MONAURAL ADAPTIVE MECHANISMS IN HUMAN SOUND LOCALIZATION

Een wetenschappelijke proeve op het gebied van de Natuurwetenschappen, Wiskunde en Informatica

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,
vvolgens besluit van het College van Decanen
in het openbaar te verdedigen op maandag 26 februari 2007
om 13.30 uur precies

door

MARC MATHIJS VAN WANROOIJ

geboren op 24 november 1976
te Veghel
CONTENTS

CHAPTER I  
GENERAL INTRODUCTION  
  1.1 General Introduction  
  1.2 Audition versus Vision  
  1.3 Sound Localization Cues  
  1.4 Interaural Time Differences  
  1.5 Interaural Level Differences and Head Shadow  
  1.6 Spectral Cues  
  1.7 From Cues to Location Percept  
  1.8 The monaural duplex theory  
  1.9 Azimuth Cue Integration and Weighting  
  1.10 Elevation Cue Integration and Weighting  
  1.11 Calibration of Cues  
  1.12 Multi-sensory Integration  
  1.13 General Overview  

CHAPTER II  
CONTRIBUTION OF HEAD SHADOW AND PINNA CUES TO CHRONIC MONOURAL SOUND LOCALIZATION  
  2.1 Introduction  
  2.2 Methods  
    Monaural and binaural listeners  
    Apparatus  
    Auditory stimuli  
    Paradigms  
    Data analysis  
  2.3 Results  
    The influence of intensity on monaural sound-source azimuth localization  
    The influence of spectral cues on monaural sound-source azimuth localization  
    Monaural sound-elevation localization  
    Training to localize azimuth monaurally  
  2.4 Discussion  
  Acknowledgements  

CHAPTER III  
SOUND LOCALIZATION UNDER PERTURBED BINAURAL HEARING  
  3.1 Introduction  
  3.2 Methods  
    Participants  
    Apparatus  
    Auditory stimuli  
    Plugs  

1  11  33

2  2  12

2  2  12

4  4  14

5  5  14

5  5  15

15

18

18

22

25

27

28

31

34

35

35

35

36

37
Measurement of audiograms  37  
Paradigms  37  
Hearing Conditions  37  
Data analysis  37  
3.3 Results  39  
Acute effects of a unilateral plug  39  
Influence of sound intensity and source azimuth on azimuth localization  40  
Influence of intensity and azimuth on elevation localization  45  
Weighting of cues  48  
3.4 Discussion  50  
Spectral cues for azimuth  50  
Azimuth cues for elevation  51  
Monaural Listening  51  
Integration of acoustic cues  52  
Neural mechanisms  52  
Acknowledgements  53  

**Chapter IV**  55  
Relearning Sound Localization With A New Ear  
4.1 Introduction  56  
4.2 Methods  58  
Participants  58  
Apparatus  58  
Auditory stimuli  58  
Molds  59  
Measurements of directional transfer functions  60  
DTFs and their correlations  60  
DTF similarity index  60  
Paradigms  61  
Hearing conditions  61  
Front-back reversals and confusions  62  
Data analysis  62  
4.3 Results  63  
Acute effects of a unilateral spectral perturbation  63  
Conflicting spectral shape cues  67  
Adaptation to a unilateral mold  68  
Front-back reversals  72  
Test of the adaptation models  72  
4.4 Discussion  75  
Summary  75  
Comparison to other studies  75  
Learning behavior  75  
Potential mechanisms  76  
Other factors  76  
Calibration  77  

Outlook and applications 77
Acknowledgements 77

CHAPTER V
AUDITORY-VISUAL INTERACTIONS SUBSERVING GOAL-DIRECTED SACCADIES IN A COMPLEX SCENE

5.1 Introduction 80
5.2 Methods 82
Subjects 82
Apparatus 82
Stimulus Generation 82
Paradigms 83
Data analysis 84
5.3 Results 86
Properties of unimodal V-saccades and A-saccades 86
Properties of AV-saccades (no temporal asynchronies) 89
AV interactions as a function of stimulus timing 93
AV-control experiment 97
5.4 Discussion 97
Consideration of mechanisms other than neural integration 98
Rules for multisensory integration of bimodal signals and comparison to previous work 100
Conceptual model of auditory-visual interactions in a complex scene 101
Conclusions 103
Acknowledgements 103

SUMMARY 105

SAMENVATTING 107

BIBLIOGRAPHY 111

DANKWOORD 121

CURRICULUM VITAE 123
CHAPTER I
GENERAL INTRODUCTION
1.1 General Introduction

Most of us localize sounds without giving a second thought about it. For example, when hearing your name shouted in a crowded room, you might turn around. Your name is called again, and this time you perceive the voice from coming straight ahead, indicating that your first attempt to localize the person shouting at you has been quite successful. Another shout and your eyes are directed immediately to your friend. Without much consideration you have performed a complicated process, whereby various localization cues (e.g. arrival time and level differences between both ears, and spectral shape properties of the perceived sound) are used to attempt to form a single accurate percept of a sound source location. This attempt is not always successful as most people with monaural hearing problems might attest (see Chapter 2). In this PhD thesis several behavioral experiments are described, in which binaural and monaural localization cues are perturbed, in order to investigate the contribution of each of the cues to sound localization. In addition to the acute effects (Chapters 3, 4, 5), the long-term effects of an abnormal cue arrangement are investigated in search for monaural adaptive plasticity in both normal-hearing people (Chapter 4) and unilaterally deaf patients (Chapter 2). Each of the experiments is described in their own chapter in this thesis, accompanied by their own specific introduction. This general introduction expands upon those introductions to broaden the theoretical basis needed for the appreciation of these experiments combined and summarizes the primary results of the experiments.

1.2 Audition versus Vision

The above example (localization of a shout in a crowded room) does not explain why sound localization is a complicated process. It does illustrate however that sound localization might often be used to direct our eyes towards an interesting event. Sounds can be perceived from all around us, while vision is limited to a narrow field. Nevertheless, those eyes can distinguish very small differences in position of roughly 5-10 seconds of arc (Westheimer and McKee 1977), while our ears are “only” able to distinguish between positions differing about 1-2° (Mills 1958). Thus, it does not seem to be too far-fetched to assume that an important facet of sound localization is telling the eyes where to look. The fact that animals with narrow fields of best vision (like humans, macaques, and cats) are considerably more precise in localizing sounds compared with animals with broad visual streaks (like horses and rabbits), whose field of best vision encompasses much of the entire panorama (Heffner 2004), seems to support this assumption.

One can better appreciate the accuracy of the human sound localization system, even though it is worse than the visual localization system (within the limited confines of the visual field, of course!), if one understands the complexity of sound localization versus the simplicity of visual localization. In fact, visual localization is almost as simple and as straightforward as it can get. The retina (the visual sensory organ) is topographically organized: each neighboring point in the visual world corresponds to a neighboring point on the retina. No complicated processing has to take place in order to obtain position information. The cochlea (the auditory sensory organ), on the other hand is tonotopically organized: a point on the basilar membrane of the cochlea will peak for a tone of a certain frequency, regardless of its position in space. The cochlea acts as a frequency analyzer, and any explicit position information of the sound is lost.

1.3 Sound Localization Cues

Fortunately, implicit position information is encoded, and can be made explicit by combining
the information from both ears and by integrating information from the different frequency channels. Since humans are most sensitive for sound frequencies between 2 and 4 kHz, it might come as a surprise, then, that humans have the best localization performance for frequencies below 1.5 kHz and above 5 kHz (Stevens and Newman 1936). Almost a century ago, Lord Rayleigh (1907) already proposed the duplex theory of binaural hearing, stating that two different mechanisms existed for localizing azimuth (the horizontal direction of a sound); one operating at low frequencies, the other at high frequencies. This dichotomy arises simply due to physical characteristics of the sound, as will be described in the following two sections.

1.4 **Interaural Time Differences**

The low-frequency mechanism in the duplex theory is mediated by timing differences in arrival time of the sounds at both ears (interaural time differences, or ITDs); sound arrives at the ear nearest to the sound-source earlier than at the far ear, due to the path-length difference between both ears and the finite sound velocity (~340 m/s). The maximum ITD in humans is ~700 μs for extreme lateral sound-source positions (Wightman and Kistler 1992). While differences in onset time of a sound might yield a clue to the possible sound location, for ongoing sounds the auditory system relies on differences in phase (interaural phase differences, IPDs). However, for pure tones, IPDs can act as an appropriate localization cue only for frequencies up to 1500 Hz, whereas for higher frequencies the wave-length will become smaller than head diameter. Indeed, the auditory system seems to ignore IPDs as a potential localization cue for high-frequency sounds (for complex sounds: Macpherson and Middlebrooks 2002; for pure tones: Zwislocki and Feldman 1956).

1.5 **Interaural Level Differences and Head Shadow**

The mechanism for azimuth localization of high-frequency sounds according to the duplex theory is mediated by interaural level differences (ILDs). These differences in sound level between both ears are partly due to the reflection of sound waves by the head, essentially casting a shadow around the head. Relative to the ear facing the sound, this head shadow effect (HSE) can attenuate sounds at the far ear as much as 25 dB, depending on frequency (Irvine 1986). Figure 1.1 shows the directional effects of the HSE measured in the left ear of one listener; a sound 50° of the midline will not only be attenuated in the far ear by ~10 dB, due to the sound-collecting capabilities of the pinnae sounds at the near ear will be amplified. Since the HSE magnitude depends in a systematic way on azimuth, it could be used as a monaural intensity cue for azimuth location. However, it is an invalid cue when sound levels are varied; e.g. a perceived level of 60 dB in the right ear might simply arise from a sound straight ahead with an actual sound level of 60 dB, or from a 50 dB sound 50° to the right of the midline, or from a 65 dB sound 30° to the left of the midline, etc. Only when listeners have foreknowledge about the absolute intensity of the sound (e.g. regular speech), can the head shadow be used as a valid localization cue. For sounds that vary widely in level, listeners can compare the sound level information in both ears to obtain ILDs. These ILDs are a valid binaural azimuth localization cue regardless of the absolute sound level of the sound-source. In contrast to the ITDs, the ILDs are effective localization cues only for higher frequencies (above 1500-2000 Hz). At lower frequencies, the head causes only little attenuation.
1.6 Spectral Cues

The duplex theory might resolve the problem of localizing the azimuth of a sound-source; however the interaural differences are negligible when trying to determine the elevation (vertical direction) of the sound-source (i.e. both ears of humans are usually placed at the same height). Furthermore, the ITDs and ILDs cannot distinguish between the front and back direction. This leads to front-back confusions as observed in many behavioral experiments. Fortunately, the strangely-shaped pinna (Fig. 1.2) is shaped nicely to differentially reflect and diffract sounds from different directions. The spectra of the sounds will contain complex direction-dependent features, referred to as spectral cues (Fig. 1.3), at frequencies above 3-4 kHz. And these cues enable listeners to determine elevation and front-back direction. These cues are available monaurally as both ears reshape the sound spectrum separately and furthermore they are, contrary to the HSE, level-independent due to their frequency-dependency.

1.7 From Cues to Location Percept

To obtain an accurate and coherent percept of sound location, the localization cues need to
Figure 1.2. The human pinna or auricula and the nomenclature of its various features. Taken from Gray’s Anatomy 1918.

Figure 1.3. Elevation Directional Transfer Functions (DTFs). A. DTFs of listener JO’s right ear for several elevations with different shades of gray coding for elevation. B. Contour plot of the elevation DTFs of the same ear with gray-scale coding for power (in dB). Note the clear direction- and frequency-dependent features, especially the prominent notch running from elevation -50 deg and frequency 5 kHz to elevation 50 deg and frequency 10 kHz.
General Introduction

be mapped and calibrated to veridical sound-source positions. In the initial stages of the auditory pathways, this processing is taking place separately for each of the cues in their own brainstem nuclei (ITDs in medial superior olive [MSO], ILDs in lateral superior olive [LSO] and spectral cues in the dorsal cochlear nucleus [DCN]). This separate processing does not withhold the auditory localization system to obtain a single spatial percept (e.g. Hafter et al. 1989). Thus, the localization cues have to be fused at higher-level brain structures to form an integrated percept. The inferior colliculus (IC) seems to be a prime candidate for this integration, as all efferent pathways from the cochlear nucleus, the MSO and the LSO converge on this midbrain structure. And there is indeed neurophysiological evidence that the IC is an important initial stage for combining the cues (e.g. Chase and Young 2005), although convergence also seems to take place in the higher-level auditory cortex (Yamada et al. 1997). The experiments in this thesis take a behavioral approach to elucidate several aspects of this cue convergence and mapping. This was done by observing the localization behavior of monaural listeners or by actively manipulating the various localization cues of normal-hearing binaural listeners, as will be discussed in the following sections.

1.8 THE MONOAURAL DUPLEX THEORY

The duplex theory simply cannot hold for the unilateral deaf, as they do not have access to any binaural information. Still, there are many reports on an astounding performance of some long-term monaurally deaf people to localize sound-source azimuth (Humes et al. 1980; Newton 1983; Newton and Hickson 1981; Slattery and Middlebrooks 1994; Viehweg and Campbell 1960). In the literature it has been implicitly or explicitly assumed that for these listeners the monaural spectral cues have taken over the role of the binaural difference cues. In Chapter 2 of this thesis, another possibility is forwarded, namely that the unilateral deaf use the HSE as an indicator of azimuth. The experiment utilized a large range of intensities (extending the limited range of intensities in previous monaural studies) confounding the HSE as a reliable cue (as detailed above). All unilateral deaf showed very poor localization capabilities, and responses were influenced by sound level; low sound levels elicited listener responses towards the deaf ear, and higher sound levels elicited listener responses towards the hearing ear. Several of the monaural listeners in this study even commented that the difference between sound level and sound direction is moot; a dull sound had to originate from a source near the deaf ear.

The benefit of using the ambiguous intensity cues over the correct spectral cues is that the intensity cues can easily be learned in a simple environment, as shown in this thesis. Still, for some of the monaural listeners spectral cues do contribute to azimuth localization, and their performance on the hearing side exceeds those of the other listeners.

Another localization cue, one that was not investigated in this thesis, is the dynamic position cue provided by head movements. For static conditions, low-frequency and tonal sounds possess little to no usable spectral cues. Listeners will therefore have trouble determining elevation and front-back direction for these sounds. Yet, by turning their heads during sound-presentation, the acoustic input will change as a result of these head-movements (Vliegen et al. 2004). This can aid normal-hearing binaural listeners in resolving front-back ambiguities (Perrett and Noble, 1997a) and determine elevation (Perrett and Noble, 1997b). In theory, monaural listeners might also benefit from these dynamic position cues. In fact, some of the monaural listeners that participated in the experiments described in chapter 2 commented that they use head movements to localize sounds in every-day life. In this thesis, however, the limited presentation time of the stimuli restricted them from doing so.

1.9 Azimuth Cue Integration and Weighting
As suggested earlier, the three major localization cues (ITDs, ILDs and spectral cues) need to be integrated to form a single percept. Still, some of the cues seem to contribute more to the percept than others. For example, for broad-band sounds an azimuth percept can be derived from both ITDs and ILDs. Still, in the presence of both low-frequency and high-frequency information, ITDs seem to be favored over ILDs, when these two cues conflict (Macpherson and Middlebrooks 2002; Wightman and Kistler 1992). Spectral cues, while containing some azimuth-information (Fig 1.1), do not seem to contribute towards localizing sound-source azimuth; many behavioral studies found that disturbing spectral cues did not lead to an appreciable decline in azimuth localization performance (Hofman and Van Opstal 2003; Hofman et al. 1998; Morimoto 2001; Oldfield and Parker 1984). And humans are not capable of determining sound-source azimuth with the binaural cues removed, even with spectral cues present (Wightman and Kistler 1997). Nevertheless, at extreme lateral positions the resolution of the binaural difference cues degrades, and it would be expected that spectral cues can contribute to an accurate azimuth percept.

Simply put, the auditory system seems to make choices between the various cues, and weighs the information obtained from the various cues. In Chapter 3 of this thesis, this weighting process is influenced by changing the accuracy and reliability of the binaural cues by inserting a plug in one of the listeners’ ears. This plug attenuates sounds (by up to ~45 dB), and profoundly affects the ILDs, while presumably leaving low-frequency ITDs almost and monaural spectral cues of the unplugged ear completely unscathed (Wightman and Kistler 1997). This kind of manipulation leads to a change in localization behavior, which is mostly characterized by a shift in localization responses towards the unplugged ear (e.g. Flannery and Butler 1988; Musicant and Butler 1984b; Oldfield and Parker 1986; Slattery and Middlebrooks 1994), as would be expected from the attenuation provided by the plug. The study in this thesis elaborates on these studies by using a larger range of sound levels. At low levels, binaural level differences are practically non-existent due to the attenuation, while higher levels contain more reliable ILDs. Since the plugged listeners localize sound-source azimuth surprisingly well for low-level sounds on the unplugged side, where spectral cues are strong and valid, it seems that the auditory system copes with this difference in reliability of cues, re-weights the various cues accordingly, and seems to ignore faulty binaural information in favor of valid spectral cues.

Another weighting-process is also seen in Chapter 2 for the monaural listeners; instead of weighting spectral and binaural cues, these listeners need to weight the spectral and HSE cues. Both the binaural and monaural weighting, despite the qualitative similarities between listeners, seem to be highly individual; some listeners outperform others by relying more on the valid spectral cues. Still, all listeners with no or perturbed binaural cues cannot completely ignore the faulty intensity cues (be they binaural, i.e. ILDs, or monaural, i.e. HSE). This illustrates the relative importance and dominance of the binaural difference cues over the spectral cues in sound azimuth localization.

1.10 Elevation Cue Integration and Weighting

Cue weighting with respect to the elevation percept is essentially limited to the reconciliation of the two sets of spectral cues from each ear, as the interaural differences in elevation are minimal. Humanski and Butler (1988) showed that spectral cues of the near ear contribute more to the elevation percept than the far ear. Morimoto (2001) and Hofman and Van Opstal (2003) extended their findings by showing that the final elevation percept is indeed a weighted average of the two perceived spectral sets, and that the relative contribution varies gradually with azimuth. No distinction was made between actual sound-source azimuth and perceived azimuth, as they were coincident in these studies. This distinction is relevant. Weighting based on actual azimuth can simply be explained by lower signal-to-noise ratios
in the far ear making it harder for listeners to extract directional information from the spectral cues of this ear. Unilaterally deaf (Chapter 2) and plugged listeners (Chapter 3) with a disturbed azimuth percept, that was different from the actual azimuth percept, would in this case still be able to localize elevation on the hearing side. The results in these chapters show that this is not the case and so show that weighting is influenced by perceived azimuth rather than by the differential acoustics of the two sets of spectral cues. When listeners (both monaural, Chapter 2, and binaural, Chapter 3) incorrectly perceive sound-source azimuth, they also incorrectly judge the sound-source elevation, while a higher accuracy in the azimuth-domain corresponds to better performance in elevation.

1.11 Calibration of Cues

The calibration for accurate sound localization seems to be highly individual, since trying to localize with another’s ears (Middlebrooks 1999) and different localization cues degrades performance severely (Hofman et al. 1998). Still head and ear size change throughout life accompanied by changes in localization cues, so recalibration also needs to be possible at a later age. In fact, the adult auditory system does seem to be able to cope with the changes in an adaptive and plastic way. Adult owls (Bergan et al. 2005; Knudsen et al. 1994) and ferrets (King et al. 2001) can adjust their localization capabilities to accommodate changes late in development. Also humans can cope with late-onset changes: changes in spectral cues (Hofman et al. 1998), binaural cues (Shinn-Cunningham et al. 1998a, 1998b) and vision (Zwiers et al. 2003) all lead to changes in auditory localization behavior.

In the Hofman et al. study (1998) listeners were equipped with molds in their conchas (Fig. 1.2) that basically changed the shape of the spectral cues (Fig. 1.3), while leaving the ear canals open and the binaural cues undisturbed. Performance in azimuth localization was unaffected by this manipulation, but elevation localization deteriorated immediately. After wearing these molds for several weeks, the listeners regained their ability to localize sound-source elevation.

Adaptation could have taken place by recalibration of the spectral cues (e.g. Fig. 1.4, mapping stage $M_\varepsilon$); listeners basically learning to correlate new spectral cues to familiar sound-source locations. Alternatively, listeners might learn to favor the ear with the least amount of changes to its spectral cues (i.e. readjusting their binaural weighting of the spectral cues [Fig. 1.4, weighting stage $W_\varepsilon$]). Simply by ignoring the new and unfamiliar spectral cues, listeners then would not have to learn a completely different spectral cue set. The experiments described in Chapter 4 suggest the former: when the spectral cues of a single ear are perturbed, listeners will learn these new spectral cues and the other ear will not take over.

1.12 Multi-sensory Integration

The attention of the final experimental chapter of this thesis (Chapter 5) is slightly expanded to include the visual system and localization of a multisensory event. Traditionally, visual information is regarded as superior and dominant over the other sensory systems. In fact, when visual information about an event is paired with conflicting auditory information, the auditory percept is often drawn to the visually perceived location, a phenomenon referred to as “visual capture” or the “ventriloquist effect”. This makes sense since visual spatial information is usually far more reliable than the auditory information; optical signals provide high-resolution topological information, while the auditory tonotopic information is easily distorted by objects in the environment.

Another view is taken in this thesis; both the visual and auditory system yield a location percept of a multisensory event, and rather than starting a conflict, both systems can help each other in creating a fast and accurate orienting response towards that event, even (or
especially) under noisy conditions. This integration of multisensory signals is far from trivial, as the audiovisual signals are encoded initially in different reference frames (e.g. Sparks and Mays 1990; Vliegen et al. 2004). Still, multisensory integration has been observed in neurons of the mammalian superior colliculus, (Stein and Meredith 1993), and behavioral experiments show that saccadic reaction times towards audiovisual targets are reduced (for review see Colonius and Arndt 2001). The experiments described in Chapter 5 show that under complex conditions the auditory and visual information of a single event can be combined to form a more accurate and faster response. The actual response behavior depends on the timing of the two signals and the signal-to-noise ratio of the auditory signal.

1.13 General Overview

Figure 1.4 presents a schematic overview of all the experiments and their results in this thesis. Basically, in every experiment listeners were exposed to deplorable localization conditions (downwards arrows in figure 1.4). In chapter 2 the binaural cues were absent, in chapter 3 they were perturbed, in chapter 4 the spectral cues were altered, and in chapter 5 noise was added to auditory and visual signals. And basically, in every experiment, the human localization system corrected for these disturbances (upwards arrows in figure 1.4). To regain sound localization performance listeners favor the best cues available and combine information from various cues (Chapters 2, 3 and 5). Listeners essentially use “the best of all worlds”. And even when the best is not good enough, listeners can adapt and learn to use new cues (Chapter 4 and 2).
Figure 1.4. Schematic overview. Azimuth percept (P_α) and elevation percept (P_ε) are obtained by mapping (M_α and M_ε, respectively) and weighting (W_α and W_ε, respectively) various localization cues. The auditory percept (P_{aud}) is combined with the visual percept (P_{vis}) to obtain a multisensory percept (P_{av}). The experimental perturbations described in the various chapters (Ch) of this thesis are represented by the downwards arrows in the gray areas. The upwards arrows represent the increase in contribution of the perturbed cues or improvement in localization percept. When binaural cues (interaural time differences, ITD; interaural level differences, ILD) become unreliable (Ch 2 and 3) the contribution of the head-shadow effect (HSE, Ch 2) and the spectral cues of the hearing ear (Sc, spectral cues for the ear contralateral to the perturbation/deafness, Ch 2 and 3) increases (upwards arrows). When the spectral cues of one ear (S_i) are altered (Ch 4), the mapping is relearned for that ear, without affecting the binaural weighting stage. W_ε is affected by the P_α (Ch 2 and 3). The P_{aud} and P_{vis} are integrated to obtain an enhanced P_{av} compared to the unimodal percepts (Ch 5).
CHAPTER II

CONTRIBUTION OF HEAD SHADOW AND PINNA CUES TO CHRONIC MONAURAL SOUND LOCALIZATION

Monaural Sound Localization

2.1 INTRODUCTION

The auditory system relies on implicit acoustic cues to encode sound location. Interaural level and time differences relate to the left-right angle of sounds (azimuth). The up-down and front-back angles (elevation) are determined by spectral pinna cues (Shaw, 1966; Musicant and Butler, 1984a; Wightman and Kistler, 1989; Blauert, 1997). Independent neural pathways in the brainstem process the different acoustic cues (Irvin, 1986; Yin, 2002; Young and Davis, 2002).

The independent encoding of azimuth and elevation has interesting implications. For example, altering the spectral cues with pinna molds abolishes elevation localization but leaves azimuth performance intact (Oldfield and Parker, 1984b; Hofman et al., 1998). Moreover, human listeners can adapt to modified spectral cues and relearn to localize elevation without changing their azimuth performance (Hofman et al., 1998). Narrow-band sounds distort elevation localization but not azimuth (Middlebrooks, 1992; Goossens and Van Opstal, 1999). Conversely, reversing binaural inputs with hearing aids reverses the azimuth responses while leaving elevation unaffected (Hofman et al., 2002).

However, spectral cues may also determine sound-source azimuth. After plugging one ear, young ferrets relearn to localize sounds, presumably by using spectral cues of the intact ear (King et al., 2000). It has been argued, however, that because plugging perturbs binaural cues in a frequency-dependent way, listeners could in principle maintain a binaural strategy by relying on low-frequency information (Wightman and Kistler, 1997).

A binaural strategy for azimuth localization is impossible for unilaterally deaf listeners. Slattery and Middlebrooks (1994) proposed that some of their monaural listeners had learned to use spectral cues of their intact ear. Yet, in a real monaural situation, perceived intensity of the sound source also relates to its azimuth because of the head-shadow effect (HSE). Although the HSE is ambiguous for unknown intensities, monaural listeners might have adopted the HSE to cope with familiar acoustic environments in daily life. So far, localization studies with monaural participants only used a small range in sound intensities (Humes et al., 1980; Newton and Hickson, 1981; Newton, 1983; Slattery and Middlebrooks, 1994), leaving it unclear to what extent the HSE may have played a role.

Here, we study to what degree the unilateral deaf rely on intensity and spectral cues to localize sounds. Listeners made rapid head movements to sounds with varying intensities and locations within the frontal hemifield. By modifying the pinna geometry of the intact ear, we determined the contribution of spectral cues. We also assessed their ability to localize azimuth in a simple single-intensity condition.

Our data show that all monaural listeners strongly relied on the HSE, whereas this cue is entirely ignored by binaural control listeners. Multiple linear regression indicated that some monaural listeners did extract azimuth information, regardless of sound level. These listeners based their responses partly on spectral cues. Moreover, the stronger the contribution of spectral cues to azimuth, the better monaural listeners could also localize elevation. We conclude that, despite its ambiguity, the HSE dominates monaural sound localization.

