra, and with the listener's DTFs. Note the differences in scale of the elevation axes in the different panels. Same format as Fig. 3.7.*

bandwidths, we tested different models of how the auditory system may deal with the ill-posed problem of elevation extraction. Second, in the reconstruction algorithm we used an estimate of the sensory spectrum at the eardrum, rather than the free-field stimulus spectrum. And third, by also positioning the speaker at an overhead location, we reduced the influence of the speaker-induced DTF on the sensory spectrum and the elevation responses.

Our experiments showed that for different ripple bandwidths, randomly shaped

stimulus spectra were systematically mislocalized. The correlation between responses to identical stimuli within each listener was typically high. Even though the stimuli were always presented from a fixed location, and listeners were aware of this fact, their responses covered a broad range of elevations. Despite the large variation between rippled spectra in the different stimulus sets (peak-to-peak amplitudes were up to about 25 dB, but peak widths were very different for the different ripple bandwidths; see e g Fig 3.2), none of the listeners reported front-back confusions when stimuli emanated from the straight-ahead direction This is probably due to the fact that the DTF associated with straight-ahead imposed sufficient acoustic power in the 3 to 7 kHz band, which is a major cue for distinguishing frontal (high power in this band) vs. rear (low power) locations (Blauert, 1997) Indeed, when the speaker was moved to an upward location, for which the DTF is less pronounced in its peaks and notches, a considerable number of rear responses and front-back confusions resulted for the same stimuli.

We computed estimates of elevation response distributions that were typically based on only two localization responses per individual stimulus to estimate the importance of a given stimulus to the perceived elevation. The idea behind the reconstruction algorithm was that a listener's responses scatter around the true percept according to a Gaussian distribution. This assumption is supported by a large body of free-field localization studies (e.g. Oldfield and Parker, 1984; Makous and Middlebrooks, 1990; Middlebrooks and Green, 1991, Hofman and Van Opstal, 1998), that typically indicate a spatial resolution for sounds in the frontal elevation domain of about 4°. Here we took $\sigma_\epsilon = 2°$, in agreement with these experimental observations.

Interestingly, a simple linear weighting of each stimulus spectrum by its maximum-likelihood estimate (MLE) yielded spectral shape functions that bear a remarkable resemblance to the listener's DTFs. Many of the prominent features in the DTF also appeared in the reconstructed perceptual spectra. Moreover, the reconstructed spectra were very similar for the three different ripple bandwidths. These experiments therefore strongly support models in which the auditory system performs a cross-correlation analysis between the shape of the sensory spectrum and the entire set of stored HRTFs

Kulkarni and Colburn (1998) presented listeners with virtual sounds, filtered with HRTFs, in which the spectral details were systematically varied by the amount of smoothing, and compared their localization percepts with those to unfiltered sounds. The amplitude spectra of the HRTFs could be smoothed significantly without affecting the perceived location of the sound stimulus. Apparently, fine spectral detail is not required for sounds to be perceived at the

correct elevation.

The method of Hofman and Van Opstal (1998) provides an independent psychophysical technique to estimate the perceptual spectral shapes underlying elevation analysis within the auditory system, which requires no, or only limited, information about the listener's actual HRTFs. In our reconstructions, only the HRTF associated with the actual speaker's location was used. Indeed, when the speaker was positioned at an upward location, where the HRTFs possess relatively little spectral detail in comparison to the rippled test spectra, no prior information about the HRTFs was required to reconstruct the spectral features (Fig. 3.11B).

Hofman and Van Opstal (1998) proposed a spectral cross-correlation model that makes no a-priori assumptions about the shape of the source spectrum. In their model, the sensory spectrum is compared to a library of neurally stored HRTFs for all directions. Two basic assumptions underly the model. the first is that HRTFs are unique, and do not resemble each other. This can be readily tested by correlating the HRTFs with each other. Such an analysis indeed shows that HRTFs contain unique information about the sound-source's elevation (see e.g. Hofman and Van Opstal, 1998, Van Wanrooij and Van Opstal, 2005). The second assumption is that the free-field sound spectrum does not correlate well to any of the HRTFs in this stored representation. If true, it can be readily shown that the correlation between the sensory spectrum (the result of Eqn. 3.2) and the library of stored HRTFs will always peak at the elevation angle of the sound source. The simplest version of the cross-correlation model would therefore be to detect the peak in the cross-correlation, and to assign the perceived elevation to the location of that peak

To test this idea, we calculated the spectral correlations between the sensory spectra of all stimuli and all DTFs over a 4 to 14 kHz bandwidth for each of the listeners and each of the three stimulus sets. We then determined the elevation for which the correlation reached the maximum value, and compared this predicted elevation to the actual response of the listener. Figure 3.12A shows the measured response elevation vs. the predicted response elevation together with a regression line and the correlation for listener TE for the 3.0 c/o stimuli. Although the slopes of the regression lines were generally low (mean value of 0 17 over all listeners and stimulus sets), correlations were all highly significant ($p < 0.01$), varying between 0.32 and 0.64 (mean: 0.50). Therefore, this model does a reasonable job in explaining the variation of the observed behavior.

Yet, the model is too simple to account for the observed responses, as it falls clearly short in explaining the observed resolution (i.e the slope). There are at least two important points not accounted for by this simple algorithm:

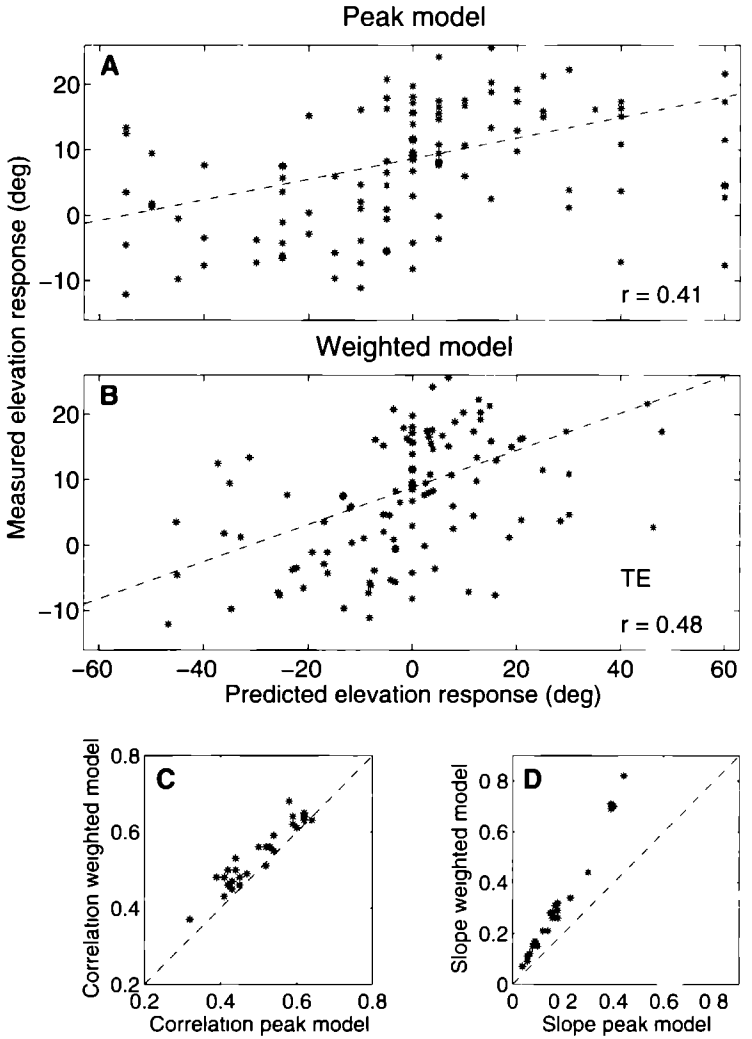


Figure 3.12: **(A)** *Measured elevation responses vs predicted elevation responses according to the maximum correlation model of Hofman and Van Opstal (1998) The correlation is indicated in the lower-right corner Data for the 3 0 c/o stimuli of listener TE* **(B)** *Same data, predicted elevation responses according to the weighted correlation model (see text)* **(C)** *Correlation values between measured and predicted elevation responses for all listeners and all stimulus sets Values for the weighted correlation model are plotted against the values for the peak correlation model Note that all data points are at or above the unity line* **(D)** *Same as in C but for the slope of the regression line through the measured vs predicted elevation responses Note that all data points are above the unity line*

first, rippled amplitude spectra are likely to yield several peaks in the cross-correlations. In our experiments, the DTF corresponding to the straight-ahead speaker position had a large influence on the sensory spectrum, and therefore always induced a peak in the cross-correlation. The presence of this peak could explain why the range of the listeners' responses was compressed toward the straight-ahead direction, but none of the models predicts the amount by which such a compression should occur. Second, the simple model works well if the stimulus does not correlate with the HRTFs *at all*, since in that case the peak in the cross-correlation will have a height close to one at the correct elevation. However, the random spectral-shape stimuli applied in this study yielded relatively low correlations with the DTFs (values typically around 0.40, or less, not shown) The value of this correlation could be indicative to the auditory system for the reliability of that particular elevation angle, and may hence determine the willingness of the system to move away from its default horizon (e.g. a preset bias, as seen in the median values in the response data of Tab. 3.1).

In a recent study we proposed that the value of this correlation, together with other sources of information like pre-knowledge of the environment, might thus determine the gain of the elevation responses (Vliegen and Van Opstal, 2004). Interestingly, several studies have indicated that a systematic variation of acoustic parameters may lead to a systematic decrease of the slope of the stimulus-response relationship for elevation. For example, reducing the duration of a noise burst systematically reduces the elevation gain (Hofman and Van Opstal, 1998; Vliegen and Van Opstal, 2004). In addition, varying the stimulus level for short-duration stimuli leads to a variation of the slopes of the stimulus-response relation (Hartmann and Rakerd, 1993; Macpherson and Middlebrooks, 2000; Vliegen and Van Opstal, 2004). Also the introduction of a noisy background, to manipulate the overall signal-to-noise ratio of the stimulus, systematically reduces the elevation response gain (Good and Gilkey, 1996; Zwiers et al., 2001). All these acoustic manipulations have in common that they affect the reliability of the spectral cues, albeit in different ways: background noise masks the spectral peaks and notches; loud, short-duration clicks may saturate the cochlear response, and hence abolish spectral detail, and soft brief noise bursts may deny the auditory system sufficient time to integrate the acoustic input signal in order to extract sufficient fine spectral detail

Thus, a simple extension of the cross-correlation model would incorporate the reliability of the spectral shape for a given elevation by multiplying the predicted elevation at the site of maximum cross-correlation with the value of that correlation The result of this analysis is shown in Figure 3.12B. Both the correlation between the predicted and measured elevation responses and the

slope of the regression line increase, while the correlation improves considerably, the value for the slope almost doubles This is exemplary for all listeners on average the correlation improved from 0 50 to 0 54, whereas the slope increased from 0 17 to 0 30 In Figures 3 12C and 3 12D the correlation and slope of the weighted correlation model are plotted against the values for the peak correlation model for all listeners and all stimulus sets Note that practically all data points are above the unity line Clearly, this simple extension can already improve the predictions of the cross-correlation model

Taken together, we conjecture that the auditory system somehow incorporates the possibility of multiple peaks by weighting the different peaks (i e candidate elevation angles) to extract the perceived elevation One possible weighting scheme could rely on taking the center of gravity of the different candidate elevations, in which the correlations between the stimulus spectrum and the HRTFs act as the weighting factors Since in the present experiment for most stimuli the DTF for straight ahead will typically yield the highest correlation, this might account for the reduced elevation response range

Other, more elaborate possibilities could be based on a Bayesian approach, in which the weighting may also depend on the prior distribution of spectral ripples This prior may be influenced by past experience, or by expectations on the distribution of spectral ripples Further work will be needed to study these different possibilities

Acknowledgments

We are grateful to Paul Hofman for the excellent stimulus generation and data analysis software, and to Gunther Windau, Ger van Lingen, Ton van Dreumel and Hans Kleijnen for technical support This research was supported by the Radboud University Nijmegen (AJVO, TVE) and the Netherlands Organization for Scientific Research (NWO - section Maatschappij- en gedragswetenschappen, MaGW, project nr 410-20-301, JV)

Chapter 4

Dynamic sound localization during rapid eye-head gaze shifts

Human sound localization relies on implicit head-centered acoustic cues However, to create a stable and accurate representation of sounds despite intervening head movements, the acoustic input should be continuously combined with feedback signals about changes in head orientation Alternatively, the auditory target coordinates could be updated in advance by using either the preprogrammed gaze-motor command, or the sensory target coordinates to which the intervening gaze shift is made ("predictive remapping") So far, previous experiments cannot dissociate these alternatives

Here we study whether the auditory system compensates for ongoing saccadic eye and head movements in two dimensions (2D) that occur during target presentation. In this case, the system has to deal with dynamic changes of the acoustic cues, as well as with rapid changes in relative eye and head orientation that cannot be preprogrammed by the audiomotor system. We performed visual-auditory double-step experiments in 2D in which a brief sound burst was presented while subjects made a saccadic eye-head gaze shift toward a previously flashed visual target.

Our results show that localization responses under these dynamic conditions remain accurate Multiple linear regression analysis revealed that the intervening eye and head movements are fully accounted for Moreover, elevation response components were more accurate for longer-duration sounds (50 ms) than for extremely brief (3 ms) sounds, for all localization conditions Taken together, these results cannot be explained by a predictive remapping scheme Rather, we conclude that the human auditory system adequately processes dynamically varying acoustic cues that result from self-initiated rapid head movements to construct a stable representation of the target in world coordinates This signal is subsequently used to program accurate eye and head localization responses.

Adapted from Vliegen J, Van Grootel TJ, and Van Opstal AJ (2004) *Dynamic sound localization during rapid eye-head gaze shifts* **J Neurosci** 24· 9291-9302

4.1 Introduction

Unlike the eye, the ear does not possess a topographical representation of the external world Instead, points on the basilar membrane respond to specific sound frequencies, thus providing a tonotopic code of sounds To localize sounds, the auditory system relies on implicit cues in the sound-pressure wave Binaural differences in sound arrival time and sound level vary systematically in the horizontal plane (azimuth), whereas direction-dependent spectral filtering by the head and pinnae (head-related transfer functions, or HRTFs) encodes positions in the vertical plane (elevation, Oldfield and Parker, 1984, Wightman and Kistler, 1989a, Middlebrooks, 1992, Blauert, 1997, Hofman and Van Opstal, 1998)

However, adequate sound localization behavior cannot rely exclusively on acoustic input (Poppel, 1973) In humans, the acoustic cues define a head-centered reference frame Therefore, accurate eye movements toward sounds require a coordinate transformation of the target into eye-centered motor commands, which necessitates information about eye position in the head (Jay and Sparks, 1984, 1987) Furthermore, in everyday life eye and head positions change continuously, both relative to the target sound and to each other To ensure accurate acoustic orienting of eyes and head, the audiomotor system should account for these changes (Goossens and Van Opstal, 1999)

In typical free-field localization experiments eyes and head start pointing straight-ahead Under such conditions, eye-centered, head-centered, and world-coordinate reference frames coincide, and a craniocentric target representation suffices to localize sounds and guide eye-head movements To dissociate the different reference frames, Goossens and Van Opstal (1999) employed an open-loop double-step paradigm (Fig 4 1A), in which the auditory gaze shift was made after an intervening eye-head saccade toward a visual target (ΔG_1) Saccades toward the sound reached the actual spatial target location (vector FB, Fig 4 1A), suggesting that the initial craniocentric target coordinates (T_H) were combined with the first eye-head movement Although this supports the hypothesis of a reference frame in world coordinates for sounds, an important alternative explanation, advanced in the visuomotor literature (Duhamel et al , 1992, Colby et al , 1995, Walker et al , 1995, Umeno and Goldberg, 1997), cannot be ruled out In this so-called "predictive remapping scheme" the craniocentric target location is updated either by prior efference information of the primary gaze shift (G_1, motor predictive vector MP, Fig 4 1A), or by the visual target vector (FV, visual predictive vector VP, Fig 4 1A) Note that these three different hypotheses yield nearly equivalent performance in the classical double-step task

The present study extends these experiments in two important ways First, by presenting the sound during eye-head gaze shifts, the binaural and spectral

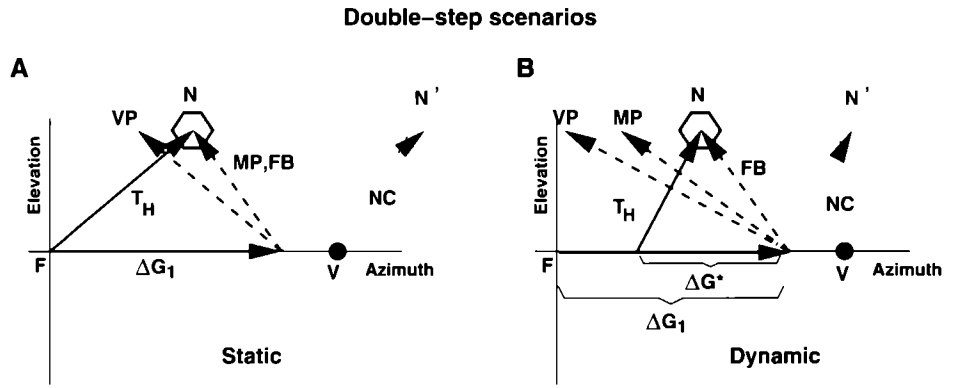


Figure 4.1: *Four models for how the audiomotor system could behave in the double-step paradigm* **(A)** *Static double-step trial in which the sound (N) is presented prior to the first gaze shift (ΔG_1) The non-compensation model (NC) predicts that the auditory target (N) is kept in a fixed craniocentric reference frame (T_H) Thus, after making the first gaze shift to V (the visual target), the second movement is directed to the location at N' In the dynamic feedback model (FB), the eye-head motor response to the sound fully accounts for the actual intervening gaze shift, ΔG_1 The response is given by $\Delta G_2 = T_H - \Delta G_1$ and is directed to N In the visual-predictive remapping model (VP), the system uses the predicted first gaze shift, specified by the required movement FV The second saccade is preprogrammed as $\Delta G_2 = VN = T_H - FV$ Any localization error of the first movement will not be accounted for Thus, the response is directed to P, rather than to N The motor-predictive model (MP) predicts accurate localization (update based on ΔG_1) in this condition* **(B)** *Predictions of the same four models for the dynamic double step, in which the sound is presented during the first gaze shift This yields a different head-centered target location, T_H In this condition, the MP model still uses the pre-programmed full first gaze shift to update the head-centered target location, instead of the partial gaze displacement following sound presentation (ΔG^*, as does the FB model), thereby directing the response to P, rather than to N Note that, in contrast to the FB model, the NC, VP, and MP models predict different responses than in the static paradigm, and that the predictions of the FB and MP models are now better dissociated.*

acoustic cues are no longer static, but vary in an extremely complex way with head velocity.

Second, the audiomotor system is denied any prior information about either the upcoming target location, or the subsequent changes in eye and head orientation, which renders the acoustic cue dynamics entirely unpredictable. This poses a serious problem for the predictive remapping model, according to which the craniocentric target location is updated on the basis of the (pre-programmed) full first gaze-displacement vector, rather than on the partial gaze shift following target presentation. This allows for a clear dissociation of the different schemes (Fig. 4.1B, cf. vectors FB, MP, and VP).

Our data show that, in contrast to the prediction of the predictive remapping

models, the audiomotor system remains accurate, also under these dynamic conditions. These results demonstrate that the system is capable to create, and adequately use, a stable representation of sounds in world coordinates.

4.2 Methods

Subjects

Five subjects (one female, four males; ages 25 to 46) participated in the experiments. All had normal hearing and were experienced in the type of sound-localization experiments conducted in our laboratory All subjects had normal vision except for JO (an author), who is amblyopic in his right, recorded, eye. Oculomotor and head-motor responses of subjects were all within the normal range. Subjects MW and RK were kept naive about the purpose of this study. Subjects JO, JV, and TG participated in all experiments; subject RK only participated in the first target configuration (see below); subject MW only in the second target configuration, so that each experiment contains data from four subjects.

Apparatus

Experiments were conducted in a completely dark, sound-attenuated room (3 x 3 x 3 m^3) in which the four walls, floor, ceiling and all other large objects were covered with black sound-absorbing foam that eliminated acoustic reflections down to 500 Hz (Schulpen Schuim, The Netherlands). The ambient background noise level in the room was 35 dBA SPL (measured with BK-414 microphone and BK-2610 amplifier, Brüel & Kjær, Norcross, GA) Subjects were seated comfortably on a chair in the center of the room with support in their back and lower neck. They faced an acoustically transparent thin-wire hemisphere with a radius of 0.85 m, the center of which coincided with the center of the subject's head. On this hemisphere 85 red/green light-emitting diodes (LEDs) were mounted at seven visual eccentricities $R = [0, 2, 5, 9, 14, 20, 27, 35]^\circ$ relative to the straight-ahead viewing direction (defined in polar coordinates as $[R.\phi] = [0, 0]^\circ$), and at twelve different directions, given by $\phi = [0, 30, ..., 330]^\circ$, where $\phi = 0^\circ$ is rightward from the center location, and $\phi = 90^\circ$ is upward. The hemisphere was covered with thin black silk to hide the speaker completely from view (Hofman and Van Opstal, 1998; Corneil et al., 2001).

Auditory stimuli emanated from a mid-range speaker that was attached to the end of a two-link robot, which consisted of a base with two nested L-shaped arms, each driven by a stepping motor (Berger-Lahr, Lahr, Germany,

type VRDM5). The speaker could move quickly (within 3 s) and accurately (within 0.5°) to practically any point on a virtual hemisphere at a radius of 0 90 m from the subject's head. To prevent sounds generated by the stepping motors from providing potential clues to the subject about either the location or displacement of the speaker, the robot always made a random dummy movement of at least 20° away from the previous location, before moving to its next target position. Earlier studies in our group have verified that this procedure guaranteed that sounds from the stepping motors did not provide any consistent localization cues (Frens and Van Opstal, 1995; Goossens and Van Opstal, 1997b).

Stimuli

Auditory stimuli were digitally generated with Matlab software (The Mathworks, Natick, MA) Signals consisted of 50 ms duration broad-band (0.2 to 25 kHz) Gaussian white noise, with 0.5 ms sine-squared on- and offset ramps, and were stored on disk at a 50 kHz sampling rate. Upon receiving a trigger, the stimulus was passed through a 12-bit DA-converter (Data Translation DT2821, output sampling rate 50 kHz), band pass filtered (Krohn-Hite model 3343, 0.2 to 20 kHz), and passed to an audio amplifier (Luxman A-331) that fed the signal to the robot's speaker (Philips Eindhoven, The Netherlands, type AD-44725). The intensity of the auditory stimuli was fixed at 55 dBA SPL (measured at the position of the subject's head). Visual stimuli consisted of red LEDs with a diameter of 2.5 mm (which subtended a visual angle of 0 2° at 0 85 m viewing distance) and an intensity of 0.15 cd/m^2.

Measurements

Head and eye movements were measured with the magnetic search-coil induction technique (Collewijn et al., 1975). Subjects wore a lightweight helmet (about 150 g), consisting of a narrow strap above the ears, which could be adjusted to fit around the head, and a second strap that ran over the head. A small coil was mounted on the latter. Subjects also wore a scleral search coil on one of their eyes. In the room two orthogonal pairs of 3 x 3 m^2 square coils were attached to the side walls, floor, and ceiling to create the horizontal (30 kHz) and vertical (40 kHz) oscillating magnetic fields that are required for this recording technique. Horizontal and vertical components of head and eye movements were detected by phase-lock amplifiers (Princeton Applied Research, models 128A and 120), low pass filtered (150 Hz), and sampled at 500 Hz per channel before being stored on disk.

Two PCs controlled the experiment. One PC-486 was equipped with the

hardware for data acquisition (Metrabyte DAS16), stimulus timing (Data Translation DT2817), and digital control of the LEDs (Philips I2C). The other PC-486 controlled the robot movements and generated the acoustic stimuli upon receiving a trigger from the DT2817.

Experimental paradigms

Calibration of eye and head Each experimental session started with three runs to calibrate the eye and head coils (Goossens and Van Opstal, 1997b). Prior to the calibration, subjects were asked to keep their head in a neutral, comfortable straight-ahead position and adjust a dim red LED mounted at the end of a thin pliable aluminum rod that was attached to their helmet (at a distance of ~0 40 m in front of the subject's eyes) such that it was approximately aligned with the center LED of the hemisphere. This rod LED was illuminated only in the second and third calibration sessions, and was off during the actual localization experiments.

First, eye position in space ("gaze") was determined. During this calibration, subjects kept their head still in the straight-ahead position and fixated the LEDs on the hemisphere with their eyes. Targets (N = 37) were presented once, in a fixed counterclockwise order, at the center location ($R = 0$), followed by three different eccentricities: $R = [9, 20, 35]^\circ$, and all twelve directions. When subjects fixated the target, they pushed a button to start data acquisition, while keeping their eyes at that location for at least 1000 ms

In the second calibration run, the eye-in-head offset position was determined. To that end, subjects fixated the dim red LED on the helmet rod (rather than the LED on the hemisphere) while keeping their head in the straight-ahead position. This procedure kept their eyes at a fixed orientation in the head. When the subject assumed the neutral head posture, he or she pushed a button to start 1000 ms of data acquisition. This procedure was repeated ten times. In between trials subjects were asked to freely move their head before re-assuming the neutral position.

The third calibration run served to calibrate the coil on the head. Now subjects had to fixate the dim red LED at the end of the head-fixed rod with their eyes and align this rod LED with the same 37 LED targets on the hemisphere as in the eye calibration run. In this way, the eyes remained at the same fixed offset position in the head as in the second calibration. When the subject pointed to the target, he or she started 1000 ms of data acquisition by pushing a button.

After the calibration runs were completed, the experimental localization sessions started. One experimental session consisted of at least four different blocks of trials: *1*) visual single-step localization; *2*) visual-visual double-step local-

ization; *3*) auditory single-step localization, and *4*) visual-auditory double-step localization. Blocks of one modality were always presented together, and the single-step block was always presented first. After these four blocks, additional visual-auditory double-step blocks could be performed until the subject wanted to stop. In this paper, we will focus on the auditory single and double-step experiments only. Results of the visual eye-coordination experiments will be presented elsewhere. All calibration and experimental sessions were performed in complete darkness.

Auditory single-step paradigm To determine a subject's baseline localization behavior, a single-step localization experiment was performed. Each trial started with the presentation of a fixation LED. During fixation, subjects had their eyes and head approximately aligned. After 800 ms, this LED was switched off and 50 ms later an auditory stimulus was presented at a peripheral location. Subjects were asked to point to the apparent location of the stimulus as quickly and as accurately as possible by redirecting their gaze line to the perceived peripheral target location. As stimuli were always extinguished well before the initiation of the eye and head movement, the subject performed under completely open-loop conditions.

To enable a direct comparison of the single-step responses with the second gaze shifts from the double-step paradigms (see below), we designed the single-step experiment such that the initial visual fixation targets of this experiment were the same as the first peripheral visual targets in the double-step experiments. Also the sound locations of the single step experiment were the same as those in the double-step experiments.

There were two different stimulus configurations. The first consisted of a central visual fixation target at $[R, \phi] = [0, 0]^\circ$, and ten different auditory target positions (relative to the straight-ahead direction) with $[R, \phi] = [14, 0]$, $[14, 180]$, $[20, 0]$, $[20, 90]$, $[20, 180]$, $[20, 270]$, $[27, 60]$, $[27, 120]$, $[27, 240]$, or $[27, 300]^\circ$. Target locations were selected in random order. One block consisted of 20 trials. In the second configuration the initial fixation target was at either $[R, \phi] = [20, 90]$ or $[20, 270]^\circ$ (pseudo-randomly chosen with both fixation targets occurring equally often). Auditory targets were presented at a randomly selected position within a circle of $R = 35^\circ$ around the straight-ahead direction, but always at least 10° away from the initial fixation target. A total of 24 trials were presented in one block.

Visual-auditory double-step paradigms We used both a static double-step target condition, in which the second target was presented before initiation of

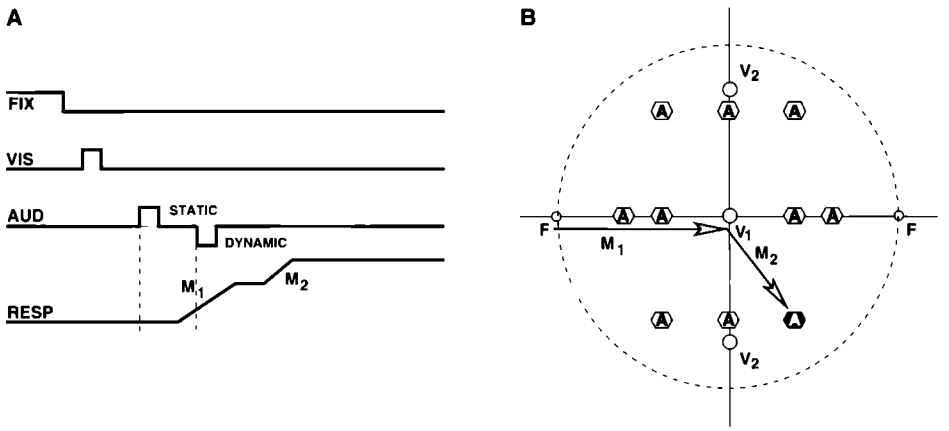


Figure 4.2. *Double-step paradigms* **(A)** *Temporal order of the different targets in the static and dynamic double-step trials* M_1, M_2 *first and second eye-head movement* **(B)** *Spatial layout of the target configurations* F *initial fixation positions* V_1 *visual target in the first double-step series, where* M_1 *is a purely horizontal movement* V_2 *visual target in the second double-step series in which* M_1 *is an oblique gaze shift* A *potential auditory target locations in the first target configuration. Dashed circle area within which auditory targets were selected for the second target configuration*

the first eye-head movement, and a dynamic condition in which the second target was presented during the first eye-head movement. The latter paradigm is adopted from the classical saccade-triggered visuomotor paradigm of Hallett and Lightstone (1976). The visual-auditory double-step paradigm is illustrated in Figure 4.2. First, a fixation target (F) is presented for 800 ms. Then, after 50 ms of compete darkness, a visual target (V) is flashed for 50 ms (Fig. 4.2A).

The timing of the second, auditory, target (N) was varied, resulting in three different conditions:

1. Non-triggered (static) condition, in which the auditory target was presented after a fixed delay of 50 ms after extinction of the peripheral visual target. In this condition both targets were presented before the first gaze-shift onset, which typically started at a latency of about 200 ms after the visual stimulus flash.

2. Early-triggered (dynamic) condition, in which the auditory target was triggered as soon as the head velocity in the direction of the visual target exceeded 40°/s. In this way, the timing of the auditory stimulus fell early in the first head movement, and often was presented while the gaze line (the eye in space) was still moving.

3. Late-triggered (dynamic) condition, in which the auditory target was triggered 50 ms after head velocity in the direction of the visual target exceeded 60°/s. In this way, stimulus presentation fell approximately halfway through the first head movement, and typically close to the moment of the head's peak velocity (e.g. Goossens and Van Opstal, 1997b).

Two different stimulus configurations were used (Fig. 4.2B). The first configuration (subjects JO, JV, RK, TG) consisted of an eccentric fixation target at $[R.\phi]$ = [35, 0]° or [35, 180]° (pseudo-randomly chosen as above), a visual target at $[R.\phi]$ = [0, 0]°, and an auditory target at ten possible target positions at polar coordinates $[R.\phi]$ = [14, 0], [14, 180], [20, 0], [20, 90], [20, 180], [20, 270], [27, 60], [27, 120], [27, 240], or [27, 300]°. Target locations were selected in random order. One block consisted of 20 non-triggered and 20 early-triggered trials (randomly interleaved).

Because in this configuration the peripheral visual target was always at the same position, eight additional catch trials were included in the experiment to prevent the subject from making a predictive movement to the visual target position. In catch trials the visual target was at either $[R.\phi]$ = [35, 30], [35, 150], [35, 210], or [35, 330]°, and the second (auditory) target was presented at either $[R.\phi]$ = [20, 90] or [20, 270]° (pseudo-randomly chosen with all positions occurring equally often).

In the second double-step configuration (subjects JO, JV, MW, TG) the initial fixation target was again at $[R.\phi]$ = [35, 0] or [35, 180]°, but now the peripheral visual target was at either $[R.\phi]$ = [20, 90] or [20, 270]° (both pseudo-randomly chosen as above). This resulted in a first gaze shift with a horizontal as well as a considerable vertical component, in contrast to the first target configuration, in which the first gaze shift was always purely horizontal. The auditory target was presented at a randomly selected position within a homogeneous area of $R = 35°$ around straight-ahead, but always at least 10° away from the visual target. This block consisted of 48 late-triggered trials, but if, after four experimental blocks, the subject was capable of doing additional experiments, we repeated this visual-auditory block with a reduced number of 24 trials.

In all experimental localization sessions, subjects were free to move their head and eyes to localize both targets. They were asked to localize the stimulus as quickly and as accurately as possible, by fixating the perceived stimulus location with their eyes, but they were not given specific instructions about the movements of their head.

Data analysis

After calibration, the coordinates of auditory and visual target locations, as well as the eye and head positions and movement displacement vectors, were all expressed in a double-pole azimuth-elevation coordinate system in which the origin coincides with the center of the head (Knudsen and Konishi, 1979) In this system, the azimuth angle, α, is defined as the angle within the horizontal plane with the vertical midsaggital plane, whereas the elevation angle, ε, is defined as the direction within a vertical plane with the horizontal plane through the subject's ears The straight-ahead direction is defined by $[\alpha\ \varepsilon] = [0, 0]^\circ$ The relation between the $[\alpha\ \varepsilon]$ coordinates and the polar $[R.\phi]$ coordinates defined by the LED hemisphere (see above) is given in Hofman and Van Opstal (1998)

Calibration of the data The raw eye-position data and the corresponding known LED positions from the first calibration run were used to train two three-layer back propagation neural networks that mapped the raw eye-position signals to calibrated azimuth/elevation angles of eye position in space (gaze) The networks compensated for minor cross talk between the horizontal and vertical channels, and for small non-homogeneities and nonlinearities in the magnetic fields

Calibration of the head-coil fixations was obtained in the following way First, the calibrated eye-position data from the second calibration session, with the head in the neutral position, were determined and averaged, to yield an average eye-in-head offset gaze position, G_0 Then, the raw eye-position data obtained from the head-coil calibration run were calibrated with the eye-coil calibration networks from the first calibration run Subsequently, the static head position data were corrected for the mean offset in eye-in-head position according to $H = G - G_0$, where H represents the position of the head in space, as measured with the eye coil Finally, the head-coil data were calibrated by mapping the raw head position signals on the calibrated eye-coil data with an additional set of two neural networks (Goossens and Van Opstal, 1997b) In the calibrated response data, we identified head and gaze saccades with a custom-written computer algorithm that applied separate velocity and mean-acceleration criteria to vectorial saccade onset and offset, respectively Markings were visually checked and corrected, if deemed necessary To ensure unbiased detection criteria, the experimenter was denied any information about the stimulus Responses with a first-saccade latency shorter than 80 ms (considered to be predictive) or longer than 800 ms (potentially caused by inattentiveness of the subject) were discarded from further analysis To ensure that the static trials were indeed static, we checked whether first head-saccade latency in those trials exceeded 150 ms

(offset time of auditory target re onset visual target) This requirement was met for all trials (see Fig 4 5, for an example)

Regression Analysis and Statistics To evaluate to what extent the audiomotor system compensates for the occurrence of intervening eye and head movements we analyzed the second gaze shift and the second head movement by applying a multiple linear regression analysis to the azimuth and elevation response components, respectively Parameters were determined on the basis of the least-squares error criterion

The bootstrap-method was applied to obtain confidence limits for the optimal fit parameters in the regression analyses To that end, 100 data sets were generated by random selections of data points from the original data Bootstrapping thus yielded a set of 100 different fit parameters The standard deviations in these parameters were taken as an estimate for the confidence levels of the parameter values obtained in the original data set (Press et al , 1992)

To determine whether two (non-Gaussian) data distributions were statistically different, we applied the Kolmogorov-Smirnov (KS) test This test provides a measure (d-statistic) for the maximum distance between the two distributions, for which the significance level, p, that the distributions are the same, can be readily computed If $p < 0\ 05$ the two data sets were considered to correspond to different distributions For data expressed as 2D distributions (e g the azimuth-elevation end-points in Fig 4 7), we computed the 2D KS-statistic to measure their mutual distance and its significance level (Press et al , 1992)

The bin-width (BW) of histograms (Figs 4 5 and 4 7) was determined by $BW = Range/\sqrt{N}$, with $Range$ the difference between the largest and smallest values (excluding the two most extreme points), and N the number of included points

4.3 Results

Double-step response behavior Figure 4 3 shows three typical examples of head and gaze traces as a function of time of subject JV elicited in the double-step experiments, one for the static condition (Fig 4 3A) and two for the dynamic conditions (early-triggered Fig 4 3B, late-triggered Fig 4 3C) In the static double-step condition the visual and the auditory target are both presented and extinguished before the initiation of the visually evoked head and gaze movement For the two dynamic conditions the auditory target, which is triggered by the head movement, falls either early in (Fig 4 3B), or halfway through (Fig 4 3C), the first head saccade For all three conditions, gaze sac-

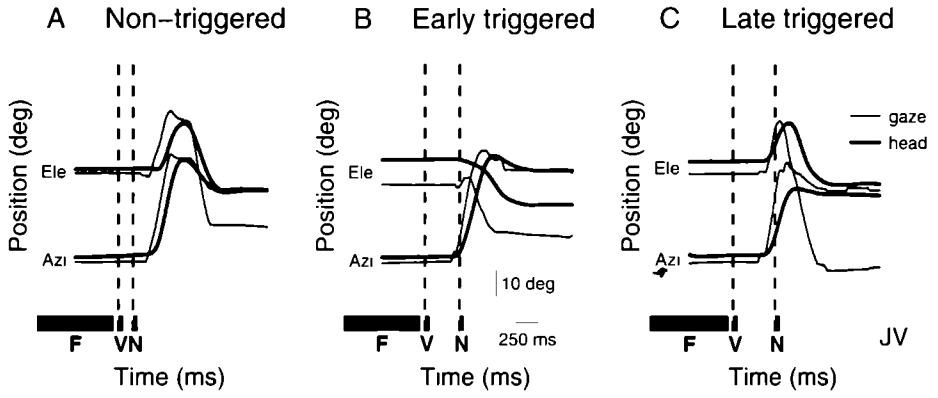


Figure 4.3: *Head (thick lines) and gaze (thin lines) double-step responses as a function of time for azimuth and elevation components. F, V and N indicate the time of presentation of the fixation target, the visual target, and the auditory target, respectively.* **(A)** *Trial from the non-triggered static condition, where the second, auditory, target is presented before initiation of the primary head and gaze movement.* **(B)** *Trial from the early-triggered dynamic double-step condition. Here, the auditory target is presented early in the saccade.* **(C)** *Trial in the late-triggered dynamic condition. Here the auditory target falls halfway through the first head saccade. Data from subject JV.*

cades are faster and larger than head saccades, which is a typical pattern for eye-head coordination (e.g. Goossens and Van Opstal, 1997b). At the end of the second gaze shift, the vestibulo-ocular reflex (VOR) ensures that gaze-in-space remains stable, despite the ongoing movement of the head.

Figure 4.4 shows six typical examples of 2D spatial head and gaze trajectories of subject JV for the static condition (Fig. 4.4A), for the early-triggered condition (Fig. 4.4B), and for the late-triggered condition (Fig. 4.4C). The dashed squares (N') indicate the spatial locations to which the second gaze shift would be directed if it were only based on the initial head-centered acoustic input. However, these examples show that head and gaze responses are both directed toward the actual stimulus location. Gaze approaches the auditory target more closely than the head, which tends to undershoot the vertical target component (top row of Fig. 4.4).

Head and eye movements during sound stimuli The aim of the triggered double-step experiments was to ensure considerable and variable head movements during the presentation of brief acoustic stimuli. To verify that head and eye were indeed moving substantially during sound presentation, Figure 4.5 shows all 2D head (left) and eye-movement (right) traces of subject JO during

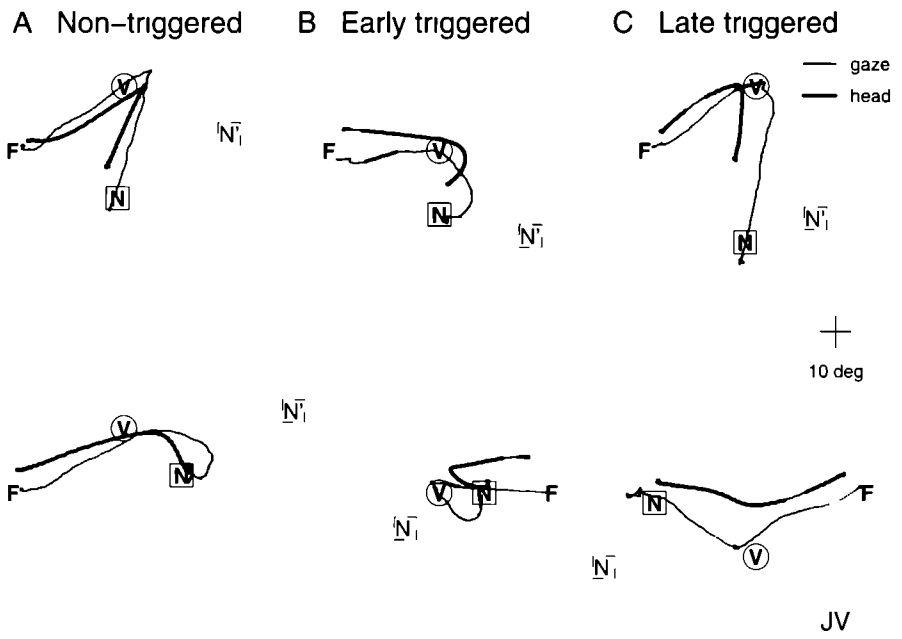


Figure 4 4 *Head (thick lines) and gaze (thin lines) double-step response traces in space F V and N indicate the positions of the fixation target the visual target and the auditory target respectively* **(A)** *Two representative trials from the non-triggered condition* **(B)** *Two trials from the early-triggered condition The target presentation epoch is indicated by a change in line thickness* **(C)** *Two trials from the late triggered condition If the second saccade would be based purely on the initial head-motor error the responses would be directed toward the dashed square (N') For the dynamic conditions the initial target re-head position is defined as the target position relative to the head at sound onset Note that the responses are directed toward the veridical location of the sound Data from subject JV*

the 50 ms acoustic noise burst pooled for the two dynamic triggering conditions (Fig 4 5A) The onsets of all movements are aligned in (0,0)° for ease of comparison Note that the majority of head displacements during the brief stimulus were on the order of 10° or more (Fig 4 5A, left) Typically, the eye moved much faster in an eye-head gaze shift (see Fig 4 3) Therefore, in the late-triggered double-steps the eye often reached the visual target location, while the head was still moving In those cases, the VOR kept gaze at its new position Yet, for the majority of dynamic trials, also the eye-in-space moved substantially during sound presentation (Fig 4 5A, right), especially for the early-triggered condition (horizontal traces) The head and eye-movement amplitude in the dynamic condition, averaged across subjects, was 8 0° ± 3 0° and 5 0° ± 3 0°, respectively

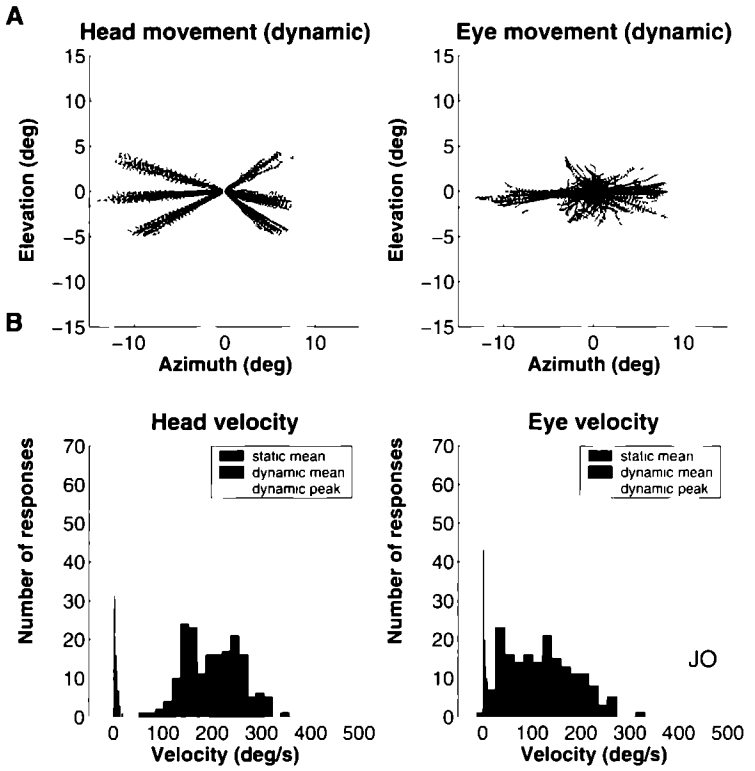


Figure 4.5: *Properties of ongoing head and eye movements during presentation of the auditory target* **(A)** *Two-dimensional head (left) and eye-movement traces (right) during stimulus presentation in the dynamic condition (early- and late-triggered trials pooled)* **(B)** *Head and eye mean and peak velocity during the 50 ms stimulus presentation for both the static (dark-gray histogram, only mean velocity shown) and the dynamic condition (black and light-gray histograms for mean and peak velocities, respectively) Note large trial-to-trial variability in eye and head movement kinematics for the dynamic double-step trials Data from subject JO*

Figure 4.5B shows histograms of the mean (black) and peak (light-gray) head (left) and eye (right) velocities during sound presentation in both the dynamic and static (dark-gray; only mean velocity shown) double-step conditions for this subject. As required, eyes and head were not moving in the static double-step trials. In the dynamic conditions, however, there is a large range of both the mean and peak head velocities. The mean head velocity is about 150°/s; peak head velocity is on average 200°/s. As a result, the acoustic cues vary considerably from trial to trial, and in an unpredictable way. Moreover, in many trials the eyes also moved substantially with respect to the sound.

Although at the start of a double-step trial the eyes and head were ap-

proximately aligned, this is no longer the case after the first gaze shift. To illustrate the trial-to-trial variability in eye-head misalignment at the onset of the auditory-evoked gaze shift, Figure 4.6 shows the distribution of eye-in-head positions pooled for all subjects across trials. The shaded central square indicates trials for which both the horizontal and vertical eye-position eccentricity was less than 10° (see also Fig. 4.9). Note that the misalignment between eye and head can be as large as 30°, although for the majority of trials the eye stays within 10° of the center of the oculomotor range.

Figure 4.6: *Eye-in-head positions at the onset of the second, auditory-evoked gaze shift (E_0 in Eqn 4.1). The eye is typically eccentric in the head, so that gaze-in-space and head-in-space are not aligned at the start of the second gaze shift. Points within the square correspond to eye positions with azimuth and elevation components $< 10°$*

Sound-localization errors To compare response accuracy for the different stimulus conditions, Figure 4.7 shows the 2D distributions of the endpoints of second gaze saccades for static (filled circles) and dynamic (gray triangles) double-step trials (early- and late-triggered data pooled, as they were statistically indistinguishable), as well as for the single-step localization responses (open dots). In this figure, all auditory target locations have been aligned with the origin of the azimuth-elevation coordinate system. Gaze end-positions are plotted as undershoots (azimuth, elevation < 0) or overshoots (azimuth, elevation > 0) with respect to the target coordinates. The static double-step data are summarized by the black histograms, and the corresponding dashed lines indicate their medians. The dynamic double-step data are represented by the gray histograms, and the continuous lines show their median values. The medians of the single-step condition are indicated by black dotted lines.

Quite remarkably, the response distributions for the single-step localization

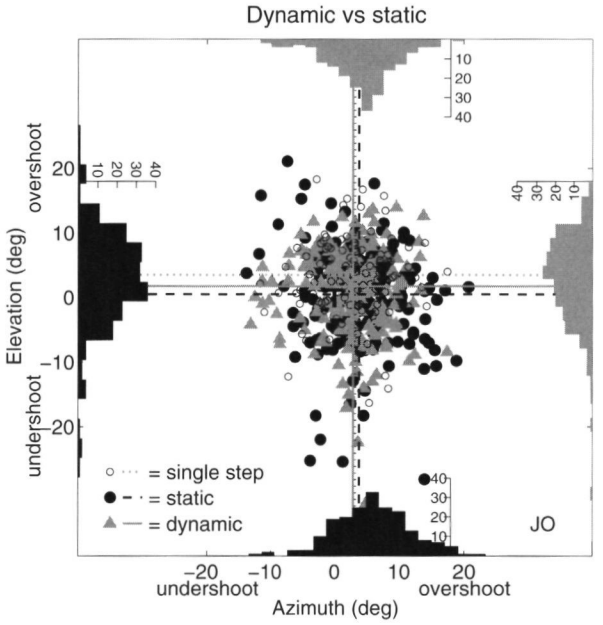


Figure 4.7: *Endpoints of second gaze saccades in azimuth and elevation plotted relative to the acoustic target position. The latter (T) is aligned with (0,0)°; gaze responses are expressed as undershoots or overshoots with respect to the target location. Histograms show the respective response distributions for the static (black, filled black circles) and triggered dynamic (gray, gray triangles) double-step responses. Dashed lines indicate means of the static double steps; solid lines of the dynamic double steps. Note similarities in the distributions. Open dots correspond to gaze endpoints toward single-step targets, dotted lines indicate their means. Data from subject JO.*

trials, and the static and dynamic double-step trials are very similar. The mean unsigned errors and standard deviations for the three conditions are virtually the same. The three pair-wise 2D Kolmogorov-Smirnov tests (Press et al., 1992) indicated that the end-point distributions were statistically indistinguishable, except for the single-step vs. the non-triggered double-step comparison (single step vs. non-triggered double-steps: $p < 0.05$, $d = 0.25$; single step vs. triggered double-steps: $p = 0.09$, $d = 0.17$; non-triggered vs. triggered double-steps: $p = 0.10$, $d = 0.14$). Table 4.1 summarizes the mean unsigned errors for the different conditions, pooled for all subjects. Note also that for all conditions the response distributions are broader for elevation than for azimuth response components (all three KS-tests on azimuth vs. elevation: $p < 0.001$). Such a difference in response accuracy is typical for human sound localization performance to single steps, and underlines the different neural mechanisms for the extraction of the

Condition	Azimuth (degrees)	Elevation (degrees)	KS test	n
Single steps	1.8 ± 5 8	3.1 ± 9 8	$p < 0.001$	432
Non-triggered double steps	3.4 ± 6 4	3.0 ± 12 1	$p < 0\ 001$	657
Triggered double steps	1.2 ± 6 3	3.3 ± 11 6	$p < 0\ 001$	651

Table 4.1: *Mean and standard deviation of saccade endpoint errors for the single-step and double-step paradigms The 1D KS-test was performed on ranked azimuth vs elevation distributions within each stimulus condition Data pooled for all five subjects, and recording sessions*

spatial acoustic cues This feature appears to be preserved also in the static and dynamic double-step localization trials.

Regression analysis: sound reference frame To test in a quantitative way to what extent the intervening eye and head movements of the first gaze shift are accounted for in planning the eye-head saccade to the auditory targets, we performed multiple linear regression on the second, auditory-guided gaze displacement. In this analysis, ΔG_2, which is the displacement of the eye in space from its starting position at the end of the first gaze shift, was described by a linear combination of the initial sound location in head-centered coordinates, $T_{H\,ini}$, the subsequent displacement of the head during the first gaze shift, ΔH_1, and the position of the eye in the head after the first gaze shift, E_0, according to:

$$\Delta G_2 = a \cdot T_{H\,ini} + b \cdot \Delta H_1 + c \cdot E_0 + d \tag{4.1}$$

in which (a, b, c) are dimensionless response gains, and d (in deg) is the response bias. Eqn. 4.1 was applied separately to the azimuth and elevation response components.