2.2 METHODS

Monaural and binaural listeners

Nine listeners with chronic unilateral hearing loss (18-50 years of age) participated in the free-field localization experiments (Table 2.1). They were given a short practice session before the start of the actual experiment under open-loop conditions (i.e., no feedback was given to the actual performance of the listeners). The monaural listeners’ good ear had normal hearing [within 20 dB of audiometric zero] as determined by an audiogram obtained
Table 2.1. Monaural Listeners

<table>
<thead>
<tr>
<th>Patient</th>
<th>Deaf side</th>
<th>Age onset</th>
<th>Age exp</th>
<th>Intensity</th>
<th>Spectral</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>Left</td>
<td>12</td>
<td>23</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>CD</td>
<td>Right</td>
<td>Birth</td>
<td>23</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>GK</td>
<td>Right</td>
<td>&lt; 1</td>
<td>24</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IE</td>
<td>Left</td>
<td>Birth</td>
<td>18</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>JP</td>
<td>Right</td>
<td>Birth</td>
<td>24</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>LD</td>
<td>Right</td>
<td>Birth</td>
<td>20</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>PO</td>
<td>Right</td>
<td>48</td>
<td>50</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RH</td>
<td>Left</td>
<td>Birth</td>
<td>23</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SB</td>
<td>Right</td>
<td>Birth</td>
<td>21</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Monaural Listener Group. **Deaf Side,** Each monaural listener’s deaf side; **Age Onset,** presumed onset age of one-sided deafness; **Age Exp,** the age at which the monaural listener participated in the experiments; **Experiment,** indicates in which paradigms the monaural listener participated; **Intensity,** all possible targets at seven different intensities under open-loop conditions; **Spectral,** all possible targets at four different intensities with or without spectral modification by means of a mold under open-loop conditions; **Training,** horizontal targets at one fixed intensity under open-loop conditions or with LEDs.

Figure 2.1. Audiograms for monaural and binaural listeners. A, Thresholds (dBA) for 10 tones (kHz) were measured for all monaural listeners and both ears, except for monaural listener PO. Thresholds on the deaf side are plotted individually for each monaural listener. They are clearly worse than the thresholds of their good ear (white line, mean threshold; gray patch, SD). Some monaural listeners did not detect some tones at the deaf side. B, For binaural listeners, mean thresholds of the left ear (white line, light gray patch, SD) are indistinguishable from the mean thresholds of the right ear (black line, dark gray patch, SD). Note that the good ear of the monaural patients is as good as the ears of the binaural listeners.
with a standard staircase procedure (10 tone pips, 0.5 octave separation, between 500 Hz and 11.3 kHz), but thresholds were ~60 dB sound pressure level, A-weighted (dBA) higher in the impaired ear (Fig. 2.1). Eight monaural listeners were diagnosed with a unilateral hearing loss at a young age (CD, GK, IE, JP, RH, and SB: younger than 4 years of age; BN and LD: diagnosed at age 12), presumably because of a congenitally underdeveloped cochlea, except for participant BN who had his left cochlea removed and participant GK who lost her hearing because of meningitis. Participant PO (50 years of age) had a sudden hearing loss of unknown origin at the age of 48. None of the monaural listeners had any known uncorrected visual disorder. Six binaural listeners (21-40 years of age) also participated in these experiments and acted as a control reference. None had any auditory or uncorrected visual disorder. Three control listeners (JV, HV, MW) had previous experience with sound localization studies; participant MW is an author of this study. The other binaural control and monaural listeners were inexperienced and were kept naive about the purpose of this study.

**Apparatus**

During the experiments, the listener was seated comfortably in a chair in the center of a completely dark, sound-attenuated room (height times width times length = 2.45 x 2.45 x 3.5 m). The walls, ceiling, floor, and every large object present were covered with acoustic foam that eliminated echoes of sound frequencies >500 Hz. The room had an ambient background noise level of 20 dBA. A total of 58 light-emitting diodes (LEDs) attached to the center of small broad-range loudspeakers (MSP-30; Monacor International GmbH, Bremen, Germany) was mounted on a thin wooden frame that formed a hemispheric surface 100 cm in front of the listener. Stimulus coordinates ranged from -75 to +75° in both azimuth and elevation, as defined in a double-pole coordinate system (Knudsen and Konishi, 1979). In this system, azimuth (α) is defined as the angle between the sound source or response location, the center of the head, and the midsagittal plane, and elevation (ε) is defined as the angle between the sound source, the center of the head, and the horizontal plane (Hofman and Van Opstal, 1998). The origin of the (α, ε) coordinate system corresponds to the straight-ahead speaker location. Head movements were recorded with the magnetic search-coil induction technique. The listener wore a lightweight (150 g) “helmet” consisting of two perpendicular 4 cm wide straps that could be adjusted to fit around the listener’s head without interfering with the ears. A small coil was attached to the top of this helmet. From the left side of the helmet, a 40 cm long, thin, aluminum rod protruded forward with a dim (0.15 Cd/m²) red LED attached to its end, which could be positioned in front of the listener’s eyes. Two orthogonal pairs of 2.45 x 2.45 m coils were attached to the edges of the room to generate the horizontal (60 kHz) and vertical (80 kHz) magnetic fields. The head-coil signal was amplified and demodulated (Remmel Labs, Ashland, MA), after which it was low-pass filtered at 150 Hz (model 3343; Krohn-Hite, Brockton, MA) before being stored on hard disk at a sampling rate of 500 Hz per channel for off-line analysis.

**Auditory stimuli**

Acoustic stimuli were digitally generated using Tucker-Davis System II hardware (Tucker-Davis Technologies, Gainesville, FL) with a 16 bit digital-to-analog converter (TDT, model DA1; 50 kHz sampling rate). A programmable attenuator (TDT, model PA4) controlled sound level, after which the stimuli were passed to a buffer (TDT, model HB6) and finally to one of the speakers in the experimental room. All acoustic stimuli consisted of Gaussian white noise and had 0.5 msec sine-squared onset and offset ramps. The auditory stimuli were either broadband (BB; flat broadband characteristic between 1 and 20 kHz) or high-pass (HP; high-pass filtered at 3 kHz) stimuli (see below) with a duration
of 150 msec. Sound intensities ranged from 30 to 60 dBA (see below). Absolute free-field sound levels were measured at the position of the listener’s head with a calibrated sound amplifier and microphone (BK2610/BK4144; Brüel and Kjær, Norcross, GA).

Paradigms
Calibration experiment. Head-position data for the calibration procedure were obtained by instructing the listener to make an accurate head movement while redirecting the dim rod LED from the central fixation LED to one of the 57 peripheral LEDs that was illuminated as soon as the fixation point extinguished. Each experimental session started with a calibration run.

Auditory localization. The listener started a trial by fixating the central LED. After a pseudo-random period of 1.5-2.0 sec, this fixation LED disappeared, and an auditory stimulus was presented 400 msec later. The listener was asked to redirect the head by pointing the dim rod LED as accurately and as fast as possible to the perceived location of the sound stimulus. Because the response reaction times typically exceeded 200 msec, all responses were made under complete open-loop conditions. To investigate the role of different auditory and non-auditory cues in monaural sound localization, the localization experiments were run according to the three paradigms described below. The paradigms were run on separate days. Monaural listeners participated in the various paradigms as indicated in Table 2.1.

Intensity paradigm. This paradigm was used to investigate the role of stimulus intensity on binaural and monaural sound localization. Both monaural listeners and binaural controls participated in this experiment. HP stimuli were presented at 57 locations (excluding the straight-ahead speaker) and seven intensities (30, 35, 40, 45, 50, 55, and 60 dBA), making a total of 399 stimuli.

Spectral paradigm. The stimuli in this paradigm were HP, presented at 57 locations and four different intensities (30, 40, 50, and 60 dBA; total of 228 stimuli) and were run with the monaural listeners only. To verify whether monaural listeners used monaural spectral cues, this paradigm was run twice in the same session. In the first run, participants performed the standard localization task. Before the second run, the concha of their intact ear was filled with wax to perturb the spectral cues of that ear without occluding the ear canal.

Training paradigm. In the third paradigm, monaural listeners were trained to localize a single-intensity stimulus. During training, a BB stimulus of 60 dBA was presented on the horizontal meridian at 1 of 10 locations \( \alpha \in (-75, 60, ..., 60, 75)^\circ; \varepsilon = 0^\circ, \) excluding the fixation target at \( \alpha = 0^\circ \). Listeners were explicitly told that the stimulus had one fixed intensity. Each location was presented five times in pseudorandom order \((n = 50 \text{ stimuli})\). A similar block of stimuli then followed this first block of stimuli, but this time the LED at the speaker location was also illuminated to provide the subject with visual feedback. This auditory-visual block was then followed by another pure-auditory block and so on, until three blocks of each type were presented.

Acoustic head-shadow measurements. To quantify the HSE, a silicone tube attached to a miniature microphone (EA1842; Knowles, Itasca, IL) was placed near the entrance of the listener’s ear canal to record the acoustic signal. The head was restrained to face the center speaker while the sound stimuli (BB, 60 dBA) were presented from all 58 speaker locations. Signals were stored on a hard disk at a sampling frequency of 50 KHz for off-line analysis. The recordings were made for two binaural listeners \((JV \text{ and } MW)\) and two monaural listeners \((JP \text{ and } RH)\).

Data analysis
Data calibration. The 58 fixation points obtained from the calibration experiment were used to train two three-layer back-propagation neural networks that served to calibrate the head-movement data. Both networks received the raw horizontal and vertical head-position sig-
Monaural Sound Localization

nals as inputs and yielded the desired azimuth and elevation angles (in degrees), respectively, as their output. The trained networks were subsequently used to map the raw data to calibrated two-dimensional head positions with an absolute accuracy within 4% over the entire response range (for details, see Goossens and Van Opstal, 1997). Response coordinates were defined in the same double-pole azimuth-elevation coordinates as the stimuli (see above). For binaural listeners, a positive azimuth angle refers to targets and responses located on the right-hand side. For ease of comparison between the monaural listeners, we defined azimuth as positive when targets and responses were located on the side of the good ear.

Head movement detection. Saccadic head movements were detected from the calibrated head-movement signals by setting thresholds to the head velocity for onset and offset, respectively, using a custom-made program (onset velocity = 20 °/sec; offset velocity = 15 °/sec). Detection markings from the program were visually checked by the experimenter and could be adjusted manually when deemed necessary. Head movements with reaction times <80 or >1000 msec were discarded, because responses with extremely short latencies may be regarded as anticipatory and responses with excessive latencies are usually the result of inattentiveness of the listener.

Statistics. All responses were analyzed separately for each listener by determining the optimal linear fit for the following stimulus-response relationship:

\[
\alpha_R = a + b \cdot \alpha_T \quad \text{and} \quad \varepsilon_R = c + d \cdot \varepsilon_T
\]

for the azimuth and the elevation components, respectively, by minimizing the least-squares error. In Eq. 2.1, \(\alpha_R\) and \(\varepsilon_R\) are the azimuth and elevation response components, and \(\alpha_T\) and \(\varepsilon_T\) are the actual azimuth and elevation coordinates of the stimulus. Fit parameters, \(a\) and \(c\), are the biases (offsets; in degrees), whereas \(b\) and \(d\) are the gains (slopes, dimensionless) of the azimuth and elevation responses, respectively. Note that an ideal listener should yield gains of 1.0 and offsets of 0.0°. Also, Pearson’s linear correlation coefficient, the residual error (SD around the fitted line), and the mean absolute localization error were calculated.

As described in the Introduction, we hypothesized that the HSE might potentially underlie the localization behavior of the monaural listeners. To quantify the acoustic effect of the head on sound level, proximal to the ear, as a function of sound azimuth, we used the following model to describe the HSE:

\[
\text{HSE}(\alpha_T) = e \cdot \sin(f \cdot \alpha_T + g) + h
\]

with \(\alpha_T\) (in degrees) target azimuth. Parameters \(e\) and \(h\) (in dBA), \(f\) (in degrees-1), and \(g\) (dimensionless) were found by minimizing the mean-squared error (Gauss-Newton method). Because the differences between various listeners were small, measured HSE data were pooled across four listeners to determine the optimal fit parameters of Eq. 2.2 (see Fig. 2.5A).

Azimuth. To evaluate the potential role of both stimulus azimuth and sound level in determining the subject’s responses, the data were analyzed by applying multiple linear regression. However, because azimuth and sound level are measured in different units, a direct regression does not allow for a quantitative comparison of the relative contributions of these two stimulus factors. To deal with this problem, we normalized the relevant variables and performed a standardized multiple linear regression analysis by fitting the following relationship:
\[ \hat{\alpha}_T = k \cdot \hat{\alpha}_T + m \cdot \hat{I}_p \]  

(2.3)

where \( \alpha_T, \alpha_R, \) and \( \hat{I}_p \) are now dimensionless variables [where \( x = (x - \mu x)/\sigma_x \) is the so-called z-score of variable \( x \), with \( \mu x \) as the mean, and \( \sigma_x \) as the variance of variable \( x \)], \( k \) and \( m \) are the (dimensionless) partial correlation coefficients that result from the fit, and \( I_p \) is the stimulus intensity at the good ear, which in this study will be termed “proximal stimulus intensity.”

The latter was determined by the following equation:

\[ I_p(\alpha_T) = HSE(\alpha_T) + I_F \]  

(2.4)

In Eq. 2.4, \( I_F \) is the free-field (absolute) stimulus level (in dBA).

The partial correlation parameters \( k \) and \( m \) provide a measure for relative importance of the associated variable (azimuth and intensity, respectively) to explain the subject’s responses. By definition, \( k \) and \( m \) are constrained to values between -1 and 1. When \( k = 1 \) and \( m = 0 \), the subject’s responses are entirely described by changes in stimulus azimuth and are insensitive to changes in sound level. Conversely, when \( k = 0 \) and \( m = 1 \), the azimuth responses of the subject are entirely determined by changes in sound level, regardless of the actual stimulus azimuth. The squared values of \( k \) and \( m \) quantify how much of the variance in the data is explained by the respective variable.

Because azimuth is defined as positive for monaural listeners when locations are on their hearing side, a positive value of \( m \) indicates that monaural listeners orient their responses to the side of the good ear when the proximal sound level \( I_p \) is high and toward the deaf ear when these intensities are low. Although there is no a priori reason for a linear effect of \( I_p \) on perceived azimuth, the results show that this first-order approximation is quite reasonable.

Elevation. Previous research has indicated that each ear contributes to elevation localization on the contralateral side in an azimuth-dependent way (Morimoto, 2001; Hofman and Van Opstal, 2003). Because the results in this study show that unilaterally deaf listeners rely on both target azimuth and proximal intensity for their azimuth responses, these two stimulus parameters (effectively determining the listener’s perceived azimuth) were included in the analysis of the elevation responses. To that end, elevation localization behavior was quantified by fitting the following standardized multiple linear regression:

\[ \hat{e}_R = n \cdot \hat{e}_T + o \cdot \hat{\alpha}_T + q \cdot \hat{I}_p \]  

(2.5)

where the normalized elevation response \( \hat{e}_R \) may depend not only on target elevation, \( \hat{e}_T \), but also on target azimuth, \( \hat{\alpha}_T \), and proximal intensity, \( \hat{I}_p \). Proximal intensity, \( \hat{I}_p \), was replaced by the free-field intensity, \( I_F \), in Eqs. 2.3 and 2.5 when these regressions were performed on the responses from binaural control listeners.

The bootstrap method was applied to obtain confidence limits for the optimal fit parameters in the regression analyses 2.1, 2.2, 2.3, and 2.5. To that end, 100 data sets were generated by randomly selecting (with replacement) data points from the original data. Bootstrapping thus yielded a set of 100 different fit parameters. The SDs in these parameters were taken as estimates for the confidence levels of the parameter values obtained in the original data set (Press et al., 1992).
2.3 Results

The influence of intensity on monaural sound-source azimuth localization

Figure 2.2 exemplifies the azimuth and elevation responses of a typical binaural listener (Fig. 2.2A,B) and a typical monaural participant (Fig. 2.2C,D) to one of the stimulus types in the intensity paradigm (45 dBA; high-pass noise). The binaural listener was quite precise in localizing these stimuli as demonstrated by the near-optimal regression lines and the small amount of scatter for both sound-source azimuth (Fig. 2.2A) and elevation (Fig. 2.2B). Although the scatter is clearly larger for the monaural listener, localization of the 45 dBA stimulus appeared to be remarkably good despite the absence of binaural cues (Fig. 2.2C,D). Data such as these are therefore in line with a previous report (Slattery and Middlebrooks, 1994).

Figure 2.3 shows the responses of the same two listeners to all stimuli in the intensity paradigm (pooled intensities from 30-60 dBA; see Materials and Methods). Note that the responses for the binaural listener were insensitive to the large range in sound levels. The regression lines through the pooled data were indistinguishable from the regression on the 45 dBA data in Figure 2.2. Also, the scatter around the regression lines was quite modest. In contrast, the monaural listener appeared to be unable to localize the sound source, because her responses hardly correlated with the actual stimulus coordinates. Note, however, that this listener still appeared to perceive a large range of sound-source azimuths and elevations, which was comparable with that of the binaural listener. The responses of the monaural listener clearly were not directed merely to the side of the intact ear as reported in acute
monaural studies (Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997). Yet, the responses did not seem to be driven by spatial information either. Localization capabilities of the other unilateral deaf in the intensity paradigm were also poor (Table 2.2), as evidenced by the poor correlation between stimulus and response azimuth ($r^2 < 0.3$), the low response gains, $b$ ($0.42 \pm 0.18$, median $\pm$ SD), and the high biases, $a$ ($11 \pm 11^\circ$, median $\pm$ SD), in the linear regression analysis (Eq. 2.1) and the large mean unsigned errors, which were typically $>30^\circ$.

To further illustrate how the azimuth responses of the two listeners were influenced by absolute sound level, we collected each listener’s azimuth responses into small target azimuth-stimulus intensity bins ($3.75^\circ \times 3$ dBA). Response angles were averaged if a bin contained more than one response. Subsequently, the average response azimuth was gray coded and plotted in the corresponding stimulus bin [with dark gray indicating responses into the far left (binaural) or the deaf (monaural) hemifield, and light gray into the right or hearing hemifield]. If listeners responded spatially accurate, regardless of sound level, a uniform pattern of vertical iso-gray bands should emerge. The overall performance of the binaural listener in Figure 2.4A approached this ideal situation quite well.

In contrast, performance of the unilaterally deaf listener IE across the entire range of free-field sound intensities was quite different (Fig. 2.4B). The data clearly show that free-field stimulus level dominates her localization behavior. In particular, note that the loudest stimuli were located exclusively in the unaffected hemifield (Fig. 2.4B, light shading), whereas the weakest stimuli all appeared to arrive from the deaf side (dark shading). Monaural listeners often did have trouble hearing the weakest stimuli on their far-deaf side, so that a response was sometimes not made at all (Fig. 2.4B, indicated by the white bins). Stimuli with midrange

![Figure 2.3. Sound localization responses of the same two listeners as in Figure 2.2, pooled for all HP stimuli in the intensity paradigm (30, 35, 40, 45, 50, 55, and 60 dBA; binaural listener JV, A and B; monaural listener IE, C and D). Thick solid lines denote the linear regression lines for both azimuth (A, C) and elevation (B, D) components. Note the near-veridical performance of the binaural listener and the absence of a clear stimulus-response relationship for the monaural listener.](image-url)
Monaural Sound Localization

Table 2.2: Results of linear regression on azimuth responses

<table>
<thead>
<tr>
<th></th>
<th>( b(\alpha_T) )</th>
<th>( a )</th>
<th>( r^2 )</th>
<th>( N )</th>
<th>MUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>0.67±0.06</td>
<td>20.44±2.31</td>
<td>0.20</td>
<td>364</td>
<td>39.1±28.7</td>
</tr>
<tr>
<td>CD</td>
<td>0.35±0.04</td>
<td>13.35±1.02</td>
<td>0.25</td>
<td>374</td>
<td>20.8±23.2</td>
</tr>
<tr>
<td>GK</td>
<td>0.77±0.07</td>
<td>0.09±2.11</td>
<td>0.28</td>
<td>348</td>
<td>31.0±24.6</td>
</tr>
<tr>
<td>IE</td>
<td>0.34±0.09</td>
<td>4.48±2.73</td>
<td>0.06</td>
<td>234</td>
<td>35.8±27.8</td>
</tr>
<tr>
<td>JP</td>
<td>0.42±0.07</td>
<td>21.11±2.24</td>
<td>0.10</td>
<td>304</td>
<td>39.5±28.2</td>
</tr>
<tr>
<td>LD</td>
<td>0.64±0.07</td>
<td>10.98±2.01</td>
<td>0.22</td>
<td>393</td>
<td>30.8±29.0</td>
</tr>
<tr>
<td>PO</td>
<td>0.42±0.10</td>
<td>-4.96±3.41</td>
<td>0.04</td>
<td>263</td>
<td>53.8±30.5</td>
</tr>
<tr>
<td>RH</td>
<td>0.33±0.04</td>
<td>7.59±0.97</td>
<td>0.30</td>
<td>374</td>
<td>19.6±15.2</td>
</tr>
<tr>
<td>SB</td>
<td>0.23±0.04</td>
<td>29.73±1.33</td>
<td>0.09</td>
<td>363</td>
<td>31.9±30.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( b(\alpha_T) )</th>
<th>( a )</th>
<th>( r^2 )</th>
<th>( N )</th>
<th>MUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV</td>
<td>1.21±0.02</td>
<td>2.14±0.49</td>
<td>0.94</td>
<td>399</td>
<td>9.8±7.3</td>
</tr>
<tr>
<td>JV</td>
<td>1.09±0.01</td>
<td>-0.54±0.41</td>
<td>0.95</td>
<td>399</td>
<td>6.2±5.8</td>
</tr>
<tr>
<td>MW</td>
<td>1.01±0.01</td>
<td>7.70±0.24</td>
<td>0.98</td>
<td>399</td>
<td>7.8±3.9</td>
</tr>
<tr>
<td>MT</td>
<td>1.23±0.02</td>
<td>-1.49±0.61</td>
<td>0.91</td>
<td>392</td>
<td>11.0±9.4</td>
</tr>
<tr>
<td>SW</td>
<td>1.10±0.02</td>
<td>-4.37±0.57</td>
<td>0.91</td>
<td>359</td>
<td>8.9±8.0</td>
</tr>
<tr>
<td>TE</td>
<td>1.12±0.02</td>
<td>1.88±0.51</td>
<td>0.93</td>
<td>394</td>
<td>8.4±7.0</td>
</tr>
</tbody>
</table>

Parameters from linear fits of the response azimuths according to Eq. 2.1. Stimuli are from the intensity-paradigm with levels ranging from 30 to 60 dBA (all pooled). Column 1: monaural patient or binaural subject. Columns 2-3: fit parameters and estimated error: \( b(\alpha_T) \) is target azimuth gain, \( a \) is azimuth bias. Column 4: \( r^2 \) - between the fitted responses and the actual data. Column 5: number of responses. Column 6: mean unsigned error, MUE, and standard deviation.

Figure 2.4. The influence of free-field stimulus intensity on binaural and monaural sound-source azimuth localization. Gray shading encodes averaged, interpolated response location (°) for each of the stimulus azimuth-intensity bins. Bright areas refer to responses made into the right hemifield (binaural listener) or into the hearing side (monaural listener), and dark-shaded bins correspond to responses made into the (binaural) left hemifield or (monaural) deaf side (white bins indicate that no responses were made). A. Sound localization responses of binaural listener JV are independent of the free-field intensity. B. Free-field intensity is an important factor to explain the localization behavior of monaural listener IE.
intensities typically elicited a larger range of azimuth responses, so that listeners seemed to make near-normal localization responses (Fig. 2.2C,D).

However, because of the HSE, the free-field sound is filtered and attenuated in an azimuth-dependent way by the acoustic properties of the head. To incorporate this acoustic effect in our analysis, we first measured the average sound level at each ear as a function of sound-source azimuth. The result is shown in Figure 2.5A (data pooled for four listeners). Across the range of target azimuths (from -75 to 75°), sound level at each ear varied over a range of 20 dBA. The analysis presented thus far has not accounted for this potential localization cue.

In Figure 2.5B, the data are shown as a function of proximal sound level in the same format as in Figure 2.4B after incorporating the HSE (Eq. 2.4) (see Materials and Methods). Analyzed in this manner, the data clearly show that responses of monaural listener IE are determined solely by the proximal intensity.

**Figure 2.5. The influence of the HSE on monaural sound azimuth localization.**

- **A,** The proximal intensities (in dBA) of a 60 dBA Gaussian white noise sound in the free field measured at the right ear of listeners JV and MW and monaural listeners JP and RH for 58 speaker locations. The free-field intensity was subtracted from the right-ear intensity to obtain the average pooled HSE. The HSE depends on azimuth location in a systematic manner (thick line, best fit according to Eq. 2.2; parameters, e = 9.7 dBA; f = 0.02 degrees-1; g = 0.27 radians; h = 0.00 dBA).
- **B,** The format of this illustration is the same as in Fig. 4 B, but free-field intensity ($I_F$) is now replaced by the proximal intensity at the hearing ear (Eq. 2.4) ($I_P$; see Materials and Methods). Analyzed in this manner, the data clearly show that responses of monaural listener IE are determined solely by the proximal intensity.
Monaural Sound Localization

Figure 2.6. Multiple linear regression analysis of binaural and monaural azimuth localization performance. The partial correlation coefficients for intensity \( m \) and azimuth \( k \) are plotted against one another for each binaural listener (gray triangles) and monaural listener (black circles). Note that all binaural listeners have an azimuth coefficient close to 1 (with a nonsignificant deviation from 0 of the intensity coefficient, indicated by **) whereas most monaural listeners have a low azimuth coefficient (**, denotes statistically nonsignificant deviations from 0; \( p > 0.05 \)) and a high intensity coefficient.

ers (IE, JP, and PO), the azimuth coefficient \( k \) was insignificant \( (p > 0.05) \), whereas the coefficient for proximal sound level was high \( (m \approx 0.75) \). These monaural listeners therefore made no distinction between the proximal intensity of a sound and its location in the horizontal plane. Interestingly, for five other monaural listeners, the partial correlation coefficients for stimulus azimuth were significant \( (p < 0.05) \), although also in these monaural listeners proximal intensity dominated their responses \( (m > k) \). Only for monaural listener CD both coefficients were approximately equal but still low \((\sim 0.35)\). As seen in this figure, the more monaural listeners relied on actual target azimuth, the lower the partial correlation for intensity \( (r = -0.74; \ p < 0.05) \). The data therefore indicate that monaural azimuth localization is nearly impossible for monaural listeners when they are subjected to an acoustic environment with a large range of potential stimulus intensities. All monaural listeners relied heavily on the proximal sound intensity cue provided by the HSE, even though this cue is ambiguous for sound localization. Therefore, actual performance of these listeners in such an environment was quite poor, because their stimulus-response relationships nearly completely broke down (Fig. 2.3).