Note that if the audiomotor system would not compensate for the intervening eye-head gaze shift but instead would keep the sound in the initial head-centered coordinates determined by the acoustic cues, the regression should yield $a = 1$, and $b = c = d = 0$ (indicated by model I in Fig. 4.1A). Full compensation for the first gaze shift requires that $a = 1$, $b = c = -1$ and d =0 (FB model in Fig. 4.1A), in which case Eqn. 4.1 simply reduces to $\Delta G_2 = T_{H\,ini} - \Delta G_1$. For the static, non-triggered double-step responses, the first head displacement (ΔH_1) is defined as the entire head displacement, whereas for the triggered double-step trials it is the portion of the head displacement that followed the sound onset (see Fig. 4.1B). The head-centered location of the sound is determined by the

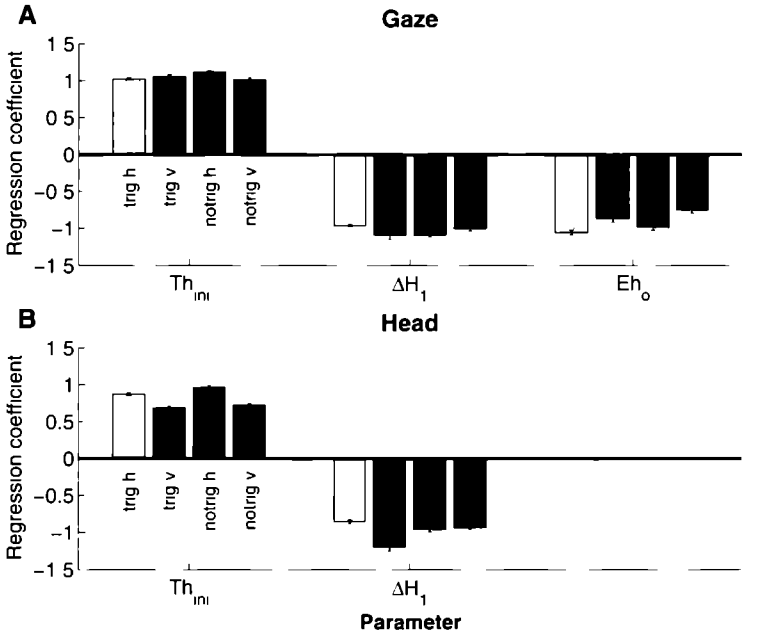


Figure 4.8. **(A)** *Regression coefficients of Eqn 4 1) for second gaze saccades (ΔG_2), averaged across subjects and recording sessions* **(B)** *Regression coefficients of Eqn 4 2) for second head saccades (ΔH_2) Different double-step conditions (dynamic/static) and response directions (horizontal/vertical) are represented by the different gray-coded bars Dotted lines at the values of +1 0 and −1 0 correspond to ideal compensation for the intervening movements*

head position in space at sound onset. Data from the early-triggered and late-triggered experiments were pooled. The resulting gains (a, b, c) of the regression, averaged across subjects, are summarized in Figure 4.8A for the different conditions and response components (results for individual subjects are provided in Supplementary Table 4.2). The gain-coefficient (a) for the craniocentric target location is close to +1.0 for all conditions and response components. Moreover, the response gains for head displacement, as well as for eye-in-head position, are close to the optimal values of −1.0. The coefficient for eye-in-head tends to be slightly lower than −1.0. As we did not systematically control eye-position offset, it varied between subjects; some subjects made relatively large head movements, causing their eyes to remain closer to the center of the oculomotor range. Since there were no subjects who over-compensated eye-in-head position, the average across subjects tended to be lower than −1.0. The offsets (d) were always close to zero deg and are not shown

This result implies that subjects fully compensate for the intervening eye-

head movement, even under dynamic localization conditions.

A similar multiple regression analysis was performed on the second head-movement vector, ΔH_2, in response to the auditory target. In that case, the head response was described by:

$$\Delta H_2 = a \cdot T_{H\,ini} + b \cdot \Delta H_1 + d \tag{4.2}$$

The results (averaged across subjects) are shown in Figure 4.8B. Note that also for the head, the fitted gains $(a.b)$ are close to the ideal values of $+1.0$ and -1.0, respectively, The target elevation gain (a, in Eqn. 4.2) for the head responses was found to be lower than for the eye (Eqn. 4.1) This probably reflects a robust motor strategy to withhold the head from making large movements against gravity (André-Deshays et al., 1988). Results of this analysis for the individual subjects are provided in Supplementary Table 4.3.

Regression analysis: motor error frames In generating a gaze shift toward an auditory target, it is not trivial that eye and head both move toward the target, especially if eye and head are not aligned. For that to happen, the world target coordinates need to be transformed into oculocentric and head-centered coordinates, respectively. Alternatively, both could be driven by the same motor-error signal, either by an oculomotor gaze-error signal (as in the so-called common-gaze control model for eye and head; Vidal et al., 1982; Guitton, 1992; Galiana and Guitton, 1992), or by a (acoustically-defined) head motor-error signal. The difference between these two reference frames is determined by the position of the eye in the head, which varies considerably and unpredictably from trial to trial, and can be as large as 30° (Fig. 4.4). To investigate this point, we subjected the data to a normalized multiple linear regression in which the auditory-evoked head movement, ΔH_2, and the gaze shift, ΔG_2, are each described as a function of gaze motor error, GM, and head motor error, HM:

$$\Delta H_2' = p \cdot GM' + q \cdot HM' \tag{4.3}$$

$$\Delta G_2' = p \cdot GM' + q \cdot HM' \tag{4.4}$$

In Eqn. 4.3 and 4.4 head motor error (HM) was determined as the difference between the auditory target in space and the head position in space at the start of the second gaze shift. Gaze motor error (GM) was taken as the difference between the auditory target location and the eye position in space at the start of the gaze shift (i.e. the sound's retinal error). These response variables were transformed into their (dimensionless) z-scores: : $x' = (x - \mu_x)/\sigma_x$, with μ_x the

mean of variable x, and σ_x its variance. In this way, the variables can be directly compared, and p and q are the (dimensionless) partial correlation coefficients for gaze motor-error and head motor-error, respectively. If $p > q$, the head (or eye) is driven predominantly by an oculocentric gaze-error signal. If $q > p$, the head (or eye) rather follows the head-centered motor error signal In case $p > q$ (or $p < q$), for both equations, eye and head are considered to be driven by the same error signal. To allow for a meaningful dissociation of the oculocentric and head-centered reference frames, we only incorporated trials for which the absolute azimuth or elevation component of eye-in-head position exceeded 10° (those positions falling outside the square in Fig. 4.6), and the directional angle between the head and gaze motor-error vectors was at least 15°. In this way, we incorporated a sufficient number of data points for three subjects.

Figure 4.9 shows the regression coefficients on the pooled data from all subjects for all conditions. It can be seen (Fig. 4.9A) that for head movement, the coefficients for head motor-error are larger than those for gaze motor-error (for all conditions $p < 0.01$, apart from the triggered vertical condition, where the difference failed to reach significance). This suggests that the head is indeed driven by a craniocentric motor command. Conversely, the eye-in-space is clearly driven by gaze motor-error, as for all conditions $p > q$ (Fig. 4.9B) (for all conditions $p < 0.01$). These data therefore show that the audiomotor system is capable to dynamically transform the auditory target coordinates into the appropriate motor reference frames. Data for individual subjects are summarized in Supplementary Tables 4.4 and 4.5. The values for p and q vary somewhat between subjects and conditions, especially for the head movements, where for 2/16 conditions $p > q$. We have no obvious explanation for this variability.

Quantitative Model Tests In the Introduction we described four different models to predict the coordinates of the second gaze shift in a visual-auditory double-step paradigm (Fig. 4.1). In particular, it was argued that the results from the non-triggered double-step trials could be explained equally well by two conceptual models: In the dynamic feedback scheme (FB model) the instantaneous head and eye movements are incorporated in the computation of the auditory spatial coordinates. In contrast, the motor-predictive remapping scheme (MP model) employs prior (static) information of the upcoming gaze shift to update the auditory target location. To test whether the results from the triggered double-step experiments could indeed dissociate these models, we computed the predicted second gaze displacement for the different schemes from the recordings.

The predictive remapping model was tested in two different ways: in the first

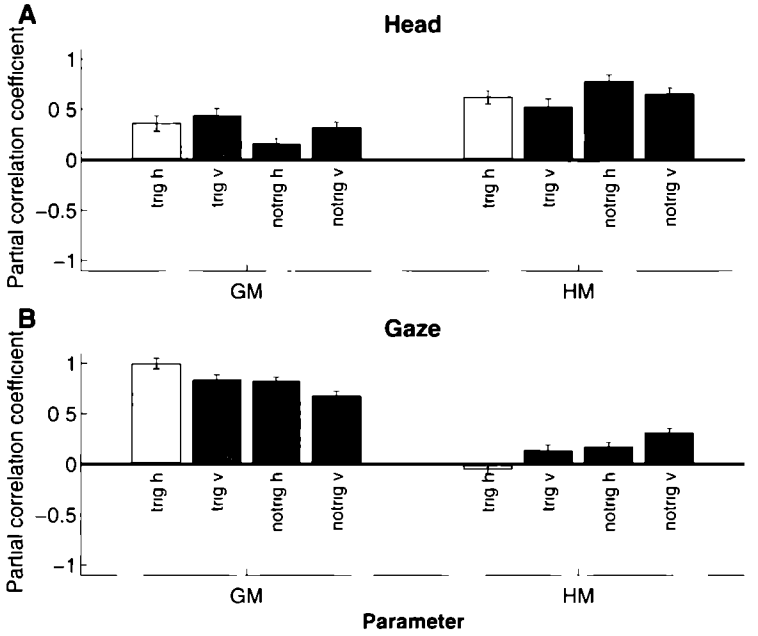


Figure 4.9: *Partial correlation coefficients for the regression on the second head saccade (ΔH_2)* **(A)** *and the second gaze saccade (ΔG_2)* **(B)**, *which are described as a function of the gaze (GM) and head motor errors (HM) (Eqns 4 3 and 4 4) The different gray-coded bars represent the two different conditions (dynamic/static) and response components (azimuth/elevation) Data are pooled across all subjects and recording sessions Note that eye and head are mainly driven by motor commands expressed in their own reference frames*

version (visual predictive) we used the initial retinal error vector for the first gaze shift, FV, as the predictive signal for remapping (indicated by the VP model in Fig. 4.1A). In the second version (motor predictive), we instead took the actual first gaze displacement (ΔG_1) to update the auditory target (the MP model in Fig. 4.1A). This leads to the following two predictive remapping models:

$$\Delta G_2 = a \cdot T_{H\,ini} + b \cdot FV + c \tag{4.5}$$

$$\Delta G_2 = a \cdot T_{H\,ini} + b \cdot \Delta G_1 + c \tag{4.6}$$

Note that Eqns. 4.1 and 4.6 predict the same gaze shift for the non-triggered double-step experiment if the first gaze shift is fully accounted for (i.e. when $b = c = -1$ in Eqn. 4.1). Also, when the first gaze shift brings the eye close to the extinguished visual target location, vectors FV and ΔG_1 will be very similar,

as will Eqns 4 5 and 4 6 (Fig 4 1A) However, if the first gaze shift misses the visual target location, Eqns 4 5 and 4 6 yield different predictions

For the triggered double-step experiments, the head-centered auditory target coordinates were taken relative to the position of the head in space at sound onset (see Fig 4 1B) The head-displacement signal for the model of Eqn 4 1 is then given by the subsequent displacement after sound onset Note, however, that for the predictive remapping schemes the preprogrammed signals in Eqns 4 5 and 4 6 are the same for the non-triggered and triggered double-step conditions, since they relate to information about the first gaze-displacement before it was actually generated

Figure 4 10 shows the predicted gaze displacement, ΔG_2, for each of the four models, plotted against the measured gaze shift for the azimuth and elevation response components (pooled for all subjects and sessions), together with the R^2 values Figure 4 10A shows the results for the non-triggered double-step conditions, while Figure 4 10B gives the predictions for the triggered double-steps As expected, the non-compensation model (left column) does not yield a good description of the measured data for neither double-step condition The predictive remapping model based on retinal error (visual predictive, Eqn 4 5, second column) performs slightly better, but is clearly inferior to the predictive remapping model that is based on the actually programmed first gaze shift (motor predictive, Eqn 4 6, third column) In the non-triggered condition performance of the motor-predictive model is equal to the dynamic feedback model (right-hand column in Fig 4 10A) In the triggered double-step condition however, the motor predictive model bases the updated craniocentric target location on the pre-programmed, full, first gaze shift, whereas the dynamic feedback hypothesis updates the craniocentric target location with the partial gaze shift following the auditory target presentation (Fig 4 1) In this condition the dynamic feedback model provides the best prediction of the measurements (Fig 4 10B)

Short- vs. Long-Duration Sounds Recent experiments have indicated that the auditory system needs a minimum duration (about 20 to 40 ms) of broadband input to build a stable percept of sound-source elevation For shorter sound durations the elevation gain decreases systematically with either decreasing stimulus duration (Hofman and Van Opstal, 1998, Vliegen and Van Opstal, 2004), or increasing stimulus level (Macpherson and Middlebrooks, 2000, Vliegen and Van Opstal, 2004) The former phenomenon was proposed to be due to a neural integration process that improves its elevation estimate by accumulating spectral evidence about the current HRTF through consecutive short-term (few ms) "looks" at the acoustic input

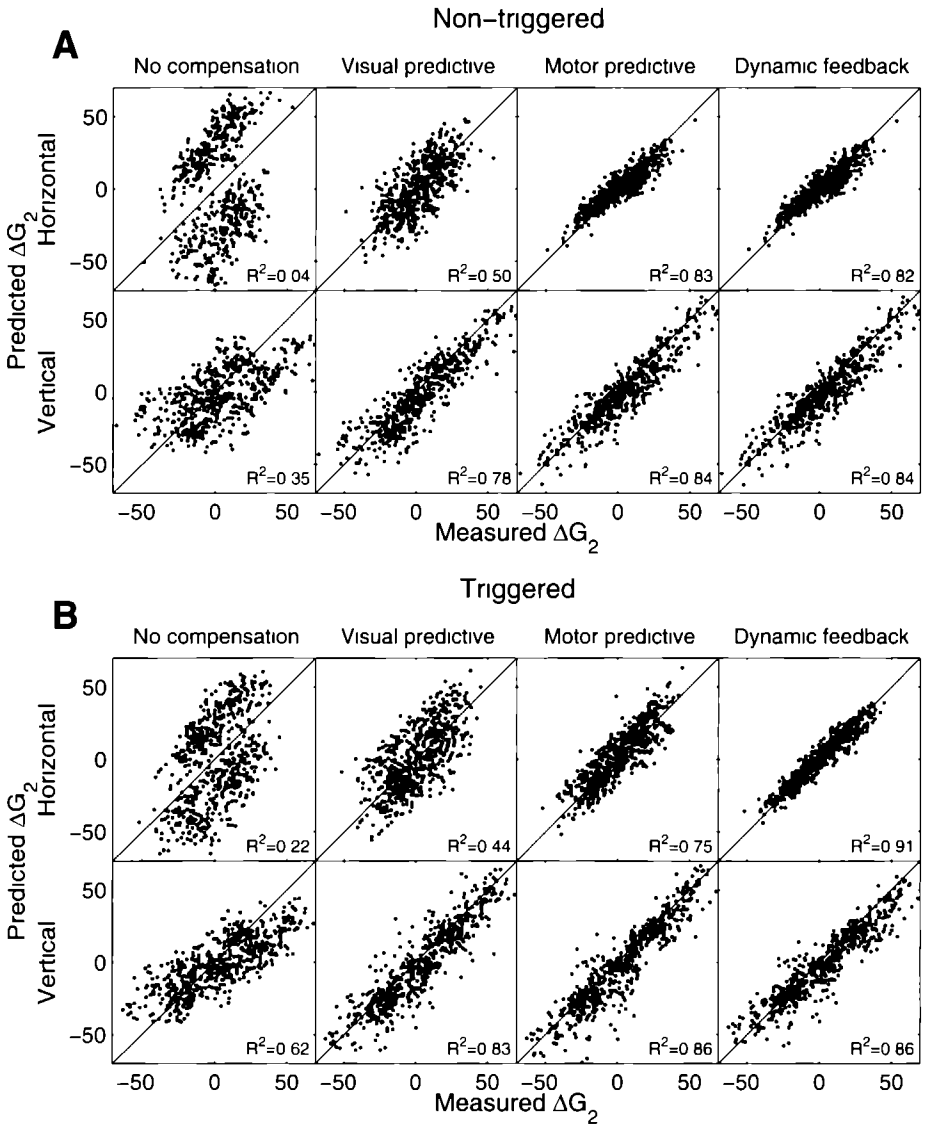


Figure 4 10 *Predicted auditory-evoked gaze shifts (ΔG_2 ordinate) for the four models described in the text plotted against measured responses (abscissa) Data are pooled across subjects and recording sessions* **(A)** *Static double-step condition for horizontal (top row) and vertical (bottom row) response components* **(B)** *Dynamic double-step condition for both response components If the model would predict ΔG_2 perfectly, data points would fall on the unity line, and R^2 would be 1 R^2 values are given in the lower right corner of all panels The predictions of the dynamic feedback model are superior to the other models*

So far, experiments that have studied the influence of sound duration have been performed with a stationary head during stimulus presentation. Since high-velocity ($> 200°/s$) 2D head movements sweep the acoustic input across a multitude of different HRTFs on a short time scale, it is conceivable that the resulting dynamic changes in spectral input could interfere with the integrity of the neural integration process.

Suppose, however, that self-generated head movements would somehow enhance the performance of the short-term cue-extracting mechanisms Accurate localization of elevation during rapid eye-head movements could then also be explained by a strategy that incorporates only a brief portion of the sound, say the first few ms, while bypassing the neural integration stage. If true, short stimuli (<10 ms) should be localized better when presented during rapid head movements than without head movements. Moreover, there should be no benefit of longer stimulus durations during head movements.

To test these predictions we repeated the single-step, and static and dynamic double-step experiments with four subjects by presenting very short (3 ms) and longer (50 ms) acoustic stimuli (randomly interleaved across trials; late-triggered conditions only). Figure 4.11 summarizes the results as cumulative error distributions for the elevation response components for the two different stimulus durations (short: solid lines; long: lines through symbols) and three spatial-temporal target configurations (different gray codes: single-step. black; static double step: dark-gray; dynamic double step: light-gray). The figure shows that localization performance is quite comparable for the three conditions (single step, static and dynamic double steps), although the single-step trials yielded slightly more accurate responses than the two double-step conditions ($p < 0.05$, KS-test). Thus, the self-generated head movements for short- and long-duration stimuli did clearly not enhance localization performance in the double-step experiments.

More importantly, however, for all three conditions the short stimuli yielded larger localization errors than the longer stimuli (KS-statistic for 3 ms vs. 50 ms data: single step· $p = 0.008$, $d = 0\ 21$; non-triggered double steps: $p = 0.02$, $d = 0.20$, triggered double steps: $p = 0.002$, $d = 0.25$). The median differences in absolute response errors were 2.1° for the single-step data, 4.0° for the non-triggered condition and 4.3° for the triggered condition (indicated by arrowheads). The differences were negligible for the azimuth response components for all six conditions ($p > 0.05$, not shown).

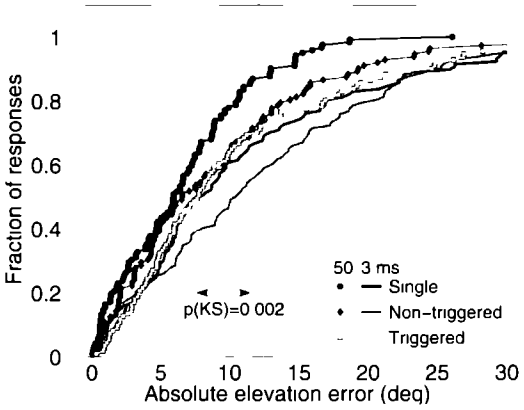


Figure 4.11: *Comparison of gaze-elevation responses for single-step, static and dynamic double-step conditions for very brief (3 ms) and longer (50 ms) sound durations Data are pooled for all four subjects (JO, JV, MW, RK), and ranked, to create cumulative distributions of absolute elevation response-errors for each of the three conditions and two stimulus durations (see legend for explanation of symbols and line styles) Note the larger errors in the 3 ms data, when compared to the 50 ms data for all three conditions ($p < 0\ 025$, KS-test) Thus, responses to the 50 ms stimuli are more accurate than to the 3 ms stimuli The result for the triggered double-step responses is highlighted ($p = 0\ 002$) Horizontal dashed line at 50 median of the response distributions For both stimulus durations, responses to the triggered and non-triggered double-steps are slightly more inaccurate than to the single-step targets ($p < 0\ 05$)*

4.4 Discussion

Summary Our results show that eye-head localization responses to brief acoustic noise bursts are equally accurate under fundamentally different stimulation conditions (Fig. 4.7, Tab. 4.1). In the (open-loop) dynamic double-step experiment, eye and head position, as well as the acoustic localization cues varied at a high and variable speed during sound presentation (Fig. 4.5). Nevertheless, the intervening eye and head movements were, on average, fully compensated under all conditions tested (Fig 4.8). Considering the complexities of the underlying coordinate transformations, this is quite a remarkable result. We also found that for all localization conditions, eyes and head both made goal-directed movements toward the sound (Fig. 4.9), which further strengthens the idea that they are each driven by motor commands in their own appropriate reference frame, rather than by a common signal. Finally, improved response accuracy for longer sound stimuli is preserved during rapid head movements (Fig 4.11).

Comparison to other studies In a recent study, Kopinska and Harris (2003) measured the lateralization percept of dichotic auditory targets. First, a sound was presented with head and body upright and eyes looking straight ahead. Then, subjects reproduced the memorized sound location in the head, by adjusting the binaural level difference after adopting a new horizontal eye, head or body posture. Neither eye position in the head, nor whole-body orientation in space affected localization judgments. In contrast, head orientation on the body, and body orientation with the head fixed, had a small but systematic effect on response accuracy. The authors concluded that acoustic stimuli are expressed in a body-centered frame of reference.

Lewald and Ehrenstein (1998) and Lewald et al. (2000) came to a similar conclusion, when reporting small shifts in the localization of sounds after changes in horizontal head orientation.

Lewald and colleagues (1998, 2000) also reported systematic shifts in the perceived midline after changes in eye position. However, in dichotic (Kopinska and Harris, 1998) and free-field localization tasks (Goossens and Van Opstal, 1999; this study) effects of eye position were absent.

The studies by Kopinska and Harris (1998) and Lewald and colleagues (1998, 2000) all suggested that the sound-localization errors were caused by an inaccurate representation of the head-on-body signal. Because these experiments did not vary body and head orientation with respect to gravity, the static posture changes did not result in a tonic stimulation of the otoliths. In contrast, Goossens and Van Opstal (1999) found that saccadic eye movements made to free-field noise bursts were not systematically affected by static changes in vertical head tilt. The present data further extend these findings to dynamic localization conditions.

Note that the localization responses to pure tones under static head tilts did vary with head orientation. However, this effect was shown to depend strongly on the sound's frequency, which suggested that a signal about head orientation is incorporated at a level where acoustic input is still tonotopically represented (Goossens and Van Opstal, 1999).

Lewald et al. (2000) observed systematic undershoots for horizontal head pointing to free-field sounds. However, since they did not measure eye position, it is possible that their subjects actually pointed with their eyes, even when asked to point with their nose. Indeed, undershoots disappeared when subjects used a visual reference.

An important difference between our study and the previous studies resides in immediate, reflexive open-loop gaze orienting to brief sounds in our experiments, versus voluntary, perceptual and closed-loop localization tasks to long-duration

stimuli in the other studies. It is conceivable that the mechanisms underlying action (rapid orienting) and perception (voluntary judgments) employ different computational strategies, weightings, and neural pathways to update the frames of reference (see also below).

Note that because the subject's body was stationary in our experiments, we cannot distinguish body-centered from world-coordinate representations. Yet, we expect that rapid sound localization behavior will compensate for intervening changes in body orientation too.

Implications for models Single-step localization performance can be explained by at least three different mechanisms (Fig. 4.1). The first model does not account for the intervening gaze shift under double-step conditions, and can be readily dismissed because of the static double-step results. The latter condition still allows alternative possibilities to explain accurate localization behavior.

The predictive remapping interpretation is inspired by the neurophysiology underlying visuomotor behavior, and was proposed by Goldberg and colleagues. Their studies convincingly demonstrated predictive visual responses in neurons within posterior parietal cortex (Duhamel et al.,1992; Colby et al., 1995), frontal eye fields (Umeno and Goldberg, 1997), and superior colliculus (Walker et al., 1995). This activity preceded the occurrence of a saccade that would, after its completion, bring the stimulus into the cell's visual receptive field. These predictive visual responses can be regarded to update the retinal coordinates of visual stimuli through an impending eye movement (efference copy) or, alternatively, by the retinal stimulus location evoking that eye movement. Predictive transformations could underlie the percept of a stable visual environment despite intervening saccades (transsaccadic integration), but could also explain the fast and accurate programming of subsequent eye movements in e.g. open-loop double-steps or remembered target sequences.

Here, we propose that a similar mechanism might be used for the sound-localization system. During head movements, the system should be able to dissociate changes in acoustic cues due to self-motion, from those that arise as a result of target motion. Furthermore, the localization percept of sound-sources needs to incorporate changes in head orientation to maintain spatial constancy and accuracy. A predictive mechanism that remaps perceived sound locations by impending head movements (or, alternatively, prior sound-source locations) could thus underlie the percept of a stable acoustic environment.

An alternative explanation for accurate double-steps to visual or auditory targets, however, is that the target location is continuously updated, rather than beforehand. In this proposal, the target is mapped into a reference frame in world

coordinates as soon as stimulus information becomes available, and is kept in memory for as long as this information is required (Fig. 4.12). This transformation requires, in its simplest form, dynamic feedback about absolute eye position in the head, about head orientation on the body, and body orientation in space, rather than about impending displacement signals.

The predictive models and the dynamic feedback mechanism yield near-identical predictions for the static double-step trials (Figs. 4.1A, 4.10A). However, in the dynamic double-step paradigm, these schemes predict quite different updated target locations (Fig. 4.1B) Hallett and Lightstone (1976) showed that saccadic eye movements toward visual targets, flashed in mid-flight during an intervening saccade, remain accurate. While the predictive schemes yield an updated target location on the basis of wrong gaze-displacement information (Fig. 4.1B), only the dynamic feedback scheme predicts such accurate localization behavior. This is indeed the result of our sound-localization experiments (Fig. 4.10B)

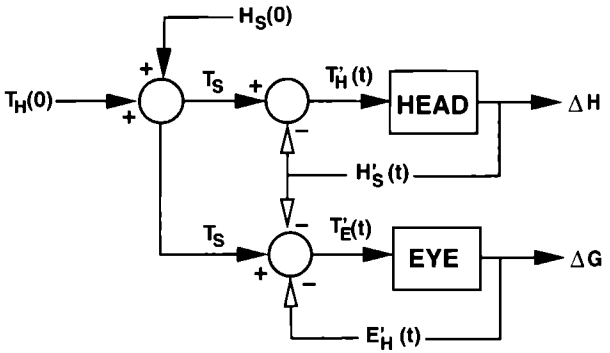


Figure 4.12: *Conceptual scheme underlying accurate dynamic gaze control toward acoustic targets Two stages are discerned First, head-centered target location at target onset, $T_H(0)$, is added to head orientation in space at sound onset, $H_S(0)$, to create a stable representation of the sound in world coordinates, TS This target representation is kept in memory until a new target is selected In the second stage, adequate motor commands for eyes and head are generated, by remapping the sound into head-centered ($T'_H(t)$) and eye-centered ($T'_I(t)$) target locations, respectively This latter process requires signals about instantaneous eye-position in the head, $E'_H(t)$, and about head orientation in space, $H'_S(t)$*

Interestingly, responses for early-triggered and late-triggered double-steps were equally accurate. This finding contrasts with experiments from the visuomotor literature showing that visual stimuli, presented around the onset of a saccadic eye movement, are systematically miss-localized (Dassonville et al., 1995; Schlag and Schlag-Rey, 2002).

Our data suggest that although predictive remapping could underlie the pro-

cess of transsaccadic integration to form a stable spatial percept of the (visual and acoustic) environment, it is not adequate to update the target coordinates under dynamic spatial orienting tasks Thus, we propose that the mechanisms underlying transsaccadic integration (presumably subserving spatial perception of the sensory scene) and target updating (dedicated to goal-directed actions toward specific stimuli) are quite different Whereas the former may rely on an upcoming motor command, the latter needs to account for instantaneous changes in eye and head orientation This is in agreement with recent visuomotor studies that show an effect of saccadic adaptation on the perceived target location, but not on the actual eye saccades toward the target (Bahcall and Kowler, 1999, Tanaka, 2003)

Dynamic localization cues Perrett and Noble (1997a,b) observed that in the absence of pinna cues elevation localization improved with horizontal head movements, provided low frequencies were present in the signal They suggested that the system uses dynamic changes in binaural timing differences to localize sound-source elevation Other studies showed that slow head movements (Hofman et al , 2002) and self-induced stimulus motion (Wightman and Kistler, 1999) for long-duration broadband sounds can resolve front-back confusions In contrast, Goossens and Van Opstal (1999) reported that rapid 2D head movements did neither improve nor deteriorate the localization of pure tones with a duration of 500 ms, despite the systematic head-movement related changes in sound level that result when the tone sweeps across the different HRTFs Thus, head movements are beneficial for some localization conditions, but not for others

Fast head movements during broadband sounds may also pose potential problems to the auditory system, because of the resulting rapid variations of spectral localization cues These changes could interfere with the need to improve the elevation estimate through the integration of multiple short-term "looks" of the otherwise stable stimulus spectrum

The results of the stimulus-duration experiments (Fig 4 11) show, however, that there was neither an advantage nor a disadvantage of head movements, as there was no difference between static and dynamic conditions Moreover, the results indicate that the neural integration process also functions during fast head movements, since elevation performance was significantly better for long vs short stimulus durations

Taken together, the data from the current experiments support the possibility that the high-velocity and variable changes in head orientation are already incorporated at the stage of dynamic neural integration, and that the outcome of

this process could be the target in world-centered, rather than in head-centered, coordinates.

Acknowledgments

We thank G. Van Lingen, H. Kleijnen, G. Windau and T. Van Dreumel for technical assistance. This research was supported by the Radboud University Nijmegen (AJVO, TVG), and the Netherlands Organization for Scientific Research (NWO - section Maatschappij- en gedragswetenschappen, MaGW, project nr. 410-20-301; JV).

Appendix

Supplementary tables

Table 4.2 provides the linear regression results of Eqn. 4.1 on the *gaze-displacement* data from the five individual subjects who participated in the experiments (JO, JV, TG, RK, MW), as well as for the pooled data.

a = gain (dimensionless) for the initial sound location relative to the head (measured at the time of sound onset)
b = gain for the head displacement following sound onset
c = gain for the eye position in the head at the start of the second gaze shift
d = offset (in deg)
N = number of data points
R = Pearson's linear correlation coefficient between fit and data
Standard deviations in the parameters were obtained by bootstrapping the data 100 times (see Methods).

Ra = partial correlation coefficient for parameter a
Rb = partial correlation coefficient for parameter b
Rc = partial correlation coefficient for parameter c
Rd = partial correlation coefficient for parameter d

h = horizontal response components (azimuth)
v = vertical response components (elevation)
tr = triggered double-step condition
nt = non-triggered double-step condition

Table 4.3 provides the linear regression results of Eqn. 4.2 on the *head-movement* data from the same subjects, as well as on the pooled data. Same conventions as in Table 4.2.

subj	trig	h/v	a	Ra	b	Rb	c	Rc	d	Rd	R	N
jo	tr	h	1 06 ± 0 02	0 96	−1 01 ± 0 02	−0 96	−1 18 ± 0 07	−0 80	−1 13 ± 0 39	−0 21	0 96	178
		v	0 89 ± 0 03	0 91	−1 21 ± 0 06	−0 83	−1 00 ± 0 09	−0 64	2 14 ± 0 72	0 22	0 97	178
	nt	h	1 19 ± 0 04	0 90	−1 13 ± 0 04	−0 90	−0 97 ± 0 08	−0 67	−0 33 ± 0 47	−0 06	0 90	164
		v	0 79 ± 0 04	0 87	−1 01 ± 0 05	−0 87	−0 59 ± 0 12	−0 36	3 70 ± 0 85	0 33	0 95	164
jv	tr	h	1 09 ± 0 02	0 98	−0 94 ± 0 02	−0 97	−1 06 ± 0 06	−0 87	−0 33 ± 0 34	−0 09	0 98	121
		v	1 25 ± 0 04	0 94	−1 45 ± 0 17	−0 63	−1 06 ± 0 12	−0 63	4 05 ± 0 88	0 39	0 95	121
	nt	h	1 23 ± 0 05	0 94	−1 25 ± 0 05	−0 94	−1 02 ± 0 08	−0 80	0 22 ± 0 58	0 04	0 93	92
		v	1 36 ± 0 09	0 86	−1 30 ± 0 14	−0 69	−0 95 ± 0 17	−0 52	6 23 ± 1 38	0 43	0 87	92
tg	tr	h	0 94 ± 0 03	0 95	−0 83 ± 0 03	−0 90	−1 06 ± 0 07	−0 80	0 92 ± 0 50	0 16	0 96	130
		v	0 99 ± 0 06	0 80	−0 61 ± 0 12	−0 42	−0 58 ± 0 13	−0 38	4 65 ± 1 23	0 32	0 83	130
	nt	h	1 03 ± 0 03	0 95	−0 97 ± 0 04	−0 92	−0 97 ± 0 07	−0 78	−0 49 ± 0 67	−0 06	0 94	131
		v	0 93 ± 0 05	0 87	−0 84 ± 0 06	−0 80	−0 56 ± 0 10	−0 44	5 07 ± 1 06	0 39	0 88	131
rk	tr	h	0 89 ± 0 03	0 97	−0 88 ± 0 03	−0 97	−0 88 ± 0 08	−0 84	1 16 ± 0 45	0 33	0 97	59
		v	0 92 ± 0 04	0 94	−0 37 ± 0 42	−0 12	−1 03 ± 0 18	−0 60	4 05 ± 0 94	0 50	0 96	59
	nt	h	1 08 ± 0 06	0 93	−1 20 ± 0 06	−0 94	−1 01 ± 0 15	−0 69	1 20 ± 0 71	0 23	0 95	53
		v	0 95 ± 0 05	0 94	−0 39 ± 0 23	−0 24	−0 92 ± 0 09	−0 81	4 71 ± 0 69	0 70	0 97	53
mw	tr	h	1 07 ± 0 03	0 97	−1 09 ± 0 05	−0 91	−0 89 ± 0 13	−0 56	2 82 ± 0 62	0 41	0 96	105
		v	0 94 ± 0 06	0 85	−1 46 ± 0 09	−0 85	−1 02 ± 0 11	−0 67	11 29 ± 1 21	0 68	0 97	105
	nt	h	1 28 ± 0 04	0 95	−1 28 ± 0 06	−0 91	−1 20 ± 0 13	−0 68	4 23 ± 0 63	0 56	0 93	101
		v	1 18 ± 0 05	0 93	−1 09 ± 0 05	−0 92	−1 19 ± 0 10	−0 76	8 00 ± 1 07	0 60	0 97	101
all	tr	h	1 02 ± 0 01	0 96	−0 96 ± 0 01	−0 95	−1 05 ± 0 03	−0 80	0 38 ± 0 22	0 07	0 96	593
		v	1 06 ± 0 02	0 90	−1 09 ± 0 06	−0 62	−0 87 ± 0 05	−0 60	4 83 ± 0 41	0 44	0 93	593
	nt	h	1 12 ± 0 02	0 93	−1 09 ± 0 02	−0 92	−0 99 ± 0 04	−0 76	0 47 ± 0 29	0 07	0 91	541
		v	1 02 ± 0 02	0 89	−1 00 ± 0 03	−0 82	−0 75 ± 0 05	−0 57	4 52 ± 0 39	0 44	0 92	541

Table 4 2 *Gaze displacement* $\Delta G_2 = a\ T_{H\,ml} + b\quad \Delta H_1 + c\ E_0 + d$

subj	trig	h/v	a	Ra	b	Rb	d	Rd	R	N
jo	tr	h	1 10 ± 0 03	0 94	−1 11 ± 0 03	−0 94	−1 33 ± 0 54	−0 18	0 94	178
		v	0 77 ± 0 03	0 90	−1 29 ± 0 05	−0 89	8 42 ± 0 39	0 85	0 96	178
	nt	h	1 31 ± 0 04	0 92	−1 27 ± 0 04	−0 91	−1 01 ± 0 55	−0 14	0 90	164
		v	0 68 ± 0 03	0 85	−1 00 ± 0 04	−0 89	7 88 ± 0 50	0 78	0 93	164
jv	tr	h	0 92 ± 0 03	0 95	−0 87 ± 0 03	−0 93	−1 85 ± 0 53	−0 31	0 94	121
		v	0 71 ± 0 04	0 87	−1 08 ± 0 11	−0 66	5 29 ± 0 82	0 51	0 89	121
	nt	h	1 09 ± 0 06	0 89	−1 22 ± 0 06	−0 91	−0 82 ± 0 67	−0 13	0 88	92
		v	0 92 ± 0 07	0 80	−1 07 ± 0 12	−0 70	6 89 ± 1 02	0 58	0 84	92
tg	tr	h	0 80 ± 0 02	0 95	−0 70 ± 0 03	−0 89	−3 40 ± 0 45	−0 56	0 94	130
		v	0 78 ± 0 05	0 82	−0 80 ± 0 10	−0 57	2 92 ± 0 91	0 27	0 86	130
	nt	h	0 89 ± 0 03	0 94	−0 85 ± 0 04	−0 90	−4 19 ± 0 60	−0 53	0 93	131
		v	0 70 ± 0 04	0 84	−0 84 ± 0 04	−0 86	3 11 ± 0 78	0 33	0 89	131
rk	tr	h	0 74 ± 0 04	0 91	−0 75 ± 0 04	−0 92	4 51 ± 0 62	0 70	0 92	59
		v	0 59 ± 0 03	0 92	0 13 ± 0 31	0 06	2 08 ± 0 79	0 33	0 93	59
	nt	h	0 90 ± 0 05	0 92	−1 01 ± 0 04	−0 96	5 04 ± 0 77	0 68	0 94	53
		v	0 61 ± 0 05	0 85	−0 24 ± 0 30	−0 11	1 58 ± 0 68	0 31	0 93	53
mw	tr	h	0 73 ± 0 03	0 90	−0 69 ± 0 05	−0 79	1 64 ± 0 53	0 29	0 93	105
		v	0 56 ± 0 06	0 71	−1 44 ± 0 09	−0 85	0 81 ± 0 78	0 10	0 96	105
	nt	h	0 91 ± 0 05	0 88	−0 88 ± 0 07	−0 81	3 05 ± 0 64	0 44	0 90	101
		v	0 84 ± 0 04	0 89	−0 95 ± 0 04	−0 92	1 65 ± 0 77	−0 21	0 95	101
all	tr	h	0 87 ± 0 02	0 89	−0 85 ± 0 02	−0 86	−0 51 ± 0 29	−0 07	0 91	593
		v	0 69 ± 0 02	0 84	−1 20 ± 0 05	−0 71	4 01 ± 0 35	0 42	0 91	593
	nt	h	0 97 ± 0 02	0 86	−0 96 ± 0 03	−0 84	−0 09 ± 0 33	−0 01	0 87	541
		v	0 73 ± 0 02	0 85	−0 93 ± 0 03	−0 84	3 99 ± 0 36	0 43	0 90	541

Table 4 3 *Head displacement* $\Delta H_2 = a\ T_{H\,ini} + b\quad \Delta H_1 + d$

Tables 4 4 and 4 5 provide the normalized linear regression results of Eqns 4 3 and 4 4 on the head and gaze-displacement data respectively, from three individual subjects (JO, JV, TG), as well as on the pooled data of all subjects The subjects were selected on the basis of a sufficient number of data points that satisfied both the eye position offset criterion ($> 10°$) and target direction re eye and head criterion (i e the angle between eye and head vectors $> 15°$) As a result, subjects RK and MW did not yield enough data points to perform meaningful regressions on both motor responses

$p =$ normalized gain (dimensionless z-score) for the head motor error at the start of the second gaze shift
$q =$ normalized gain for gaze motor error at the start of the second gaze shift
$N =$ number of included data points for which eye eccentricity after the first gaze shift exceeds $10°$
$R =$ Pearson's linear correlation coefficient between fit and data
Standard deviations in the parameters were obtained by bootstrapping the data 100 times

$Rp =$ partial correlation coefficient for parameter p
$Rq =$ partial correlation coefficient for parameter q

$h =$ horizontal response components (azimuth)
$v =$ vertical response components (elevation)
$tr =$ triggered double-step condition
$nt =$ non-triggered double-step condition

subj	trig	h/v	p	Rp	q	Rq	R	N
jo	tr	h	0 55 ± 0 74	0 16	0 42 ± 0 73	0 12	0 97	24
		v	0 96 ± 0 41	0 37	0 00 ± 0 41	0 00	0 96	36
	nt	h	−0 16 ± 0 07	−0 40	0 98 ± 0 08	0 93	0 93	26
		v	−0 32 ± 0 57	−0 11	1 25 ± 0 58	0 40	0 93	27
jv	tr	h	0 26 ± 0 13	0 38	0 72 ± 0 11	0 79	0 92	26
		v	0 38 ± 0 10	0 57	0 59 ± 0 10	0 73	0 89	31
	nt	h	0 39 ± 0 13	0 60	0 65 ± 0 12	0 80	0 88	19
		v	0 10 ± 0 12	0 21	0 81 ± 0 15	0 81	0 86	18
tg	tr	h	0 33 ± 0 30	0 22	0 63 ± 0 32	0 37	0 94	25
		v	0 17 ± 0 12	0 23	0 75 ± 0 15	0 64	0 89	38
	nt	h	0 35 ± 0 19	0 33	0 63 ± 0 18	0 55	0 95	29
		v	0 13 ± 0 10	0 19	0 80 ± 0 11	0 75	0 89	43
all	tr	h	0 36 ± 0 07	0 47	0 62 ± 0 06	0 73	0 91	84
		v	0 44 ± 0 07	0 47	0 53 ± 0 08	0 48	0 93	135
	nt	h	0 16 ± 0 05	0 30	0 78 ± 0 06	0 79	0 89	97
		v	0 32 ± 0 06	0 46	0 65 ± 0 06	0 70	0 93	123

Table 4 4 *Head Displacement* $\Delta H_2' = p\ GM' + q\ HM'$ *Note that* $q > p$, *indicating that the head is predominantly driven by the head motor error*

subj	trig	h/v	p	Rp	q	Rq	R	N
jo	tr	h	1 11 ± 0 48	0 45	−0 13 ± 0 46	−0 06	0 98	24
		v	1 31 ± 0 42	0 47	−0 35 ± 0 43	−0 14	0 96	36
	nt	h	0 61 ± 0 09	0 82	0 46 ± 0 10	0 68	0 88	26
		v	−0 20 ± 0 34	−0 12	1 17 ± 0 33	0 58	0 97	27
jv	tr	h	0 90 ± 0 09	0 90	0 10 ± 0 08	0 25	0 97	26
		v	0 77 ± 0 08	0 87	0 23 ± 0 08	0 49	0 94	31
	nt	h	0 90 ± 0 12	0 88	0 09 ± 0 08	0 25	0 94	19
		v	0 21 ± 0 22	0 24	0 68 ± 0 24	0 58	0 80	18
tg	tr	h	1 12 ± 0 24	0 69	−0 18 ± 0 27	−0 14	0 94	25
		v	0 89 ± 0 15	0 70	0 01 ± 0 16	0 02	0 90	38
	nt	h	0 88 ± 0 15	0 74	0 09 ± 0 14	0 12	0 96	29
		v	0 63 ± 0 13	0 60	0 32 ± 0 11	0 42	0 91	43
all	tr	h	1 00 ± 0 05	0 90	−0 05 ± 0 04	−0 12	0 96	84
		v	0 84 ± 0 05	0 81	0 13 ± 0 06	0 20	0 95	135
	nt	h	0 82 ± 0 04	0 89	0 17 ± 0 04	0 37	0 93	97
		v	0 67 ± 0 05	0 77	0 31 ± 0 05	0 52	0 95	123

Table 4 5 *Gaze Displacement* $\Delta G_2' = p\ GM' + q\ HM'$ *Note that* $p > q$, *indicating that gaze (= eye position in space) is predominantly driven by the gaze motor error (= the oculocentric error of the sound)*

Chapter 5

Gaze orienting in dynamic visual double steps

Visual stimuli are initially represented in a retinotopic reference frame. To maintain spatial accuracy of gaze (i.e. eye in space) despite intervening eye and head movements, the visual input could be combined with dynamic feedback about ongoing gaze shifts. Alternatively, target coordinates could be updated in advance by using the *preprogrammed* gaze-motor command ("predictive remapping") So far, previous experiments have not dissociated these possibilities

Here we study whether the visuomotor system accounts for saccadic eye-head movements that occur during target presentation. In this case, the system has to deal with fast dynamic changes of the retinal input, and with highly variable changes in relative eye and head movements that cannot be preprogrammed by the gaze control system. We performed visual-visual double-step experiments in which a brief (50 ms) stimulus was presented during a saccadic eye-head gaze shift toward a previously flashed visual target Our results show that gaze shifts remain accurate under these dynamic conditions, even for stimuli presented near saccade onset, and that eyes and head are driven in oculocentric and craniocentric coordinates, respectively.

These results cannot be explained by a predictive remapping scheme We propose that the visuomotor system adequately processes dynamic changes in visual input that result from self-initiated gaze shifts, to construct a stable representation of visual targets in an absolute, supraretinal (e.g. world) reference frame. Predictive remapping may subserve transsaccadic integration, thus enabling perception of a stable visual scene despite eye movements, while dynamic feedback ensures accurate actions (e.g. eye-head orienting) to a selected goal

---

Adapted from. Vliegen J, Van Grootel TJ, and Van Opstal AJ (2005) *Gaze orienting in dynamic visual double steps* **J Neurophysiol** 94 4300-4313

5.1 Introduction

This paper concerns the transformations underlying the programming of two-dimensional (2D) head-free gaze shifts to visual targets. Gaze is the orientation of the visual axis in space, defined by the sum of the orientations of the eye in the head and the head in space.

In studies of the gaze control system the typical situation is one in which eye and head orientations are initially aligned (exceptions are e.g. Volle and Guitton, 1993, Goossens and Van Opstal, 1997b, and Stahl, 2001). Under such conditions, there is a one-to-one correspondence between the retinal location of a briefly flashed visual stimulus and the motor commands for eyes and head to acquire the target. However, under more natural conditions, the eyes are not fixed in the head. Eyes and head may then not point in the same direction, and make different intervening movements prior to the orienting response. As illustrated in Figure 5.1, in such cases the initial retinal error ($T_{E\,0}$) no longer suffices as a valid motor command for eyes and head (e.g. Goossens and Van Opstal, 1997b). Instead, the correct motor errors ($T_{E\,2}$ and $T_{H\,2}$, respectively) require different transformations of the target location that incorporate both the eye-head misalignment and the intervening eye-head movements.

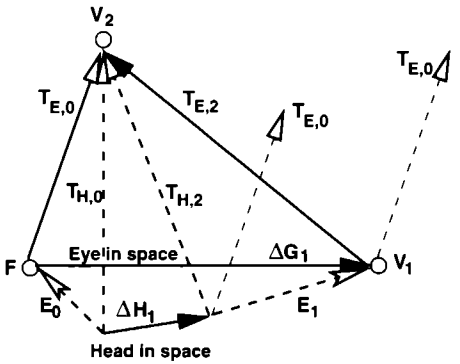


Figure 5.1: *In head-free gaze shifts, eye and head are typically unaligned, and make movements of different amplitudes, thus creating different error signals for a visual target The scheme shows a fixation point (F), a first visual stimulus (V_1) and a second visual target (V_2), both presented before the eye (ΔG_1) - head (ΔH_1) movement toward V_1 At the start of the trial, the eye-in-head position is E_0 The initial retinal error for V_2 is $T_{E\,0}$, but for the head it is different $T_{H\,0}$ After the first gaze shift, the eye-in-head position has changed (E_1), and the motor errors for eye ($T_{E\,2}$) and head ($T_{H\,2}$) are very different from the original retinal error, $T_{E\,0}$ They depend on the intervening eye-head gaze shift for the eye-in-space ($T_{E\,2} = T_{E\,0} - \Delta G_1$), and on the eye-head misalignment and head (or eye-head) movement for the head-in-space ($T_{H\,2} = T_{E\,0} + E_0 - \Delta H_1 = T_{E\,0} + E_1 - \Delta G_1$, respectively*

The gaze control system could implement these transformations in a variety of ways. For example, it could use preprogrammed (feedforward) information about the upcoming gaze shift. Alternatively, it could rely on continuous dynamic feedback about the actual movements of eyes and head.

To study these different transformations we have elicited eye-head saccades to visual stimuli that were briefly flashed *during* an intervening eye-head gaze shift

Static double-steps

Our paradigm contrasts with the classic saccade double-step experiment, which we here denote as the *static* double-step (Fig. 5.2A). In that experiment two peripheral targets are presented shortly after each other, but *before* the initiation of the first gaze shift. The subject is instructed to foveate both targets at the remembered spatial locations in the order of their appearance. The double-step paradigm has been used in a number of (head-fixed) saccadic eye-movement studies (Becker and Jürgens 1979; Ottes et al., 1984; Goossens and Van Opstal 1997a), that all showed that saccades toward the second target fully account for the size and direction of the first eye movement. Several theories to explain this result have been forwarded in the literature.

Position vs. displacement feedback

Neurophysiological studies have demonstrated that the primate visuomotor system also compensates for an intervening saccade evoked by microstimulation of the midbrain superior colliculus (SC; Mays and Sparks 1980; Sparks and Mays 1983). These studies suggested that the retinal location of the target is transformed into a head-centered reference frame, in which the system can readily program the future saccade by incorporating eye-in-head position (Van Gisbergen et al., 1981; Sparks and Mays 1983; Zipser and Andersen 1988). However, lack of evidence for a head-centered representation of visual targets, together with the idea that the SC represents saccades in an eye-centered, rather than in a head-centered motor map (Robinson 1972; Sparks and Mays 1983), has prompted others to propose an eye-displacement updating scheme to explain these results. In this scheme, the visuomotor system keeps targets in an eye-centered reference frame, while updating the saccade plan with feedback about intervening eye-displacement (Jürgens et al., 1981). Neurophysiological recordings in the primate frontal eye fields (FEF) provided support for this alternative model by demonstrating that FEF cells carry the signals required for this

transformation (Goldberg and Bruce 1990). Note that both schemes incorporate dynamic feedback of actual motor performance to update the future response.

Feedforward vs. feedback

More recent studies, however, have suggested that at different stages within the visuomotor pathway a coordinate transformation may already be performed *before* the initiation of the first saccade (so-called predictive remapping; posterior parietal cortex, PPC: Duhamel et al., 1992; Colby et al., 1995; FEF: Umeno and Goldberg 1997; SC: Walker et al., 1995). Such a feedforward mechanism could provide a neural correlate for trans-saccadic integration, which is thought to underlie the percept of a stable visual environment, despite saccadic eye movements that sweep the visual scene across the retina (Duhamel et al., 1992). However, such a mechanism could also underlie spatially accurate performance of saccades in the double-step paradigm

In Figure 5.2, we have outlined two slightly different versions of the predictive remapping idea: in the *visual-predictive* (VP) model the initial retinal target representation ($T_{E\,0}$) is updated on the basis of the retinal coordinates of the first visual target (FV_1), while the *motor-predictive* (MP) model relies on an efference copy of the planned first gaze shift (ΔG_1) Because the VP scheme is based on visual information only, it does not account for a possible localization error of the first target, in contrast to the MP model. This results in slightly different predictions for the response to the second target. So far, neurophysiological recordings do not allow a distinction between these two alternatives.