The influence of spectral cues on monaural sound-source azimuth localization
Despite the poor localization behavior and strong contribution of proximal sound level, six monaural listeners were able to extract a significant amount of information about the ve-
Parameters from multiple linear regression on the response azimuths according to Eq. 2.3. Stimuli are from the intensity-paradigm with levels ranging from 30 to 60 dBA (all pooled). Column 1: monaural patient or binaural subject. Columns 2-3: model parameters and estimated error (determined by bootstrapping): $k(\alpha_T)$ is target azimuth gain, $m(I_F)$ is free-field intensity gain (for binaural subjects), $m(I_P)$ is perceived intensity gain (for monaural patients). Column 4: $r^2$ between fit and actual data. Column 5: number of responses.

<table>
<thead>
<tr>
<th>Monaural</th>
<th>$k(\alpha_T)$</th>
<th>$m(I_F)$</th>
<th>$r^2$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BN</td>
<td>0.11±0.04</td>
<td>0.75±0.03</td>
<td>0.65</td>
<td>364</td>
</tr>
<tr>
<td>CD</td>
<td>0.37±0.06</td>
<td>0.30±0.05</td>
<td>0.32</td>
<td>374</td>
</tr>
<tr>
<td>GK</td>
<td>0.16±0.05</td>
<td>0.73±0.04</td>
<td>0.68</td>
<td>348</td>
</tr>
<tr>
<td>IE</td>
<td>-0.09±0.04</td>
<td>0.88±0.03</td>
<td>0.72</td>
<td>234</td>
</tr>
<tr>
<td>JP</td>
<td>-0.04±0.05</td>
<td>0.79±0.05</td>
<td>0.59</td>
<td>304</td>
</tr>
<tr>
<td>LD</td>
<td>0.20±0.06</td>
<td>0.57±0.04</td>
<td>0.48</td>
<td>393</td>
</tr>
<tr>
<td>PO</td>
<td>-0.18±0.05</td>
<td>0.72±0.06</td>
<td>0.66</td>
<td>263</td>
</tr>
<tr>
<td>RH</td>
<td>0.28±0.05</td>
<td>0.61±0.05</td>
<td>0.60</td>
<td>374</td>
</tr>
<tr>
<td>SB</td>
<td>0.15±0.05</td>
<td>0.35±0.06</td>
<td>0.19</td>
<td>363</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Binaural</th>
<th>$k(\alpha_T)$</th>
<th>$m(I_F)$</th>
<th>$r^2$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV</td>
<td>0.97±0.01</td>
<td>0.01±0.01</td>
<td>0.94</td>
<td>399</td>
</tr>
<tr>
<td>JV</td>
<td>0.98±0.01</td>
<td>0.00±0.01</td>
<td>0.95</td>
<td>399</td>
</tr>
<tr>
<td>MW</td>
<td>0.99±0.01</td>
<td>-0.01±0.01</td>
<td>0.98</td>
<td>399</td>
</tr>
<tr>
<td>MT</td>
<td>0.96±0.01</td>
<td>0.01±0.02</td>
<td>0.92</td>
<td>392</td>
</tr>
<tr>
<td>SW</td>
<td>0.96±0.02</td>
<td>0.00±0.01</td>
<td>0.91</td>
<td>359</td>
</tr>
<tr>
<td>TE</td>
<td>0.97±0.02</td>
<td>-0.03±0.01</td>
<td>0.93</td>
<td>394</td>
</tr>
</tbody>
</table>

To test for this possibility, we subjected seven of our monaural listeners to the spectral paradigm (see Materials and Methods) in which they localized sounds of variable intensities with and without a wax mold applied to the pinna of their good ear. This wax mold perturbed the spectral cues while leaving the HSE unaffected. As expected, the mold produced a severe impairment in the elevation localization behavior of all monaural listeners (data not shown). The azimuth data from this experiment were analyzed in the same way as in the intensity experiment. The effects of the wax mold on the partial correlation coefficients for both proximal intensity and azimuth are plotted in Figure 2.7. Note that when the azimuth partial correlation, $k$, obtained from responses during the free-ear condition was non-zero ($p < 0.05$; monaural listeners BN, CD, LD, and RH), they were significantly lower for the mold condition. For monaural listeners GK, JP, and SB, the azimuth coefficients in both conditions did not differ from zero in this experiment. Intensity coefficients increased significantly for three of the monaural listeners in the mold condition (CD, LD, and RH) but did not change for the
Figure 2.7. Spectral cue extraction for sound azimuth localization. Both the azimuth \((k)\) and intensity \((m)\) partial correlations were obtained for the free-ear and mold conditions. The coefficients for the different conditions are plotted against one another. Note the decrease of the azimuth coefficients resulting from the insertion of a mold. Intensity coefficients are either similar or slightly larger.

Figure 2.8. Spectral cue extraction for azimuth localization on the hearing side and on the deaf side. Azimuth coefficients \((k)\) obtained for the hearing and deaf side or, for binaural listeners, for the right and left hemifield \((\alpha_T > 5^\circ)\) are plotted against one another for the monaural listeners (black circles) and the binaural listeners (gray triangles), respectively. Note the negligible azimuth coefficients on the deaf side and the higher azimuth coefficients on the hearing side for some of the unilateral deaf. No such interaural asymmetry was obtained for the binaural listeners.
other monaural listeners. These data support the possibility that monaural listeners can use monaural spectral cues to extract information about sound-source azimuth. However, only half of the monaural listeners did so and even then performance was severely hampered when compared with normal binaural localization.

We next investigated whether the spectral cues of the hearing ear could also contribute to localization of sound-source azimuth on the deaf side. To that end, the multiple linear regression analysis was repeated on the data from the intensity paradigm (Eqs. 2.3 and 2.4) (see Materials and Methods) but now separate for the hearing ($\alpha_T > 5^\circ$) and deaf sides ($\alpha_T < -5^\circ$) (Fig. 2.8). The results indicate that the partial correlations for azimuth on the deaf side were much lower than those on the hearing side. Furthermore, the intensity coefficients did not differ between hearing and deaf sides ($p > 0.05$; data not shown).

In summary, our data suggest that (1) monaural spectral cues are only used by approximately half of the monaural listeners to localize azimuth, and (2) these cues cannot be used to extract azimuth information over the entire azimuth range.

**Monaural sound-elevation localization**

As shown in Figures 2.3 and 2.4, the binaural listener was quite accurate at localizing sound elevation for the different sound intensities, whereas performance of the monaural listener was clearly compromised. Previous studies with binaural listeners have shown that disruption of the spectral cues of one ear systematically affects elevation localization as a function of sound-source azimuth on the disrupted side (Musicant and Butler, 1984a; Morimoto, 2001; Hofman and Van Opstal, 2002).
Monaural Sound Localization

Our data thus far show that the monaural listeners localize azimuth on the basis of both spectral cues and proximal sound level (HSE) cues with different relative contributions (Fig. 2.6). We therefore hypothesized that the percept of sound-source elevation in these monaural listeners would not only rely on actual target elevation (spectral cues) but also on perceived sound-source azimuth, which was shown to be a function of target azimuth and proximal sound level (Fig. 2.6).

As exemplified for one of our binaural control listeners in Figure 2.9A and B, the binaural localization responses solely depended on the actual sound elevation, regardless of either target azimuth or free-field intensity. However, quite different response patterns were obtained for the monaural listeners. For example, the elevation responses of monaural listener IE depended slightly on target azimuth, which is apparent from the systematic gradient in the azimuth direction in Figure 2.9C. The monaural listener’s elevation responses also depended on proximal sound level. At low proximal intensities, responses were located at a small range of low elevations; however, at higher proximal intensities, her responses were based toward more veridical elevation angles (Fig. 2.9D). In contrast, monaural listener JP responded almost exclusively to proximal intensity, because the responses did not systematically depend on either target elevation or target azimuth (Fig. 2.9E,F). Low proximal intensities elicited low response elevations, whereas high proximal sound levels yielded high response elevations in this participant.

TABLE 2.4: Results of multiple linear regression on elevation responses.

<table>
<thead>
<tr>
<th>Monaural</th>
<th>$n(\varepsilon_T)$</th>
<th>$o(\alpha_T)$</th>
<th>$q(I_P)$</th>
<th>$r^2$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GK</td>
<td>0.56±0.05</td>
<td>-0.06±0.05</td>
<td>0.03±0.04</td>
<td>0.26</td>
<td>348</td>
</tr>
<tr>
<td>RH</td>
<td>0.83±0.05</td>
<td>0.06±0.06</td>
<td>0.05±0.06</td>
<td>0.69</td>
<td>374</td>
</tr>
<tr>
<td>SB</td>
<td>0.54±0.05</td>
<td>0.00±0.04</td>
<td>0.06±0.04</td>
<td>0.29</td>
<td>363</td>
</tr>
<tr>
<td>BN</td>
<td>0.40±0.05</td>
<td>-0.32±0.06</td>
<td>0.14±0.05</td>
<td>0.23</td>
<td>364</td>
</tr>
<tr>
<td>CD</td>
<td>0.67±0.05</td>
<td>-0.05±0.05</td>
<td>0.10±0.04</td>
<td>0.44</td>
<td>374</td>
</tr>
<tr>
<td>IE</td>
<td>0.56±0.05</td>
<td>0.27±0.06</td>
<td>-0.10±0.05</td>
<td>0.36</td>
<td>234</td>
</tr>
<tr>
<td>JP</td>
<td>0.20±0.05</td>
<td>-0.19±0.05</td>
<td>0.58±0.05</td>
<td>0.30</td>
<td>304</td>
</tr>
<tr>
<td>LD</td>
<td>0.71±0.04</td>
<td>0.16±0.03</td>
<td>0.20±0.03</td>
<td>0.57</td>
<td>393</td>
</tr>
<tr>
<td>PO</td>
<td>0.08±0.06</td>
<td>-0.14±0.07</td>
<td>0.34±0.06</td>
<td>0.09</td>
<td>263</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Binaural</th>
<th>$n(\varepsilon_T)$</th>
<th>$o(\alpha_T)$</th>
<th>$q(I_F)$</th>
<th>$r^2$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV</td>
<td>0.93±0.02</td>
<td>-0.07±0.01</td>
<td>0.02±0.02</td>
<td>0.89</td>
<td>399</td>
</tr>
<tr>
<td>JV</td>
<td>0.93±0.02</td>
<td>-0.03±0.02</td>
<td>0.04±0.02</td>
<td>0.87</td>
<td>399</td>
</tr>
<tr>
<td>MW</td>
<td>0.96±0.02</td>
<td>0.01±0.02</td>
<td>-0.03±0.02</td>
<td>0.92</td>
<td>399</td>
</tr>
<tr>
<td>MT</td>
<td>0.78±0.04</td>
<td>-0.08±0.02</td>
<td>0.02±0.03</td>
<td>0.62</td>
<td>392</td>
</tr>
<tr>
<td>SW</td>
<td>0.73±0.05</td>
<td>0.03±0.02</td>
<td>0.07±0.04</td>
<td>0.53</td>
<td>359</td>
</tr>
<tr>
<td>TE</td>
<td>0.86±0.05</td>
<td>-0.01±0.02</td>
<td>0.03±0.03</td>
<td>0.65</td>
<td>394</td>
</tr>
</tbody>
</table>

Parameters from fits of the response elevations with Eq. 2.5. Stimuli are from the intensity-paradigm with levels ranging from 30 to 60 dBA (all pooled). See Table 2.3 for details, parameters are $n(\varepsilon_T)$, elevation gain, $o(\alpha_T)$, azimuth gain and $q(I_P/I_F)$, intensity gain.

2003). Our data thus far show that the monaural listeners localize azimuth on the basis of both spectral cues and proximal sound level (HSE) cues with different relative contributions (Fig. 2.6). We therefore hypothesized that the percept of sound-source elevation in these monaural listeners would not only rely on actual target elevation (spectral cues) but also on perceived sound-source azimuth, which was shown to be a function of target azimuth and proximal sound level (Fig. 2.6).
To quantify the influence of target elevation, target azimuth, and proximal sound intensity for all unilaterally deaf listeners and binaural control listeners, we performed multiple linear regression on the elevation response data (Eq. 2.5) (Table 2.4). As expected, all binaural listeners responded exclusively to actual target elevation (median value of $n \pm SD$, 0.90 ± 0.09), regardless of the changes in target azimuth ($\alpha \sim 0$) and free-field intensity ($q \sim 0$). However, the results for the monaural listeners were more idiosyncratic. For most monaural listeners, actual target elevation played a much smaller role than in the binaural controls (median value of $n \pm SD$, 0.56 ± 0.24), with the notable exception of monaural listener RH, whose performance ($n = 0.83$) equaled that of the naive binaural listeners ($MT$, $SW$, and $TE$). For monaural listeners $GK$, $RH$, and $SB$, responses depended exclusively on sound-source elevation. For the other six monaural listeners, responses also depended on proximal sound level and target azimuth (Table 2.4). Interestingly, monaural listeners with a high partial correlation coefficient for elevation, $n$, also had a high partial correlation coefficient for azimuth in their azimuth responses, $k$ (Fig. 2.10A). These two coefficients were highly correlated ($r = 0.79$; $p < 0.05$), which indicates that unilaterally deaf listeners who were able to make use of their spectral cues to localize sound-source azimuth were also successful at using the spectral cues to localize elevation. Some of the monaural listeners could also localize elevation on their deaf side (Fig. 2.10B).

**Training to localize azimuth monaurally**

Most monaural listeners responded with head movements covering a range of azimuths that
Monaural Sound Localization

varied with the applied stimulus intensities (Fig. 2.4). Low-intensity stimuli typically elicited responses far into the deaf hemifield, whereas loud stimuli tended to elicit responses far into the hearing side. Apparent near-normal localization capabilities were seen for intermediate intensities. This finding suggests that these monaural listeners may have learned to use the ambiguous HSE cue to cope with the acoustic environment encountered in daily life. We therefore wondered whether monaural listeners could rapidly learn to improve their azimuth localization performance to a novel stimulus of one fixed intensity. To that end, five monaural listeners (CD, GK, JP, RH, and SB) were trained to localize loud BB noise bursts of 60 dBA under open-loop (auditory) and closed-loop (auditory-visual) stimulus conditions. Six blocks of stimuli were presented, alternating between auditory and auditory-visual blocks, with the first block being purely auditory (see Materials and Methods). Monaural listener JP (Fig. 2.11A) was able to reduce her mean-squared localization error already within the first open-loop block by 18°. In the subsequent two open-loop blocks, these errors were reduced even further. The other monaural listeners started out with smaller initial errors (Fig. 2.11B, dark circles), but performance increased for all monaural listeners tested to an average mean unsigned localization error of 18° (Fig. 2.11B, white circles). This high performance rate was already achieved in the first open-loop block (Fig. 2.11B, gray circles), so the subsequent visual feedback blocks and open-loop blocks did not add much to the improvement. Thus, the improvements shown in these data suggest that the HSE can be easily and rapidly remapped, even in the absence of visual feedback, for a simple fixed-intensity sound.

2.4 Discussion

This study investigated the role of monaural spectral pinna cues and the HSE on two-dimen-
sional sound localization performance of listeners with a complete unilateral hearing loss and compared their response behavior to that of normal-hearing binaural control listeners. We found that none of the binaural controls relied on stimulus intensity as a potential cue for sound location (Figs. 2.3, 2.4, 2.6).

In contrast, all monaural listeners depended heavily on the HSE to localize sound-source azimuth (Figs. 2.4B, 2.5B, 2.6). This is quite remarkable, given that the HSE is in principle an ambiguous signal for azimuth location. Indeed, because monaural listeners heavily relied on this cue, azimuth localization performance in the randomized intensity paradigm was quite poor (Fig. 2.3C,D).

Previous free-field localization studies have also shown substantial localization deficits for listeners with a unilateral hearing loss (Angel and Fite, 1901; Jongkees and Van der Veer, 1957; Viehweg and Campbell, 1960; Gatehouse and Cox, 1972; Gatehouse, 1976; Humes et al., 1980; Newton and Hickson, 1981; Newton, 1983; Slattery and Middlebrooks, 1994; Bosman et al., 2003). Some of these studies reported near-normal localization performance in these monaural listeners (Newton, 1983; Slattery and Middlebrooks, 1994), a finding that is not supported by our results (Fig. 2.3C,D). The discrepancy is likely attributable to the limited range of intensities used in these previous studies when compared with our study. For sound levels in the midrange (40-50 dBA), adequate response behavior often can be observed (Fig. 2.2C,D), even though the responses have to be entirely explained by the HSE (Fig. 2.5B).

Measurements of the head-related transfer functions indicate that the monaural spectral cues can provide unique information about sound-source location for both the azimuth and elevation directions (data not shown) so that, at least in principle, monaural listeners might have regained adequate localization performance by relying on the spectral cues of their normal ear. Note that the normal-hearing control listeners do not appear to use these more subtle spectral cues for azimuth localization. After application of a unilateral or binaural mold, their azimuth localization responses are not affected (data not shown in this study, but see Oldfield and Parker, 1984; Hofman and Van Opstal, 2003), whereas manipulation of spectral cues in a virtual set-up has little influence on lateral angle judgments (Macpherson and Middlebrooks, 2002). Applying a monaural plug to normal-hearing listeners leads to an immediate and long-lasting shift of the azimuth percept toward the side of the free ear both in humans (Oldfield and Parker, 1986; Butler et al., 1990; Slattery and Middlebrooks, 1994) and in experimental animals (Knudsen et al., 1982; King et al., 2000).

A majority of the monaural listeners in our study showed evidence of using spectral cues to localize sound-source azimuth. This is in line with previous suggestions made by Newton (1983) and Slattery and Middlebrooks (1994). The present study extends these findings by quantifying how this contribution varied from listener to listener. Furthermore, the use of spectral cues was restricted to the hearing hemifield for the majority of monaural participants (Fig. 2.8). Interestingly, the strength of the spectral-cue contribution to a listener’s azimuth percept provided a good predictor for the monaural listener’s ability to localize sound-source elevation also (Fig. 2.10A). However, some monaural listeners did not use the spectral cues at all. Their azimuth localization responses were entirely dominated by proximal stimulus intensity.

The unilateral deaf adapted to their loss of binaural information by incorporating monaural cues into their sound-azimuth localization behavior. In human plasticity studies, the evidence for an adaptive shift in sound-azimuth response behavior after a modification of binaural information is sparse. By introducing a monaural plug in binaural listeners, Slattery and Middlebrooks (1994) did not see a change in azimuth localization behavior within 24 hr. Listeners exposed to reversed binaural cues did not even adapt over a period of up to 19 d (Hofman et al., 2002). Of course, our unilateral deaf were subjected to their binaural loss during a much greater span of time (typically >18 years). Auditory-evoked potentials ob-
Monaural Sound Localization

tained from a late-onset unilaterally deaf listener suggest that the changes in cortical activity, evidenced by an increased inter-hemispheric symmetry, occur gradually and may continue for at least 2 years (Ponton et al., 2001).

This long-term adaptive shift is likely to be very different from the quick changes in response behavior during the training paradigm. Although the incorporation of monaural cues for azimuth localization appears to be very difficult, if not impossible, learning the relationship between proximal sound intensity and sound location may be quite simple for familiar acoustic situations. Under such conditions, listeners may adopt a strategy to respond to the proximal intensity by mapping the midrange intensities to central locations. The training paradigm showed that the unilateral deaf could apply this strategy within a single block of trials when told that the stimulus is of a fixed intensity.

Our results show a clear degradation of both azimuth and elevation response performance when compared with binaural listeners. Also, the unilateral deaf in the study by Slattery and Middlebrooks (1994) had problems localizing sound elevation, especially on the deaf side. In binaural listeners, spectral disruption of one ear degrades elevation localization on the side ipsilateral to the mold (Morimoto, 2001; Hofman and Van Opstal, 2003). Our results extend these findings by showing that degradation in azimuth performance also induces a similar degradation in elevation performance (Fig. 2.10A). Therefore, incorporation of the spectral cues to guide azimuth localization behavior is linked to the ability to use spectral cues from the good ear for elevation localization on the deaf side.

Our observation that the unilateral deaf may have residual localization abilities that could be attributed to their use of spectral cues may seem to disagree with results from monaural sound localization with a virtual acoustic set-up in normal-hearing listeners (Wightman and Kistler, 1997). The difference between the two studies is probably explained by the much longer exposure to a binaural information loss in the unilateral deaf, in contrast with the immediate monauralization of the normal-hearing listeners. Thus, when listeners have not learned to use the monaural spectral cues, their ability to monaurally localize sound-source azimuth as well as elevation is abolished. Indeed, some of our monaural listeners were entirely unable to localize sounds in both azimuth and elevation (Fig. 2.10).

In conclusion, the apparent conflict in results from monaural listeners across different studies in the literature is probably attributable to two factors. First, a significant fraction of the listeners has not learned to incorporate spectral cues to extract azimuth location. Second, most studies did not use sufficient variation of stimulus intensities to enable a dissociation of the different contributions of the HSE and spectral cues.

Note that the complex spectral cues, although veridical, will contribute almost exclusively to localization on the hearing side and at most a small amount on the deaf side (Fig. 2.8). The signal-to-noise ratio in these cues for low-intensity sounds will deteriorate rapidly for stimuli on the deaf side. Thus, these complex cues can only be applied successfully for a limited class of sound sources that are both loud enough and contain a relatively flat broadband spectrum.

In contrast, intensity cues can easily be learned in simple acoustic environments for a variety of sounds (compare Fig. 2.11). Thus, the unilateral deaf might have adopted a pragmatic strategy by incorporating the relatively straightforward monaural intensity cue to localize sounds while neglecting the veridical but limited spectral cues. However, monaural listeners who neglect the spectral cues altogether also lack the ability to localize sounds in elevation. A recent study with normal-hearing listeners showed that adult listeners could relearn new spectral cues within a period of only a few weeks (Hofman et al., 1998). Given the importance of adequate sound-localization performance in the highly dynamic and complex acoustic environments of everyday life, it would be worthwhile to explore the possibility of training monaural listeners to use their spectral cues and thus to radically improve their overall localization behavior.
Acknowledgements
This research was supported by the University of Nijmegen (A.J.V.O., M.M.V.W.) and the Human Frontiers Science Program Grant RG 0174-1998/B (M.M.V.W.). We thank G. Van Lingen, H. Kleijnen, G. Windau, and T. Van Dreumel for technical assistance. We also express our gratitude to the nine monaural and five binaural volunteers for repeatedly participating in our experiments.
CHAPTER III
SOUND LOCALIZATION UNDER PERTURBED BINAURAL HEARING

Perturbed Binaural Sound Localization

3.1 INTRODUCTION

Sound localization relies on the neural processing of acoustic cues that result from the interaction of sound waves with the torso, head and ears. Directional hearing in the horizontal plane (azimuth) depends on binaural differences in sound arrival time and ongoing phase for relatively low (<1.5 kHz) frequencies (so-called interaural time differences, or ITDs). At higher frequencies (>3 kHz) the head-shadow effect causes differences in sound level (interaural level differences, ILDs). Localization in the vertical plane (elevation) and front-back discrimination require an analysis of spectral shape cues that arise from direction-dependent reflections within the pinna, the effect of which is described by so-called head-related transfer functions, or HRTFs. The latter mechanism thus essentially constitutes a monaural localization cue for sound frequencies exceeding about 3-4 kHz. However, several studies have suggested that the computation of sound-source elevation also involves binaural interactions, as a monaural perturbation of the spectral cues (e.g. by inserting a mold in one pinna) has a systematic detrimental effect on elevation performance contralateral to the manipulated ear (Hofman and Van Opstal 1998; Humanski and Butler 1988; Morimoto 2001; Van Wanrooij and Van Opstal 2005).

A large body of experimental evidence supports the notion that the processing of the azimuth and elevation components of a sound’s location is embedded in independent neural pathways. In mammals, the ITDs emerge in the medial superior olive (MSO), while the ILDs are extracted in another nucleus of the superior olivary complex the lateral superior olive (LSO; see Yin 2002, for a recent review). Evidence also suggests that the first neural correlates of spectral shape analysis may be found in the dorsal cochlear nucleus, which receives monaural input from the ipsilateral ear (Young and Davis 2002).

Psychophysical evidence supports the hypothesis of independent processing of the acoustic cues. Experimental manipulations can degrade elevation localization performance considerably, while azimuth localization is far more robust: e.g. by inserting molds, either binaurally (Hofman et al. 1998; Oldfield and Parker 1984b), or monaurally (Hofman and Van Opstal 2003; Morimoto 2001; Van Wanrooij and Van Opstal 2005) by introducing background noise (Good and Gilkey 1996; Zwiers et al. 2001b), or by varying sound levels and sound duration considerably (Hartmann and Rakerd 1993; Hofman and Van Opstal 1998; MacPherson and Middlebrooks 2000; Vliegen and Van Opstal 2004).

Yet, the idea of independent pathways for the processing of a sound’s azimuth and elevation coordinates may be too simple. Clearly, at more central neural stages, like the midbrain Inferior Colliculus (IC) and beyond, the outputs of the different cue-processing pathways converge (Chase and Young 2006), and are therefore likely to interact. So far, however, little is known about these higher processing stages.

Recent psychophysical evidence has suggested that the computations underlying the extraction of azimuth and elevation may indeed interact. In a study with unilateral deaf listeners we have demonstrated that only listeners who used the spectral-shape cues from their intact ear to localize azimuth could also localize elevation (Van Wanrooij and Van Opstal 2004). Especially the failure of the other listeners to localize elevation was remarkable, as these results indicated that under chronic monaural conditions the azimuth and elevation components are not processed independently.

To study the mechanisms underlying the integration of the different acoustic cues, this paper reports on the acute effects of a monaural plug on localization performance of normal-hearing listeners in the two-dimensional frontal hemifield. Plugging the ear of an otherwise normal-hearing subject is qualitatively different from the situation of the monaurally deaf, and it is therefore difficult to readily extrapolate findings from monaurally deaf listeners to plugged hearing, and vice versa (Wightman and Kistler 1997).

Earlier studies have assessed the effect of a monaural plug on human sound localization
performance (Flannery and Butler 1981; Musicant and Butler 1984b; Oldfield and Parker 1986; Slattery and Middlebrooks 1994). All studies reported as a major effect a horizontal shift (‘bias’) of localization to the side of the unplugged ear. However, it has remained unclear whether or not the spectral cues contribute to the localization of azimuth. Although Musicant and Butler (1984a; 1984b; 1985) reported that localization of far-lateral targets relied on monaural spectral cues, none of the plugged listeners in the study of Slattery and Middlebrooks (1994) were able to localize in azimuth. Moreover, Wightman and Kistler (1997) showed that a complete removal of the binaural cues in a dichotic setup abolished sound localization performance altogether. That study therefore suggested that the spectral cues are not sufficient for sound localization in the horizontal plane.

In this study we took a different approach, by studying the effect of a plug on localization responses to a variety of acoustic stimuli that varied both in bandwidth and over a considerable range in sound level. The insertion of a plug perturbs the binaural difference cues in normal listeners in a frequency- and level-dependent way, and is therefore expected to affect the ITDs, ILDs, and HRTFs in different ways. We measured localization across the two-dimensional frontal hemifield immediately after inserting the plug, and quantified the changes in localization responses as a function of the acoustic parameters. In our experiments we specifically employed the possibility that listeners could use remaining binaural difference cues.