According to the feedback model (FB), the retinal error of the second target (T_L) is continuously updated with information about eye and head movements through *dynamic feedback* from the gaze control system. In this way, the motor errors that will drive the eye and head are always accurate.

The static double-step paradigm (Fig. 5.2A), in which both targets are presented before the eye-head movement onset, cannot dissociate updating schemes based on dynamic feedback from those based on predictive remapping. Although the VP model predicts a localization error that depends on the error for the first target, both the MP and the FB models predict equally accurate localization responses.

Dynamic double steps

In the present study we have applied a dynamic double-step paradigm, in which the second target is presented in mid-flight of the first, intervening gaze shift (ΔG_1, Fig. 5.2B). If predictive remapping would underlie the programming of

the future gaze shift, systematic errors are expected in this paradigm (VP, MP vectors), because in these models, the system is supposed to update the initial retinal input on the basis of prior (i.e. preprogrammed) information about the entire first gaze shift. According to the MP model, the target is missed by the difference between the full gaze shift and the partial movement following the onset of the second target, ΔG^*. For the VP model, this localization error corresponds to the difference between vector FV_1 and ΔG^*. The feedback model (FB), however, predicts accurate performance under all stimulus conditions.

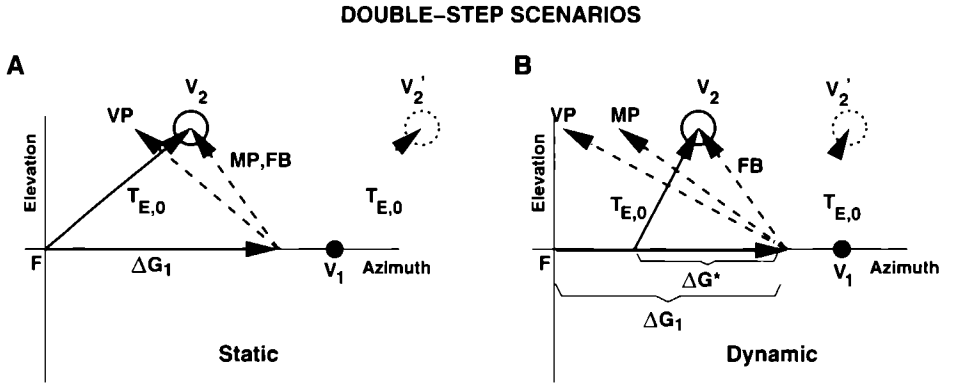


Figure 5.2: *Schematic outline of the two different double-step experiments in this study, showing a fixation point (F), two visual targets (V_1 and V_2), the first gaze shift (ΔG_1, which is here assumed to undershoot the first target), and the predicted second gaze shift according to three models The head movement is not shown, for clarity $T_{E,0}$ is the initial retinal error of V_2 V_2' is the retinal position of V_2 after ΔG_1* **(A)** *Static double step Targets are presented before the first gaze shift onset Vector VP represents the response of the visual predictive model (update based on vector FV_1) The motor predictive model (MP) and the dynamic feedback model (FB) both predict accurate localization (update based on actual ΔG_1)* **(B)** *Dynamic double step V_2 is presented during the first gaze shift, which dissociates the MP and FB models The MP model still uses the full first gaze shift to update the target, while the FB model incorporates only the partial gaze shift after the onset of the second target (ΔG^*)*

Related studies

Interestingly, Hallet and Lightstone (1976) applied a dynamic paradigm for horizontal eye movements, and reported accurate secondary saccades. However, later studies showed that both visual perception and oculomotor programming can be strongly influenced by concomitant saccades. For example, systematic inaccuracies in stimulus localization arise when dim targets are briefly flashed in darkness around the onset of a saccade (Dassonville et al., 1995; Schlag and Schlag-Rey 2002). These errors, which could be as large as 70% of the saccade

amplitude, were attributed to the use of a low-pass filtered representation of eye position in updating target locations (Schlag et al., 1989; Dassonville et al , 1992). According to this idea, the response endpoints in Figure 5.2B would systematically shift from V_2' (no compensation) to V_2 (full compensation) as a function of the stimulus timing relative to first-saccade onset.

Moreover, under photopic stimulus conditions saccades cause a considerable perceptual deformation of the visual field that leads to systematic misjudgments about stimulus locations (Bremmer and Krekelberg, 2003; Lappe et al., 2000; Ross et al., 1997).

These latter studies suggest that the observed mislocalizations could be due to visual-perceptual factors, rather than to a property of sensorimotor integration. Note, that an inaccurate representation of oculomotor feedback signals would also induce mislocalizations of non-visual stimuli, like e.g. sounds. We recently studied eye-head gaze shifts to brief sounds presented under static and dynamic visual-auditory double-steps, and showed that sound localization behavior remained accurate for all conditions (Vliegen et al., 2004). That study therefore suggested that the motor feedback signals are accurate.

The present paper extends these previous studies in a number of ways. First, with the notable exception of a few studies (Volle and Guitton, 1993; Ron et al., 1993, 1994; Goossens and Van Opstal 1997b), eye-head coordination performance in double steps has hardly been studied so far. Moreover, these studies were all confined to static double steps. As may be appreciated from the schematic outline in Figure 5.1, head-free gaze shifts in dynamic double steps poses far from trivial problems to the gaze-control system. Because the eyes and head move toward the initial visual target at highly variable and different velocities, at different relative latencies, and with different amplitudes, the eye and head motor errors at the time of stimulus presentation are essentially unpredictable. This property provides a serious challenge for gaze-control strategies based on predictive remapping.

Second, our experiments also relate to the ongoing dispute, described above, about absolute position coding of visual targets in an supraretinal reference frame (e.g. head-centered; Mays and Sparks 1980; Van Gisbergen et al., 1981; Sparks and Mays 1983; Zipser and Andersen 1988; or even body-centered; Andersen, 1997; Xing and Andersen, 2001; Kopinska and Harris, 2003), vs. relative displacement coding of target locations in an eye-centered reference frame (Jürgens et al., 1981; Goldberg and Bruce 1990; Duhamel et al., 1992; Colby et al., 1995; Nichols and Sparks 1995). Third, visual and auditory targets are initially encoded in different reference frames. As auditory targets are encoded in a head-centered reference frame, the sound-localization cues become dynamic as a result of

head movements. In contrast, visual targets are encoded in a retinotopic reference frame, inducing fast dynamic changes of the retinal input as a result of *eye* movements. Thus, for accurate movements of eyes and head to audio-visual stimuli, the modality-specific reference frames need to be updated dynamically on extremely short time scales, but with different transformation rules for eyes and head (Goossens and Van Opstal 1997b).

Our experiments show that 2D eye-head coordination is equally accurate under static and dynamic visual double steps, although we noted subtle differences with our recent auditory localization study. Further, we obtained no large systematic localization errors when the visual stimulus fell around the onset of the first saccadic gaze shift. We propose a target-updating scheme in which the programming stage of the gaze control system operates under continuous feedback, using accurate information about instantaneous eye and head movements. We discuss our findings in terms of current models of gaze control and target updating, and argue that our data are hard to reconcile with a pure displacement feedback scheme in which visual targets remain in an eye-centered reference frame.

5.2 Materials and Method

Subjects

Nine subjects (four females, five males; ages 20 to 46) participated in the experiments All had normal vision except for JO, who is amblyopic in his right, recorded, eye. All subjects had oculomotor and head-motor responses within the normal range. Subjects MK, ML, MV, MW, RK, and SP were kept naive about the exact goal of this study. The authors (JO, JV, TG) participated in all parts of the experiments. Subject RK participated only in the first target configuration (see below) and subjects MK, ML, MV, MW, and SP only in the second target configuration. Informed consent was obtained from all participants. Experiments adhered to the principles of the Declaration of Helsinki, and the US federal regulations for the Protection of Human Subjects.

Apparatus and stimuli

Experiments were conducted in a completely dark (< 0.001 cd/m^2, measured with a Minolta LS-100 Luminance meter), sound-attenuated room (3 x 3 x 3 m^3). The subject was seated comfortably on a chair in the center of the room with firm support in the back and lower neck. Viewing was binocular. The subject faced a thin-wire hemisphere with a radius of 0.85 m, the center of

which coincided with the center of the subject's head. On this hemisphere 85 red/green light-emitting diodes (LEDs; Knightbright Electronics, L59EGW/CA) were mounted at seven visual eccentricities $R = [0, 2, 5, 9, 14, 20, 27, 35]^\circ$ relative to the straight-ahead viewing direction (defined in polar coordinates as $[R.\phi] = [0, 0]^\circ$), and at twelve different directions, given by $\phi = [0, 30, ..., 330]^\circ$, where $\phi = 0^\circ$ is rightward from the center and $\phi = 90^\circ$ is upward. The hemisphere was covered with thin black silk.

Stimuli were delivered by the red LEDs (= 625 nm), that had a diameter of 2.5 mm, which corresponded to a viewing angle of 0.2° at the position of the subject's head The LEDs were powered with current pulses (frequency 150 Hz), and set at a luminance of approximately 2 cd/m^2.

Measurements

Head and eye movements were measured with the magnetic search-coil induction technique (Robinson, 1963). Subjects wore a lightweight helmet (about 150 g), consisting of a narrow strap above the ears, which could be adjusted to fit around the head, and a second strap that ran over the head. A small coil was mounted on the latter. Subjects also wore a scleral search coil on one of their eyes (Collewijn et al., 1975). In the room two orthogonal pairs of 3 x 3 m^2 square coils were attached to the walls, floor, and ceiling to create the horizontal (30 kHz) and vertical (40 kHz) oscillating magnetic fields that are required for this recording technique. Horizontal and vertical components of head and eye movements were detected by phase-lock amplifiers (Princeton Applied Research, models 128A and 120), low pass filtered (150 Hz), and sampled at 500 Hz per channel before being stored on disk.

A PC-486 was equipped with the hardware for data acquisition (Metrabyte DAS16), stimulus timing (Data Translation DT2817), and digital control of the LEDs (Philips I2C)

Experimental paradigms

Each experimental session started with three calibration runs to calibrate the eye and head coils (Goossens and Van Opstal 1997b). Before calibration, subjects were asked to keep their head in a comfortable straight-ahead position and adjust a dim red LED mounted at the end of a pliable rod that was attached to the helmet, such that it was aligned with the center LED of the hemisphere. This rod LED was only illuminated during the eye-in-head and head calibration sessions, and was extinguished during the actual experiments

First, eye position in space ("gaze") was determined by calibrating the eye coil. Subjects kept their head still in the straight-ahead position and refixated with their eyes the LED targets on the hemisphere. The targets (n = 37) were presented once, in a fixed counterclockwise order, at the center location ($R = 0$), followed by three different eccentricities, $R = [9, 20, 35]^\circ$, and all 12 directions. When subjects fixated the target, they pushed a button to start data acquisition, while keeping their eyes at the target location for at least 1000 ms.

In the second calibration run, the eye-in-head offset position was measured to account for the potential fixed misalignment of the eye and head coils. To that end, subjects fixated the rod LED to keep their eyes fixed in the head. Subjects were asked to assume the neutral, straight-ahead head position and push a button to start 1000 ms of data acquisition. This procedure was repeated 10 times.

The third calibration run served to calibrate the head orientation in space. Again, subjects were asked to fixate the rod LED with their eyes and to align it with the same 37 LED targets on the hemisphere as in the eye calibration run. In this way, the eyes remained at the same fixed offset position in the head as in the second calibration run. When the subject pointed to the target, he or she started 1000 ms of data acquisition by pushing a button.

After the calibration runs were completed, the aluminum rod was removed, and four different localization blocks were performed: *1*) visual single step, *2*) visual-visual double step, *3*) auditory single step, and *4*) visual-auditory double step. Blocks of one modality were always presented together and the single-step block was always presented first. In this paper we will focus on the visual experiments only. Subjects MK, ML, MV, and SP only performed in the visual experiments. The results of the auditory experiments are described in Vliegen et al. (2004). All calibration and experimental sessions were performed in complete darkness.

Visual single-step paradigm

To determine a subject's baseline localization behavior, a single-step experiment was performed. Each trial started with the presentation of a fixation LED. During fixation subjects had their head and eyes approximately aligned. After 800 ms the fixation LED was extinguished and after a 50 ms gap of complete darkness a target LED was flashed for 50 ms at a different location. Subjects were asked to look at the remembered location of the target LED as quickly and accurately as possible. As stimuli were always well extinguished before the initiation of the eye-head movement, subjects performed under completely open-loop conditions. Moreover, during stimulus presentation the hemisphere and other objects in the

room were invisible, so that no exocentric localization cues were available to the subject.

We used two different stimulus configurations. The first consisted of a central fixation target at $[R, \phi] = [0, 0]^\circ$ and ten visual target positions with $[R, \phi] =$ [14, 0], [14, 180], [20, 0], [20, 90], [20, 180], [20, 270], [27, 60], [27, 120], [27, 240], or [27, 300]°. Target locations were selected in random order One block consisted of 20 trials. Three blocks were run on separate days

In the second configuration the initial fixation position was at either $[R, \phi]$ = [20, 90] or [20, 270]° (pseudo-randomly chosen with both fixation targets occurring equally often). Visual targets were presented at randomly selected positions within a circle of $R = 35^\circ$ around the straight-ahead direction, but always at least 10° away from the fixation point. A total of 12 trials were presented in one block and at least two blocks were run on separate days. Subjects MK, ML, MV, and SP participated in one block of 96 trials of the second configuration.

Visual-visual double-step paradigm

In the double-step experiment, two visual targets were presented shortly after each other. First, a fixation target was presented for 800 ms. After 50 ms of darkness a first visual target was presented for 50 ms. The second target was also presented for 50 ms, but the timing varied, which resulted in three conditions:

1. *non-triggered (static) condition*, in which the second target was presented after a fixed delay of 50 or 100 ms (first or second target configuration respectively) after extinction of the first visual target. In this condition, both targets were presented and extinguished before initiation of the first eye-head movement.

2. *head-triggered (dynamic) condition* (first target configuration), in which the second target was triggered as soon as *head* velocity toward the first visual target exceeded 40°/sec.

3. *gaze-triggered (dynamic) condition* (second target configuration), in which the second target was presented 20 ms after the *gaze* velocity to the first visual stimulus exceeded 60°/sec.

In both dynamic conditions, the second stimulus was presented while eyes and head were moving. Because in visually-evoked gaze shifts the head-movement onset typically follows the eye-movement (Goossens and Van Opstal, 1997b), the second visual target was typically presented in mid-flight of the gaze shift

for both triggering conditions Due to the considerable variability in eye-head onset disparities, we also obtained a large range of second target onsets relative to gaze saccade (see Results, e g Fig 5 5)

As in the single-step experiment, two target configurations were used To enable direct comparison of the results of the single-step and the double-step experiments, the locations of the first and second visual targets of the double-step paradigm corresponded to the positions of the fixation targets and visual targets of the single-step paradigm Fixation targets in the double-step experiment were at $[R\ \phi] = [35, 0]^\circ$ or $[35, 180]^\circ$ in both target configurations Because in the first configuration the first target was always at $[R\ \phi] = [0, 0]^\circ$, eight catch trials were included in the experiment to prevent predictive response behavior In these catch trials the first target was at either $[R\ \phi] = [35, 30]$, $[35, 150]$, $[35, 210]$, or $[35, 330]^\circ$, and the second target was presented at either $[R\ \phi] = [20, 90]$ or $[20, 270]^\circ$ (pseudo-randomly chosen with all positions occurring equally often see Fig 5 2 in Vliegen et al , 2004, for an illustration)

Note that in the first target configuration the gaze shift to the first target was always horizontal, whereas in the second target configuration the first gaze shift also had a considerable vertical component The double-step block consisted of 48 trials (eight of which were catch trials for the first configuration) Half of the trials were dynamic trials, which were randomly interleaved with the static trials Three blocks were run on separate days in the first configuration and at least two in the second configuration Subjects MK, ML, MV, and SP did three to four blocks of 96 trials on one day In all experimental sessions, subjects were free to move their head and eyes to localize the target The specific instruction given to the subject was *"fixate both visual targets with your eyes as quickly and as accurately as possible"*

Data analysis

The raw position data from the three calibration sessions were mapped to calibrated azimuth/elevation angles of eye and head position in space by means of two neural networks (for details, see Goossens and Van Opstal 1997b) From the calibrated response data of the localization experiments, head and gaze saccades were identified off-line with a custom-written computer algorithm that detected saccades on the basis of separate onset and offset velocity and acceleration criteria (Goossens and Van Opstal 1997b) Saccade boundaries were visually checked by the experimenter and corrected if needed Responses with a first-saccade latency shorter than 80 ms or longer than 800 ms were discarded from further analysis All static double-step trials in which the first-saccade latency fell between 80 ms and the second stimulus offset were considered dynamic

trials They were included in the dynamic double-step database, provided mean gaze velocity during stimulus presentation exceeded 50°/s

The coordinates of targets and of calibrated eye and head positions were expressed in a double-pole azimuth-elevation coordinate system, in which the origin coincides with the center of the head (Knudsen and Konishi, 1979) In this system, the azimuth angle, α, is defined as the angle within the horizontal plane with the vertical midsaggital plane, whereas the elevation angle, ε, is defined as the direction within a vertical plane with the horizontal plane through the subject's ears The straight-ahead direction is defined by $[\alpha\ \varepsilon] = [0,\ 0]^{\circ}$ The relation between the $[\alpha.\varepsilon]$ coordinates and the polar $[R.\phi]$ coordinates defined by the LED hemisphere (see above) is given by Hofman and Van Opstal (1998)

Statistics

To evaluate to what extent the visuomotor system compensates for intervening eye-head movements, we performed a multiple linear regression on the azimuth and elevation components of the second gaze and head displacement vectors Regression parameters were determined on the basis of the least-squares error criterion

The bootstrap method was applied to obtain confidence limits for the optimal fit parameters in the regression analyses To that end, 1000 data sets were generated by random selections of data points from the original data Bootstrapping thus yielded a set of 1000 different fit parameters The sds in these parameters were taken as an estimate for the confidence levels of the parameter values obtained in the original data set (Press et al , 1992) To test whether two fit parameters (gaze and head-motor error, Fig 5 8, and the parameters for the different models, Eqns 5 5 to 5 7) differed significantly, we performed a t-test for two independent regression coefficients (Howell, 1997)

To determine whether the azimuth-elevation endpoint data for the different conditions were statistically different, we applied the 2D Kolmogorov-Smirnov (KS) test This test provides a measure (d-statistic) for the maximum distance between the two distributions, for which the significance level, p, that the distributions are the same, can be readily computed (Press et al , 1992) If $p < 0\ 05$ the two data sets were considered to correspond to different distributions

The bin-width (BW) of histograms (Figs 5 4 and 5 6) was determined by $BW = Range/\sqrt{N}$, with $Range$ the difference between the largest and smallest values (excluding the two most extreme points), and N the number of included data points

5.3 Results

Figure 5.3 shows two examples of gaze and head traces of subject MW as a function of time (top row), as well as the corresponding spatial trajectories (bottom row) for a static (Fig. 5.3A) and a dynamic trial (Fig. 5.3B). Note that in the static condition both targets were indeed presented before gaze and head movement onsets. In the dynamic condition, the second target was presented during the first gaze shift. Typically, gaze starts to move well before the head, which is apparent in both conditions (Fig. 5.3A and 5.3B). The dashed square in the spatial plots indicates the location to which responses would be directed if the system would not compensate for the intervening gaze shift. Instead, the eye and head saccades are clearly directed to the actual spatial location of the target.

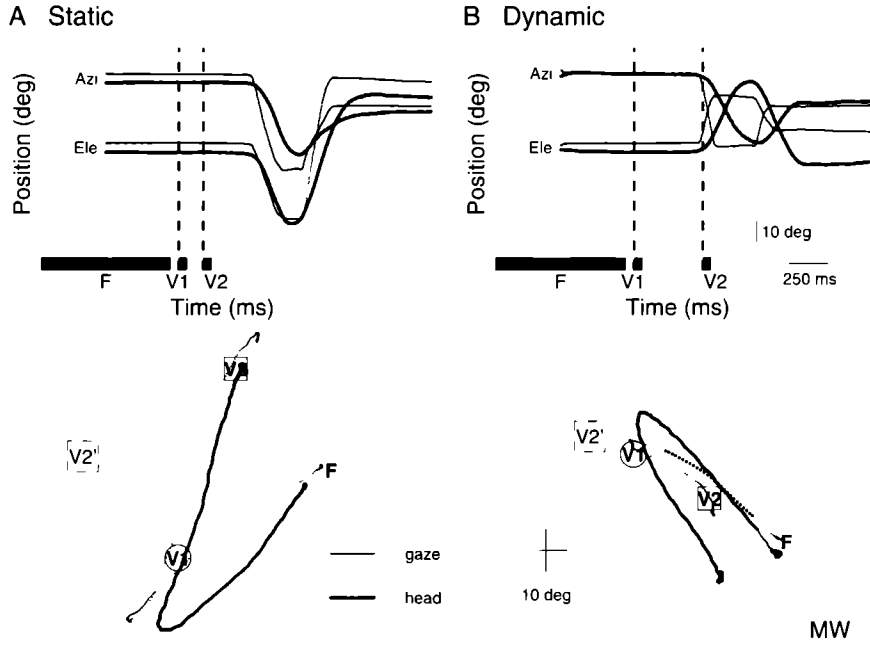


Figure 5.3: *Temporal and spatial gaze (thin lines) and head (thick lines) traces for one static trial* **(A)** *and one dynamic trial* **(B)** *of subject MW. Top panels: Temporal traces for gaze and head for both azimuth and elevation. Black bars: target timings for F, V_1, V_2 (target onset shown as vertical dashed lines). Bottom panels: Spatial trajectories of head and gaze, and target locations. The time of target presentation is indicated as a change in line thickness. Dashed square: hypothetical response location in case of no compensation for intervening eye-head movements.*

Figure 5.4A illustrates the spatial trajectories and kinematics of the eye and

head movements during the 50 ms that the visual stimulus was presented in the dynamic double steps, for all experimental sessions of subject JV. Note that the eye movements (and hence, the retinal "smearing") can be as large as 15°, while the head movements are typically somewhat smaller. The latter is due to the fact that eye velocity is typically higher at the time of stimulus presentation than head velocity. This point is further illustrated in Figure 5.4B. In the dynamic condition the mean and peak velocities (dark gray and light gray histograms, respectively) show a large trial-to-trial variability. This illustrates the fact that the saccade kinematics varied considerably between trials. Gaze latencies varied between 150 and 668 ms (mean: 258 ms) in the static condition, and between 88 and 582 ms (mean: 240 ms) in the dynamic condition. Head latencies varied between 130 and 750 ms (mean: 249 ms) in the static condition, and between 122 and 470 ms (mean· 235 ms) in the dynamic condition. Mean amplitude for the second gaze shift was 14°, with a maximum of 59°, in azimuth and 17° with a maximum of 66° in elevation. For the head the mean azimuth amplitude was 14°, with a maximum of 50°, and the mean elevation amplitude was 15° with a maximum of 72°.

To further demonstrate that the first gaze shift was well underway at the onset of the second target, and thus that ΔG^* differed appreciably from ΔG_1 (see Fig. 5.2B), Figure 5.5 shows these two variables plotted against each other, pooled for all subjects. Note that all data points lie clearly below the identity line. Moreover, there is a considerable variability in the $\Delta G^* - \Delta G_1$ differences, indicating that the second target was triggered at various gaze orientations during the first saccade The two symbols represent the different target configurations. As the head typically lagged the eyes, and had a more variable latency, the ΔG^* values for the first target configuration (head triggered) were typically smaller and more variable than those for the second target configuration (gaze triggered).

In Figure 5.6 the 2D distributions of the gaze endpoints to the second target are shown for the static (black filled dots) and the dynamic (gray triangles) double-step condition, as well as for the single-step (open circles) responses for six subjects. In this figure, all target positions (T) were aligned with the origin of the azimuth-elevation coordinate system, and the gaze endpoint coordinates are plotted as an undershoot or overshoot relative to the target location. The black histograms show the distributions of the static double-step data with the means indicated by the black dashed lines. The gray histograms and continuous lines represent the dynamic double-step data. The dotted lines indicate the means of the single step data. For all subjects, the data cluster around the actual target location, and the means for all three conditions are close together and to the

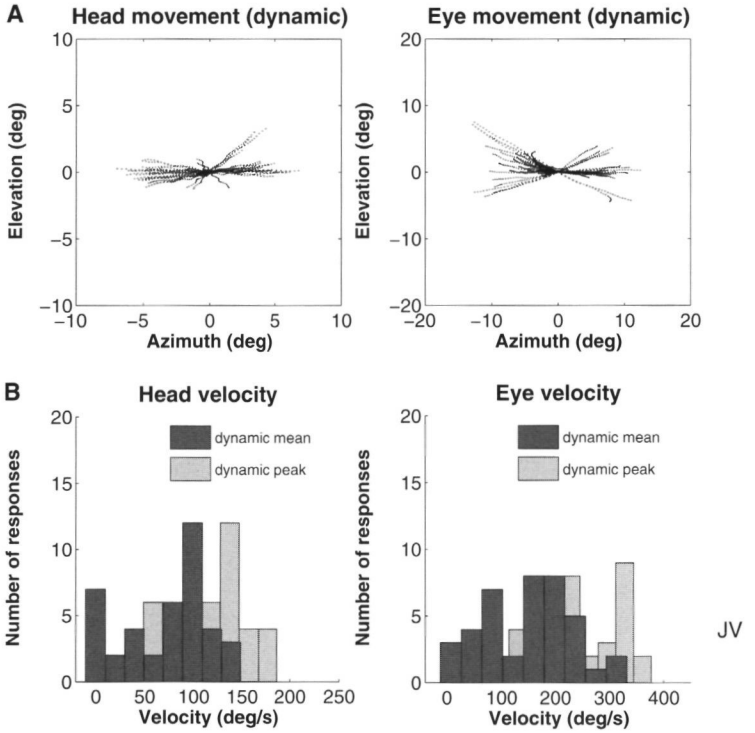


Figure 5.4: *Head (left) and eye (right) movement properties during second target presentation.* **(A)** *Two-dimensional head and eye movements during target presentation in the dynamic condition.* **(B)** *Histograms of mean and peak head and eye velocity profiles during target presentation. Dark gray histograms: mean velocity during dynamic double steps. Light gray histograms: peak velocity during dynamic double steps. Note large velocity range. Data for all sessions of subject JV.*

target location. The static and dynamic double-step responses show similar distributions and indeed for seven out of nine subjects (JO, JV, MK, ML, MV, SP, TG) the difference between these conditions was not significant. The single step data distributions, however, had a smaller variance and did differ significantly from both double-step conditions for all subjects (KS test; $p < 0.01$).