Our analysis shows that the shift in azimuth responses depends on sound-source location, the sound spectrum, and on sound level. Moreover, performance in elevation is influenced by both sound level, and the perceived azimuth location, rather than by the actual stimulus azimuth. Our data therefore support the hypothesis that the processing of both sound-source azimuth and elevation involve weighted contributions from binaural difference cues, as well as from spatial-shape cues. The relative weights of the acoustic cues are rapidly adjusted under acoustic perturbations that render a given cue unreliable.

3.2 METHODS

Participants
Five listeners (ages 25-47) participated in the experiments (including both authors, listeners MW and JO). All listeners were experienced with the type of sound localization studies carried out in the laboratory, and all had normal hearing (within 20 dB of audiometric zero) as determined by an audiogram obtained with a standard staircase procedure (10 tone pips, 0.5 octave separation, between 500 Hz and 11.3 kHz). None of these listeners had any auditory or uncorrected visual disorder, except for listener JO who is amblyopic in his right eye.

Apparatus
During the experiments, the listener was seated comfortably in a chair in the center of a completely dark, sound-attenuated room (HxWxL = 2.45x2.45x3.5 m³). The walls, ceiling, floor, and every large object present were covered with black acoustic foam that eliminated echoes for sound frequencies above 500 Hz. The room had an ambient background noise level of 25 dB SPL, A-weighted (dBA).

The seated listener faced an array of 58 small broad-range loudspeakers (MSP-30; Monacor International GmbH, Bremen, Germany) containing light-emitting diodes (LEDs) in their center. These speakers were mounted on a thin wooden frame that formed a hemispheric surface 100 cm in front of the listener, at polar coordinates \( R = [0, 15, 30, 45, 60, 75] \) deg and \( \Phi = [0, 30, \ldots, 300, 330] \) deg. \( R \) is the eccentricity relative to the straight-ahead viewing direction (defined in polar coordinates as \([R,\Phi] = [0,0]\) deg), and \( \Phi \) is the angular coordinate where \( \Phi = 0 \) deg is rightward from the center location, and \( \Phi = 90 \) deg is upward. The lower three speakers (at \( R = 75 \) deg, and \( \Phi = [240, 270, 300] \) deg) were left out to allow room for
Perturbed Binaural Sound Localization

Head movements were recorded with the magnetic search-coil induction technique (Robinson 1963). To that end, the listener wore a lightweight (150 g) “helmet” consisting of two perpendicular 4 cm wide straps that could be adjusted to fit around the listener’s head without interfering with the ears. On top of this helmet, a small coil was attached. From the left side of the helmet a 40 cm long, thin aluminum rod protruded forward with a dim (0.15 Cd/m²) red LED attached to its end, which could be positioned in front of the listener’s eyes. Two orthogonal pairs of 2.45x2.45 m² coils and one pair of 2.45x3.5 m² coils were attached to the room’s edges to generate the left-right (60 kHz), up-down (80 kHz) and front-back (40 kHz) magnetic fields, respectively. This arrangement allows for a precise recording of head orientations in all directions, including the rear hemifield. The head-coil signal was amplified and demodulated (Remmel Labs, Katy, TX), after which it was low-pass filtered at 150 Hz (model 3343; Krohn-Hite, Brockton, MA) before being stored on hard disk at a sampling rate of 500 Hz/channel for off-line analysis.

Auditory stimuli

Acoustic stimuli were digitally generated using Tucker-Davis System II hardware (Tucker-Davis Technologies, Alachua, FL), with a TDT DA1 16-bit digital-to-analog converter (50-kHz sampling rate). A TDT PA4 programmable attenuator controlled sound level, after which the stimuli were passed to the TDT HB6 buffer and finally to one of the speakers in the experimental room. All acoustic stimuli consisted of Gaussian noise, and had 0.5 ms sine-squared on- and offset ramps. The auditory stimuli were either broadband (BB, flat characteristic between 1-20 kHz) or high-pass (HP, high-pass filtered at 3 kHz) stimuli with a duration of 150 ms. Sound levels ranged from 30 to 60 dBA (see below). Absolute free-field sound levels were mea-
sured at the position of the listener’s head with a calibrated sound amplifier and microphone (BK2610/BK4144; Bruel & Kjaer, Norcross, GA).

**Plugs**
Listeners were equipped with a precisely fitting plug in their left ear-canal to perturb their binaural cues. The plugs were manufactured by filling the ear-canal with rubber casting material (Otoform Otoplastik-K/c; Dreve GMBH, Unna, Germany).

**Measurement of audiograms**
To determine the attenuation provided by the custom-made plugs, audiograms (10 tone pips, 0.5 octave separation, between 500 Hz and 11.3 kHz) were taken of the listeners’ ears, with and without the plug (Fig. 3.1A). Although some plugs attenuated more than others, the attenuation was always considerable (> 20 dBA). For high frequencies the mean attenuation provided by the plugs (>3 kHz: 25-50 dBA) was equal to or higher than for low frequencies (<3 kHz: ~25 dBA, Fig. 3.1B).

**Paradigms**
*Calibration experiment.* Head-position data for the calibration procedure were obtained by instructing the listener to make an accurate head movement while redirecting the dim rod LED in front of the eyes from the central fixation LED to each of the 57 peripheral LEDs that was illuminated as soon as the fixation point extinguished. Each experimental session started with a calibration run.

*Auditory localization.* The listener started a trial by fixating the central LED with the head-fixed LED pointer. After a pseudo-random period of 1.5 to 2.0 s, this fixation LED disappeared and an auditory stimulus was presented 400 ms later. The listener was asked to redirect the head by pointing the dim rod LED, which was on continuously throughout the experiment, as accurately and as fast as possible to the perceived location of the sound stimulus. As the response reaction times typically exceeded 200 ms, all responses were made under open loop conditions.

**Hearing Conditions**
Sound localization experiments were run under two different hearing conditions. In the free condition, both ears had normal hearing. In the plug condition, the left ear was sealed with a plug. Listeners wore their plug only once, i.e. during the experiment, which started immediately after insertion of the plug. Care was taken not to provide plugged listeners with any acoustic input other than the experimental stimuli.

In one experimental session (either with plug or without), two different stimuli of different bandwidths were tested, in two subsequent runs. During one run listeners had to localize BB stimuli of various intensities (40, 50 and 60 dBA). One such BB-run consisted of 3 stimulus intensities x 57 locations = 171 targets.

In the other run HP stimuli were presented, which were typically more attenuated by the plug than the lower frequencies (see Fig. 3.1B). In the HP-run a larger range of stimulus intensities ([30, 35,…, 55, 60] dBA) than in the BB-run was used. In this way, one complete experimental HP-run consisted of 7 stimulus intensities x 57 locations = 399 targets that were randomized across trials. After 200 trials a short break was introduced in which the lights in the experimental room were turned on.

**Data analysis**
*Data calibration.* The calibration experiment provided a set of 58 LED/speaker locations
Perturbed Binaural Sound Localization

and raw head position signals. These locations were all transformed into the double-pole azimuth-elevation coordinate system (Knudsen and Konishi 1979). In this system, azimuth, \( \alpha \), is defined as the angle between the sound source (or response direction), the center of the head, and the midsagittal plane. Elevation, \( \varepsilon \), is defined as the angle between the sound source, the center of the head, and the horizontal plane. The origin of the \((\alpha, \varepsilon)\) coordinate system corresponds to the straight-ahead speaker location. Azimuth and elevation can be calculated from the polar coordinates, \((R, \Phi)\) by:

\[
(3.1) \quad \alpha = \arcsin(\sin R \cos \Phi) \quad \text{and} \quad \varepsilon = \arcsin(\sin R \sin \Phi)
\]

These 58 fixation points and raw head-position signals were used to train two three-layer neural networks that served to calibrate the head-movement data, using a back-propagation algorithm based on the gradient descent method of Levenberg-Marquardt (Matlab; The Mathworks Inc, Natick, MA). The networks corrected for small inhomogeneities in the magnetic fields and could adequately cope with minor cross-talk between the channels that result from small deviations from orthogonality of the magnetic field coils. The trained networks were subsequently used to map the raw data to calibrated two-dimensional head-positions, yielding azimuth- and elevation response components with an absolute accuracy within 4% over the entire response range.

Head movement detection. Saccadic head movements were detected from the calibrated head-movement signals by setting thresholds to the vectorial head velocity for on- and offset, respectively, using a custom-made program (onset velocity=20 deg/s, offset velocity=15 deg/s). Detection markings from the program were visually checked by the experimenter, and could be adjusted manually, when deemed necessary. Head movements with reaction times below 80 ms, or above 1000 ms, were discarded, because responses with extremely short latencies may be regarded as anticipatory, and responses with excessive latencies are usually due to inattentiveness of the listener.

Statistics. Each listener’s responses were quantified by determining the optimal linear fit for the following stimulus-response relations:

\[
(3.2) \quad \alpha_R = a + b \cdot \alpha_T \quad \text{and} \quad \varepsilon_R = c + d \cdot \varepsilon_T
\]

for the azimuth and the elevation components, respectively, by minimizing the least-squares error (Press et al. 1992). In Eq. 3.2, \( \alpha_R \) and \( \varepsilon_R \) are the azimuth and elevation response components, \( \alpha_T \) and \( \varepsilon_T \) are the azimuth and elevation coordinates of the target. Fit parameters, \( a \) and \( c \), are the response biases (offsets, in degrees), while \( b \) and \( d \) are the overall response gains (slopes, dimensionless) of the azimuth and elevation response components, respectively. An ideal listener yields a gain of one and an offset of zero deg. Also, Pearson’s linear correlation coefficient and residual errors around the regression line were calculated. Because in the plugged hearing condition the regression results heavily depended on the applied sound level, regressions were performed separately for each sound level. Localization performance was also quantified by determining the mean absolute error (MAE) of the responses:

\[
(3.3) \quad \text{MAE}(\alpha) = \frac{1}{N} \sum_{i=1}^{N} |\alpha_R - \alpha_T| \quad \text{and} \quad \text{MAE}(\varepsilon) = \frac{1}{N} \sum_{i=1}^{N} |\varepsilon_R - \varepsilon_T|
\]

for the azimuth and the elevation components, respectively, with \( N \) the total number of trials.
In order to account for the strong azimuth-dependence of the azimuth and elevation response components in the plugged hearing conditions (see Results), data analysis was also performed within restricted regions of azimuth space (local regression/local MAE). To that end, responses were collected within 25 deg-wide azimuth bins (pooled across elevation), each shifted in 5 deg steps (thus with 20 deg overlap between adjacent bins), from which we determined a smooth estimate of the local azimuth and elevation regression parameters and the MAE (e.g. Figure 3.4; Zwiers et al., 2003).

We also compared the gain and bias between the plug and the control session by computing the relative gain:

\[
\begin{align*}
    b_{rel} &= \frac{b_{plug}}{b_{control}} \quad \text{and} \quad d_{rel} = \frac{d_{plug}}{d_{control}}
\end{align*}
\]

and the change in bias:

\[
\begin{align*}
    \Delta a &= a_{plug} - a_{control} \quad \text{and} \quad \Delta c = c_{plug} - c_{control}
\end{align*}
\]

for azimuth \((b_{rel}, \Delta a)\) and elevation \((d_{rel}, \Delta c)\) components, respectively.

The bootstrap-method was applied to determine confidence limits for the optimal fit parameters in the regression analyses. To that end, 100 data sets were generated by randomly selecting (with replacement) data points from the original data set. 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).

### 3.3 Results

**Acute effects of a unilateral plug**

The attenuation of sounds with a plug in the left ear (Fig. 3.1), while leaving the right ear unperturbed, changed a listener’s sound localization ability quite profoundly. In figure 3.2 the acute effect of the plug is exemplified by comparing the localization performance of a typical listener (RK) to HP stimuli during normal hearing (Fig. 3.2A and B; pooled across intensities), with the acute plug condition (Fig. 3.2C and D). With normal binaural hearing, this listener was quite precise, accurately localizing both sound azimuth (Fig. 3.2A) and elevation (Fig. 3.2B), regardless of the stimulus intensity. Despite the large range in stimulus levels, including some as low as 30 dBA, regression lines were near-optimal (as indicated by an overall regression gain near 1.0 and a bias near 0.0 deg), with modest scatter of the data around the regression line (residual errors: ±10.8 and ±11.4 deg, respectively).

The plugged condition, however, introduced a large shift in the listener’s azimuth localization responses towards the unplugged ear (Fig. 3.2C; bias \(a = 45.2\) deg). In addition, the response gain \((b = 0.18)\) and correlation between target and response location degraded substantially. Despite this clear detriment in sound localization (residual error: ±16.3 deg), however, the stimulus-response correlation was still positive and significant \((r = 0.3, p<0.05)\). The listener’s elevation responses, on the other hand, were affected much less, as there was only a slight decrease in both the gain \((d = 0.76)\) and the correlation \((r = 0.85, p<<0.01; \text{residual error: ±14.5 deg})\). As such, these data are in line with previous reports on sound localization to fixed-intensity stimuli under acutely plugged listening conditions (Oldfield and Parker 1986; Slattery and Middlebrooks 1994). The observed effects also underline the strong dominance of the binaural cues for normal-hearing sound localization in the horizontal plane.
Because the plug appeared to have a dominant effect on the azimuth response components, we will first focus on the results of azimuth localization. The effects on elevation localization will be dealt with later in this section.

**Influence of sound intensity and source azimuth on azimuth localization**

While pooling the data across intensities clearly demonstrates the deterioration of azimuth localization, it ignores a potential systematic effect of sound level. As an illustration, figure 3.3 shows the localization responses of listener RK for three different intensities (30, 45 and 60 dBA, HP stimuli). For normal binaural hearing, no noticeable effect of intensity appeared (Fig. 3.3A, B, C). In the acute plug condition, however, changes in sound level influenced the listener’s responses considerably (Fig. 3.3D, E, F). For all intensities, sound azimuth localization was clearly perturbed, but the overall localization bias towards the unaffected ear increased strongly with intensity. Higher intensities induced a larger shift in the acute localization bias.
plug condition (e.g. at 60 dBA: bias $a = 56$ deg; Fig. 3.3D) than lower intensities (e.g. at 30 dBA: $a = 36$ deg; Fig. 3.3F). Note that the azimuth response gain was highest for the low intensity stimulus ($b = 0.14$ at 60 dBA, but $b = 0.32$ for the 30 dBA sounds). This is remarkable since at low intensities the listener was effectively monaural, while for the higher intensities the plugged ear still received acoustic input, albeit strongly attenuated (see Fig. 3.1). Thus, although the strongest perturbations of the ILDs were obtained for low-intensity sounds, localization performance was better than for the higher sound levels.

Figure 3.4A shows the systematic influence of sound level on the azimuth localization bias of HP stimuli. Note that all listeners demonstrated a similar effect: the bias is at a minimum for the lowest sound level, and increased monotonically with intensity. Although the actual bias values varied considerably from listener to listener, the variation of the bias with sound level was quite similar for each of the listeners: over a 30 dBA intensity range, the response bias shifted by approximately 20 deg.

Also the spatial gain (quantified by the azimuth regression slope) changed in a systematic way with sound level. Figure 3.4B shows for each individual listener the gain as function of sound level, after subtracting each listener’s mean azimuth gain (which is shown in the inset in Fig. 3.4B). The data show that azimuth gain tended to be highest for the lowest stimulus levels, with a dip at intermediate levels, for all listeners, except listener MW (whose mean azimuth gain was not significantly different from 0, see inset).

Interestingly, the azimuth gain and bias in the plugged condition were highly negatively cor-
related, as shown in figure 3.4C (BB and HP stimuli, and listeners pooled; for HP stimuli: \( r^2=0.58, p<0.01; \) for BB stimuli: \( r^2=0.31, p<0.05 \)). This indicates that a smaller shift in the azimuth responses co-varied with a larger spatial gain.

To verify whether the observed differences in the azimuth regression parameters across listeners could be attributed to inter-subject differences in the perceptual attenuation of the plug we compared the regression results with each listener’s audiogram (compare with Fig. 3.1A). The result is shown in figure 3.4D, in which for each listener the azimuth bias (averaged across intensity) is plotted against the subjective attenuation of the plug (averaged across the HP and LP frequency bands; see Fig. 3.1A). Interestingly, we obtained no correlation for either the HP stimuli (containing only potential ILDs), or the BB sounds (both ITDs and ILDs may be present). A similar result was obtained for the azimuth gains (not shown). Thus, the plug’s attenuation, which corresponds to a fixed, but frequency-dependent perturbation of the ILDs, does not predict the value of the shift in perceived sound-source azimuth.

We next studied whether listeners with a monaural plug could have relied on spectral-shape cues to localize azimuth. Note that the plug caused an immediate effect on the quality of the spectral cues across the azimuth domain: on the side ipsilateral to the plug the spectral cues were strongly attenuated (and often abolished altogether), while on the side of the normal-
Thus, if listeners would only rely on spectral cues to extract the sound-source azimuth, response accuracy should depend on the sound’s azimuth location. To exemplify our analysis, figure 3.5 quantifies the dependence of azimuth performance on the acoustic parameters for all HP stimuli for listener MW, for both the control (top) and plugged (bottom) hearing conditions. In each 7x10 matrix, an entry corresponds to a particular combination of sound level (abscissa) and target azimuth (ordinate), whereas the average response azimuth (Fig. 3.5A, B, D) and the mean absolute error (MAE, Eq. 3.3; Fig. 3.5C, E) values are grayscale encoded. Figure 3.5A illustrates the appearance of the localization matrix for an ideal listener, whose responses do not depend on sound level, and whose response azimuth corresponds exactly to target azimuth ($\alpha_R = \alpha_T$). Note that in the control hearing condition (Fig. 3.5B) the responses of listener MW corresponded quite well to those of the ideal listener. As a result,
Perturbed Binaural Sound Localization

the azimuth MAE was small (Fig. 3.5C, about 5-10 deg), and did not vary systematically across the azimuth-intensity parameter space.

In the plugged condition, however, the listener’s responses showed a marked shift towards the unplugged ear (as the light-gray color indicates far-rightward responses). Note, however, that the shift is slightly less (darker gray) for the low-intensity stimuli on the side of the unplugged ear (Fig. 3.5D). This systematic difference becomes more evident when the MAE in azimuth is plotted as a function of stimulus intensity and target azimuth (Fig. 3.5E). The MAE is minimal for low-intensity stimuli on the unplugged hearing side (lower-right, dark-gray voxels), while it increased systematically for targets toward the plugged side, and for the higher sound levels (left and upper voxels).

Despite some quantitative inter-subject variability, all listeners demonstrated a similar effect for the HP stimuli (Fig. 3.6): the MAE reached a minimum for low-intensity sounds presented on the side of the normal-hearing ear (dark-gray, lower-right voxels), and increased both with increasing intensity and distance from the unplugged ear (light-gray, upper-left voxels). The results for the BB stimuli were indistinguishable from the HP data (not shown). We also performed a local regression analysis on these data (see Methods), which revealed a similar trend: the local azimuth bias was lower, and the local azimuth gain was highest for the low-intensity stimuli contralateral to the plug (data not shown).

In summary, in the acute monaural plug condition listeners localized low-intensity sounds
better than high-intensity sounds on the unplugged hearing side (evidenced by a lower MAE for azimuth), despite a total absence of the binaural difference cues for the low-intensity stimuli. On the plugged side, azimuth localization was always poor regardless of the stimulus intensity. These data therefore strongly suggest that plugged listeners relied on the monaural spectral cues from their unplugged ear, especially for the low-intensity stimuli for which the ILDs and ITDs were either poor or non-existent. Furthermore, these spectral cues appeared to be binaurally weighted, as their contribution depended in a systematic and gradual way on azimuth.

**Influence of intensity and azimuth on elevation localization**

Figure 3.2D suggested a relative robustness of a listener’s ability to localize sound-source elevation when one ear is plugged. Yet, performance was affected, as evidenced by a larger amount of scatter and a lower gain, even for the pooled responses. A detrimental effect on the pooled data was to be expected, as no spectral cues survived at the plugged side for nearly all sounds. In what follows, we will quantify the effect of sound level and azimuth on the subject’s elevation performance in more detail.

As a first step in our analysis, Figure 3.7 plots the linear regression results on elevation for listener RK for three sound levels for control hearing (top row), and for the plugged hearing condition (bottom). The data are pooled for azimuth locations across the frontal hemifield. The insertion of a plug immediately reduced the overall gain for all three sound levels. The
Perturbed Binaural Sound Localization

The elevation response bias appeared to change slightly, but systematically, with sound level when compared to the control experiment (change in bias \( \Delta c \), Eq. 3.5, equaled -8.2, -4.0, and +1.5 for 60, 45, and 30 dBA stimuli, respectively).

Consistent changes in bias and gain are better observed for the relative elevation gain \( d_{\text{rel}} \) (Eq. 3.4) and the change in elevation bias \( \Delta c \), Eq. 3.5) for all sound levels and all participants (Fig. 3.8). To account for a potential systematic azimuth-dependence, we first performed the regression-analysis for two separate regions: for stimuli on the side of the plug \( \alpha_T < -20 \) deg, Fig. 3.8A and C), vs. stimuli on the normal-hearing side \( \alpha_T > 20 \) deg, Fig. 3.8B and D). The mean relative elevation gain (averaged across sound levels) on the plugged side varied somewhat from listener to listener, although it was clearly worse than in the control condition for all listeners \( d_{\text{rel}}<1 \), Fig. 3.8A, inset). All listeners, except for JO (open circles), exhibited a similar effect on the elevation gain as function of sound level (Fig. 3.8A): higher sound levels
elicited a higher response gain. Note that listener JO was not able to localize elevation at all on the plugged side (see inset Fig. 3.8A, white bar). Also on the normal-hearing side, the mean relative elevation gain decreased for all listeners (Fig. 3.8B, inset), although not as much as on the plugged side. Surprisingly, however, the effect of sound level on elevation gain was now reversed (Fig. 3.8B): on the hearing side, low stimulus intensities elicited higher response gains. Apparently, both sound level and azimuth determined gain in the elevation direction. The overall change in elevation bias (Eq. 3.5, averaged across sound levels) was characterized by a substantial decrease on the plugged side (Fig. 3.8C, inset), while for all listeners, except MW, it was barely affected on the normal-hearing side (Fig. 3.8D, inset). Furthermore, the change in bias was influenced by sound level in a similar way on both the plugged and the normal-hearing side: lower stimulus intensities elicited a higher (i.e. more upward) change in elevation bias than louder sounds.

To quantify the influence of sound-source azimuth on the localization of elevation with a monaural plug in more detail, we adopted a similar local regression analysis as for the azimuth responses. Figure 3.9 (A-D) shows, in a similar format as figures 5 and 6, the local elevation gain (left) and bias (right) as a function of sound azimuth and level for listener MW. The top row shows the data for the control hearing condition. Note that local elevation gain is uniformly high (light-grey) across the frontal hemifield (Fig. 3.9A), while the local bias is around +15 deg, with a tendency to slightly increase for higher HP sound levels (Fig. 3.9B). For the plugged hearing condition, the response bias decreased to about 5 deg (Fig. 3.9D). The local elevation gain (Fig. 3.9C) depended systematically on both the azimuth and intensity parameters: the maximal local elevation gain was obtained for sounds on the normal-hearing side, at the lowest intensities, while it decreased for higher sound levels and

**Figure 3.9. The influence of sound intensity and target azimuth on plugged elevation localization.**

The format is similar to figure 3.5, but grayscale now codes for elevation gain (A, C), bias (deg)(B, D), relative elevation gain (E) and change in elevation bias (deg)(F) for the non-plugged (A, B), plugged (C, D), and the plugged vs. control condition (E, F). Note the high relative elevation gain for low-intensity stimuli on the normal-hearing side.
for stimuli on the plugged side. The bottom row shows the same data, now expressed as the local relative elevation gain (Eq. 3.4, Fig. 3.9E), and the local change in bias (Eq. 3.5; Fig. 3.9F). Whereas the relative elevation gain showed a similar trend as the absolute elevation gain, the overall change in bias was negative, and did not systematically depend on the stimulus parameters for this listener.

Figure 3.10 shows the results of this analysis for the other four listeners. Despite the variability in absolute values, listeners yielded quite consistent results; elevation localization had its highest spatial gain at locations far into the normal-hearing hemifield, and for the lowest intensity stimuli (lower-right voxels). Higher stimulus intensities, as well as locations on the plugged side yielded lower elevation gains. The plug also influenced the elevation bias (Fig. 3.10E-H), in a pattern that was qualitatively similar across listeners (with the exception of MW. Fig. 3.9F). Listeners responded more upwards for the low-intensity stimuli on their normal-hearing side, while systematically pointing to more downward locations for high-intensity stimuli and for stimuli on the plugged side. Results for the BB-stimuli were very similar (data not shown).

**Weighting of cues**

Our results on the localization of sound-source azimuth (Figs. 3.2-6) give rise to the idea that the computation of azimuth is not solely determined by binaural difference cues, but also by the spectral-shape cues of the normal-hearing ear. In turn, the elevation data (Figs. 3.7-10) suggest that apart from spectral-shape cues, also azimuth is an important factor that contributes to the elevation percept. Here, one should distinguish the actual source azimuth from the perceived azimuth location, as under plugged hearing these locations are very different (e.g. Fig. 3.2C). Previous studies with molds have shown that the computation of elevation involves binaural interactions (see Introduction), but it remains unclear whether this concerned the actual azimuth, or the perceived azimuth location as a mold does not perturb the azimuth percept.

Note that the spectral cues on the hearing side were unperturbed under monaural plugged
hearing. In other words, if the actual source azimuth would be the determining factor, elevation would be perceived at its veridical location for all azimuths on the hearing side. As Figs. 3.8-10 show, this was not the case, and we therefore wondered whether in fact the perceived azimuth location might have determined elevation. To assess this point, we selected all responses for HP-targets presented on the far-lateral hearing side (source azimuths between 40-60 degs), as for these locations the spectral cues would be optimal for all hearing conditions. Separately for each listener, and each stimulus level, we then took the azimuth MAE of these responses as a measure for the perceived azimuth location at these nearly fixed azimuth positions: if the MAE is low, the perceived azimuth approaches the veridical azimuth, whereas for large MAEs the perceived azimuth is very different. If the actual azimuth would determine the perceived elevation, elevation gain should not depend on the MAE. Instead, figure 3.11A shows that for all listeners the elevation gain varied with the MAE: the larger the MAE, the lower the elevation gain, which suggests that perceived azimuth is the

**Figure 3.11. The influence of spectral cue saliency and azimuth percept on plugged localization.** (A) Relative elevation gain vs. change in azimuth MAE (both parameters corrected for individual differences by subtracting their means) for stimuli between 40 and 60 deg azimuth, for all listeners and sound levels... (B) Relative elevation gain vs. stimulus intensity (dBA). Note that the highest levels (indicative for high signal-to-noise ratios) elicit the lowest gains. (C) Change in azimuth MAE (deg) vs. relative elevation gain (means subtracted) for all listeners and different azimuth sectors. Sound level: 30 dBA. The lower the elevation gain, the higher the MAE, and vice versa.
Perturbed Binaural Sound Localization

relevant parameter.

Note, that as the change in stimulus intensity (and hence, perhaps, the integrity of the spectral cues) is implicit in Fig. 3.11A, one might suspect that the low elevation gains could have resulted from a poor signal-to-noise ratio of the spectral shape cues at low intensities. However, figure 3.11B shows that in fact the opposite is true: the lower stimulus intensities produced the highest elevation gains for all listeners (see also Fig. 3.9-10). Therefore, the spatial resolution in elevation depended on the MAE, and hence was determined by the perceived azimuth location, rather than by the actual azimuth location.

In figure 3.11C we tested whether the perceived azimuth angle was in turn determined by the spectral-shape cues. To that end, we plotted azimuth performance (i.e. the local MAE for each of the ten different azimuth regions) as a function of the relative elevation gain, but now for the low-intensity stimuli only (30 dBA). Note that for all listeners, the MAEs for these stimuli were consistently correlated with the elevation gain: the higher the relative elevation gain (i.e. the better the spectral cues were resolved by the auditory system), the lower the azimuth error. These data therefore strongly support the idea that for plugged hearing, the spectral cues do indeed contribute to azimuth localization.

3.4 Discussion

We studied the effect of an acute perturbation of the binaural localization cues on 2D sound localization behaviour. All listeners immediately mislocalized sounds predominantly in the azimuth direction (e.g. Fig. 3.2C). Localization of target elevation was also affected, albeit in a more subtle way (Fig. 3.2D). We found that listeners were best at localizing low-intensity sounds on the hearing side for both the azimuth and elevation directions. This is remarkable, as at higher intensities the spectral-shape cues have a better signal-to-noise ratio (SNR). However, at the higher intensities binaural inputs will overcome the plug’s attenuation, although they are highly perturbed (Fig. 3.1). The plug thus creates ILDs that far exceed the normal physiological range provided by the head-shadow. Yet, these erroneous cues were incorporated in forming the azimuth percept, giving rise to larger localization errors than for low-intensity stimuli for which the binaural cues were absent (Figs. 3.3-6). The presence of perturbed binaural difference cues also affected elevation localization at higher intensities (Figs. 3.8-10).

We conclude from these findings that under plugged hearing conditions, listeners employ spectral-shape cues from the normal-hearing ear to localize sound-source azimuth (Fig. 3.11C). We also conclude that target elevation is not only determined by the spectral-shape cues of the ear ipsilateral to the sound, but also results from a binaural weighting process. However, this weighting is strongly influenced by the perceived azimuth location of the sound, rather than by the actual azimuth (Fig. 3.11A,B).

Spectral cues for azimuth

Several arguments support our conclusion that plugged azimuth localization depends on spectral cues. First, for low-intensity sounds binaural cues were virtually absent (Fig. 3.1A), suggesting that listeners had no other possibility but to use the spectral cues. Furthermore, the extent of the plug’s attenuation (20-55 dBA) suggests that listeners did not exclusively employ the perturbed ILDs to generate their responses. The ILDs of a human head reach a maximum of about 20 dBA for the extreme lateral positions (e.g. Blauert, 1997; Van Wanrooij and Van Opstal, 2004). Thus, an ILD of 20 dBA corresponds to a location beyond 60 deg from the midline. However, plugged listeners responded to far less extreme locations, even for the HP stimuli (around 30 deg; e.g. Figs. 3.2C, 4A), suggesting that they also relied on other cues. Indeed, the insertion of a mold in the normal-hearing ear, when a plug plus additional muff attenuated sounds by at least 40 dBA, completely abolished sound localization.
performance in azimuth and elevation (Hofman, 2003), showing the importance of spectral cues for monaural hearing.

Our findings show that for plugged hearing the spectral cues dominated for the lowest sound levels (e.g. Figs. 3.5 and 3.6). Indeed, azimuth localization performance gradually decreased toward the plugged side, in concordance with a decrease in elevation performance (Fig. 3.11C). Thus, when binaural difference cues were severely perturbed, or became ambiguous, the spectral localization cues became increasingly important.

We believe that the weighting of the different cues changes instantaneously, as in our experiments the stimuli were randomized. Thus, although the plug introduces a large and frequency-dependent conflict between the binaural localization cues at higher sound levels, they were still favoured over monaural spectral-shape cues. Under normal binaural hearing the difference cues dominate entirely, as azimuth localization is robust against large variations in sound level and SNRs (e.g. figure 3.3A-C, see also Good and Gilkey 1996; Hofman and Van Opstal 1998; MacPherson and Middlebrooks 2000; Van Wanrooij and Van Opstal 2004; Vliegen and Van Opstal 2004; Zwiers et al. 2001b), and does not depend on the integrity of the spectral cues (Hofman and Van Opstal 2003; Morimoto 2001; Oldfield and Parker 1984; Van Wanrooij and Van Opstal 2005; Wightman and Kistler 1997).

**Azimuth cues for elevation**

Our results show that a monaural plug had a systematic effect on the localization of sound-source elevation, not only on the side of the plug, but also on the normal-hearing side. Although the effects were smaller than the dramatic localization deficits for azimuth, they were systematic and consistent across listeners. Butler and coworkers (1990) have shown that under plugged hearing, advance knowledge of the sound’s azimuth may enhance elevation performance. In line with this, findings obtained with a monaural mold (Hofman and Van Opstal 2003; Humanski and Butler 1988; Morimoto 2001; Van Wanrooij and Van Opstal 2005) indicated that the localization of elevation involves binaural interactions, the strength of which varies gradually with azimuth. The present study extends these results by showing that the binaural weighting depends on the perceived azimuth location, which in case of plugged hearing may be quite erroneous (Fig. 3.11A,B). This is a remarkable finding as the spectral shape cues, the sound level, and the SNR of the peaks and notches in the sound spectra at the normal-hearing side were all unaffected by the plug. When localizing elevation, acutely plugged listeners showed a marked decrease in performance on the plugged side for the lower sound intensities (Fig. 3.8A). In contrast, listeners performed better for low-intensity sounds at their normal-hearing side (Fig. 3.8B). We believe that two different factors underlie these seemingly different behaviors. First, the spectral cues of the plugged ear have a low SNR, which affects low-intensity sounds more than high-level sounds (Fig. 3.8A). Second, the elevation percept comes about by fusing the spectral cues from each ear with a weight determined by the perceived azimuth (see above), which on the normal-hearing side is perturbed more at higher intensities than at low intensities (e.g. Figs. 3.5-3.6, 3.8B).

**Monaural Listening**

Our results extend recent findings obtained from monaurally deaf listeners (Van Wanrooij and Van Opstal 2004). That study showed that all monaural listeners relied heavily on the head-shadow effect (i.e. absolute sound level at the hearing ear) to localize sounds in the horizontal plane, whereas half of the listeners also incorporated the spectral-shape cues of their intact ear to estimate azimuth. Interestingly, only listeners who had learned to use spectral cues to localize azimuth, could also localize elevation on the side of their intact ear. Monaural listeners who did not employ the spectral cues for azimuth could not localize elevation either, which hinted at the possibility that the ability to localize elevation strongly depended on the performance in azimuth.
Perturbed Binaural Sound Localization

Plugging the ear of an otherwise normal-hearing binaural listener is very different from real monaural hearing of the unilaterally deaf for a number of reasons: First, although the plug strongly attenuates sounds, the acoustic input will not be entirely abolished as in the monaurally deaf. Thus, for sounds at a sufficiently high intensity, all localization cues are still present, albeit heavily perturbed. Second, the acoustic effect of the plug typically depends on frequency (Fig. 3.1), yielding ambiguous localization cues when compared to normal hearing. For example, not only will the ILDs differ for different frequency bands (and hence point to different azimuth locations), also the ITDs and ILDs are affected in a different way. Thus, the outputs of the two binaural localization streams will often not agree on sound-source azimuth either. For sufficiently low sound levels, however, plugged hearing approaches real monaural hearing. The plug’s intensity- and frequency dependence therefore poses an interesting and non-trivial challenge to the sound-localization system. Third, the monaurally deaf have had long-term exposure to their hearing condition, allowing ample time for adaptive processes to reshape their localization behavior. This contrasts with the immediate and complex effect of a plug on localization of the normal-hearing listener. Conversely, binaural listeners have had ample experience in employing the binaural difference cues and the detailed complex spectral shape cues from either ear.

Taken together, it is reasonable to expect that the central organization of the sound localization systems of the monaurally deaf and of binaural listeners may be quite different (see also Bilecen et al. 2000; Ponton et al. 2001; Scheffler et al. 1998).

Integration of acoustic cues

We propose that azimuth and elevation are both computed on the basis of evidence from all available acoustic cues, but that their relative weights depend on the acoustic conditions. Figure 3.12 depicts this conceptual model, which extends the classical idea that the stimulus coordinates are determined by independent, non-interacting pathways. The localization of azimuth is based on a weighting of binaural difference cues as well as of spectral shape cues. When the binaural cues become unreliable or ambiguous (e.g. for weak sounds at far-lateral locations, or after plugging one ear) their weights are reduced, and the contribution of the spectral cues from the contralateral ear increases. In turn, the computation of sound-source elevation involves a weighting of the spectral shape cues from both ears. The strength of the binaural weighting is modulated by the perceived azimuth location.

Neural mechanisms

At the initial stages in the auditory system, the localization cues are processed by independent brainstem pathways, both in birds (e.g. the barn owl; Takahashi 1989, for review) and in mammals. In mammals, the medial superior olive (MSO) constitutes the ITD pathway, while the ILD pathway is processed in the lateral superior olive (LSO; Irvine 1986; Yin 2002, for reviews). The elevation pathway has yet to be identified, but recent evidence suggests that the first stages of spectral shape analysis may already occur at the level of the dorsal cochlear nucleus (Reiss and Young 2005; Young and Davis 2002).

Our finding that the percept of sound-source azimuth is determined by spectral cues, when the binaural cues become unreliable, could be due to pre-attentive mechanisms that rely entirely on the acoustic properties of the signal. As the different brainstem pathways all converge in the midbrain, the inferior colliculus (IC) could play a role in such pre-attentive computations. Although an explicit map of auditory space has not been demonstrated in the mammalian IC, unlike in the barn owl (Knudsen and Konishi 1978), recent evidence suggests that auditory space may be represented in the IC by space-specific modulations within a large population of neurons. For example, the majority of monkey IC cells are sharply tuned to sound frequency, but their firing rates are also modulated by sound level, the location of the sound in azimuth and elevation, and by non-acoustic signals like eye position.
On the other hand, our result that the elevation of a sound is determined by the perceived azimuth, rather than by the acoustically defined azimuth (Fig. 3.11), might suggest that higher, perhaps cortical, mechanisms are involved. Recent recordings have indicated that populations of cells in the primate auditory cortex may encode sound locations in a similar way as the IC (Recanzone 2000; Werner-Reiss et al. 2003). Whether the auditory cortex may be involved in the perceptual integration of acoustic cues, has to be established by future recording studies.

**Acknowledgements**

This work was supported by the Radboud University Nijmegen (AJVO, MMVW) and by the Human Frontiers Science Program (MMVW; RG 0174-1998/B). We thank Ger Van Lingen, Hans Kleijnne, Gunther Windau, Huib Versnel and Ton Van Dreumel for technical assis-
Perturbed Binaural Sound Localization

tance.
CHAPTER IV
RELEARNING SOUND LOCALIZATION WITH A NEW EAR

4.1 INTRODUCTION

From the basilar membrane up to the cortex the auditory system is tonotopically, rather than spatially, organized. As a result, sound localization relies on implicit cues in the sound-pressure wave (Blauert, 1997). Head size and resulting acoustic shadow impose interaural differences in sound-arrival time and level that relate to locations in the horizontal plane (azimuth; Blauert, 1997). In addition, the pinnae distort the high-frequency sound-spectrum, yielding spectral-shape filters that vary uniquely with locations in the horizontal and vertical plane (elevation; Blauert, 1997; Oldfield and Parker, 1984a; Middlebrooks and Green, 1991; Middlebrooks, 1992; Wightman and Kistler, 1989). It is generally thought that independent binaural and monaural neural pathways process the different localization cues (Oldfield and Parker, 1986; Yin, 2002; Young and Davis, 2002).

Data suggest that correct azimuth localization can only be achieved through binaural hearing, since the unilaterally deaf (Van Wanrooij and Van Opstal, 2004), listeners exposed to ear-plugging (Oldfield and Parker, 1986), or dichotically simulated monaural hearing (Wightman and Kistler, 1997) are very poor in azimuthal localization. Some monaurally deaf listeners, however, can localize elevation with their normal ear (Slattery and Middlebrooks, 1994; Van Wanrooij and Van Opstal, 2004). Although spectral shape constitutes a monaural localization cue, studies have shown that perceived elevation is partly due to binaural interactions, in which each ear’s contribution is weighted by stimulus laterality (Fig. 4.1A, B). For example, monaural pinna occlusion causes elevation deficits, not only on the side of the perturbed ear, but also well into its contralateral hemifield (Humanski and Butler, 1988; Oldfield and Parker, 1984b; Morimoto, 2001; Hofman and Van Opstal, 2003). The complex relationship between pinna geometry and spectral-shape cues, together with the fact that human ears change size and shape throughout life, suggest that adaptive mechanisms in the auditory system should preserve calibration of sound-source elevation. Indeed, a study in which binaural molds perturbed the spectral-shape cues demonstrated that listeners relearned localization within a few weeks (Hofman et al., 1998). Interestingly, listeners learned the new spectral cues without interfering with the original ones, as they could localize with both their own, and with occluded ears.

As outlined in figure 4.1A, B, this remarkable adaptive response can in principle be explained by two different mechanisms. First, the auditory system could have learned new spectral representations, separate for each ear (adaptation at a spectral-to-spatial mapping stage). Alternatively, the binaural weighting could have been adjusted such that one ear dominates both ipsi- and contralateral space. To dissociate these possibilities, we investigated sound-localization behavior after adaptation to a unilateral mold (FM listening) under four different hearing conditions: normal-ear listening (FF), a unilateral mold left or right (MF, or FM), and listening with binaural molds (MM; Fig. 4.1C). Since the normal mappings are not lost, no effect on FF or MF performance would be expected. The performance in the FM case improves on the mold side due to the adaptation for both stages in either scheme. Adjustments at the spatial mapping stage of either scheme in figure 4.1A, B predicts that localization on the adapted side remains accurate under MM listening. In contrast, changes at the binaural weighting stage preclude accurate MM localization, as the increased weighting of the left ear is then based on unknown spectra. Our results demonstrate that the system can cope with monaural spectral perturbations, and that plasticity acts predominantly at the spectral-to-spatial mapping stage.
Conceptually, elevation processing consists of two stages (Hofman and Van Opstal, 2003): a spectral-to-spatial mapping stage, and a binaural weighting stage (W) that combines information from both ears. Since both processes are nonlinear, the order in which they occur, matters and adaptation at either stage leads to different predictions for our control experiments. In scheme A binaural weighting acts on monaural elevation estimates, whereas in scheme B it combines spectral shape inputs from the ears into binaural spectral shapes. (C) Prediction of elevation performance on the adapted side after complete right-ear mold (FM) adaptation for four different hearing conditions, if adaptation occurred at either the spatial mapping stage (left) or the binaural weighting stage (right) of either scheme. In the latter case, the contribution of the non-perturbed, left ear has increased and accounts for improved performance on its contralateral, right side. Consequently, an additional mold in the left ear (MM hearing) abolishes elevation performance on the right (right-hand panel). If adaptation occurred at the spatial mapping stage, MM hearing will not affect localization on the adapted side (left-hand panel).
4.2 METHODS

Participants
Eleven listeners (ages 21-47) participated in the experiments. Three participants (JO, JV, and MW) are from the lab and are experienced with sound localization studies. Two of these listeners (JO and MW, the authors) each participated in two adaptation series with different molds that were conducted two years apart. The other eight listeners were naive paid volunteers and were given a short practice session to localize sounds with head movements under open-loop conditions (i.e., no feedback was given to the actual performance of the listener), before entering these experiments. All listeners had normal hearing (within 20 dB of audiometric zero) as determined by an audiogram obtained with a standard staircase procedure (10 tone pips, 0.5 octave separation, between 500 Hz and 11.3 kHz). None of them had any auditory or uncorrected visual disorder, except for listener JO who is amblyopic in his right eye.

Apparatus
During the experiments, the listener was seated comfortably in a chair in the center of a completely dark, sound-attenuated room (HxWxL = 2.45x2.45x3.5 m³). The walls, ceiling and floor, and every large object present were covered with black acoustic foam that eliminated echoes for sound frequencies above 500 Hz. The room had an ambient background noise level of 25 dB sound pressure level, A-weighted (dBA).

A total of 58 small broad-range loudspeakers (MSP-30; Monacor International GmbH, Bremen, Germany) containing light-emitting diodes (LEDs) were mounted on a thin wooden frame that formed a hemispheric surface 100 cm in front of the listener at six eccentricities, R = [0, 15, 30, 45, 60, 75] deg, relative to the straight-ahead viewing direction (defined in polar coordinates as [R, Φ] = [0, 0] deg), and at twelve different directions, Φ=[0, 30, ..., 330] deg, where Φ = 0 deg is rightward from the center location, and Φ = 90 deg is upward (see gray background in Fig. 4.2, 12A,B and 13 for a simple illustration of the stimulus lay-out). The lower three speakers (at R = 75 deg, and Φ = [240, 270, 300] deg) were left out to create room for the listener’s legs.

Head movements were recorded with the magnetic search-coil induction technique (Robinson, 1963). To that end, the listener wore a lightweight (150 g) “helmet” consisting of two perpendicular 4cm-wide straps that could be adjusted to fit around the listener’s head without interfering with the ears. On top of this helmet, a small coil was attached. From the left side of the helmet a 40 cm long, thin aluminum rod protruded forward with a dim (0.15 Cd/m²) red LED-attached to its end, which could be positioned in front of the listener’s eyes. Two orthogonal pairs of 2.45x2.45 m² coils and one pair of 2.45x3.5 m² coils were attached to the room’s edges to generate the left-right (60 kHz), up-down (80 kHz) and front-back (40 kHz) magnetic fields. The head-coil signal was amplified and demodulated (Remmel Labs, Ashland, MA), after which it was low-pass filtered at 150 Hz (model 3343; Krohn-Hite, Brockton, MA) before being stored on hard disk at a sampling rate of 500 Hz/channel for off-line analysis.

Auditory stimuli
Acoustic stimuli were digitally generated using Tucker-Davis System II hardware (Tucker-Davis Technologies, Alachua, FL), with a TDT DA1 16-bit digital-to-analog converter (50-kHz sampling rate). A TDT PA4 programmable attenuator controlled sound level, after which the stimuli were passed to the TDT HB6 buffer and finally to one of the speakers in the experimental room.

All acoustic stimuli consisted of Gaussian noise, and had 0.5 ms sine-squared on- and off-
set ramps. The auditory stimuli were either broadband (BB, flat characteristic between 1-20 kHz) or high-pass (HP, high-pass filtered at 3 kHz) stimuli with a duration of 150 ms. Sound intensities ranged from 30 to 60 dBA (see below). Absolute free-field sound levels were measured at the position of the listener's head with a calibrated sound amplifier and microphone (BK2610/BK4144; Bruel & Kjaer, Norcross, GA).

**Molds**

Listeners participated in a long-term adaptation experiment, which required them to wear a lightweight and precisely fitting custom-made mold in the concha cavity of either their left or right pinna for an extended period of time. The molds were manufactured by filling the concha with rubber casting material (Otoform Otoplastik-K/c; Dreve GMBH, Unna, Germany). Molds for both ears were produced, as control experiments were run with a mold in either ear (see below). The ear canals were kept free, and before its hardening, the material could be easily shaped in situ to exactly fit the pinna cavity.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (years)</th>
<th>Mold</th>
<th>Time (days)</th>
<th>Number of Tests</th>
<th>Bandwidth</th>
<th>Intensities (dBA)</th>
<th>DTFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>JO1</td>
<td>45</td>
<td>Right</td>
<td>23</td>
<td>10</td>
<td>HP, BB</td>
<td>30, 40, 50</td>
<td>-</td>
</tr>
<tr>
<td>MW1</td>
<td>25</td>
<td>Right</td>
<td>49</td>
<td>16</td>
<td></td>
<td>30, 40, 50</td>
<td>-</td>
</tr>
<tr>
<td>DH</td>
<td>24</td>
<td>Left</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>JO</td>
<td>47</td>
<td>Right</td>
<td>7</td>
<td>6</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>JV</td>
<td>27</td>
<td>Right</td>
<td>14</td>
<td>19</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>KA</td>
<td>21</td>
<td>Right</td>
<td>11</td>
<td>8</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>LC</td>
<td>26</td>
<td>Left</td>
<td>11</td>
<td>8</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>MV</td>
<td>23</td>
<td>Right</td>
<td>11</td>
<td>8</td>
<td>BB</td>
<td>40, 50, 60</td>
<td>+</td>
</tr>
<tr>
<td>MW</td>
<td>27</td>
<td>Right</td>
<td>36</td>
<td>39</td>
<td></td>
<td>40, 50, 60</td>
<td>+</td>
</tr>
<tr>
<td>RK</td>
<td>27</td>
<td>Right</td>
<td>9</td>
<td>8</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>SC</td>
<td>23</td>
<td>Right</td>
<td>11</td>
<td>8</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>SW</td>
<td>26</td>
<td>Right</td>
<td>11</td>
<td>8</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>TG</td>
<td>27</td>
<td>Left</td>
<td>11</td>
<td>8</td>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

**Subject Database.** **Age:** listener's age during experiments. **Mold:** side where mold was worn during the adaptation period. **Time:** length of adaptation period in days. **Number of Tests:** number of tests during the adaptation period (excluding the pre-and post-adaptation tests). **Bandwidth:** bandwidth of the stimuli presented during tests (BB: broadband noise, HP: high-pass noise). **Intensities:** stimulus levels tested (in dBA). **DTFs:** + indicates DTFs were measured for that listener.
Measurements of directional transfer functions

Directional transfer functions (DTFs) of the ears with and without the molds were measured to verify that the ear with the mold still received specific elevation-dependent spectral features for six listeners (e.g. Fig. 4.5, Table 4.1) for 25 elevations in the midsaggital plane (cf. Fig. 4.5). This was done by presenting a minimum-peak broadband frequency-modulated sweep (0.2-20kHz, flat amplitude spectrum, Schröder phase, 20 ms duration, 25 sweep repetitions per stimulus, of which the first and the last sweep had an on- and offset ramp, respectively; Schröder, 1970; Wightman and Kistler, 1989; Hofman and Van Opstal, 1998) and recording the sound-pressure level through a thin silicone tube attached to a miniature microphone (model EA1842; Knowles, Itasca, IL) at the entrance of the subject's ear canal.

DTFs and their correlations

From the measured frequency-modulated Schröder-sweep responses at the ear-canal entrance, we applied a Fast Fourier Transform (FFT) on the sound-pressure signal, averaged across 23 sweeps (excluding the first and last sweep that contained the on- and offset ramps) with the speaker at each of the 25 elevations. From the FFT we then determined the signal's amplitude spectrum for each elevation angle. The DTF for, say, elevation $\varepsilon_0$, was obtained by dividing each amplitude spectrum by the grand average across all 25 elevations:

$$DTF(\varepsilon_0, f) = N \cdot \frac{|\text{FFT}(\varepsilon_0, f)|}{\sum_{i=1}^{N} |\text{FFT}(\varepsilon_i, f)|}$$

(4.1)

with $N$ the number of elevation angles (in our experiments: $N=25$; e.g. Fig. 4.5). Each DTF contained 512 frequency bins (from 0 Hz to 25 kHz at 48.8 Hz intervals). For graphical purposes, the DTFs were smoothed by a simple Gaussian filter with a constant Q-factor of 8. To determine a quantitative measure for the resemblance between different DTFs, we only incorporated the frequency bins between 4-20 kHz, as in this frequency domain the head-related transfer functions contain most of the directional information. In this way, each DTF could be considered as a 328-dimensional vector, for which we computed the correlations with other DTF vectors. For two different sets of 25 DTFs (e.g. left ear vs. right ear, or the free right ear vs. the same ear with a mold) this procedure results in a 25x25 correlation matrix, $C(\varepsilon_1, \varepsilon_2)$, in which each entry contains the correlation between the DTF from the first set at elevation $\varepsilon_1$, with the DTF from the second set at elevation $\varepsilon_2$ (e.g. Fig. 4.5C,F). Note, that this matrix is not symmetric when computed for different DTF sets, i.e. in general $C(\varepsilon_1, \varepsilon_2) \neq C(\varepsilon_2, \varepsilon_1)$. Also, the maximum correlation is not necessarily found on the main diagonal (i.e. at $\varepsilon_1 = \varepsilon_2$).

The autocorrelation matrix of a DTF set is symmetric, and measures how well the DTFs of a given set discriminate the different elevation angles. A typical ear yields a correlation of 1.0 only on the main diagonal, indicating that the DTF for a given elevation is unique (e.g. Hofman and Van Opstal, 1998).

DTF similarity index

To quantify with one scalar the overall similarity of the sets of DTFs from the free ear and the ear with the mold, we constructed a similarity index, $I_{\text{sim}}$, in the following way: first, for a given elevation angle, $\varepsilon_m$, the standard deviation in the correlation coefficients, $C(\varepsilon_m, \varepsilon_i)$ for
this mold DTF with all free-ear DTFs was computed ($\sigma_m$). The similarity index was then taken as the average standard deviation for all mold DTFs:

$$I_{sim} = \frac{1}{N} \sum_{m=1}^{N} \sigma_m$$

for $N = 25$ elevation angles. For example, if the mold and the free-ear DTFs have a high correlation along the diagonal and low correlations for off-diagonal angles (e.g. Fig. 4.5F), there will be a high standard deviation. Consequently, $I_{sim}$ is high. When mold and free-ear DTFs do not correlate well for any position (e.g. Fig. 4.5C) the standard deviations, $\sigma_m$, and the resulting $I_{sim}$ will be low.

**Paradigms**

*Calibration experiment.* Head-position data for the calibration procedure were obtained by instructing the listener to make an accurate head movement while redirecting the dim rod LED in front of the eyes from the central fixation LED to each of the 57 peripheral LEDs that was illuminated as soon as the fixation point extinguished. Each experimental session started with a calibration run.

*Auditory localization.* The listener started a trial by fixating the central LED with the head-fixed LED pointer. After a pseudo-random period of 1.5 to 2.0 s, this fixation LED disappeared and an auditory stimulus was presented 400 ms later. The listener was asked to redirect the head by pointing the dim rod LED as accurately and as fast as possible to the perceived location of the sound stimulus. As the response reaction times typically exceeded 200 ms, all responses were made under open loop conditions.