From Figures 5.3 and 5.6 it appears that localization responses are directed on average toward the actual spatial location of the visual target for both double-step conditions. To test in a quantitative way to what extent the intervening eye and head movements are accounted for in the second gaze shift, we performed a multiple linear regression analysis on the azimuth and elevation components of the second gaze displacement vector (ΔG_2), which is expressed as a weighted vector sum of the initial target position relative to the eye ($T_{E,0}$), and the first

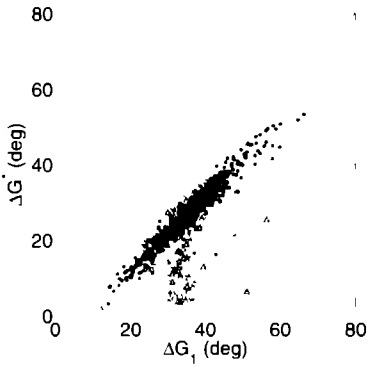


Figure 5.5: *Partial gaze shift after second target onset (ΔG^*) plotted against the full first gaze shift (ΔG_1). Data pooled for all subjects. Triangles represent data for the first target configuration, in which head movement triggered the second target. Asterisks correspond to the second target configuration, where target triggering was based on eye movement. Note, that all data points fall below the identity line, and that there is a considerable range of ΔG^* vs ΔG_1 differences.*

gaze shift (ΔG_1; see also Fig. 5.1):

$$\Delta G_2 = a \cdot T_{E,0} + b \cdot \Delta G_1 + c \tag{5.1}$$

If the second gaze movement would be based only on the initial retinal target location, the slope a should equal one, and slope b and offset c zero. This case corresponds to the right-hand $T_{E,0}$ response vector (dashed) in Figure 5.1. However, in case the gaze control system does fully compensate for the intervening gaze shift, the slope b will be exactly minus one (the response then corresponds to arrow $T_{E,2}$ in Fig. 5.1). Figure 5.7A shows the values of the actual regression coefficients for the static and dynamic stimulation conditions, and for the horizontal and vertical response components (pooled data of all subjects and sessions). Apparently, the gaze control system does account for the previous movement, as coefficient a was found to be around $+1.0$ and b close to -1.0 for all conditions. The offsets (c) were close to 0° and are not shown.

We performed a similar regression on the second head displacement (ΔH_2), to check whether the head also made goal-directed movements. Thus, ΔH_2 was described as a function of the initial target position relative to the eye ($T_{E,0}$), the first gaze shift (ΔG_1), and the eye-in-head offset position at the start of the second gaze shift (E_1):

$$\Delta H_2 = a \cdot T_{E,0} + b \cdot \Delta G_1 + c \cdot E_1 + d \tag{5.2}$$

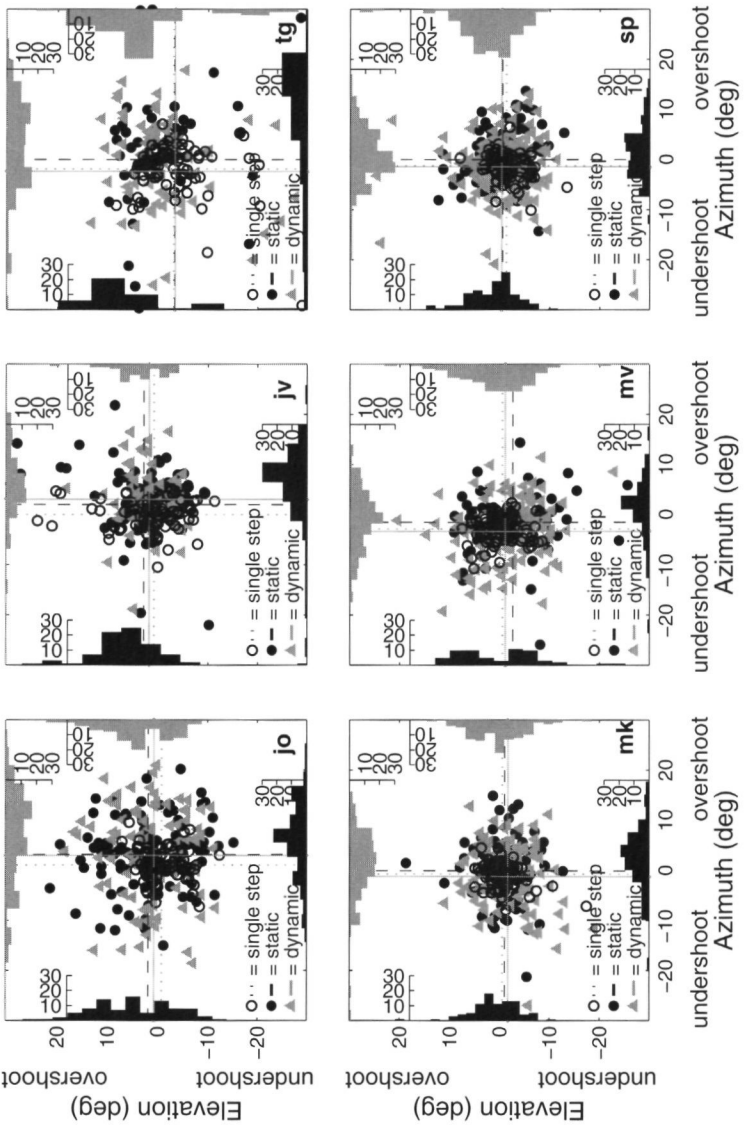


Figure 5.6: *Endpoints of the second gaze saccades for both double-step conditions (static data: filled black dots; dynamic data: gray triangles) and single steps (open circles) for six subjects. All second target positions (T) are aligned with the origin and localization responses are shown as undershoots or overshoots relative to target position. Distributions of the double-step responses, together with their means, are shown as black histograms and a black dashed line for the static condition and gray histograms and a gray solid line for the dynamic condition. Single-step means are indicated by the dotted lines.*

If the head would move toward the spatial target position, parameters a and c should be one, b should be minus one, and d should be zero (corresponding to response arrow $T_{H,2}$ in Fig. 5.1; see also the legend of Fig. 5.1). The results are summarized in Figure 5.7B. For all conditions a was found to be fairly close to $+1.0$, and b close to -1.0. The value of c differed between conditions, but was always significantly different from zero and positive. Note that subjects did not receive any specific instructions about their head movements. As a result, the eye position coefficient varied across subjects and sessions, and was less than the ideal value. The offsets (d) were around zero and are not shown.

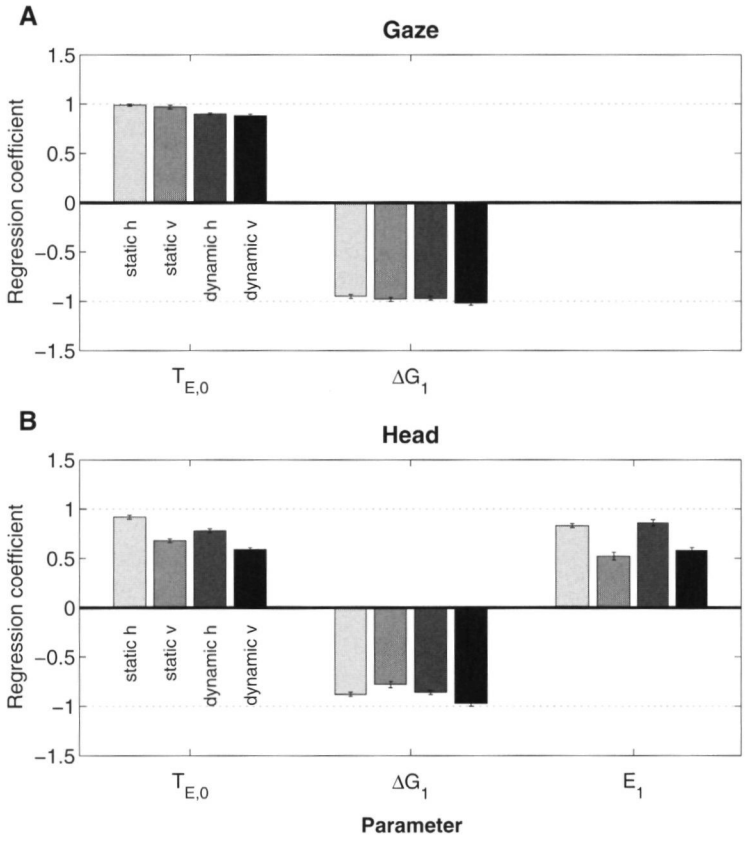


Figure 5.7: **(A)** *Regression coefficients of Eqn. 5.1 for second gaze shifts (ΔG_2), on pooled data of all subjects and sessions.* **(B)** *Regression coefficients of Eqn. 5.2 for second head shifts (ΔH_2). The different gray-colored bars indicate different directions (azimuth, elevation) and double-step conditions (static, dynamic). Error bars indicate standard deviation. Dotted lines at $+1.0$ and -1.0 correspond to the ideal values for full compensation.*

If eyes and head are both free to move, it is not trivial that both move toward the target, especially if they are not aligned at the onset of the gaze shift. In that case they have to move in different directions to reach the target (e.g. Fig. 5.1). For that to happen, the respective motor commands need to be transformed into oculocentric and craniocentric coordinates, respectively Alternatively, eyes and head could both be driven by a common signal, like the gaze motor error, as has been proposed for the common-gaze control model (Vidal et al., 1982; Guitton, 1992; Galiana and Guitton, 1992). To further quantify whether the eyes and head were indeed driven by a common error signal, or by signals expressed in their own reference frame, we performed a normalized multiple regression on the data in which the second gaze movement, ΔG_2, and head movement, ΔH_2, were each described as a function of both the gaze motor error, GM, and the head motor error, HM, at movement onset:

$$\Delta G_2 = p \cdot GM' + q \cdot HM' \tag{5.3}$$

$$\Delta H_2 = p \cdot GM' + q \cdot HM' \tag{5.4}$$

In Eqn. 5.3 and 5.4 the gaze motor error (GM) and head motor error (HM) vectors were determined as the difference between the spatial location of the second target and the gaze and head position in space at the second gaze shift onset. These variables were then transformed into their (dimensionless) z-scores: $x' = (x - \mu_x)/\sigma_x$, with μ_x the mean of variable x, and σ_x its variance. In this way, the variables are dimensionless, and p and q are the (dimensionless) partial correlation coefficients for gaze motor-error and head motor-error, respectively. If $p > q$, the eye (or head) is driven predominantly by an oculocentric gaze-error signal If $q > p$, the eye (or head) rather follows the head-centered motor error signal. In case $p > q$ (or $p < q$), for both equations, eye and head are driven by the same error signal. To allow for a meaningful dissociation of the oculocentric and craniocentric reference frames, we only incorporated trials for which the absolute azimuth or elevation component of eye-in-head position exceeded 10°, and the directional angle between the head and gaze motor-error vectors was at least 15° (we thus obtained 200 to 270 trials, depending on condition).

Figure 5.8 shows the regression coefficients on the pooled data from all subjects for all conditions. Eye-in-space (Fig. 5.8A) is clearly driven by the eye motor-error, as the coefficients for gaze motor-error are much larger than those for head motor-error (t-test: for all conditions $p < 0.001$). Conversely, the head movements (Fig. 5.8B) appear to be driven by the head-motor error. The difference between p and q was significant for the horizontal, but not for the vertical conditions ($p < 0.001$).

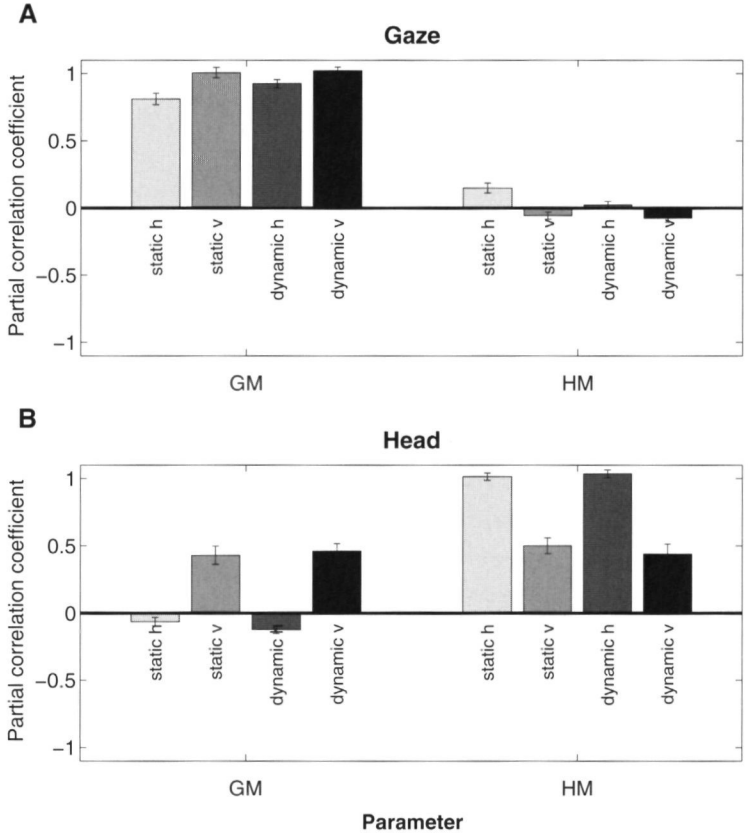


Figure 5.8: *Partial correlation coefficients for the regression on second gaze saccade (ΔG_2)* **(A)** *and second head saccade (ΔH_2)* **(B)**, *which are described as a function of gaze-motor error and head-motor error (Eqn. 5.3 and 5.4). Data pooled across subjects and sessions. Same format as Figure 5.7.*

In the Introduction we described three different models to predict the second gaze shift in a double-step experiment. All models account for intervening eye-head movements, but they differ in the type of information used to update the initial retinotopic target location ($T_{E,0}$). According to the predictive remapping models, information about planned eye-head movements is used; this could be either based on purely visual information (visual predictive, VP), or on the actual planned movement (motor predictive, MP). The dynamic feedback model (FB) states that target position is updated based on continuous information of eye and head movements. The three different models can thus be quantified by:

$$\Delta G_{VP} = a \cdot T_{L\,0} + b \cdot FV_1 + c \tag{5.5}$$

$$\Delta G_{MP} = a \cdot T_{E\,0} + b \cdot \Delta G_1 + c \tag{5.6}$$

$$\Delta G_{FB} = a \cdot T_{L,0} + b \cdot \Delta G^* + c \tag{5.7}$$

where FV_1 is the retinal error of the first target, ΔG_1 is the first saccade vector, and ΔG^* is the partial first gaze shift following stimulus presentation. In the static double step, where both targets are presented before the eye-head movement onsets, the motor-predictive model (Eqn. 5.6) makes the same prediction as the dynamic feedback model (Eqn. 5.7). To calculate the predictions for these different models, we substituted the ideal values of $a = 1.0$, $b = -1.0$ and $c = 0.0$ in Eqns. 5.5 to 5.7. For the static double steps we thus found that the MP/FB models outperformed the VP model for both the horizontal (R^2: 0.89 vs. 0.70, respectively) and vertical (R^2: 0.91 vs. 0.80, respectively) response components (data not shown). This result therefore demonstrates that the visuomotor system accounts for the actual intervening gaze shift, ΔG_1, rather than for the initial retinal error (FV_1; see Fig. 5.2).

However, for the dynamic condition the predictions of the MP and FB models are dissociated, because ΔG_1 and ΔG^* are different (see Fig. 5.5). Figure 5.9 shows the predictions of the three different models plotted against the actually measured second gaze shift for the dynamic double steps. Figure 5.9A shows the data for the head-triggered paradigm. As is apparent from Figure 5 5, the differences between ΔG_1 and ΔG^* are largest and more variable for this experiment, which is a key point in discriminating between the predictive remapping models and the dynamic feedback model. Here, we only show the horizontal components of the second gaze shifts, because in this paradigm the initial gaze shift was always purely horizontal (see Methods). In Figure 5.9B all data are pooled across subjects and sessions, and are shown for both horizontal and vertical response components. In all cases, the predictions for the dynamic feedback model outperform the predictive remapping models. This is most strongly demonstrated for the head-triggered data in Figure 5.9A.

We also performed linear fits on the measured ΔG_2 for all three models, given in Eqns 5.5 to 5.7 to determine the actual regression coefficients. We applied these fits on the data of the dynamic double steps, both for the head-triggered data separately, and for all data pooled, to quantify the dissociation between the predictive remapping models and the dynamic feedback model. The optimal fit parameters are given in Table 5.1, together with the R^2 values of the fits. If

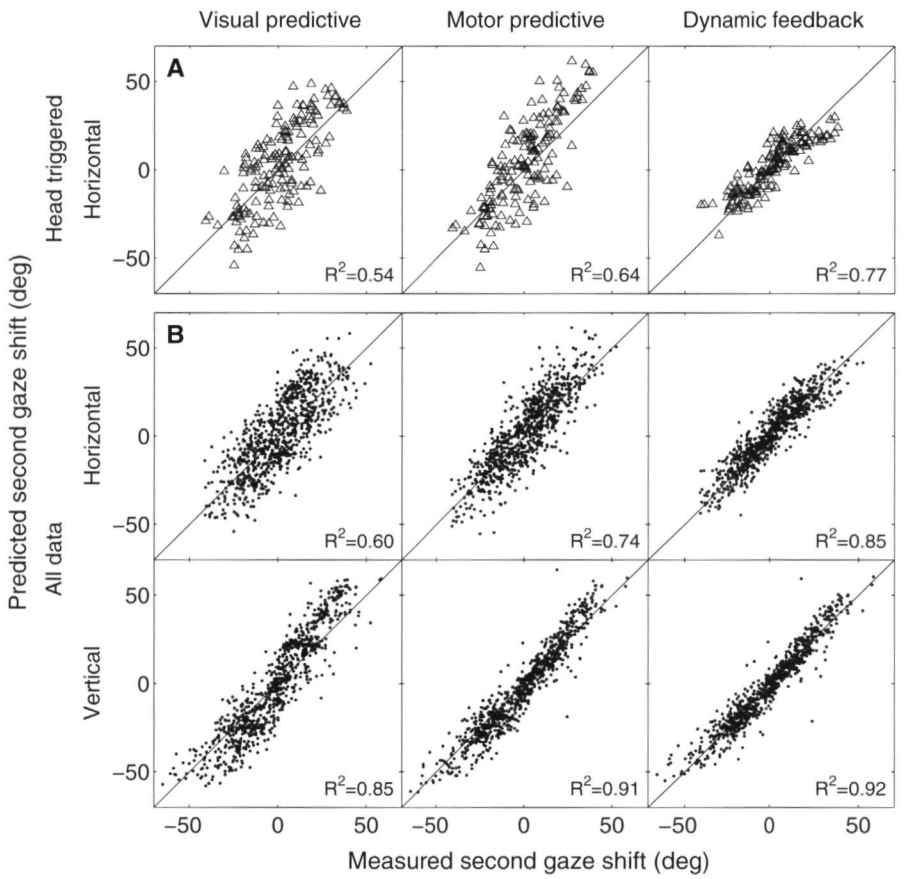


Figure 5.9: *Predictions for the second gaze shift (ΔG_2) for three different models (columns), as a function of the measured second gaze shift.* **(A)** *Data for the dynamic condition of the head-triggered paradigm (horizontal data only, pooled across subjects).* **(B)** *Horizontal and vertical response components for all data, pooled across subjects and sessions. If a model would predict ΔG_2 perfectly, the data would fall on the identity line and R^2 would be 1. R^2 values are given in the lower right corner of all panels. In all cases, the dynamic feedback model provides the best prediction for the data.*

a model provides a good prediction of the measured ΔG_2, not only should R^2 be 1, but also the parameter values should be close to the ideal values of $a = 1$, $b = -1$ and $c = 0$ (see Figs. 5.1 and 5.2). The two predictive remapping models give a rather good prediction, with R^2 values of 0.77 or higher, but their optimal parameter values clearly differ from one and minus one. The dynamic feedback model however, yields parameter values that are close to these ideal values, and R^2 values that are closest to one. As expected, the difference is largest for the

		a		b		c, deg		R^2	
	N	153	796	153	796	153	796	153	796
Model	Condition								
VP	Dyn hor	0 72	0 82	−0 55	−0 61	−1 09	−0 06	0 77	0 85
	Dyn ver	-	0 87	-	−0 55	-	0 70	-	0.93
MP	Dyn hor	0 77	0 85	−0 57	−0 67	−2 15	−0 65	0 85	0 91
	Dyn ver	-	0 87	-	−0 78	-	0 86	-	0 95
FB	Dyn hor	1 00	0 90	−1 24	−0 97	−0 87	−0 44	0 92	0 92
	Dyn ver	-	0 88	-	−1 02	-	0 45	-	0 96
Ideal		1		−1		0		1	

Table 5.1: *Results of fitting Eqns 5 5 to 5 7 to the dynamic double-step data for the horizontal and vertical response components of measured gaze displacements, ΔG_2 Results for the horizontal head-triggered data are given on the left-hand side of each column, results for all data are shown on the right N number of data points included in the regression analysis Note that although the motor-predictive model yields a reasonable fit to the data, the dynamic feedback model is clearly superior in describing the data first, the goodness of fit is higher for the FB model and, second, the fitted parameters for the FB model are significantly closer to the ideal values (bottom row) than those for the MP model This is most apparent for the head-triggered data, for which the difference between ΔG_1 and ΔG^* - and thus the difference between the two predictive remapping models and the dynamic feedback model - is largest (see also Figs 5 5 and 5 9)*

head-triggered data The a and b values of the dynamic feedback model are significantly different from both predictive models for the horizontal and vertical response components ($p < 0.001$ for all conditions, except for the a values for all data in the vertical condition).

Accuracy around saccade onset

In contrast to the accurate localization responses found by Hallett and Lightstone (1976) in their dynamic double steps, Dassonville et al. (1995) reported systematic errors when short-duration (2 ms) visual targets were presented near saccade onset (see also Schlag and Schlag-Rey, 2002, for a review). To test whether this discrepancy in results might have been caused by the presence or absence of allocentric localization cues (e.g. the difference vector between the target locations on the retina), Dassonville et al. (1995) systematically varied the dark gap between the two targets between 45 and 495 ms. However, only part of the discrepancy with Hallett and Lightstone's (1976) study could be explained by a potential allocentric cue, as errors diminished only slightly for the shortest gaps. Also Honda (1990, 1991) obtained systematic localization errors

for stimuli presented before saccade onset, like Dassonville et al. (1995), but he also reported errors in the opposite direction after saccade onset. In our experiments also a considerable variation arose in the duration of the dark period between the first and second visual target that ranged from 50 to 590 ms (non-triggered: 50 ms; triggered: 80 to 590 ms). In Figure 5.10A, we have therefore plotted our data in the same way as Dassonville et al. (1995; their figure 2), with mean azimuth localization error plotted against stimulus onset relative to gaze shift onset. Localization error is plotted relative to the direction of the first gaze shift, with positive values indicating an error in the same direction as the first gaze shift. The dashed line is the prediction of the damped eye-position model of Dassonville et al. (1995) and Schlag et al. (1989; time constant 65 ms) for a 35° horizontal gaze shift. Figure 5.10B plots the measured and the predicted errors against each other. In contrast to Dassonville et al. (1995) and Honda (1990, 1991), but in line with Hallett and Lightstone (1976), we did not obtain a large systematic increase in the localization error around gaze-shift onset in the direction of the first saccade.