**Hearing conditions**

Before the adaptation period, acute sound localization experiments were run under four different hearing conditions. In the FF condition, both ears were free from spectral manipulations, and had normal hearing. In the MM condition, both ears received a mold. In the FM condition the listener wore the mold (right or left ear) that would also be used during the adaptation period, in the MF condition, a mold was applied to the other ear only (left or right; Hofman and Van Opstal, 2003). Conditions were changed between experimental runs with the participant in complete darkness, so that he/she received no visual feedback to sounds made e.g. by the experimenter when entering or leaving the room.

In the adaptation condition, the participant moved and acted in his normal living environment, with the unilateral mold continuously in either the right (ten adaptation experiments in eight subjects) or left ear (three series in three subjects; see Table 4.I). Sound localization performance with the mold was repeatedly tested during this adaptation period (on a daily basis, excluding weekends and holidays). The mold was worn by most listeners for at least 11 days (Table 4.I), except for listeners JO and RK whose experiments were terminated as soon as they reached stable performance. Listeners JO, MW and MW consented to wear the mold for longer periods (more than 22 days).

Listeners either improved and reached a stable performance level, or they did not improve during their adaptation period (see also Results). After this adaptation period, the listeners were again tested for the four different hearing conditions.

The stimuli during the experimental runs were BB noise at 40, 50 and 60 dBA (one complete run consisted of 3 stimulus intensities x 57 locations = 171 targets, randomized across trials). During the pre- and post-adaptation experiments, all conditions were tested each in two runs in one day (thus listeners made 4 conditions x 2 runs x 171 targets = 1368 re-
sponses, both on the pre-adaptation and the post-adaptation experimental day). During the adaptation period, the daily experiments consisted of 3 consecutive runs (1 FM condition x 3 runs x 171 targets = 513 responses). In the analysis, even though lower stimulus intensities elicited slightly worse localization performance, all intensities were pooled to obtain more robust parameter estimates for the regression lines.

Two listeners (JO1 and MW1) participated also in an earlier adaptation series in which the localization responses were made to HP and BB noise stimuli with levels of 30, 40 and 50 dBA (one run consisted of 2 spectral bandwidths x 3 intensities x 57 locations = 342 targets, randomized across trials, presented once in each session). Their molds differed for the two adaptation series, and these two series were separated by more than two years. The responses to HP and BB stimuli were pooled in the data analysis, since they did not differ significantly.

Participants will be referred to by their initials; MW1 and JO1 will be used for listeners MW and JO in their first adaptation series.

Front-back reversals and confusions
Several listeners reported front-back (F-B) confusions and F-B reversals with the mold in situ, especially during the early stages of adaptation. In those cases, stimuli presented ipsilateral to the mold were sometimes (in case of an F-B confusion) or consistently (in case of a reversal) localized at a rear location on the same side. To avoid having the subject make large (and inaccurate) head movements toward these rear locations, we instead instructed the subject to indicate the occurrence of such F-B reversals by a button press, and to generate a head movement toward the location in the frontal hemifield that was mirrored with respect to the coronal plane through the two ears.

Data analysis
Data calibration. The calibration experiment provided a set of 58 LED/speaker locations and raw head position signals. These locations were defined in a double-pole coordinate system (Knudsen and Konishi, 1979). In this system, azimuth, $\alpha$, is defined as the angle between the sound source (or response direction), the center of the head, and the midsagittal plane. Elevation, $\varepsilon$, is defined as the angle between the sound source, the center of the head, and the horizontal plane. The origin of the ($\alpha$, $\varepsilon$) coordinate system corresponds to the straight-ahead speaker location. Azimuth and elevation can be calculated from the polar coordinates, $(R, \Phi)$ by:

$$\alpha = \arcsin(\sin R \cos \Phi) \quad \text{and} \quad \varepsilon = \arcsin(\sin R \sin \Phi)$$

These 58 fixation points and raw head-position signals were used to train two three-layer neural networks that served to calibrate the head-movement data, using a back-propagation algorithm based on the gradient descent method of Levenberg-Marquardt (Matlab; The Mathworks Inc, Natick, MA).

The networks corrected for small inhomogeneities in the magnetic fields and could adequately cope with minor cross-talk between channels. The trained networks were subsequently used to map the raw data to calibrated two-dimensional head-positions with an absolute accuracy within 4% over the entire response range. For illustrative purposes, stimulus and response coordinates were sometimes plotted in polar coordinates (i.e. Fig. 4.2, Fig. 4.11A,B, Fig. 4.12). However, in the analysis these coordinates were always transformed into the double-pole azimuth-elevation coordinates.

Head movement detection. Saccadic head movements were detected from the calibrated head-movement signals by setting thresholds to the vectorial head velocity for on- and off-
set, respectively, using a custom-made program (onset velocity=20 deg/s, offset velocity=15 deg/s). Detection markings from the program were shown to the experimenter without revealing any stimulus information, and could be adjusted manually, when deemed necessary.

**Statistics.** Each listener’s responses were quantified by determining the optimal linear fit for the following stimulus-response relations:

\[
\alpha_R = a + b \cdot \alpha_T \quad \text{and} \quad \epsilon_R = c + d \cdot \epsilon_T \quad (4.3)
\]

for the azimuth and the elevation components, respectively, by minimizing the least-squares error (Press et al., 1992). In Eq. 4.3, \( \alpha_R \) and \( \epsilon_R \) are the azimuth and elevation response components, \( \alpha_T \) and \( \epsilon_T \) are the azimuth and elevation coordinates of the target. Fit parameters, \( a \) and \( c \), are the response biases (offsets, in degrees), while \( b \) and \( d \) are the overall response gains (slopes, dimensionless) of the azimuth and elevation response components, respectively. Note that an ideal listener should yield gains of 1.0 and offsets of 0.0 deg. Also, Pearson’s linear correlation coefficient, the residual error (standard deviation around the fitted line), and the mean absolute localization error, were calculated.

In order to account for the strong azimuth-dependence of the elevation responses in some of the hearing conditions (see Results), regressions were also performed within restricted regions of azimuth space. By dividing the elevation responses into 30 deg-wide azimuth bins, each shifted in 5 deg steps (thus 25 deg overlap between adjacent bins), we determined the so-called local elevation gain, which provides a smooth estimate of the response gain as a function of stimulus azimuth (e.g. figure 4.4; Zwiers et al., 2003).

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 randomly selecting (with replacement) data points from the original data set. 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).

### 4.3 Results

**Acute effects of a unilateral spectral perturbation**

The application of a mold could change a listener’s localization behavior quite drastically. This is illustrated in figure 4.2 (listener RK), which compares the hearing condition with both ears free (FF) to acute pre-adaptation right-ear mold (FM) listening. Head movements made towards the auditory broadband stimuli were close to the target in the FF condition, and responses were distributed across the entire hemifield (Fig. 4.2A). With the mold, however, responses into the right hemifield only covered a small range in elevations, typically near the horizontal plane. Note, that even the elevation responses into the left hemifield were limited to a smaller range than without the mold. Thus, in line with earlier reports (Morimoto, 2001; Hofman and Van Opstal, 2003), the mold affected localization of sound-source elevation well into the contralateral hemifield, which is a clear indication of a binaural interaction. In contrast, the response azimuth distribution was unaffected across the entire hemifield (Fig. 4.2B).

To quantify the overall response patterns, we performed linear regression on the azimuth (Fig. 4.3A-C) and elevation (Fig. 4.3D-F; listener RK) response components (Eq. 4.3; Methods). For the azimuth responses, we performed separate linear regressions in three non-overlapping target regions along the azimuth axis (ipsilateral to the mold: \( \alpha_T > +20 \) deg; near the midsagittal plane: \(-20 \) deg < \( \alpha_T < +20 \) deg, and contralateral to the mold: \( \alpha_T < -20 \) deg;
Relearning Sound Localization

Despite the introduction of a mold, FM sound azimuth localization (Fig. 4.3A-C, black line: FF, grey line: FM) remained accurate and was indistinguishable from the FF condition. This holds true for all three regions, even for sound-sources ipsilateral to the mold (Fig. 4.3C). This underlines the notion that sound-source azimuth is predominantly determined by interaural level and timing differences, rather than by spectral cues. The azimuth results for the other listeners were similar (data not shown).

Also for the elevation analysis we took the same target regions along the azimuth axis. Insertion of a mold had a clear detrimental effect on elevation localization in the region ipsilateral to the mold, as indicated by the regression line with a low gain (Fig. 4.3F, grey line). The contralateral data (Fig. 4.3D) were least affected by the mold, as the elevation gain seemed barely affected. However, there is also a clear localization defect of the mold for elevation responses around the midline (Fig. 4.3E).

To better quantify the systematic azimuth-dependence of perceived elevation, we refined the analysis of figure 4.3 by determining the local elevation gain (see Methods). Figure 4.4 shows the result of this analysis for all subjects (for left-ear molds, the azimuth sign was reversed). For all subjects, the effect of the mold on perceived elevation may be described by a qualitatively similar smooth function of azimuth. However, quantitatively, the subjects appeared to fall into two separate groups, here distinguished by filled grey circles (strong effect of the mold, N=8) and open squares (a smaller effect of the mold; N=5).

To check how the observed differences between the two groups of subjects is related to the way in which the mold perturbs the listener’s original spectral cues, figure 4.5 shows for two listeners the directional transfer functions (DTFs) of the right ear without (Fig. 4.5A,D) and
Figure 4.3. Linear regression analysis. (A–C) Azimuth localization responses towards broadband stimuli from listener RK in the FF (black triangles) and the FM (open circles) hearing condition (same data as Fig. 4.2). Responses are divided in three regions: contralateral to the mold (A,D: $\alpha_T<-20$ deg), central (B,E: $-20<\alpha_T<20$ deg), and ipsilateral to the mold (C,F: $\alpha_T>20$ deg). Thick black lines denote the linear regression lines (Eq. 4.3) for azimuth (A, B, C) and elevation (D, E, F) components in the FF condition; thick grey lines for responses in the FM condition. Note highly similar responses and regression lines for the azimuth components in the FM and FF condition for all regions, while elevation performance clearly depends on stimulus laterality in the FM condition.

Figure 4.4. Azimuth-dependent elevation responses in acute FM hearing. The local elevation gain in the FM condition is a smooth function of target azimuth for all 11 listeners (13 experiments). For all listeners elevation response performance gradually degrades towards ear ipsilateral to the mold (positive azimuth angles), but to a lesser extent for some listeners (open squares) than for others (grey circles).
Relearning Sound Localization

Figure 4.5. Effect of molds on DTFs. (A) DTFs of the free right ear of listener JO show a clear and distinct notch (dark-gray) that runs from about 5 to 9 kHz for elevations between -50 and +50 deg. (B) This notch almost completely disappeared after application of the mold, while a new notch formed above 9 kHz. (C) Correlation between the original right-ear’s DTFs and the mold-induced DTFs for listener JO is without much structure. A somewhat increased correlation is found for downward elevations. The superimposed open dots show elevation judgments (ordinate) versus actual elevation (abscissa) upon application of the mold: responses have a low gain and a considerable downward bias (regression line (white): gain 0.20, bias: -10.9 deg). (D) The free right ear of listener JV also shows a clear notch in her DTFs, starting near 6 kHz, while gradually increasing to about 11 kHz. (E) This notch, while slightly diminished, is still present, although it now runs up to about 8 kHz. (F) The DTFs of free ear and mold for listener JV have a high positive correlation (light-gray) just off the main diagonal, in line with the shift of the most prominent notch. Her acute elevation responses (superimposed open dots) appear to be guided by the band of high positive correlation (regression line (white): gain 0.85, bias: -5.9 deg). Note also considerable scatter in the responses. Iso-correlation contours at 0.25 intervals.

with (Fig. 4.5B,E) the mold in situ. The mold of listener JO (top) induced a strong effect on his localization responses, whereas the mold of listener JV had a modest effect on her behavior (Fig. 4.4). Indeed, a qualitative comparison of the DTFs revealed that the mold had a strong disruptive effect on the spectral patterns in listener JO (Fig. 4.5A,B), but to a much a lesser extent in listener JV (Fig. 4.5D,E). The directional transfer functions of JO’s right ear (Fig. 4.5A) show a prominent notch at a frequency near 5.5 kHz for a downward elevation at -50 deg, that systematically shifts to higher frequencies, until at an upward elevation of +50 deg it is found near 9 kHz. Note that this notch was nearly absent in his mold-induced DTFs (Fig. 4.5B). Because of these differences, the correlation matrix, $C(e_1, e_2)$, between the two DTF sets is generally low and without much structure (Fig. 4.5C; see Methods). This predicts that localization on the ipsilateral side should indeed be worse in the acute mold condition. The picture is quite different for listener JV. The notch in her normal right ear runs between 6 kHz and 11 kHz (Fig. 4.5D), but a qualitatively similar pattern was seen when that ear contained the mold (Fig. 4.5E). With the mold, the notch was somewhat less pronounced, and ran between 6.0 and 8.5 kHz. Despite these differences, there is a strong correlation between the two DTF sets near the main diagonal and a very low correlation for other locations.
FIGURE 4.5. Conflicting spectral shape cues
The introduction of a single mold changes the spectral cues for one ear, while the other ear still provides the normal, unperturbed cues. Because elevation localization also relies on binaural interactions (Morimoto 2001; Hofman et al. 2003; Fig. 4.4), the auditory system should reconcile these two conflicting sets of spectral cues. For the normal FF condition the spectral-shape cues of both ears are very similar, as illustrated by the DTF correlation map for listener JO in figure 4.7A. A high correlation between the left and right ear is found only
Relearning Sound Localization

Hofman and colleagues (1998) imposed new spectral cues on both ears, and showed that subjects adapted to this condition within a few weeks. Interestingly, the molds from both ears in that study yielded very similar DTFs ($I_{sim} = 0.5$). This is illustrated in figure 4.7B for listener JO, who also participated as a subject in that study. While the DTF correlations are not as high as in the FF condition for this subject, both molds in the MM hearing condition produced near-identical spectral cues for similar elevations ($I_{sim} = 0.5$). Note that the mold DTFs in that study did not correlate well with the listener’s ears (not shown).

The binaural spectral similarity was heavily perturbed when a mold was inserted in one ear only, resulting in low values for the DTF binaural correlation matrix ($I_{sim} = 0.3$). This is illustrated in figure 4.7C for listener JO in the present study. Thus, the unilateral mold introduced a strong conflict between both sets of spectral cues, which may have posed an additional problem for the sound-localization system when it tried to learn the new cues. The question to be studied next is whether inter-aural similarity of the spectral cues is a necessary factor to guide the learning of elevation localization.

Adaptation to a unilateral mold

After the acute localization tests with the molds, listeners wore one mold continuously for several weeks (see Table 4.1, for details). Meanwhile, they were subjected to sound localization experiments on an almost daily basis. The mold had a profound effect on elevation performance over time in all listeners, but based on the response patterns observed, we could distinguish two groups of listeners. Interestingly, the two groups consisted of the same listeners as the ones shown in figure 4.4. Figure 4.8 illustrates these two response modes for two representative listeners for three non-overlapping spatial sectors. Listener JO gradually improved not only his elevation gain for sounds ipsilateral to the mold (Fig. 4.8A, light-grey lines), but also for sounds near the midsaggital plane, and even slightly for sounds on the contralateral side (Fig. 4.8A, dark-grey and black lines, respectively). In contrast, listener JV (Fig. 4.8B) started with much higher initial gains upon application of the mold. The elevation...
Figure 4.8. Two different adaptation responses. Three distinct target azimuth sectors are selected to show the change in elevation gain as a function of time during adaptation for two representative subjects (black line: $\alpha_T < -20$ deg; light-grey line: locations ipsilateral to the mold, $\alpha_T > +20$ deg; dark grey line: locations near the midline, $-20 < \alpha_T < +20$ deg). (A) Listener JO. Sounds ipsilateral to the mold were initially badly localized (low gain), but after about 10 sessions the elevation gain reached a plateau at a value of 0.6. Also localization near the midline improved gradually over time. (B) No systematic improvement of ipsilateral, central, or contralateral responses was obtained for listener JV.

Gain on the ipsilateral side was lower than the gain near the midsaggital plane, which in turn was lower than the gain on the contralateral side (see also Fig. 4.4). Over time, however, her local elevation gains did not systematically improve, but rather appeared to oscillate between two extreme values for all three sectors.

These two different adaptive response patterns were prototypical for the entire group of subjects. One group of seven listeners (eight experiments: DH, JO and JO1, KA, MW1, RK, SC, and TG; see also Fig. 4.4, grey symbols) gradually adapted to a clearly improved performance. The other group of five listeners (JV, LC, MV, MW, and SW; Fig. 4.4, open squares) started out with a high gain that did not show a subsequent clear improvement over time, but instead tended to display an oscillatory pattern.

To quantify the gradual change in elevation gain from ipsilateral to contralateral azimuths for the first group of subjects, figure 4.9 plots for six of these listeners the local normalized elevation gain, $G_N(\alpha)$, as a function of both time (abscissa) and azimuth (ordinate), according to:

$$G_N(\alpha) = \frac{G(\alpha) - G_{FM,ipsi}}{G_{FF}(\alpha) - G_{FM,ipsi}}$$

in which $G(\alpha)$ is the measured local gain at azimuth $\alpha$, $G_{FM,ipsi}$ the local gain obtained with the pre-adaptation FM experiment for ipsilateral locations ($\alpha > 20$ deg), and $G_{FF}(\alpha)$ corresponds to the value obtained for the initial free hearing condition. In this way, $G_N(\alpha) = 0$ (dark-
Relearning Sound Localization

when the measured gain equals the gain value obtained in the FM condition on the far-ipsilateral side; if $G_N(\alpha)=1$ (light-gray), it equals normal listening. $G_N(\alpha)$ was determined by pooling the responses of one experimental run of 171 trials with its two preceding and two succeeding runs to obtain a smooth estimate of the improvement in localization behavior (number of included responses per bin is on average 400; daily tests consisted of three runs, see also Methods). All listeners started out with a local elevation gain on the far contralateral side that resembled free listening, which indicates that the mold had little effect on these locations. On the far ipsilateral side, the initial localization behavior resembled acute FM listening. All listeners from this group gradually improved their performance over an increasing range of azimuth angles (which is visible as a widening of the light-gray zone into ipsilateral space). Toward the end of the experiment (top row of each panel) localization of central sounds approached the FF listening condition. For most listeners, the far-ipsilateral ($\alpha>20$ deg) gains remained low, but even there an improvement could often be observed.

In contrast, the second group of listeners had high initial elevation gains at the start of the experiment, as their mold did not perturb the spectral cues as much as in the first group of subjects (Figs. 4.4-5). Figure 4.10 plots the normalized local elevation gain as a function of time, for this group defined as $G_N(\alpha)=G_{FM}(\alpha)/G_{FF}(\alpha)$, in the same format as figure 4.9. Although the gain of these listeners was not strongly disturbed by the mold (lowest gain value on the ipsilateral side was ~0.7)(Fig. 4.10, dark-gray), performance of these listeners still varied systematically from session to session over the entire azimuth range. The behavior of listeners JV and MW (Fig. 4.10A,B) seemed to follow an oscillatory pattern over time. Also performance of the other three listeners of this group (Fig. 4.10B,C,E) followed a similar pattern, although these listeners did not participate long enough to follow performance for longer than a complete cycle. Yet, their behavior did seem to oscillate with about the same period as the other two listeners.

FIGURE 4.9. ADAPTATION RESULTS FOR LISTENERS SHOWING IMPROVEMENT.

Normalized local elevation gain (grayscale-coded; Eq 4.4) is plotted as a function of target azimuth (ordinate) and time (# runs, abscissa). Light-gray color corresponds to normal (FF) localization behavior for the associated bin, while dark-gray equals the acute FM behavior for sounds ipsilateral to the mold (i.e. no adaptive change). Listeners DH (A), JO1 (B), KA (C), MW1 (D), RK (E), and SC (F) all adapted gradually to the new spectral cues induced by the mold, as their elevation response behavior gradually improved toward FF behavior within 16 runs. This improvement shows as a gradual expansion of light-gray bins toward the right. Similar adaptive behavior was obtained for listeners JO (Fig. 4.8A) and TG (not shown).
To quantify this oscillatory behavior we fitted a sine function through the ipsilateral elevation gains \((\alpha_T > 20 \text{ deg})\) in the following way:

\[
G(\tau) = A \cdot \sin(2 \cdot \pi \cdot f \cdot \tau + e) + g
\]

in which \(G(\tau)\) is the ipsilateral elevation gain at time (expressed as number of runs), \(A\) is the (dimensionless) amplitude of the oscillation, \(f\) is the frequency of the oscillation (in runs\(^{-1}\)), while \(e\) and \(g\) are offsets. The result of a fit through the data of listener \(JV\) (A), \(MW\) (B), \(LC\) (C), \(MV\) (D), and \(SW\) (E). Note the oscillatory behavior, with similar periods for all listeners (same scaling in all panels), and the relatively high initial gains (see colorbar, dark-gray is \(~0.7\))(compare with Fig. 4.9) Contra, Contralateral; Ipsilateral; deg, degrees.

Figure 4.10. Localization behavior for listeners with high pre-adaptation FM gains. Local elevation gain during FM adaptation (here normalized to the pre-adaptation FF gain: \(G_{FM}/G_{FF}\)) as a function of time for the responses of listener \(JV\) (A), \(MW\) (B), \(LC\) (C), \(MV\) (D), and \(SW\) (E). Note the oscillatory behavior, with similar periods for all listeners (same scaling in all panels), and the relatively high initial gains (see colorbar, dark-gray is \(~0.7\))(compare with Fig. 4.9) Contra, Contralateral; Ipsilateral; deg, degrees.
Relearning Sound Localization

...ected throughout time and appeared to follow a sinusoidal pattern with a period that had about the same duration as obtained for subjects who showed a strong adaptive response.

Front-back reversals
The acute degradation of up-down localization upon insertion of a mold (Figs. 4.2-4) was not the only localization deficit. Four listeners (in five experiments: DH, JO/JO1, LC, MV) reported consistent F-B reversals (exemplified in Fig. 4.12A for listener JO, who perceived the largest number of reversals) that were not present during normal FF hearing (see Methods). The number of F-B reversals, however, decreased drastically during the adaptation period (Fig. 4.12B), and this effect depended systematically on sound level. Fig. 4.12C shows the percentage of F-B reversals as a function of time for the different stimulus intensities for listener JO. Reversals were fewer and their occurrence decreased faster for the higher stimulus intensities. The other listeners had fewer initial reversals (DH<40%, LC<20% and MV<2%), and they no longer reported any reversals at the end of the adaptation period.

Test of the adaptation models
As explained in the Introduction (Fig. 4.1), adaptation to a monaural spectral perturbation could in principle be due to two different mechanisms: either the sound localization system has acquired a new spectral representation of the perturbed ear (adaptation at the spectral-to-spatial mapping stage), or the system ignores the conflicting spectral cues altogether,
and expands the influence of the intact ear into the contralateral hemifield (adaptation at the binaural–weighting stage). To dissociate these possibilities, we subjected our listeners before and after the adaptation period to four different hearing conditions when making their localization responses: without molds (i.e. both ears free: FF), with one mold in either the right or left ear that was worn during the adaptation period (FM), a single mold in the other, unadapted, ear (MF), and finally, with binaural molds (MM).

If learning had taken place at the binaural-weighting stage (Fig. 4.1), localization in the MM condition would be impossible, as no unperturbed spectral-shape cues would remain. On the other hand, if the auditory system had learned to calibrate the new spectral-shape cues (spatial-mapping stage), after adaptation both the FM and MM condition would show comparable improved localization accuracy on the side of the mold.

Figure 4.13 shows the raw localization responses of subject JO before (Fig. 4.13A-D) and after adaptation (Fig. 4.13E-H) for the four hearing conditions. A comparison between panels B and F indicates a clear improvement of localization accuracy after the adaptation period for the FM condition. In contrast, the response distributions for the FF and MF conditions did not appear to have changed after FM adaptation. The critical test, however, is provided by the MM condition, for which the response distribution also appeared to have expanded across the azimuth range. Thus, a qualitative assessment of the results for this listener favors adaptation at the level of spectral-to-spatial-mapping.

The other adapted listeners showed very similar behavior in the four hearing conditions as listener JO. In order to pool the results for these listeners, in figure 4.14 we have plotted the differences in local response gain between the pre-adaptation and post-adaptation localiza-
Relearning Sound Localization

**Figure 4.13. Effect of FM adaptation on localization behavior of listener JO.** Raw localization responses are plotted in R-Φ coordinates for the various pre- (A–D) and post- (E–H) adaptation conditions. Note that elevation response distributions for the FF (A and E) and MF (C and G) conditions remain similar, while there is a clear improvement for localization in the FM (B and F) and MM (D and H) conditions (compare with Fig. 4.1C). deg, degrees.

**Figure 4.14. Test of adaptation models.** Differences in local elevation gain as function of azimuth between pre- and post-adaptation listening for four different hearing conditions (FF, MM, MF and FM), pooled across the seven listeners (8 experiments) that showed adaptive improvement (DH, JO and JO1, KA, MW, RK, SC, and TG; see Figs. 4.8A, 4.9). Note similarity in improvement between the FM (circles) and MM (squares) hearing conditions, and the absence of a systematic improvement in the MF (triangles) and FF (diamonds) hearing conditions (compare with Fig. 4.1C). deg, Degrees.
tion sessions for each of the four localization conditions. The results are averaged across the seven adapted subjects (N=8 experiments), and shown as a function of sound-source azimuth. Our data clearly demonstrate that the FM condition has improved, most prominently on the side ipsilateral to the mold, but even for locations well into the contralateral hemifield (Fig. 4.14A, dark-grey lines). This result implies that sound localization plasticity does not require (near-) identical spectral inputs from the two ears.

Although the pre- and post-adaptation results for FF and MF hearing were, on average, not identical (the differences, although small, were not zero), there was no systematic azimuth-dependent change in localization performance for these conditions (see also Discussion). More importantly, however, the MM hearing condition showed a clear ipsilateral improvement in localization performance across the azimuth domain. The improvement of the MM condition was quite comparable to the improvement obtained for FM hearing (Fig. 4.14B, dark-grey lines; compare with Fig. 4.1C). From these data we conclude that adaptation has taken place at the spectral-to-spatial mapping stage for all listeners.

4.4 Discussion

Summary
We studied the response of the human sound-localization system to a long-term monaural spectral perturbation, and showed that all listeners had clear signs of adaptation. Seven subjects (8/13 experiments) regained an improved elevation performance within eleven days (Fig. 4.9), while the elevation gain of five listeners oscillated (Figs. 4.10, 4.11). Adaptation to spectral cue manipulations depended on the correlation of the resulting DTFs with the subject's own ears (Figs. 4.4, 5, 6) and extended well into the contralateral hemifield. We compared localization responses with the adapting mold (FM-hearing) to results under normal hearing conditions (FF), to a mold in the other ear (MF), and to listening with binaural molds (MM). We conclude from our experiments that the human auditory system is capable of an ear-specific spectral adaptation. This plasticity predominantly acts at the stage of neural processing where the spectral-shape cues are transformed into spatial information, rather than at the level of binaural interactions (Figs. 4.1, 4.14).

Comparison to other studies
Previous studies have reported on the acute effect of a unilateral mold on sound localization and demonstrated a contribution of each ear within its opposite hemifield (Humansky and Butler 1988; Morimoto 2001; Hofman and Van Opstal 2003). Our data further corroborate these findings (Figs. 4.2-4), and we used these results to dissociate different adaptive mechanisms.

Hofman et al. (1998) studied sound localization plasticity in response to binaural molds for up to four weeks, during which subjects gradually relearned to localize elevation. We extended these experiments in several ways: First, a unilateral mold confronts the system with radically different spectral cues for each ear (Fig. 4.7C). In the Hofman et al. (1998) study, the thin plastic molds yielded spectral cues that differed substantially from the original cues, but remained roughly similar for the two ears (Fig. 4.7B). Second, to assess binaural interactions, the measurement range was extended to ±75 deg (vs. ±30 deg in Hofman et al., 1998). Finally, by comparing four different hearing conditions, we could dissociate the different models about the locus of adaptation (Fig. 4.14).

Learning behavior
All subjects showed clear signs of adaptation to the new spectral cues. Seven listeners improved their performance and reached a plateau in their elevation gain within approximately
Relearning Sound Localization

7 days. These subjects did not reach optimal localization performance (Fig. 4.9), as the far-ipsilateral elevation gain remained below FF-performance in most subjects. This could be due to a lack of sufficient spectral detail in the mold’s DTFs, resulting in a signal-to-noise problem (Good and Gilkey, 1996; Kulkarni and Colburn, 1998), to insufficient training time, to a limit in the capacity of the central nervous system to interpret altered spectral cues, or to a combination of these factors.

Three listeners kept the mold in situ well after maximal performance was reached. Nevertheless, performance did not improve any further, suggesting that insufficient training time did not underlie imperfect overall performance in our experiments. Interestingly, although five listeners never reached a stable plateau, their elevation gains oscillated with a period that was comparable to the learning rate of the other seven subjects (Figs. 4.9-11). We interpret this behavior as follows: initially, at the acute pre-adaptation FM test, each subject’s responses are based on an interpretation of the normal DTFs, which, because of the high correlation in this group, yielded high elevation gains (Fig. 4.5-6). Later, sensory feedback from the environment signals consistent errors between perceived and actual target elevations, which induce a learning response. However, because of the high DTF correlations between the normal ear and the mold, a given DTF always corresponds to two different, but relatively nearby and equally strong elevation angles. Apparently, the sound-localization system cannot deal with this ambiguity.

Potential mechanisms

Our results indicate that adaptation to long-term spectral perturbation is ear-specific, and predominantly acts at the stage of spectral-to-spatial mapping (Fig. 4.1). The conceptual schemes in figure 4.1A,B represent the general flow of acoustic processing in the auditory system, and offer a functional, rather than anatomical, description of the mechanisms extracting elevation. Our experiments do not allow a firm dissociation of either scheme, since the predictions for the different control experiments are the same (Fig. 4.1C). Yet, the distinction is not trivial, since binaural interaction and spectral-to-spatial mapping are both nonlinear (but as yet, unknown) processes. Therefore, their order matters, but different experiments are needed to determine the order.

Signals are initially processed monaurally in the cochlear nucleus. Its dorsal part (DCN) has been implicated in the processing of spectral-shape information (Young and Davis 2002). However, adaptation relies on feedback (Knudsen 2002), either from vision, head-motion signals from proprioceptive or vestibular sources, or efference copies. There is no evidence that such feedback signals reach the DCN. In contrast, the midbrain Inferior Colliculus (IC) does receive a signal about eye position (Groh et al. 2001; Zwiers et al. 2004). In barn owls, the external nucleus of the IC (ICx) receives topographically organized visual inputs from the Superior Colliculus (SC) that are crucial for the formation of its auditory space map (Brainard and Knudsen 1993; Gutfreund et al. 2002; Debello and Knudsen 2004). Convergence of visual and auditory inputs was also demonstrated in the brachium of the ferret IC (Doubell et al 2000). The presence of different non-acoustic feedback signals in the IC could thus render it as a potential candidate for sound-localization plasticity also in mammals (King et al. 2001; King 2002).

Other factors

Monaural adaptation to a mold did not induce an aftereffect, as the FF test after removal of the mold immediately yielded accurate behavior in all subjects (Figs. 4.13-14). Similarly, Hofman et al. (1998) reported the complete absence of an aftereffect in their binaural adaptation study. Yet, the FF control condition resulted, on average, in a slightly increased gain for the post-adaptation recordings across the entire azimuth domain (Fig. 4.14A, light-grey lines). Possibly, this effect is caused by a change in the open-loop head-movement strategy
of the listeners. As, during adaptation and over many recording sessions, listeners gradually become more confident in their own localization behavior, they might tend to generate slightly larger vertical head movements. It is difficult to control for such potentially confounding factors, as our experiments involve the complete action-perception cycle.

Similarly, a small negative difference was observed for the MF hearing condition (Fig. 4.14B, light-grey lines). This effect could result from small adjustments at the binaural weighting stage. If the normal-ear weighting slightly increases during adaptation, the insertion of a mold in that ear should lead to less accurate localization responses as before the adjustments. These relatively small effects notwithstanding, the strong and azimuth-dependent improvement of localization performance in the MM condition was large for all subjects, and comparable to the FM improvement. Therefore, adaptation at the spatial mapping-stage remains by far the dominant factor to explain the response behavior across the population of subjects.

**Calibration**

Studies in the barn owl (Knudsen and Knudsen, 1985; Brainard and Knudsen, 1993) have shown that calibration of its sound-localization system requires visual input. Also in ferret (King et al., 1988) vision guides the formation of a topographic map of auditory space in the midbrain SC. Recent human studies have indicated that the congenitally blind localize well under simple acoustic conditions (Zwiers et al., 2001a), but that their elevation performance breaks down in noisy environments (Zwiers et al., 2001b). Conversely, normal-sighted subjects reduced the local gain of their sound-localization responses after wearing minifying glasses with a restricted visual field (Zwiers et al., 2003). These findings corroborate the hypothesis that also in humans, vision is used to calibrate, or fine-tune, sound localization in frontal space. Apart from vision, also auditory feedback may aid in the calibration of auditory space. For example, after removing the spectral cues in young ferrets by pinnectomy, the topographic acoustic map in their SC does not develop (Schnupp et al. 1998). In our experiments, subjects were free to move around in their usual habitat. Therefore, both active vision and the use of self-generated eye- and head movements may have contributed to the adaptive response.

**Outlook and applications**

Our results indicate that the auditory system can learn to use new spectral cues for a single ear within 10 days, provided that the cues are sufficiently different from the subject’s own ear. This has important implications for spatial hearing after surgery that involves the pinna. It should also be considered for potential applications in the field of acoustic virtual reality. Our results further suggest the possibility that monaurally deaf listeners could learn to use their available spectral cues for localization on their deaf side, provided their binaural weighting could learn to favor their intact ear. We recently demonstrated a considerable variability in the localization performance of these listeners that could be fully explained by their using (or lack of using) spectral-shape information (Van Wanrooij and Van Opstal 2004). Training dedicated to use the spectral cues should therefore be considered to help these listeners to cope with the complexities of the acoustic environment, and thereby for their safer navigation through the everyday world.

**Acknowledgements**

This work was supported by Radboud University Nijmegen (A.J.V.O., M.M.V.W.) and by Human Frontiers Science Program Grant RG 0174-1998/B (M.M.V.W.). We thank Ger Van Lingen, Hans Kleijn, Gunther Windau, Huib Versnel, and Ton Van Dreumel for technical assistance and Paul Hofman for valuable suggestions. Wannes Vogels conducted the first experimental series as a graduate student.
CHAPTER V
AUDITORY-VISUAL INTERACTIONS
SUBSERVING GOAL-DIRECTED SACCADIES IN
A COMPLEX SCENE

5.1 INTRODUCTION

Saccadic eye movements reorient gaze swiftly to a new target of interest. Much has been learned about the neural processes underlying the initiation of visually guided saccades (see Findlay and Walker 1999; Munoz et al. 2000 for review). Under natural conditions, the saccadic system is typically challenged by myriad possible targets to which gaze could be directed. Often, these potential targets emit multisensory signals that may provide different combinations of visual, auditory, and tactile inputs. The integration of multisensory signals from a single event into an orienting response is far from trivial as different sensory modalities are transduced uniquely and encoded initially in different frames of reference (see Sparks and Mays 1990 for review). The oculocentric frame of reference in which saccades are represented must be derived from retinotopic signals for visually guided saccades, and from head-centered space for aurally guided saccades. This latter transformation is particularly complex because the CNS constructs the head-centered space from different acoustic cues: sound azimuth is extracted from interaural timing and intensity disparities, and sound elevation from monaural spectral shape cues induced by the pinnae (see Blauert 1997; Irvine 1986 for review).

There is ample experimental evidence that a combined presentation of auditory and visual stimuli reduces saccadic reaction times (SRTs) (see Colonius and Arndt 2001 for a recent review). These reductions generally exceed the predictions of the so-called “race model,” which entails that combined auditory and visual stimuli are processed independently but produce shorter SRTs so long as the unimodal distributions overlap, since subjects can react to either stimulus (Raab 1962). Exceeding the race model implies that the bimodal stimuli are neurally integrated prior to saccade initiation (Hughes et al. 1994; Nozawa et al. 1994). Observed SRT reductions range usually between 10 and 50 ms and diminish as the spatial and temporal separation of the stimuli increases (Colonius and Arndt 2001; Corneil and Munoz 1996; Frens et al. 1995; Harrington and Peck 1998; Hughes et al. 1998).

The neural correlates of multisensory integration have been studied extensively in anesthetized preparations and also depend on the spatial and temporal register of the stimuli (see Stein and Meredith 1993 for review). Another important property of neurons that display multisensory integration is that of “inverse effectiveness” (Meredith and Stein 1986), whereby smaller unimodal responses from near-threshold stimulus intensities are associated with conversely stronger amounts of multisensory integration. If similar mechanisms operate in awake preparations, then the behavioral benefits afforded by multisensory integration should also be greatest with low-intensity stimuli. Accordingly, improved orienting to low-intensity multisensory stimuli has been demonstrated in cats (Stein et al. 1989). So far, human studies using low-intensity stimuli have not demonstrated the dramatic behavioral benefits expected from inverse effectiveness: the SRT reductions afforded by pairing low-intensity stimuli usually approximate the SRT reductions afforded by pairing high-intensity stimuli (Frens et al. 1995; Hughes et al. 1994). Perhaps in these studies, the low intensities were not close enough to threshold, or the limited number of potential target locations may have allowed subjects to constrain their responses prior to stimulus onset. In addition, the auditory stimulus in some of these experiments did not serve as a potential target, but acted as a distractor that could have been ignored by the subject.

The purpose of the present study is to evaluate multisensory integration in human saccades in a complex experimental environment in which both the auditory and visual stimuli serve as potential targets. To this end, low-intensity unimodal or bimodal targets were distributed over 24 possible target locations within the two-dimensional (2-D) oculomotor range and embedded within an auditory-visual background (Fig. 5.1). Both the signal-to-noise (S/N) ratio of the auditory target relative to background, and the temporal register of the auditory and visual targets on bimodal trials were systematically varied. Attesting to the difficulty of
In this task, subjects generated saccade scan patterns that consisted of anywhere between 1 to over 10 saccades before localizing the target. This report focuses exclusively on the SRT and accuracy of the first saccade in each trial as indexes of how well the subjects initially localize the target(s). Accurate saccades at short SRTs imply a well-localized target, whereas inaccurate saccades at longer SRTs imply the opposite. Our results demonstrate that the behavioral benefits of auditory-visual integration vary systematically with the S/N ratio as predicted by inverse effectiveness, and that such benefits were greater in the elevation versus azimuth response component. Moreover, the observed effects depended in a systematic way on the relative timing of the auditory and visual stimuli. These behavioral data are in good agreement with the rules extracted from multisensory-evoked responses of cells in the mammalian superior colliculus (SC) (Stein and Meredith 1993).

Abstracts describing some of these data have been published (Corneil et al. 2001; Van Wanrooij et al. 2000).
5.2 Methods

Subjects
Five male subjects (ages 23-43) participated in the experiments and provided their informed consent. Experimental procedures were approved by the local ethics committee of the University of Nijmegen. All subjects were experienced with eye-movement recording protocols. Subjects JO, BC, DM, and MW are authors of this paper, although the latter three had no prior experience with sound localization studies. Subject MZ was naive as to the purpose of the study. All subjects had normal hearing, as determined by audiograms of both ears that were obtained with a standard staircase procedure (10 tone pips, 0.5-octave separation, between 500 Hz and 11.3 kHz). With corrective glasses in the experimental setup (subjects BC, DM, and MZ), all subjects had normal binocular vision except for JO, who is amblyopic in his right (recorded) eye. The calibration procedure described below corrected for any nonlinearities from this subject.

Apparatus
Experiments were conducted in a completely dark and sound-attenuated room in which the inner walls, ceiling, and floor, as well as every large object present, were covered with black sound-absorbing acoustic foam that effectively eliminated echoes above 500 Hz. The overall background sound level within the room was approximately 30 dB SPL (A-weighted). The subject was seated comfortably on a chair with back and foot support, and the head was aligned with the center of the room. A customized neck rest, rigidly attached to the floor, prevented the head from moving. Eye movements were recorded with the scleral search coil technique (Collewijn et al. 1975). Horizontal and vertical eye position signals were demodulated by lock-in amplifiers (PAR 128A), amplified and low-pass filtered (cutoff 150 Hz), and sampled at 500 Hz per channel (Metrabyte DAS16H) before being stored on hard disk.

Stimulus Generation
Visual stimuli. Visual stimuli were generated by 85 light-emitting diodes (LEDs) that were mounted on a thin wireframe that formed a hemispheric surface 85 cm in front of the subject (the “LED sky”). LEDs were positioned at visual angles that corresponded in a 2-D polar coordinate system to seven radial eccentricities \( R \in [2; 5; 9; 14; 20; 27; 35] \) deg with respect to the center of the LED sky, and 12 directions \( \Phi \in [0; 30; 60; ... ; 330] \) deg, respectively (where \( \Phi = 0 \) deg is rightward, \( \Phi = 90 \) deg is upward, etc.; Fig. 5.1A). All LEDs could be turned green or red. The visual background was formed by turning all 85 LEDs green. The initial fixation point (FP) was presented by turning the central LED at \([R,\Phi] = [0, 0]\) red. The visual target was lit by turning one of the other green LEDs red. LED intensities were kept low to ensure that localization was difficult in the presence of the background (green LEDs: 0.25 cd/m\(^2\); red LEDs: 0.18 cd/m\(^2\)). The LED sky was backed by an acoustically transparent thin black cloth.

Acoustic stimuli. The acoustic environment consisted of an auditory background sound and a target sound. All sound intensities were measured at the position of the subject’s head with a calibrated sound amplifier and microphone (Brüel and Kjaer BK2610/BK4144), and are expressed in dB SPL (A-weighted). All auditory stimuli were generated digitally at 50 kHz (National Instruments DA board, DT2821) and tapered with a sine-squared onset and offset ramp of 5 ms duration. The signals were amplified by a Luxman A331 audio amplifier and band-pass filtered (0.2-20 kHz, Krohnhite) before being passed to the speakers. The background sound was produced by a circular array of nine small speakers (Nellcor; response characteristics flat within 5 dB between 2 and 20 kHz, not corrected), mounted on
the wireframe of the LED sky at an eccentricity of about 45° relative to center (Fig. 5.1), and consisted of broadband Gaussian white noise (bandwidth 0.2-20 kHz) that was presented at a fixed intensity of 60 dB. This acoustic environment was perceived by all subjects as a spatially diffuse sound that did not emanate from any specific location.

The auditory target sound consisted of periodic broadband noise (period 20 ms, sounding like a 50-Hz buzzer, clearly discernable from the Gaussian white noise background) that had a flat broadband characteristic between 0.2 and 20 kHz. The auditory target was emitted by a broadband lightweight speaker (Philips AD-44725; response characteristics flat within 12 dB between 0.5 and 15 kHz, not corrected) and was presented at a variable intensity relative to background (see following text). The speaker was mounted on a two-link robot, which consisted of a base with two nested L-shaped arms controlled by a PC80486 computer that drove separate stepping engines (Berger-Lahr, type VRDM5) (see Hofman and Van Opstal 1998 for details). This setup enabled rapid (within 2 s) and accurate (within 0.5°) positioning of the speaker at a fixed distance of 90 cm from the subject at any location on a virtual sphere just behind the LED sky. Earlier studies (Frens and Van Opstal 1995) verified that the sounds produced by the stepping motors did not provide any consistent localization cues to the subject. Before every trial, the speaker was moved to a random location at least 20° away from the previous location before a final positioning movement was made. In this way, speaker displacement cues could not be related to final speaker location.

Paradigms

Every subject performed three types of experiments: a calibration experiment, the primary auditory-visual (AV) experiment, and an AV-control experiment. Every session began with one block of the calibration experiment without the AV-background, then two blocks of either the primary or control AV-experiment, with the AV-background.

**Calibration Experiment.** In all experimental sessions, the subjects first performed a calibration experiment without the AV-background. Subjects were instructed to look from a central red FP to a randomly selected peripheral red LED target that was illuminated as soon as the FP was extinguished (1 block consisted of 72 targets: 12 directions × 6 eccentricities, R ≥5°, each presented once), and press a hand-held button when the target was finally fixated.

The primary purpose of the calibration experiment was to provide the final fixation positions for off-line calibration of the eye coil signals (described below). However, the first-saccade data from the calibration experiment also established the SRT and accuracy of visually guided saccades in the absence of the AV-background, which was compared with visually guided saccades generated in the presence of the AV-background. Although it might seem paradoxical to analyze data from a calibration experiment, this reduced the amount of time per session. Further, only the first-saccade data were used for comparative purposes, whereas the calibration procedures used the final fixation position, regardless of the number of preceding saccades.

**AV Experiment.** The spatial and temporal layout of the AV-experiment is depicted in Fig. 5.1. At time 0 in each trial, the AV-background was turned on. After a randomly selected interval of 100, 225, or 350 ms, the central red FP changed color from green to red, and the subject was required to fixate it. At time 1,000 ms, the central FP turned from red to green, and a peripheral target was presented 100-200 ms later (see following text). The subject was instructed to acquire the peripheral target as quickly and as accurately as possible. The location of the peripheral target was selected at random from 1 of 24 different positions. All 12 directions on the LED sky were equally likely, but for each direction only 2 of the following 3 eccentricities were selected: R = 14, 20, or 27° (Fig. 5.1). Subjects made saccades to red visual targets (V-trials), auditory targets (A-trials), or to bimodal auditory-visual targets (AV-trials). The auditory target was presented at one of four different signal-to-noise inten-
Auditory-Visual Interactions

sity ratios (S/N ratio) relative to the fixed-intensity background: –6, –12, –18, or –21 dB. For the unimodal V- or A-trials, the target was presented 200 ms after the FP turned green (i.e., at time 1,200 ms) and persisted for 3,300 ms, determining the maximal search time. In AV-trials, the auditory and visual targets were always spatially coincident. The red visual target was illuminated at time 1,200 ms; the auditory target was presented randomly at time 1,100 ms, 1,200 ms, or 1,300 ms (i.e., either synchronous, or 100 ms before or after the visual target). Note that any time a visual target was presented, it was presented 200 ms after the offset of the fixation point (i.e., a 200-ms "gap"). This was done so that our data would not have to be analyzed as a function of gap interval, given the known effects of gap interval on SRT (see Findlay and Walker 1999 for review).

For coding purposes, saccades were identified by their trial type, S/N ratio of the auditory target (if applicable), and stimulus asynchrony coded relative to stimulus onset (if applicable). Thus V-saccades denote data from unimodal V-trials. A_–6-saccades denote data from unimodal A-trials where the target intensity was set to –6 dB relative to the auditory background. A_–12V-saccades denote data from AV-trials in which the auditory target (-12 dB relative to auditory background) led the visual target by 100 ms. Twelve different AV-trials were possible (3 temporal asynchronies ×4 S/N ratios). In total, 17 different trial types were tested at each target position (1 V-trial, 4 A-trials, 12 AV-trials), making a total of 408 different trials for one complete series. All trial types were randomly interleaved. Each experimental session contained 204 trials run in two blocks of 102 trials each. A subsequent session was typically run on another day and contained the remaining 204 trials to complete the series. Each subject completed at least three full series of AV-multimodal experiments (DM, MZ, and MW: 3 series; BC: 4 series; and JO: 5 series), yielding between 72 and 120 trials per trial type.

An oversight on our part replaced the A_–6V-trials with V100A_–6 trials. Although unfortunate, our conclusions were not affected by the lack of data from A_–6V-saccades.

**AV-Control Experiment.** It is known that the onset or offset of an auditory target can lower SRTs to visual targets, presumably by a warning effect that is independent of the spatial congruity of the auditory and visual stimuli (Ross and Ross 1980, 1981). To parse out the portions of the data set from the primary AV-experiment that were caused by this nonspecific warning effect, each subject was also tested in a separate control experiment. Three trial types from the primary AV-experiment (V-only, A_–12, and A_–12V) were mixed with a new type of bimodal stimulus in which the auditory target sound was generated by the nine background speakers. For this auditory control stimulus, the acoustic signal was a linear superposition of the Gaussian broadband white noise and the periodic buzzer stimulus. A pilot test indicated good audibility of this sound when its level was at -3 dB relative to background. The subjects perceived this control sound as emanating from a single point near straight ahead, although the exact location of this percept varied between subjects. Accordingly, when the auditory control stimulus was presented, the spatial coincidence of the visual and auditory targets was lost, and the subject’s task became ambiguous because of a conflict between the location of the visual target and the perceived location of the auditory control stimulus. In this experiment, a total of 192 trials was measured (each trial type presented twice at each location, yielding 48 trials per stimulus type).

**Data analysis**

**Data calibration.** Off-line calibration of horizontal and vertical eye position was achieved by training two three-layer neural networks with the back-propagation algorithm on the 72 fixation positions from the calibration experiment (when the button was pressed) and the target coordinates (see Goossens and Van Opstal 1997 for details). The absolute accuracy of the calibration was within 3% over the entire response range. The networks were subsequently
applied to the raw data from the calibration (1st saccades only), AV-multimodal, and AV-control experiments to map the measured induction voltages onto the corresponding 2-D orientations of the eye. Target and response coordinates are expressed as azimuth (α) and elevation (ε) angles, determined by a double-pole coordinate system in which the origin coincides with the center of the head. In this reference frame, target azimuth, α_T, is defined as the angle between the target and the midsagittal plane. Target elevation, ε_T, is the angle between the target and the horizontal plane through the ears with the head in a straight-ahead orientation.

After calibration, saccades were detected by a custom-made program that set separate velocity and acceleration criteria for the identification of saccade onset and offset. All markings were checked by the experimenter and corrected if needed. SRTs below 80 ms were excluded since these saccades presumably were anticipatory (Corneil and Munoz 1996), and SRTs greater than 1,000 ms were excluded because of a presumed lack of subject alertness. SRTs are expressed with respect to the onset of the first target stimulus, regardless of the stimulus asynchrony. SRTs are plotted as cumulative percentage probabilities on a probit scale (i.e., inverted gaussian) as a function of the reciprocal SRT (-1/SRT) (see Carpenter and Williams 1995). In this format, a Gaussian distribution results in a straight line.

Statistics. For saccade accuracy, the optimal linear fit of the stimulus-response relation between saccade amplitude and target eccentricity was found by minimizing the sum-squared deviation of

\[ \alpha_R = a + b \cdot \alpha_T \quad \text{and} \quad \varepsilon_R = c + d \cdot \varepsilon_T \]  

for the azimuth and elevation components, respectively. In Eq. 5.1, a and c are the response biases in degrees (offset of the fitted line), and b and d are the dimensionless response gains (slopes). Confidence levels for Pearson’s correlation coefficients were obtained through the bootstrap method (100 regressions on randomly drawn realizations of the data set) (Press et al. 1992). The gain, bias, residual error (SD relative to the fitted line), mean absolute error (the mean sum-squared difference between the target and response coordinates), and the correlation coefficient extracted from Eq. 5.1 describe different aspects of response behavior. Response gain and bias relate to spatial accuracy, whereas the residual error and the correlation coefficient relate to the variability and spatial resolution of the system, respectively. The absolute error depends on both the accuracy and the variability of the responses. To statistically analyze the differences between two distributions, the one-dimensional (1-D) Kolmogorov-Smirnov (KS) statistic was determined (Press et al. 1992). This statistic is based on the cumulative probabilities constructed from the ranked data arrays and is particularly useful for limited numbers of data points for which the underlying probability distributions are unknown. In cases where a statistical comparison was made between two 2-D distributions of data (i.e., data described by both response accuracy and SRT; see Fig. 5.6), the 2-D KS statistic for the difference between distributions was determined (Press et al. 1992).

Calculation of the Race-Model Prediction. Previous studies of the SRT reduction afforded by presenting bimodal stimuli have utilized the concept of a race model to provide a prediction for the SRT distribution that would be expected if the subject reacted simply to whichever stimulus was perceived first (Colonius and Arndt 2001; Corneil and Munoz 1996; Harrington and Peck 1998; Hughes et al. 1994, 1998). This concept, which operates like a logical OR-gate, was originally developed to model manual reaction times and is alternatively referred to as statistical facilitation or probability summation (Gielen et al. 1983; Miller 1982; Raab 1962). The SRT distribution predicted by a race model, R(τ) (where τ is a given SRT), is derived from the normalized SRT distributions for saccades to the unimodal auditory or visual stimuli, A(τ) and V(τ), respectively, by the following equation:
If the observed SRTs to bimodal stimuli are shorter than those predicted by the race model, the bimodal signals are assumed to have been neurally integrated prior to saccade initiation. We calculated the race model prediction for all bimodal trial types in the AV-experiment, using the unimodal SRT distributions for V-saccades and A-saccades at the appropriate S/N ratio in 10-ms bins. The unimodal distributions were shifted by 100 ms for A100V and V100A-saccades.