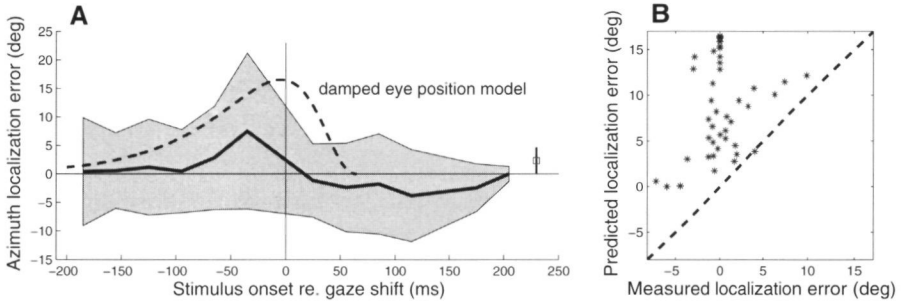


Figure 5.10: **(A)** *Mean gaze localization error (azimuth) of the second target as a function of second stimulus onset relative to saccade onset (black line), together with standard deviations (gray area). Negative values of stimulus onset indicate trials where the second stimulus is presented before gaze shift onset; for positive values the stimulus is presented after gaze shift onset. Localization error is plotted relative to the direction of the first gaze shift, with positive values indicating an error in the same direction as the first gaze shift. The mean single-step localization error (with sd) is shown as an open square at an onset value of 230 ms. Dashed line: estimate of the prediction of a damped eye position model (Schlag et al., 1989), based on the results of Dassonville et al. (1995).* **(B)** *Predicted localization error according to the damped eye position model vs. the measured localization error. Note that all points lie above the unity line.*

5.4 Discussion

Summary

Our results show that visual localization remains accurate for the three experimental paradigms in this study single steps, static double steps, and dynamic double steps Despite the fundamentally different sensorimotor transformations, on average responses were directed toward the actual spatial location of the second target for all three conditions (Figs 5 3, 5 6) Although gaze endpoint variability for double steps is higher than for single steps, we did not obtain systematic localization errors (Fig 5 6) Gaze shifts remained accurate for dynamic double steps, despite the high inter-trial variability of eye and head displacements and kinematics during target presentation (Fig 5 4) We conclude that intervening eye-head gaze shifts have not been incorporated by a preprogramming strategy (neither based on visual input, as in the visual-predictive model, nor by the intended gaze shift, as in the motor-predictive model) Instead, we propose that the gaze control system updates target locations dynamically, through continuous feedback about ongoing eye and head movements

Related studies

Mays and Sparks (1980) and Sparks and Mays (1983) have shown that monkeys redirect their gaze to previously flashed visual targets, even after microstimulation in the SC induced an intervening saccade that was not planned by the animal Schlag-Rey et al (1989) interrupted spontaneous and visually elicited saccades by SC microstimulation, and found that the saccade resumed its trajectory toward a new saccade goal imposed by the induced perturbation The authors conjectured that microstimulation created a "phosphene" that served as a new goal for the visuomotor system Such goal-directed responses were not obtained when stimulation was applied to the SC motor layers

In contrast to Dassonville et al (1995), we did not observe systematic localization errors when the target was flashed around gaze-shift onset (Fig 5 10) It is not immediately clear which factors underlie this apparent discrepancy in results, as their stimulus conditions closely resembled those of the present paper (i e complete darkness, and absence of allocentric cues), although our stimuli had a longer duration (50 ms vs 2 ms), and they were brighter (2 cd/m^2 vs 15 mcd/m^2) As a result, our stimuli induced a larger and clearer dynamic retinal "smearing", which may have provided partial information about the ongoing gaze shift (Fig 5 4) Indeed, Festinger and Holtzman (1978) demonstrated a slight benefit on saccade accuracy when saccade-like retinal smearing was applied to

the stimulus during the entire saccade Yet, retinal smearing alone cannot explain the accurate responses in our experiments, as considerable portions of the gaze shifts remained in complete darkness after target offset

Although our stimuli were much brighter than in Dassonville et al 's (1995) study, they did not induce a noticeable retinal afterimage, nor did they provide additional visual cues through spurious reflections on surfaces in the environment Even if there would be visual cues, visual landmarks are subjected to a strong perceptual deformation around a saccadic event, leading to systematic perceptual localization errors (Ross et al , 1997, Lappe et al , 2000, Bremmer and Krekelberg 2003) Yet, such systematic localization errors were not obtained in our study either (Fig 5 10) Moreover, as illustrated in Figure 5 2, retinal events alone cannot explain spatially accurate behavior, as this requires adequate use of motor signals

If gaze control relied on a sluggish, low-pass filtered eye (and possibly head) position signal, like Schlag et al (1989) and Dassonville et al (1995) propose, it is not obvious why such a (motor) effect disappears for brighter and longer-duration stimuli Indeed, our recent report that gaze shifts remain accurate also when the second stimulus is a brief *auditory* target (Vliegen et al , 2004) cannot be attributed to visual factors

A final difference with earlier studies is that our subjects were free to move eyes and head Perhaps these more natural orienting responses may have enabled the system to better use available egocentric movement cues

In an attempt to unite the apparently discrepant data sets, we conjecture that gaze control and perception may depend on the strength (or reliability) of the available sensory and motor cues Thus, the integrity of the motor feedback signals could also depend on the signal-to-noise ratio of the sensory input For very weak visual stimuli the system's feedback pathway might thus be only partially engaged, resulting in an apparently lagged eye-position signal, and error patterns along the saccade direction For more salient sensory inputs, eye and head positions could be fully incorporated in the transformations

Perception vs. action

Despite our conclusion that predictive remapping cannot account for spatially accurate performance in dynamic double-steps, there is ample neurophysiological evidence that such a mechanism is engaged in visuomotor behavior (Duhamel et al , 1992, Colby et al , 1995, Walker et al , 1995, Umeno and Goldberg 1997) Possibly, predictive information about impending saccades primarily subserves perceptual stability of the visual environment across saccades, rather than the coordination of accurate saccade sequences (action, Burr et al , 2001, Brem-

mer and Krekelberg 2003) Such an interpretation would also be consistent with saccadic adaptation experiments. In that paradigm, the saccade triggers the stimulus to rapidly jump toward a new location. Initially, the primary saccade is consistently followed by a corrective saccade to the new target location. However, during the course of many repetitions, the saccade gradually incorporates the future target jump, to eventually land on the target without the need for corrective responses. Neurophysiological evidence has indicated that adaptation acts downstream from the motor SC (Frens and Van Opstal 1997; Edelman and Goldberg 2002). Recently, Bahcall and Kowler (1999) found that perceptual localization of the target was affected by adaptation, indicating that perception did not have access to information about the adapted saccade. Rather, perception appeared to use a signal about the *intended* saccade for the initial retinal error, irrespective of the intrasaccadic target jump. More recently, however, Awater et al. (2005) have challenged this conclusion. In contrast, secondary saccades in static double steps compensate for the adapted first saccade (Tanaka, 2003), which suggests that in the double step paradigm the visuomotor system *does* have access to the actual saccade commands. These studies, together with the present data, therefore support the notion that separate neural pathways are involved in spatial perception, and in spatially accurate movement planning, or action (Burr et al , 2001).

Model implications

So far, we have not made explicit which feedback signals may be used in target updating. Here, we confront the concepts of *displacement* feedback vs. *position* feedback (see Introduction). Although these concepts were initially developed for the oculomotor system, and primarily based on static double-step results, we here apply these ideas to dynamic eye-head gaze control.

According to displacement models, the visuomotor system keeps the target in an *oculocentric* reference frame, which is updated by an ongoing gaze *displacement* signal to generate a spatially accurate dynamic gaze motor error:

$$\Delta G_2(t) = T_{E\,0} - \Delta G_1(t) \tag{5.8}$$

Alternatively, in position models the retinal target location, $T_{E\,0}$, is first transformed into an *absolute* reference frame (e.g. a world reference frame), by adding gaze position at stimulus onset, G_0:

$$T_W = T_{E,0} + G_0 \tag{5.9}$$

This target location is then updated to gaze motor error by subtracting the current gaze position, $G(t)$

$$\Delta G_2(t) = T_W - G(t) \quad (5\ 10)$$

Note that both models (Eqns 5 8 and 5 10) yield identical results, but the position model requires an additional transformation stage (Eqn 5 9) So far, neurophysiological evidence has favored the displacement scheme For example, cells in the primate FEF have discharges that relate to the three signals figuring in Eqn 5 8 (Goldberg and Bruce, 1990) Recently, Sommer and Wurtz (2002, 2004) provided evidence that the mediodorsal thalamus mediates a signal with information about the impending collicular eye-displacement command to the FEF, which might be used to update the target in a double-step paradigm Conversely, an explicit neural code of the target in a world-fixed reference frame has not been found

The neurophysiological evidence notwithstanding, we here argue that our dynamic double-step results are hard to reconcile with a displacement scheme Figure 5 11 illustrates the problem for the eye-in-space movement only, ignoring the head movement for simplicity Suppose that the eye makes a gaze shift, ΔG_1, toward the first visual target, T_1, and that T_2 is flashed in mid-flight, when gaze position is G^* According to the position model, the target is first updated to an extraretinal (world) reference frame by adding gaze position to the retinal target location $T_W = T_E + G^*$ At the end of the saccade, the gaze position is G_1 and the gaze motor error is $\Delta G_2 = T_W - G_1$ (Eqn 5 10) A gaze-displacement scheme only needs to subtract the eye movement following second target onset from the retinal error to yield the gaze motor error at saccade offset $\Delta G_2 = T_L - \Delta G^*$ (Eqn 5 8)

However, the apparent simplicity of this latter model to explain static double-step behavior now encounters a serious difficulty Given the visual delays in the system, and the need to restart the computation of gaze displacement in mid-flight at second target onset (while discarding the gaze displacement-so-far, ΔG_0), it is not at all obvious how the visuomotor system may get access to ΔG^* To generate such a signal would require either a new resettable integrator, or resetting/restarting the only resettable integrator Recent studies suggest that this process may involve a leaky process that takes at least 50 ms (Nichols and Sparks, 1995)

In the position scheme, a potential delay is not immediately fatal, as a neural estimate of eye position could be readily available from the oculomotor brainstem, and even if the computation of T_W^* might take some time, it could be finished during the continuation of the movement

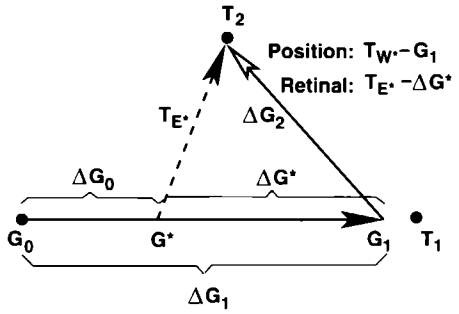


Figure 5 11: *Coordinate transformations for a visual target at T_2, presented in mid-flight of a saccadic gaze shift (gaze-displacement vector, ΔG_1) to visual target T_1 Head movements are left out, for simplicity G_0 gaze position at the start of the trial, ΔG_0 gaze displacement up to second target onset T_E target in retinal coordinates G^* gaze position at second target onset T_W target in world coordinates G_1 gaze position at saccade offset ΔG_2 gaze motor error at saccade offset ΔG^* gaze displacement after second target onset See text for explanation*

Note also that for saccades to (head-fixed) auditory targets, a signal related to eye *position*, not to eye displacement, is needed to construct the eye motor error (Jay and Sparks, 1984). We therefore propose that the updating of target locations relies on instantaneous absolute positions, rather than on relative displacements.

Figure 5.12 provides a conceptual model for the different stages within the gaze control system that transforms auditory and visual targets into a world reference frame, and subsequently into a dynamic oculocentric gaze displacement signal. For a visual stimulus, retinal error is first combined with eye (E_0) and head (H_0) position at second target onset (t_0) to construct an extraretinal target representation (T_{W_0}). For auditory localization, head position (H_0) suffices for this transformation. In line with the dynamic double step results, current eye and head positions are combined with this memorized target representation, yielding a dynamic estimate of the target in eye-centered coordinates, $\Delta G(t)$. If a saccade is planned to the target (at t_1), a fixed, local population of cells in the SC motor map is recruited that represents the desired gaze displacement, ΔG_1. This signal drives the eye and head motor systems in their own motor frames, $\Delta E_2 = \Delta G_1$, and ΔH_2, respectively (e.g. Fig. 5.8). To compute the latter signal, eye-in-head position, E_2, at movement initiation (t_2), is required.

Neurophysiological correlates

Our scheme proposes that the dynamic transformation of target location occurs upstream from the motor SC. A potential candidate for these transformations

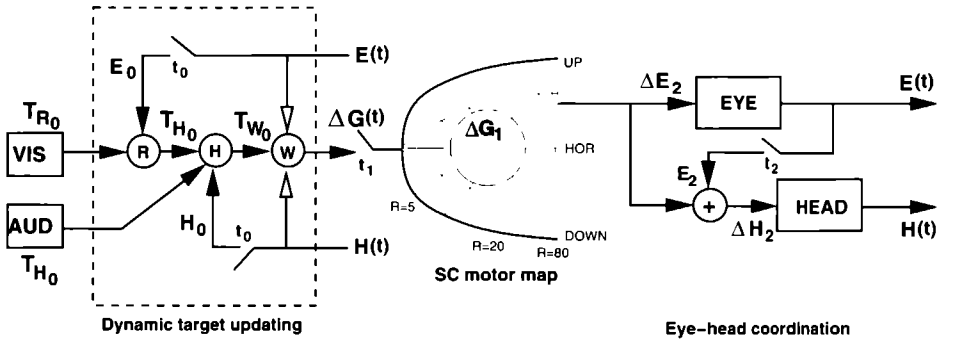


Figure 5.12: *Model for dynamic eye-head coordination to auditory and visual targets. The dynamic updating pathway builds a representation of auditory and visual targets in a world reference frame (T_{W_0}), by combining the retinal error (T_{R_0}), with eye (E_0) and head position (H_0) at second target onset (time t_0). Head-centered auditory input, (T_{H_0}), is combined with head position only. The remembered target location is continuously combined with current eye ($E(t)$) and head ($H(t)$) position, yielding a dynamic target estimate in oculocentric coordinates ($G(t)$). When the gaze shift is selected (time t_1), a local population of cells in the SC motor map is activated, representing the desired gaze displacement vector (ΔG_1). This collicular signal drives eye and head motor systems at time t_2. Local feedback circuits downstream from the SC ensure that eyes and head are driven in their own motor frames: (E_2 and ΔH_2, respectively. Details of the latter circuits are omitted for clarity. Filled arrowheads: excitatory projections; open arrowheads. inhibitory projections. E_2: eye position at gaze shift onset.*

may be the posterior parietal cortex (PPC; Andersen 1997). Although individual cells in PPC appear to possess oculocentric auditory and visual receptive fields, rather than receptive fields in world coordinates (Stricanne et al., 1996), many PPC cells have been shown to be also modulated by eye and head position signals (Snyder et al., 1998). In this way, PPC cells are better characterized by so-called "gain fields", than by a purely visual receptive field. Simulations have indicated that a distributed population of cells with gain field modulations could simultaneously represent the target in retinotopic, craniocentric, and in world coordinates (Zipser and Andersen 1988; Van Opstal and Hepp, 1995; Xing and Andersen 2001). Interestingly, PPC cells have also been shown to be involved in predictive remapping (Duhamel et al., 1992). The PPC could therefore play a crucial role in both the perception of, and the actions in, sensorimotor space. At present it is unclear whether cells displaying gain fields, and cells showing predictive remapping belong to the same population, or to different subpopulations.

Acknowledgments

We thank Ger Van Lingen, Hans Kleijnen, Günther Windau and Ton Van Dreumel for valuable technical assistance, and Jeroen Goossens for helpful discussions and suggestions. This research was supported by the Radboud University Nijmegen (AJVO, TVG) and the Netherlands Organization for Scientific Research (NWO - section Maatschappij- en gedragswetenschappen, MaGW, project nr. 410-20-301; JV).

Bibliography

Andersen RA (1997) "Multimodal integration for the representation of space in the posterior parietal cortex", Review in· *Philos Trans R Soc Lond B Biol Sci* 352: 1421-1428

André-Deshays C, Berthoz A, and Revel M (1988) "Eye-head coupling in humans. I. Simultaneous recording of isolated motor units in dorsal neck muscles and horizontal eye movements", *Exp Brain Res* 69· 399-406

Awater H, Burr D, Lappe M, Morrone MC, and Goldberg ME (2005) "Effect of saccadic adaptation on localization of visual targets", *J Neurophysiol* 93· 3605-3614

Bahcall DO and Kowler E (1999) "Illusory shifts in visual direction accompany adaptation of saccadic eye movements", *Nature* 400: 864-866

Becker W and Jürgens R (1979) "An analysis of the saccadic system by means of double-step stimuli", *Vision Res* 19: 967-983

Blauert J (1969/1970) "Sound localization in the median plane", *Acustica* 22: 205-213

Blauert J (1997) *Spatial Hearing. The Psychophysics of Human Sound Localization*, Cambridge MA: MIT Press

Bremmer F and Krekelberg B (2003) "Seeing and acting at the same time: challenges for brain (and) research", *Neuron* 38: 367-370

Burr DC, Morrone MC, and Ross J (2001) "Separate visual representations for perception and action revealed by saccadic eye movements", *Curr Biol* 11: 798-802

Butler RA (1987) "An analysis of the monaural displacement of sound in space", *Percept Psychophys* 41: 1-7

Butler RA and Musicant AD (1993) "Binaural localization: influence of stimulus frequency and the linkage to covert peak areas", *Hear Res* 67: 220-229

Colby CL, Duhamel JR, and Goldberg ME (1995) "Oculocentric spatial representation in parietal cortex", *Cereb Cortex* 5: 470-481

Collewijn H, Van Der Mark F, and Jansen TC (1975) "Precise recording of human eye movements", *Vision Res* 15. 447-450

Dassonville P, Schlag J, and Schlag-Rey M (1992) "Oculomotor localization relies on a damped representation of saccadic eye displacement in human and nonhuman primates", *Vis Neurosci* 9: 261-269

Dassonville P, Schlag J, and Schlag-Rey M (1995) "The use of egocentric and exocentric location cues in saccadic programming", *Vision Res* 35· 2191-2199

Duhamel JR, Colby CL, and Goldberg ME (1992) "The updating of the representation of visual space in parietal cortex by intended eye movements", *Science* 255: 90-92

Edelman JA and Goldberg ME (2002) "Effect of short-term saccadic adaptation on saccades evoked by electrical stimulation in the primate superior colliculus", *J Neurophysiol* 87: 1915-1923

Festinger L and Holtzman JD (1978) "Retinal image smear as a source of information about magnitude of eye movement", *J Exp Psychol Hum Percept Perform* 4· 573-585

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

Frens MA and Van Opstal AJ (1997) "Role of monkey superior colliculus in saccadic short term adapation", *Brain Res Bull* 43· 473-484

Galiana HL and Guitton D (1992) "Central organization and modeling of eye-head coordination during orienting gaze shifts", *Ann N Y Acad Sci* 656 452-471

Goldberg ME and Bruce CJ (1990) "Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal", *J Neurophysiol* 64· 489-508

Good MD and Gilkey RH (1996) "Sound localization in noise: the effect of signal-to-noise ratio", *J Acoust Soc Am* 99: 1108-1117

Goossens HHLM and Van Opstal AJ (1997a) "Local feedback signals are not distorted by prior eye movements: evidence from visually-evoked double saccades", *J Neurophysiol* 78: 533-538

Goossens HHLM and Van Opstal AJ (1997b) "Human eye-head coordination in two dimensions under different sensorimotor conditions", *Exp Brain Res* 114: 542-560

Goossens HHLM and Van Opstal AJ (1999) "Influence of head position on the spatial representation of acoustic targets", *J Neurophysiol* 81: 2720-2736

Guitton D (1992) "Control of eye-head coordination during orienting gaze shifts", *Trends Neurosci* 15: 174-179

Hallett PE and Lightstone AD (1976) "Saccadic eye movements towards stimuli triggered by prior saccades", *Vision Res* 16: 99-106

Hartmann WM and Rakerd B (1993) "Auditory spectral discrimination and the localization of clicks in the saggital plane", *J Acoust Soc Am* 94: 2083-2092

Heffner RS and Heffner HE (1992) "Visual factors in sound localization in mammals", *J Comp Neurol* 317: 219-232

Hofman PM and Van Opstal AJ (1998) "Spectro-temporal factors in two-dimensional human sound localization", *J Acoust Soc Am* 103: 2634-2648

Hofman PM and Van Opstal AJ (2002) "Bayesion reconstruction of sound localization cues from responses to random spectra", *Biol Cybern* 86: 305-316

Hofman PM, Van Riswick JG, and Van Opstal AJ (1998) "Relearning sound localization with new ears", *Nature Neurosci* 1: 417-421

Hofman PM, Vlaming MS, Termeer PJ, and Van Opstal AJ (2002) "A method to induce swapped binaural hearing", *J Neurosci Methods* 113: 167-179

Honda H (1990) "Eye movements to a visual stimulus flashed before, during, or after a saccade", In Jeannerod, M. (Ed), *Attention and performance*, vol 13, LEA, Hillsdale, NJ, pp 567-582

Honda H (1991) "The time courses of visual mislocalization and of extraretinal eye position signals at the time of vertical saccades", *Vision Res* 31: 1915-1921

Howell DC (1997) *Statistical Methods for Psychology* Duxbury Press, Belmont, CA

Jay MF and Sparks DL (1984) "Auditory receptive fields in primate superior colliculus shift with changes in eye position", *Nature* 309 345-347

Jay MF and Sparks DL (1984) "Auditory receptive fields in primate superior colliculus shift with changes in eye position", *Nature* 309 345-347

Jay MF and Sparks DL (1987) "Sensorimotor integration in the primate superior colliculus II Coordinates of auditory signals", *J Neurophysiol* 57 35-55

Jurgens R, Becker W, and Kornhuber HH (1981) "Natural and drug-induced variations of velocity and duration of human saccadic eye movements evidence for a control of the neural pulse generator by local feedback", *Biol Cybern* 39 87-96

Knudsen EI and Konishi M (1979) "Mechanisms of sound localization in the barn owl (Tyto alba)", *J Comp Physiol* 133 13-21

Kopinska A and Harris LR (2003) "Spatial representation in body coordinates evidence from errors in remembering positions of visual and auditory targets after active eye, head, and body movements", *Can J Exp Psychol* 57 23-37

Kulkarni A and Colburn HS (1998) "Role of spectral detail in sound-source localization", *Nature* 396 747-749

Lappe M, Awater H, and Krekelberg B (2000) "Postsaccadic visual references generate presaccadic compression of space", *Nature* 403 892-895

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

Lewald, J and Ehrenstein, W H (1998) "Influence of head-to-trunk position on sound lateralization", *Exp Brain Res* 121, 230-238

Lewald J, Dorrscheidt GJ, and Ehrenstein WH (2000) "Sound localization with eccentric head position", *Behav Brain Res* 108 105-125

Macpherson EA and Middlebrooks JC (2000) "Localization of brief sounds Effects of level and background noise", *J Acoust Soc Am* 108 1834-1849

Macpherson EA and Middlebrooks JC (2000) "Localization of brief sounds Effects of level and background noise", *J Acoust Soc Am* 108 1834-1849

Macpherson EA and Middlebrooks JC (2002) "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited", *J Acoust Soc Am* 111: 2219-2236

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

Mays LE and Sparks DL (1980) "Saccades are spatially, not retinocentrically, coded", *Science* 208: 1163-1165

Middlebrooks JC (1992) "Narrow-band sound localization related to external ear acoustics", *J Acoust Soc Am* 92: 2607-2624

Middlebrooks JC and Green DM (1991) "Sound localization by human listeners", *Annu Rev Psychol* 42: 135-159

Musicant AD and Butler RA (1984) "The psychophysical basis of monaural localization", *Hear Res* 14: 185-190

Nichols MJ and Sparks DL (1995) "Nonstationary properties of the saccadic system: new constraints on models of saccadic control", *J Neurophysiol* 73: 431-435

Oldfield SR and Parker SP (1984) "Acuity of sound localization: a topography of auditory space. I. Normal hearing conditions", *Perception* 13: 581-600

Ottes FP, Van Gisbergen JAM, and Eggermont JJ (1984) "Metrics of saccade responses to visual double stimuli. two different modes", *Vision Res* 24: 1169-1179

Perrett S and Noble W (1997a) "The effect of head rotations on vertical plane sound localization", *J Acoust Soc Am* 102: 2325-2332

Perrett S and Noble W (1997b) "The contribution of head motion cues to localization of low-pass noise", *Percept Psychophys* 59: 1018-1026

Pöppel E (1973) "Comment on "visual system's view of acoustic space"", *Nature* 243. 231

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP (1992) *Numerical Recipes in C*, Cambridge MA: Cambride University Press

Rayleigh, Lord (1907) "On our perception of sound direction", *Philos Mag* 13: 214-232

Robinson DA (1963) "A method of measuring eye movement using a scleral search coil in a magnetic field", *IEEE Trans BME* 10 137-145

Robinson DA (1972) "Eye-movements evoked by collicular stimulation in alert monkey", *Vision Res* 12 1795-1808

Robinson DA (1975) "Oculomotor control signals", *in Basic mechanisms of ocular motility and their clinical implications*, edited by G Lennerstrand and P Bach-y-Rita, Pergamon, Oxford, UK, p 337-374

Rogers ME and Butler RA (1992) "The linkage between stimulus frequency and covert peak areas as it relates to monaural localization", *Percept Psychophys* 52 536-546

Ron S, Berthoz A, and Gur S (1993) "Saccade-vestibulo-ocular reflex co-operation and eye-head uncoupling during orientation to flashed target", *J Physiol* 464 595-611

Ron S, Berthoz A, and Gur S (1994) "Model of coupled or dissociated eye-head coordination", *J Vestib Res* 4 383-390

Ross J, Concetta Morrone M, and Burr DC (1997) "Compression of visual space before saccades", *Nature* 386 598-601

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

Schlag J and Schlag-Rey M (2002) "Through the eye, slowly delays and localization errors in the visual system", *Nat Rev Neurosci* 3 191-215

Schlag-Rey M, Schlag J, and Shook B (1989) "Interactions between natural and electrically evoked saccades I Differences between sites carrying retinal error and motor error signals in monkey superior colliculus", *Exp Brain Res* 76 537-547

Schlag J, Schlag-Rey M, and Dassonville, P (1989) "Interactions between natural and electrically evoked saccades II At what time is eye position sampled as a reference for the localization of a target?", *Exp Brain Res* 76 548-558

Schroeder MR (1970) "Synthesis of low-peak factor signals and binary sequences with low autocorrelation", *IEEE Trans Inf Theory* 16 85-89

Snyder LH, Grieve KL, Brotchie P, and Andersen RA (1998) "Separate body- and world-referenced representations of visual space in parietal cortex", *Nature* 394 887-891

Sommer MA and Wurtz RH (2002) "A pathway in primate brain for internal monitoring of movements", *Science* 296 1480-1482

Sommer MA and Wurtz RH (2004) "What the brain stem tells the frontal cortex II Role of the SC-MD-FEF pathway in corollary discharge", *J Neurophysiol* 91 1403-1423

Sparks DL and Mays LE (1983) "Spatial localization of saccade targets I Compensation for stimulation-induced perturbations in eye position", *J Neurophysiol* 49 45-63

Stahl JS (2001) "Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity", *Exp Brain Res* 136 200-210

Stricanne B, Andersen RA, and Mazzoni P (1996) "Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP", *J Neurophysiol* 76 2071-2076

Tanaka M (2003) "Contribution of signals downstream from adaptation to saccade programming", *J Neurophysiol* 90 2080-2086

Umeno MM and Goldberg ME (1997) "Spatial processing in the monkey frontal eye field I Predictive visual responses", *J Neurophysiol* 78 1373-1383

Van Gisbergen JAM, Robinson DA, and Gielen S (1981) "A quantitative analysis of generation of saccadic eye movements by burst neurons", *J Neurophysiol* 45 417-442

Van Opstal AJ and Hepp K (1995) "A novel interpretation for the collicular role in saccade generation", *Biol Cybernet* 73 431-445

Van Wanrooij MM and Van Opstal AJ (2005) "Relearning sound localization with a new ear", *J Neurosci* 25 5413-5424

Vidal PP, Roucoux A, and Berthoz A (1982) "Horizontal eye position-related activity in neck muscles of the alert cat", *Exp Brain Res* 46 448-453

Vliegen J, Van Grootel TJ, and Van Opstal AJ (2004) "Dynamic sound localization during rapid eye-head gaze shifts", *J Neuroscience* 24 9291-9302

Vliegen J and Van Opstal AJ (2004) "The influence of duration and level on human sound localization", *J Acoust Soc Am* 115 1705-1713

Volle M and Guitton, D (1993) "Human gaze shifts in which head and eyes are not initially aligned", *Exp Brain Res* 94 463-470

Walker MF, Fitzgibbon EJ, and Goldberg ME (1995) "Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements", *J Neurophysiol* 73 1988-2003

Wightman FL and Kistler DJ (1989a) "Headphone simulation of free-field listening I Stimulus synthesis", *J Acoust Soc Am* 85 858-867

Wightman FL and Kistler DJ (1989b) "Headphone simulation of free-field listening II Psychophysical validation", *J Acoust Soc Am* 85 868-878

Wightman FL and Kistler DJ (1992) "The dominant role of low-frequency interaural time differences in sound localization", *J Acoust Soc Am* 91 1648-1661

Wightman FL and Kistler DJ (1999) "Resolution of front-back ambiguity in spatial hearing by listener and source movement", *J Acoust Soc Am* 105 2841-2853

Xing J and Andersen RA (2001) "Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames", *J Cogn Neurosci* 12 601-614

Zakarauskas P and Cynader MS (1993) "A computational theory of spectral cue localization", *J Acoust Soc Am* 94 1323-1331

Zipser D and Andersen RA (1988) "A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons", *Nature* 331 679-684

Zwiers MP, Van Opstal AJ, and Cruysberg JRM (2001) "A spatial hearing deficit in early-blind humans", *J Neurosci* 21 RC142 1-5

Zwiers MP, Van Opstal AJ, and Paige GD (2003) "Plasticity in human sound localization induced by compressed spatial vision", *Nat Neurosci* 6 175-181

Summary

Chapter 2

This chapter investigates the effect of sound duration and stimulus level on sound localization. The localization of sounds in the vertical plane (elevation) was found to deteriorate for short-duration wide-band sounds at moderate to high intensities (Hofman and Van Opstal, 1998, Macpherson and Middlebrooks, 2000). The effect is described by a systematic decrease of the elevation gain (slope of the stimulus-response relation) at short sound durations. Two hypotheses have been proposed to explain this finding. The first assumes that to compute sound-source elevation, the sound localization system integrates the sound input over a time window of at least 40-80 ms. For very short sounds, the time window is too short to accurately extract the spectral localization cues and make a reliable estimate of sound-source elevation (*neural integration hypothesis*). Alternatively, the auditory system may fail to resolve spectral details at high sound intensities due to saturation of cochlear excitation patterns (*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.

In this study we tested these predictions over a larger range of stimulus parameters than has been done so far. Stimulus durations ranged from 3 to 100 ms and sound levels ranged from 26 to 73 dB SPL. Listeners responded with rapid head movements to noise bursts in the two-dimensional frontal space.

Results showed that the elevation gain decreased for short noise-bursts at all sound levels, although the effect was most conspicuous for the lowest and highest stimulus levels. This finding supports the integration model. However, elevation gain for short noise bursts also decreased at high sound levels, which is in line with the adaptation hypothesis. Our finding that elevation gain varied as a function of sound level for all sound durations is predicted by neither model. We conclude that both mechanisms underlie the elevation gain effect and propose that the conceptual neural-integration model of Hofman and Van Opstal (1998)

can be extended to reconcile these findings

In this model, elevation gain is partly determined by the confidence level of the localization system's estimate of sound-source elevation In the absence of any information, the auditory system assumes a default elevation based on other factors like prior knowledge As evidence for the veridical elevation accumulates, the confidence level increases and the elevation estimate will be increasingly dominated by the veridical elevation The confidence level may be affected by both sound duration and sound level

Chapter 3

Human sound localization in the vertical plane (elevation) relies on an analysis of the complex spectral-shape cues provided by filtering by the pinnae However, the spectrum arriving at the eardrum (the sensory spectrum) is defined by a convolution of the sound-source signal and direction-dependent filtering by the pinnae, both of which are unknown In this chapter we study how the auditory system solves this ill-posed problem to derive an estimate of sound-source elevation from the sensory spectrum To test different spectral localization models, we presented listeners with broad-band noise stimuli with randomly shaped rippled amplitude spectra, emanating from a fixed speaker position The ripple bandwidth was varied between 1 5 and 5 0 cycles/octave

Six listeners participated in the experiments Despite the fixed speaker position, localization responses were found to vary greatly in elevation in a systematic way, depending on sound-source spectrum Based on the distributions of localization responses toward the individual stimuli, we calculated a maximum-likelihood estimate for each stimulus By weighting each sensory spectrum by its maximum-likelihood estimate, we estimated the listeners' spectral-shape cues underlying their elevation percepts

The resulting reconstructed spectral features appeared to be invariant to the considerable variation in ripple bandwidths, and for each listener they had a remarkable resemblance to his/her idiosyncratic head-related transfer functions (HRTFs) These results are not in line with models that rely on the detection of a single peak or notch in the amplitude spectrum, nor with a local analysis of first- and second-order spectral derivatives Instead, our data support a model in which the auditory system performs a cross-correlation analysis between the sensory input at the eardrum and stored representations of HRTFs to determine the perceived elevation angle

Chapter 4

Human sound localization is based on implicit acoustic cues that are head centered as the ears are fixed to the head To make an eye-head (i e gaze eye in space) response to an auditory target, eye position in the head has to be taken into account Moreover, to make a goal-directed localization movement in situations where the eyes and head have moved between target presentation and the localization response, these intervening movements have to be compensated for to create an accurate representation of the sound location This may be achieved by continuously updating the acoustic information on sound location with feedback signals about the position of the eyes and head Alternatively, the auditory target coordinates could be updated in advance by using efference information based either on the preprogrammed gaze-motor command, or on the sensory target coordinates to which the intervening gaze shift is made ("predictive remapping") So far, previous experiments cannot dissociate these alternatives

Here we use two different versions of an auditory-visual double-step paradigm In the static double-step condition (Goossens and Van Opstal, 1999) a brief visual and auditory target were presented shortly after each other, before initiation of eye-head movement The listener's task was to make gaze shifts to both targets in their order of presentation For the dynamic condition we slightly changed this paradigm and presented the second, auditory, target *while* listeners made a saccadic eye-head gaze shift toward the previously flashed first, visual, target To make an accurate eye-head gaze shift to the auditory target, the auditory system has to compensate for ongoing saccadic eye and head movements in two dimensions (2D) that occur during target presentation Moreover, the system has to deal with dynamic changes of the acoustic cues, as well as with rapid changes in relative eye and head orientation that cannot be preprogrammed by the audiomotor system

Our results show that localization responses under these dynamic conditions remain accurate The distributions of the endpoints of the saccades toward the auditory target were similar for single-step, static double-step, and dynamic double-step conditions Multiple linear regression analysis revealed that the intervening eye and head movements are fully accounted for Moreover, elevation response components were more accurate for longer-duration sounds (50 ms) than for extremely brief (3 ms) sounds, for all localization conditions, as was previously found for single-step conditions (see Chapter 2) Taken together, these results cannot be explained by a predictive remapping scheme Rather, we conclude that the human auditory system adequately processes dynamically varying acoustic cues that result from self-initiated rapid head movements to construct

a stable representation of the target in an absolute, supraretinal (e g world) reference frame This signal is subsequently used to program accurate eye and head localization responses

Chapter 5

In this chapter we replicated the experiments of Chapter 4 with a visual-visual double-step condition We investigated whether the visuomotor system, like the audiomotor system, accounts for saccadic eye-head movements that occur during target presentation In contrast to auditory stimuli, visual stimuli are initially represented in a retinotopic reference frame Moreover, in the dynamic double-step condition, the system has to deal with fast dynamic changes of the retinal input

We performed visual-visual double-step experiments in which a brief (50 ms) stimulus was presented during a saccadic eye-head gaze shift toward a previously flashed visual target Our results show that also for visual-visual double-step conditions, gaze shifts remain accurate under these dynamic conditions, even for stimuli presented near saccade onset Moreover, eyes and head appear to be driven in oculocentric and craniocentric coordinates, respectively, instead of by a common motor signal

These results cannot be explained by a predictive remapping scheme We propose that, like the auditory system, the visuomotor system adequately processes dynamic changes in input that result from self-initiated gaze shifts, to construct a stable representation of the target in an absolute, supraretinal (e g world) reference frame Predictive remapping may subserve transsaccadic integration, thus enabling the perception of a stable visual scene despite eye movements, while dynamic feedback ensures accurate actions (e g eye-head orienting) to a selected goal

Samenvatting

Hoofdstuk 2

Zowel geluidsduur als geluidsniveau zijn van invloed op de localisatie van geluiden. Uit eerdere studies is gebleken dat geluidslocalisatie in het verticale vlak (elevatie) verslechtert voor korte geluiden met een matig tot hoog geluidsniveau (Hofman and Van Opstal, 1998; Macpherson and Middlebrooks, 2000). Dit effect wordt gekenmerkt door een systematische afname van de elevatiegain (helling van de stimulus-respons relatie) voor korte geluidsduren. Om deze bevindingen te verklaren zijn in de literatuur twee verschillende hypothesen voorgesteld. De eerste hypothese is gebaseerd op de aanname dat het localisatiesysteem het binnenkomende geluid integreert over een tijdsinterval van ten minste 40-80 ms om de elevatie van de geluidsbron te berekenen. Voor zeer korte geluiden (korter dan 40-80 ms) is dit integratie-interval te kort om de spectrale localisatiecues nauwkeurig uit het signaal te kunnen halen en een betrouwbare schatting te maken van de elevatie van de geluidsbron (*neurale integratie hypothese*). De tweede hypothese veronderstelt dat het auditief systeem bij hoge geluidsniveaus niet in staat is om spectrale details uit het geluidssignaal waar te nemen door verzadiging van cochleaire excitatiepatronen (*adaptatie hypothese*). Aan de ene kant voorspelt het neurale integratiemodel dat elevatiegain onafhankelijk is van geluidsniveau, terwijl aan de andere kant volgens de adaptatie hypothese de lage elevatiegain het gevolg is van een combinatie van een korte geluidsduur en een hoog geluidsniveau.

In deze studie hebben we de voorspellingen van beide modellen getest over een grotere range van stimulus parameters dan in vorige studies gebeurd is. De stimulusduur varieerde van 3 tot 100 ms en het geluidsniveau varieerde van 26 tot 73 dB SPL. Luisteraars maakten snelle hoofdbewegingen naar de waargenomen locatie van de stimuli in het twee dimensionale frontale vlak.

De resultaten van deze experimenten laten zien dat de elevatiegain afnam voor korte ruizen bij alle geluidsniveaus, hoewel het effect het duidelijkst was voor de laagste en hoogste geluidsniveaus. Deze bevindingen ondersteunen het

integratiemodel Het feit dat de elevatiegain ook afnam voor hoge geluidsniveaus komt overeen met de voorspellingen van de adaptatie hypothese Echter, onze bevinding dat elevatiegain varieerde als functie van geluidsniveau voor alle geluidsduren wordt door geen van beide modellen voorspeld Uit deze resultaten concluderen we dat beide mechanismen (neurale integratie en adaptatie) ten grondslag liggen aan het elevatiegain effect Om de bevindingen van deze studie te verenigen stellen we een uitbreiding voor van het conceptuele neurale integratie model van Hofman and Van Opstal (1998)

In dit model wordt elevatiegain mede bepaald door de mate van zekerheid die het localisatiesysteem heeft over de schatting van de elevatie van de geluidsbron In de afwezigheid van akoestische informatie neemt het auditief systeem een default elevatie aan, gebaseerd op andere factoren, zoals voorkennis en eerdere ervaring Wanneer de evidentie voor de ware elevatie toeneemt, neemt de zekerheid over de schatting van elevatie toe en zal deze schatting in toenemende mate gedomineerd worden door de ware elevatie De mate van zekerheid over de geschatte elevatie kan beinvloed worden door zowel de geluidsduur als het geluidsniveau

Hoofdstuk 3

Mensen localiseren geluiden in het verticale vlak (elevatie) op basis van een analyse van de complexe spectrale cues die ontstaan door filtering door de oorschelpen Het geluidsspectrum dat op het trommelvlies aankomt (het sensorisch spectrum) wordt gedefinieerd door een convolutie van het bronsignaal van het geluid en de richtingsafhankelijke filtering door de oorschelpen Deze zijn beide onbekend, waardoor het voor het auditief systeem een probleem is te bepalen welke spectrale eigenschappen afkomstig zijn van het bronsignaal en welke van de richtingsafhankelijke overdrachtskarakteristiek van het oor (HRTF hoofd gerelateerde overdrachtsfunctie) In dit hoofdstuk bestuderen we hoe het auditief systeem dit probleem oplost, om een schatting van de geluidselevatie te kunnen maken op basis van het sensorisch spectrum Om verschillende modellen voor elevatielocalisatie gebaseerd op spectrale cues te testen hebben we luisteraars stimuli laten horen bestaande uit breedbandige ruis met fluctuerende amplitude spectra met een random omhullende De stimuli werden gepresenteerd door een luidspreker met een vaste positie De fluctuatiebandbreedte van de amplitude spectra varieerde tussen 1 5 en 5 0 cycli per octaaf

Zes luisteraars namen deel aan de experimenten Ondanks de vaste luidsprekerpositie bleken de localisatieresponsies aanzienlijk en op systematische wijze te varieren in elevatie, afhankelijk van het geluidsspectrum Voor elke stimulus is een maximum-likelihood schatting berekend, gebaseerd op de distributie van

localisatieresponsies naar de individuele stimuli. Door elk sensorisch spectrum te wegen met de bijbehorende maximum-likelihood schatting, hebben we een schatting gemaakt van de spectrale cues die door luisteraars gebruikt worden om tot een elevatiepercept te komen.

De resulterende gereconstrueerde spectrale eigenschappen lijken onafhankelijk te zijn van de aanzienlijke variatie in fluctuatiebandbreedte. Bovendien hebben deze spectrale vormen voor elke luisteraar een opmerkelijke overeenkomst met zijn of haar HRTFs. Deze resultaten zijn niet in overeenstemming met modellen die gebaseerd zijn op de detectie van een enkele piek of dal in het amplitude spectrum, of modellen die uitgaan van een locale analyse van de eerste of tweede orde spectrale afleiding. Onze data ondersteunen daarentegen een model waarin het auditief systeem een statistische vergelijking maakt (door middel van cross-correlatie) tussen het sensorisch spectrum op het trommelvlies en in het brein opgeslagen representaties van HRTFs om de waargenomen elevatiehoek te bepalen.

Hoofdstuk 4

Voor mensen is geluidslocalisatie gebaseerd op impliciete akoestische cues Aangezien de oren vast zitten aan het hoofd, zijn deze cues gerepresenteerd ten opzichte van het hoofd (hoofd gecentreerd). Om een oog-hoofd (oog in de ruimte: gaze genoemd) beweging naar een auditief doel te maken, moet een oog-hoofd vector tot de doelpositie berekend worden waarin de oogpositie in het hoofd verrekend wordt. Bovendien, om een doelgerichte localisatiebeweging te maken in situaties waarin oog- en hoofdpositie veranderd zijn tussen de presentatie van het auditieve doel en de localisatieresponsie, moeten deze tussenliggende bewegingen gecompenseerd worden om de locatie van de geluidsbron nauwkeurig te kunnen berekenen. Twee mogelijke manieren waarop deze compensatie tot stand zou kunnen komen zijn enerzijds een continue updating van de binnenkomende akoestische informatie over de locatie van het bronsignaal door middel van feedback signalen over oog- en hoofdpositie. Anderzijds zouden de coordinaten van het auditieve doel geupdate kunnen worden voor aanvang van de oog-hoofdbeweging door gebruik te maken van efferente informatie over het voorgeprogrammeerde neurale gaze-motorcommando (de zogenaamde efferente kopie), of over de sensorische doelcoordinaten waar de tussenliggende oog-hoofdbeweging naar is gericht ("predictive remapping"). Voorgaande experimenten deze twee alternatieve hypothesen niet kunnen dissociëren.

In deze studie testen we twee verschillende versies van een auditief-visueel dubbelstap paradigma. In de statische dubbelstap conditie (Goossens and Van Opstal, 1999) werd een kort visueel en een kort auditief doel snel na elkaar aan-

geboden, voor aanvang van de oog-hoofdbeweging De taak van de luisteraar was om snelle oog-hoofdbewegingen te maken naar beide doelen in de volgorde waarin ze aangeboden waren Voor de dynamische conditie pasten we dit paradigma enigszins aan en werd het tweede, auditieve, doel aangeboden terwijl luisteraars een snelle oog-hoofdbeweging maakten naar het eerste, visuele, doel Om een accurate oog-hoofdbeweging te maken naar het auditieve doel, moet het localisatiesysteem compenseren voor snelle oog- en hoofdbewegingen in twee dimensies (2D) naar het eerste doel, die plaatsvinden terwijl het auditieve doel aangeboden wordt Bovendien heeft het systeem door de hoofdbeweging te maken met dynamische veranderingen van de akoestische cues, en met snelle veranderingen in relatieve oog- en hoofdorientatie die niet voorgeprogrammeerd kunnen worden door het audiomotor systeem

Onze resultaten laten zien dat localisatieresponsies onder deze dynamische conditie accuraat blijven De verdeling van de eindpunten van de oog-hoofdbeweging naar het auditieve doel waren vergelijkbaar voor enkelstap, statische dubbelstap en dynamische dubbelstap condities Een multipele lineare regressie analyse toonde aan dat de tussenliggende oog- en hoofdbewegingen volledig gecompenseerd werden Bovendien waren responsies in elevatie nauwkeuriger voor geluiden met een langere geluidsduur (50 ms) dan voor extreem korte (3 ms) geluiden voor alle localisatiecondities Dit is in overeenstemming met eerdere studies met enkelstap condities (zie Hoofdstuk 2) Samenvattend kunnen deze resultaten niet verklaard worden door een model gebaseerd op predictive remapping We concluderen dat het menselijk auditief systeem dynamisch varierende akoestische cues die resulteren uit snelle hoofdbewegingen adequaat kan verwerken om tot een stabiele representatie van het auditieve doel te komen in een absoluut, supraretinaal (wereld) referentiekader Dit signaal wordt vervolgens gebruikt om accurate oog en hoofd localisatieresponsies te programmeren

Hoofdstuk 5

In dit hoofdstuk hebben we de experimenten van Hoofdstuk 4 gerepliceerd met een visueel-visuele dubbelstap conditie We hebben onderzocht of het visuomotor systeem, net als het audiomotor systeem, kan compenseren voor snelle oog-hoofdbewegingen die plaatsvinden tijdens presentatie van het te localiseren doel, in de berekening van de oog-hoofdvector om dat doel te bereiken In tegenstelling tot auditieve stimuli worden visuele stimuli initieel gerepresenteerd in een retinotopisch referentiekader Bovendien moet het localisatiesysteem in de dynamische dubbelstap conditie rekening houden met snelle dynamische veranderingen van het binnenkomende retinale signaal

We hebben visueel-visuele dubbelstap experimenten uitgevoerd waarin een

korte (50 ms) visuele stimulus gepresenteerd werd vlak voor (statisch) of tijdens (dynamisch) een snelle oog-hoofdbeweging naar een eerste visueel doel dat kort daarvoor gepresenteerd was Onze resultaten laten zien dat ook voor visueel-visuele dubbelstap condities oog-hoofdbewegingen accuraat blijven onder deze dynamische condities, zelfs voor stimuli die gepresenteerd worden rond onset van de oog-hoofdbeweging Bovendien lijken oog en hoofd gedreven te worden in oog-gecentreerde en hoofd-gecentreerde coordinaten respectievelijk, in plaats van door een gezamenlijk signaal

Deze resultaten kunnen niet verklaard worden door een model gebaseerd op predictive remapping We stellen hier daarom dat, net als het auditief systeem, het visuomotor systeem dynamische veranderingen in input die resulteren uit oog-hoofdbewegingen, adequaat kan verwerken om zo tot een stabiele representatie van het doel in een absoluut, supraretinaal (wereld) referentiekader te komen Predictive remapping zou een belangrijke rol kunnen spelen bij transsaccadische integratie, om zo de waarneming van een stabiele visuele omgeving mogelijk te maken ondanks oogbewegingen, terwijl dynamische feedback zorgt voor accurate acties (oog-hoofd orientatie) naar een geselecteerd doel

Dankwoord

Al staat alleen mijn naam op de kaft van dit boekje, meerdere mensen hebben op verschillende manieren bijgedragen aan mijn proefschrift en die wil ik hier graag bedanken

Allereerst natuurlijk John je luide enthousiasme is erg motiverend en door jouw scherpe blik zat er vaak meer in mijn data dan ik dacht Stan, jij bent op de achtergrond altijd geinteresseerd in het wel en wee binnen de afdeling en betrokken bij de voortgang van alle aio's

De data in dit proefschrift zijn deels verzameld door mijn studenten Frank en Tom, bedankt voor de goede en leuke samenwerking En Tom, ik vind het heel leuk dat jij nu letterlijk op mijn plek zit als aio bij John Thamar, hoofdstuk 3 is gebaseerd op jouw afstudeeronderzoek, heel erg bedankt voor al je werk En ik vind het erg gezellig dat we nu weer kamergenootjes zijn op het AMC!

Zonder proefpersonen zouden er geen data geweest Allemaal bedankt voor het geduldig luisteren en kijken naar vele ruisjes en lichtjes onder barre omstandigheden in een donker muf hok, met een helmpje op en soms ook nog met een oogspoel in

Mijn experimenten waren niet mogelijk geweest zonder de twee opstellingen die door Hans en Ton gebouwd zijn en werkend gehouden werden Bovendien kon ik ook altijd bij jullie terecht voor goede tips voor beginnende doe-het-zelvers en heb ik van Hans leren solderen, waar ik op het AMC nog veel profijt van heb gehad Ger en Gunter zorgden ervoor dat mijn computer meestal deed wat ik wilde zodat ik de verzamelde data kon analyseren

Judith, Margriet en Annet, zonder jullie zou alles in het honderd lopen Naast alle administratieve zaken zorgen jullie voor gezelligheid, de koffie, de nieuwste roddels, film- en boekentips, wijze raad over de dingen des levens en een luisterend oor bij een aiodip

Zonder alle collega's door de jaren heen zou mijn aiotijd een stuk minder leuk zijn geweest Biofysica is een gezellige en sportieve afdeling met veel goede tradities, zoals de gezamenlijke koffiepauzes, het fys-a-fys batavierenteam, het dagje uit en het vrijdagmiddagbiertje

Dank vooral aan mijn mede-aio's. Marc, je was van begin tot eind van mijn promotie een hele leuke kamergenoot en jouw weekendverhalen waren altijd goed om de maandagochtendsleur te doorbreken. Bovendien heb ik veel gehad aan onze gesprekken over het aio-zijn, ons onderzoek en matlab. Marjan, zonder jou was het een stuk minder gezellig geweest op de afdeling. Bovendien stond je altijd voor me klaar (en nog steeds!), zo nodig op de vreemdste uren met thee en chocola; heel erg bedankt! En ook Noël, Ronald, Martijn T, Ieke, Rens, Maaike, Sigrid, Kees, Martijn K, Joris, Onno, Marcel en alle andere aio's door de jaren heen, bedankt voor alle gezelligheid. Ik kon altijd bij jullie terecht voor matlab hulp of een goed gesprek. Maar ook voor een filmpje, samen eten, een terrasje, klimmen, hardlopen, een uitje naar de IKEA, of een the Office marathon was altijd wel iemand te vinden.

Sinds anderhalf jaar werk ik op het AMC in Amsterdam en de combinatie van baan, reizen en proefschrift is niet ideaal. Maar dankzij de leuke collega's en de goede werksfeer bij Audiologie heb ik het volgehouden. Dank vooral aan mijn kamergenootjes voor hun steun en gezelligheid. Rolph, ook bedankt voor het geduldig beantwoorden van al mijn LaTeX vragen en het gedeelde proefschriftleed.

En dan zijn er nog de mensen die niet direct hebben bijgedragen aan dit proefschrift, maar wel belangrijk zijn in mijn leven buiten het onderzoekswereldje

Zeker in de laatste 2 jaar heeft mijn sociale leven nogal te lijden gehad onder de proefschriftstress Daarom wil ik mijn vrienden bedanken voor hun begrip als ik weer eens geen tijd had omdat dit proefschrift af moest (kijk, het is er nu!). In het bijzonder Lydia, Najoua, Hilde, Ina en Saskia, bedankt voor jullie vriendschap, jullie steun als het even niet meezat, alle gezelligheid, de lange gesprekken en het gezamenlijke miepen

Edward en Eveline, jullie zijn me al voorgegaan als dr. Bedankt voor jullie steun (en de toevoeging 'in air'!) en het leuke vakantieadresje in Zürich.

Mijn ouders wil ik heel erg bedanken voor hun steun en vertrouwen. Jullie hebben me altijd gestimuleerd om na te denken en nieuwsgierig te zijn.

Kees, dankzij jou waren de laatste 2 jaar van mijn promotie behalve druk ook heel erg leuk. Bovendien zorgde jij voor ontspanning en rust temidden van alle stress.

Curriculum Vitae

Ik ben op 16 mei 1974 in Tilburg geboren. In 1992 behaalde ik mijn vwo diploma aan de Philips van Horne Scholengemeenschap in Weert. Na een propedeuse Italiaans aan de Universiteit Utrecht ben ik overgestapt op fonetiek. Deze studie heb ik afgerond met een afstudeerstage op het onderwerp auditieve streaming onder begeleiding van Andrew Oxenham, aan het Instituut voor Perceptie Onderzoek (IPO) in Eindhoven. Na mijn afstuderen in 1997 ben ik een jaar naar Cambridge gegaan om in het lab van Brian Moore mijn onderzoek voort te zetten. Na een korte omweg via de psycholinguistiek ben ik in september 2000 in de groep van John van Opstal aan mijn promotieonderzoek begonnen, bij de afdeling Biofysica aan de Radboud Universiteit Nijmegen. Daar heb ik de afgelopen 6 jaar aan dit proefschrift gewerkt. Sinds maart 2005 ben ik werkzaam bij de afdeling Klinische en Experimentele Audiologie van het AMC in Amsterdam, in de groep van Wouter Dreschler.