5.3 RESULTS

Properties of unimodal V-saccades and A-saccades
The presence of the AV-background and the S/N ratio of the acoustic environment impacted
Table 5.1. Median SRTs and linear regression results for V-saccades generated in the calibration experiment (without AV-background) and in the AV-experiment (with AV-background)

<table>
<thead>
<tr>
<th>Subject</th>
<th>SRT, ms</th>
<th>Azimuth</th>
<th>Elevation</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Gain</td>
<td>Bias</td>
<td>Corr</td>
</tr>
<tr>
<td>Calibration experiment without AV-background</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>162 ± 29</td>
<td>0.91</td>
<td>0.0</td>
<td>0.99</td>
</tr>
<tr>
<td>DM</td>
<td>226 ± 47</td>
<td>0.94</td>
<td>0.3</td>
<td>0.99</td>
</tr>
<tr>
<td>JO</td>
<td>196 ± 48</td>
<td>0.92</td>
<td>-1.1</td>
<td>0.98</td>
</tr>
<tr>
<td>MW</td>
<td>174 ± 55</td>
<td>0.88</td>
<td>0.4</td>
<td>0.99</td>
</tr>
<tr>
<td>MZ</td>
<td>168 ± 20</td>
<td>0.93</td>
<td>-0.2</td>
<td>1.00</td>
</tr>
<tr>
<td>AV-multisensory experiment with AV-background</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>352 ± 113</td>
<td>0.62</td>
<td>1.0</td>
<td>0.84</td>
</tr>
<tr>
<td>DM</td>
<td>419 ± 134</td>
<td>0.62</td>
<td>0.7</td>
<td>0.78</td>
</tr>
<tr>
<td>JO</td>
<td>329 ± 95</td>
<td>0.66</td>
<td>-1.0</td>
<td>0.83</td>
</tr>
<tr>
<td>MW</td>
<td>276 ± 82</td>
<td>0.76</td>
<td>0.9</td>
<td>0.86</td>
</tr>
<tr>
<td>MZ</td>
<td>319 ± 77</td>
<td>0.87</td>
<td>0.1</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Values in SRT are means ± SD; n is number of saccades. Note the much longer and more variable SRTs for V-saccades in the latter experiment, as well as the considerably lower accuracy, which is apparent from lower gain and correlation values and much higher residual errors (Bias and Err in deg; Gain and Corr are dimensionless). SRT, saccadic reaction time.

the SRT and accuracy of unimodal V-saccades and A-saccades. V-saccades had longer SRTs in the presence of the AV-background, as evidenced by comparing the results from the AV-experiment (with the AV-background) to the results from the calibration experiment (without the AV-background; Fig. 5.2A and Table 5.1; \( P < 10^{-8} \) for all subjects, 1-D KS test).1 Response accuracy, quantified by the parameters of the linear regression analysis between saccade amplitude and target eccentricity (see Methods), demonstrated that V-saccade accuracy in both azimuth (Fig. 5.2B) and elevation (Fig. 5.2C) was also compromised in the presence of the AV-background. These accuracy differences were significant across all subjects (\( P < 0.05 \) using the 1-D KS test; Table 5.1).

The SRT and accuracy of A-saccades depended on the S/N ratio of the auditory target relative to background. SRTs were systematically longer and more variable for lower S/N ratios, as shown for a representative subject in Fig. 5.3A (Table 5.2 for all subjects). Interestingly, the accuracy of A-saccades decreased for the lower S/N ratios, but in a manner that differed for the azimuth and elevation response components. Targets were well localized in both azimuth and elevation at the highest S/N ratios (i.e., \( A_{+6} \)-saccades), although the residual error was greater in elevation (Fig. 5.3, B and C). At the high S/N ratios, the accuracy of A-saccades was in the same range as V-saccades (compare gain and error values in Tables 1 and 2). At the lowest S/N ratio (i.e., \( A_{-21} \)-saccades), saccade accuracy in azimuth decreased only

---

1 This SRT difference cannot be due to the gap between FP offset and target onset, since the 200-ms gap in the AV-multimodal experiment would favor even shorter SRTs (see Fischer and Weber 1993 for review).
Figure 5.3. Comparison of A-saccade SRT (A) and accuracy (B and C) in the AV-multimodal experiment with different signal-to-noise (S/N) ratios for subject JO. Same format as Fig. 5.2. Open squares, light-gray circles, dark-gray triangles, and black squares denote the data from A-6, A-12, A-18, and A-21 saccades, respectively. Note in A that the distributions shift systematically to the right and cover a greater range for lower S/N ratios.

slightly when compared with A\textsubscript{6} saccades (Fig. 5.3B), yet was almost completely abolished in elevation (Fig. 5.3C). An analysis of the gain of the stimulus-response relationship (Fig. 5.4A) and the absolute response error (Fig. 5.4B) for the azimuth and elevation response components of A-saccades across S/N ratio confirmed the greater inaccuracy of aurally guided saccades in the elevation component at lower S/N ratios (i.e., for A\textsubscript{18} and A\textsubscript{21} saccades) than in the azimuth component, which was only slightly compromised for A\textsubscript{21} saccades (see also Table 5.2). These results confirm and extend earlier findings of auditory localization (Good and Gilkey 1996; Zwiers et al. 2001).

In summary, the increased SRT and decreased accuracy of V- and A-saccades (particularly at low S/N ratios) confirmed that the presence of the AV-background made the task much more difficult, although not impossible. A similar analysis on saccades generated to the dif-
<table>
<thead>
<tr>
<th>Subject</th>
<th>S/N, dB</th>
<th>SRT, ms</th>
<th>Azimuth</th>
<th>Elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gain</td>
<td>Bias</td>
</tr>
<tr>
<td>BC</td>
<td>6</td>
<td>178 ± 38</td>
<td>0.85</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>190 ± 61</td>
<td>0.82</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>212 ± 59</td>
<td>0.88</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>264 ± 128</td>
<td>0.75</td>
<td>0.7</td>
</tr>
<tr>
<td>DM</td>
<td>6</td>
<td>201 ± 57</td>
<td>0.85</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>226 ± 85</td>
<td>0.81</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>262 ± 75</td>
<td>0.90</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>360 ± 132</td>
<td>0.72</td>
<td>1.0</td>
</tr>
<tr>
<td>JO</td>
<td>6</td>
<td>186 ± 47</td>
<td>0.87</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>199 ± 49</td>
<td>0.90</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>227 ± 84</td>
<td>0.97</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>263 ± 106</td>
<td>0.79</td>
<td>1.5</td>
</tr>
<tr>
<td>MW</td>
<td>6</td>
<td>153 ± 32</td>
<td>0.80</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>170 ± 38</td>
<td>0.88</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>173 ± 47</td>
<td>0.84</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>205 ± 116</td>
<td>0.76</td>
<td>3.6</td>
</tr>
<tr>
<td>MZ</td>
<td>6</td>
<td>210 ± 53</td>
<td>0.84</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>218 ± 57</td>
<td>0.89</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>242 ± 96</td>
<td>0.91</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>299 ± 170</td>
<td>0.75</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Values in SRT are means ± SD; n is number of saccades. Note increased SRT and residual error, and decreased elevation gain and correlation coefficient with decreasing signal-to-noise (S/N) ratio. SRT, saccadic reaction time.

Different types of bimodal stimuli is now presented.

**Properties of AV-saccades (no temporal asynchronies)**
A representative example of the properties of AV-saccades is demonstrated in Fig. 5.5, in which V-, A-18, and A-V-saccades are compared. Note that the SRT distributions for A-18 and A-18V-saccades are nearly superimposed (Fig. 5.5A), equaling but not exceeding the race model prediction based on the unimodal SRT distributions (black, solid line). The relationships of AV-saccades to the race model prediction are studied more thoroughly below. A comparison of saccade accuracy demonstrated that the residual errors of A-18V-saccades were smaller than the residual errors for both V- and A-18-saccades in the elevation (Fig. 5.5C), but not in the azimuth (Fig. 5.5B), component.

The SRT and accuracy of AV-saccades is contrasted more directly with unimodal saccades using a 2-D comparison of absolute localization error (combining both azimuth and elevation) versus SRT (see Fig. 5.6 for 1 subject). Each point in Fig. 5.6 stems from an individual
Auditory-Visual Interactions

**Figure 5.4.** Response gains (A) and absolute error (B) for azimuth (circles) and elevation (squares) for A-saccades as a function of S/N ratio for all subjects (small, open symbols and dashed lines) and the sample mean (solid symbols). Note the robustness of the azimuth response component, which degrades only slightly for A-21 saccades, whereas the elevation response component degrades precipitously at higher S/N ratios.

**Figure 5.5.** Comparison of SRT (A) and saccade accuracy (B and C) in the AV-multimodal experiment for A-18-saccades (gray circles), V-saccades (gray triangles), and bimodal A-V-saccades (open squares) for subject DM. Same format as Fig. 5.2. The solid black line in A denotes the SRT distribution predicted by the race model.
Figure 5.6. Absolute localization error plotted as a function of SRT for subject MZ. Symbols denote observations from individual V-saccades (gray triangles), $A_{-18}$-saccades (gray triangles), and $A_{-19}$-V-saccades (open squares). Ellipses circumscribe 1 SD around the mean. Only data within 2 SDs of the SRT mean are shown.

saccade, and the ellipses circumscribe the mean values within one SD. Note that V-saccades were generated at longer SRTs but were more accurate than $A_{-18}$-saccades. However, the 2-D distribution of $A_{-18}$-V-saccades is clearly distinct from either unimodal distribution, as the AV-saccades attained accuracies in the range of V-saccades, but at SRTs in the range of $A_{-18}$-saccades. The 2-D KS pair-wise statistic comparing the three distributions showed that all were significantly different ($P < 10^{-5}$ for all 3 comparisons). The results of this three-way statistical comparison across all subjects and at all S/N ratios is shown in Table 5.3. When the S/N ratio was low (-18 or -21 dB), the 2-D distributions for AV-saccades differed significantly from both unimodal A-saccades and V-saccades. For the higher S/N ratios (-12 or -6 dB), the distributions for AV-saccades were often similar to the A-saccade distributions, but were always significantly different from V-saccades.

Another interesting observation from Fig. 5.6 is that the AV-saccades appeared to be distributed over a narrower accuracy-SRT range than A- and V-saccade distributions (compare the horizontal and vertical spans of the ellipses in Fig. 5.6). To quantify this point across all S/N ratios and subjects, we made two comparisons. First, we compared the SRT variance of AV-saccades to A-saccades (Fig. 5.7A) and demonstrated that the SRT variance for AV-saccades was similar to A-saccades at high S/N ratios, but had consistently lower variances at lower S/N ratios. Second, a comparison of the accuracy variance between AV-saccades and V-saccades showed that the accuracy variances for AV-saccades were consistently narrower than those for V-saccades, as the majority of data points lay below the diagonal in Fig. 5.7B. Thus, although auditory targets were barely detectable in elevation at low S/N ratios (Fig. 5.4 and Table 5.2), they were integrated effectively with the visual target to reduce both the mean and the variance of AV-saccade SRT and accuracy.

Taken together, the data suggest that the magnitude of multisensory interactions depended systematically on the S/N ratio of the auditory target relative to the background. At low S/N
**Auditory-Visual Interactions**

**Table 5.3. Statistical summary of probabilities derived from a 2-D KS statistic comparing the distributions of SRT-accuracy saccade data**

<table>
<thead>
<tr>
<th>Subject</th>
<th>S/N, dB</th>
<th>P(V = A)</th>
<th>P(V = AV)</th>
<th>P(A = AV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC</td>
<td>6</td>
<td>*</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>*</td>
<td>*</td>
<td>†</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>*</td>
<td>*</td>
<td>†</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>DM</td>
<td>6</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>*</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>*</td>
<td>*</td>
<td>‡</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>*</td>
<td>†</td>
<td>*</td>
</tr>
<tr>
<td>JO</td>
<td>6</td>
<td>*</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>*</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>*</td>
<td>*</td>
<td>†</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>MW</td>
<td>6</td>
<td>*</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>*</td>
<td>*</td>
<td>†</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>*</td>
<td>‡</td>
<td>‡</td>
</tr>
<tr>
<td>MZ</td>
<td>6</td>
<td>*</td>
<td>*</td>
<td>‡</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>*</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Note for all subjects the 2 lowest S/N ratios yielded significantly different distributions for all 3 comparisons. KS, Kolmogorov-Smirnov; n.s., not significant (P > 0.05); for other abbreviations, see Table 5.2.

* P < 0.00001.
† 0.00001 < P < 0.001.
‡ 0.001 < P < 0.05.

**Figure 5.7.** Comparison of the variances in SRT distributions between A-saccades and AV-saccades (A) and of the absolute localization error between V-saccades and AV-saccades (B). Data are shown for all 5 subjects. The dashed lines denote the unity line, where the data would cluster if the variances were equal. Note, however, that most data points lie below these lines, indicating that the AV-saccades had narrower distributions.
ratios, A-saccades were characterized by decreasing accuracy, and longer SRT and V-saccades were more accurate but were generated at much longer SRTs. When these two weak stimuli were combined, AV-saccades benefited from the “best of both worlds” in that they were as accurate as V-saccades and initiated at SRTs typical of A-saccades. Moreover, the variability of SRT and accuracy for AV-saccades was decreased compared with A-saccades and V-saccades, respectively, indicating more consistent responses.

**AV interactions as a function of stimulus timing**

In this section, we compare the properties of saccade responses to synchronous stimuli...
Auditory-Visual Interactions

(i.e., AV-saccades) to those from asynchronous stimuli (i.e., A100V- or V100A-saccades). First, the SRT distributions for each stimulus asynchrony and S/N ratio combination were compared with the distributions predicted by the race model (see Methods). To that end, the observed cumulative response distributions for bimodal stimuli were plotted as a function of the predicted cumulative race distributions (Fig. 5.8A,B,C for a representative subject). Such plots compare the relative differences between the observed and predicted cumulative distributions, regardless of absolute SRT. Note that the comparison plots for AV-saccades (solid lines in Fig. 5.8B) lay close to the unity line, implying that the observed SRT distributions were approximately equal to those predicted by the race model for all four S/N ratios, consistent with Fig. 5.5A. However, the comparison plots for V100A-saccades (dashed lines in Fig. 5.8A) lay well above the unity line, indicating that the observed SRT distributions were considerably shorter than those predicted by the race model. Conversely, the comparison plots for A100V-saccades (dashed-dotted lines in Fig. 5.8C) lay well below the unity line, meaning that the observed SRTs were much longer than predicted by the race model. This latter finding is quite striking since it implies that the delayed visual stimulus inhibits the SRTs for A100V-saccades compared with the SRTs for A-saccades.

It is not trivial to appreciate how these relationships with the race model change with the S/N ratio of the acoustic environment. To quantify this, we determined the area of the difference curve between the observed and predicted cumulative SRT distributions for those SRTs where the cumulative probabilities fell between 0.1 and 0.9. These calculated areas express the amount by which the observed data exceeded (positive values) or fell short (negative values) of the race model prediction regardless of the absolute SRTs, and are plotted for the same subject in Fig. 5.8D. Presented this way, it is clear that no systematic relationship emerged with the S/N ratio, hence the extracted areas were averaged across all S/N ratios (gray bars in Fig. 5.8D). AV-saccades did not deviate significantly from the race model \((P > 0.05)\), whereas A100V-saccades had significantly longer SRTs by about 20% \((P < 0.02)\) and V100A-saccades had significantly shorter SRTs by about 15% \((P < 0.02)\) than those predicted by the race model. This pattern of SRT responses was found for three of five subjects. In the other two subjects, there was no significant difference between the observed and predicted SRT distributions for both AV-saccades and V100A-saccades. In these two subjects, the unimodal (shifted) SRT distributions did not overlap sufficiently, so that the race model prediction equaled the shorter unimodal SRT distribution (in this case for A-saccades). Regardless, averaging across all subjects revealed that the overall patterns were consistent (Fig. 5.8E; \(P < 0.001\) for the A100V-saccades, \(P > 0.05\) for AV-saccades, and \(P < 0.05\) for V100A-saccades). Thus the relationships of the observed SRTs to those predicted by the race model depended on stimulus asynchrony.

To quantify the accuracy of bimodal saccades across stimulus asynchrony and S/N ratio, we first plotted the absolute azimuth and elevation localization errors as a function of S/N ratio for the different temporal asynchronies. Figure 5.9 shows data from one representative subject. Note that the accuracy of bimodal saccades almost always surpassed that of A-saccades, regardless of asynchrony or S/N ratio. In most cases, the accuracy of AV-saccades was also better than that of V-saccades. V100A-saccades tended to be among the most accurate, surpassing both AV- and A100V-saccades, particularly in elevation at low S/N ratios (Fig. 5.9B). Statistical analysis across all subjects confirmed that the elevation gain of bimodal saccades differed more from the gains obtained from V-saccades than A-saccades did at the lower S/N ratios (Table 5.4). However, this trend was not observed in the azimuth response component (Table 5.4).

A summary of the combined SRT-accuracy results for all bimodal stimulus conditions is shown in Fig. 5.10. These data were obtained by first normalizing the results for each stimulus condition with respect to the accuracy and SRT of V-saccades within each subject, and then averaging the normalized results for each condition across all subjects (note that data
**Figure 5.9.** Absolute azimuth (A) and elevation (B) localization error as a function of S/N ratio for subject MW. Data from different asynchronies are plotted in different series (see legend). Dashed horizontal lines denote the data from V-saccades.

**Table 5.4. Statistical summary of probabilities comparing the gains of the azimuth or elevation stimulus-response relationship between A-saccades and bimodal A100V-, AV-, or V100A-saccades**

<table>
<thead>
<tr>
<th>Response Component</th>
<th>S/N Ratio, dB</th>
<th>P (A100V = A)</th>
<th>P (AV = A)</th>
<th>P (V100A = A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azimuth</td>
<td>-6</td>
<td>n.a.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>-12</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>-18</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>-21</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Elevation</td>
<td>-6</td>
<td>n.a.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>-12</td>
<td>n.s.</td>
<td>n.s.</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>-18</td>
<td>n.s.</td>
<td>†</td>
<td>‡</td>
</tr>
<tr>
<td></td>
<td>-21</td>
<td>‡</td>
<td>‡</td>
<td>‡</td>
</tr>
</tbody>
</table>

For each response component, gains for A-saccades and bimodal saccades were subtracted from the gains for V-saccades for each completed series in each subject. This yielded a distribution of 18 data points for each S/N ratio. A 1-D KS test was then applied to compare the various distributions listed in the column headers. n.s., P > 0.05; n.a., not available; for other abbreviations, see Tables 2 and 3.

* 0.01 < P < 0.05.
† 0.001 < P < 0.01.
‡ P < 0.001.
**Figure 5.10.** Summary figure showing the mean normalized absolute localization error vs. the mean SRT averaged across all subjects. Values were first normalized for each subject to the absolute localization error and SRT for V-saccades (large circle with gray-gradient and dashed lines), and then averaged across all subjects. Data from A-saccades (black hexagrams), A100V-saccades (light-gray circles), AV-saccades (open squares), and V100A-saccades (dark-gray triangles) are plotted in different series, with the different S/N ratios denoted on the graph. Note the degradation in SRT and accuracy with decreasing S/N ratio. Also, compare the position of the bimodal saccades series with the temporal asynchronies. The largest improvements in accuracy-SRT performance compared with unimodal saccades were realized at the lower S/N ratios.

**Figure 5.11.** Comparison of SRT (A) and accuracy (B and C) in the AV-control experiment. Data pooled from all subjects. Same format as Fig. 5.2. Filled gray circles denote the data from AV-control saccades. Gray triangles denote the data for A, V-saccades, black hexagrams for A_{-12}-saccades and open squares for V-saccades.
for A<sub>100</sub>V-saccades are absent; see Methods). All bimodal data in this accuracy-SRT plane are clearly distinct from the unimodal saccades, and there were obvious patterns depending on both the synchrony and the S/N ratio. First, the normalized SRT and absolute localization error of A-saccades and bimodal saccades progressively increased with decreasing S/N ratios. Second, the position of the bimodal data in the accuracy-SRT plane depended strongly on the stimulus synchrony. Relative to AV-saccades, A100V-saccades were more inaccurate and had longer SRTs at the lower S/N ratios. This latter point is in agreement with our earlier analysis on the SRTs compared with the race model (Fig. 5.8) and again shows that the delayed visual stimulus slowed the SRT of A100V-saccades compared with unimodal A-saccades. In contrast, V100A-saccades were more accurate than AV-saccades, but had longer SRTs. Yet, V100A-saccades clearly surpassed the predictions of the race model (Fig. 5.8; recall that the unimodal distributions had to be shifted by 100 ms to determine the race model predictions). Overall, the best performances, indexed by the relative position of the bimodal saccades compared with the unimodal counterpart, were observed for AV- and V100A-saccades at the lowest S/N ratios.

**AV-control experiment**

We conducted a control experiment to test for the presence and influence of a generalized warning effect of the auditory target on both SRT and accuracy. Figure 5.11 shows the data pooled for all subjects from the AV-control experiment, which used an additional bimodal stimulus consisting of a visual target with an auditory stimulus set up by the background speakers (recall that subjects perceived this sound as emanating from a fixed location near center). We emphasize two main points from this experiment. First, although the control auditory stimulus provided some warning cue information to shorten SRTs of AV-control saccades compared with V-saccades, the SRTs for spatially coincident A<sub>12</sub>V-saccades were still shorter (Fig. 5.11A). Thus, although one component (around 60 ms) of the shorter SRTs for A<sub>12</sub>V-saccades could be attributed to a warning effect, another component (accounting for an additional 65 ms) depends on the spatial alignment of the stimuli. Second, note that AV-control saccades were much more inaccurate than spatially coincident A<sub>12</sub>V-saccades (Fig. 5.11, B and C) or either V- or A-saccades (Table 5.5). Thus, although the nonlocalizable auditory target conferred a beneficial warning effect on SRTs, it degraded saccade accuracy. These results were consistent across all subjects (Table 5.5), from which it was concluded that the combined benefits conferred by auditory-visual integration across SRT and accuracy depended on the spatial alignment of the stimuli.

**5.4 Discussion**

This study investigated the first-saccade responses to visual, auditory, and bimodal stimuli distributed throughout the 2-D oculomotor range and embedded within a complex AV-background. We believe the timing and metrics of the first saccade provide a measure for the speed and precision with which the oculomotor system can localize and orient to the stimuli. The properties of saccades to unimodal stimuli testify to the complexity of the task: the SRT and error of V-saccades increased greatly in the presence of the AV-background (Fig. 5.2, Table 5.1), and the SRT and error of A-saccades depended systematically on the S/N ratio of the acoustic scene, becoming prolonged and inaccurate, particularly in the elevation component, at lower S/N ratios (Figs. 5.3 and 5.4; Table 5.2). The properties of unimodal saccades provided wide ranges over which the benefits afforded by multisensory integration were realized. Specifically, saccades to bimodal stimuli were generated at SRTs typical of A-saccades, but at accuracies typical of V-saccades. These results depended critically on the temporal register of the stimuli and on the S/N ratio of the acoustic environment (Fig. 5.10).
The control experiment demonstrated that the spatial register of the stimuli is also important (Fig. 5.11; Table 5.5), although this variable was not systematically manipulated. In this discussion, we argue that mechanisms other than neural integration of the auditory and visual signals cannot explain all aspects of our data. Our results are then related to behavioral and neurophysiological studies. Last, we propose a conceptual neural framework.

**Consideration of mechanisms other than neural integration**

We consider three mechanisms that could underlie the observed properties of bimodal saccades: race models, aurally assisted visual search, and auditory warning-cue effects. Each predicts specific patterns of SRTs and accuracy that differ substantially from those we observed. For example, race models state that subjects respond to whichever stimulus is perceived first, and derive SRT distributions from the unimodal data (Eq. 5.2) (Colonius and Arndt 2001; Corneil and Munoz 1996; Gielen et al. 1983; Harrington and Peck 1998; Hughes et al. 1994, 1998; Nozawa et al. 1994). Since the SRTs for A-saccades were much shorter than for V-saccades, race models predict that most saccades in bimodal trials would be initiated in response to the auditory target. However, if the subjects only reacted to the auditory target on bimodal trials, then the accuracy of bimodal saccades should equal the accuracy.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Parameter</th>
<th>Azimuth</th>
<th>Elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>V</td>
<td>A12</td>
</tr>
<tr>
<td>BC</td>
<td>Corr</td>
<td>0.91</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.4</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6.1</td>
<td>4.3</td>
</tr>
<tr>
<td>DM</td>
<td>Corr</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>4.2</td>
<td>3.7</td>
</tr>
<tr>
<td>JO</td>
<td>Corr</td>
<td>0.82</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>0.83</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.1</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>8.6</td>
<td>4.1</td>
</tr>
<tr>
<td>MW</td>
<td>Corr</td>
<td>0.73</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>0.63</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.7</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>11.0</td>
<td>4.9</td>
</tr>
<tr>
<td>MZ</td>
<td>Corr</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Gain</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>1.2</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>5.7</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Median SRT, ms | 332 ± 93 | 206 ± 65 | 210 ± 56 | 275 ± 100

Values in Median SRT are means ± SD. Saccade accuracy was better when the bimodal stimuli were spatially aligned (A_{12}V) than when the auditory target was nonlocalizable (AV-cont). V-saccades had the longest SRTs (bottom row, averaged across subjects). SRT, saccadic reaction time.

The control experiment demonstrated that the spatial register of the stimuli is also important (Fig. 5.11; Table 5.5), although this variable was not systematically manipulated. In this discussion, we argue that mechanisms other than neural integration of the auditory and visual signals cannot explain all aspects of our data. Our results are then related to behavioral and neurophysiological studies. Last, we propose a conceptual neural framework.
of A-saccades. This was never observed; bimodal saccades were always more accurate than A-saccades (Fig. 5.10). Even a trial-by-trial comparison of SRT and accuracy shows that individual AV-saccades combine properties typical of both A-saccades and V-saccades (Fig. 5.6).

Whereas the observed SRTs for AV-saccades agree nicely with those predicted by the race models, the observed SRTs for A100V- were longer and the SRTs for V100A-saccades were shorter than the race model predictions (Fig. 5.8), testifying to another inadequacy of a race model mechanism. At first, it might seem surprising that the observed SRTs for AV-saccades did not exceed the predicted SRTs, given the many examples of race model violations in the literature (Colonius and Arndt 2001; Harrington and Peck 1998; Hughes et al. 1994, 1998). However, many of these race model violations stem from simple experiments in which subjects orient to the target(s) without the presence of distracting stimuli. Complicating the experiments by employing distracting stimuli, or by instructing the subjects to orient to the auditory instead of the visual target, lead to observed SRTs to bimodal stimuli that do not exceed, let alone meet, the SRTs predicted by the race model (Corneil and Munoz 1996; Hughes et al. 1994). More complex experimental paradigms, such as the one described here, presumably engage processes related to target selection and/or discrimination that elongate SRTs and demonstrate the insufficiency of a simple race model mechanism in accounting for the observed data. Below, we surmise on neural mechanisms that could account for the shorter SRT and improved accuracy of V100A-, but not A100V-saccades.

An “aurally assisted visual search” mechanism (Perrott et al. 1990, 1991) also cannot explain the combined patterns of SRT and accuracy. This mechanism proposes that the role of the auditory localization system is to bring the fovea into line with an auditory stimulus, constraining the area over which the visual system searches for a visual target, thereby expediting the time to locate and identify a visual target without necessitating auditory-visual integration. Importantly, while this mechanism considers processes beyond the generation of the first saccade and could explain the evolution of the scanning pattern, it holds that the first saccade to a bimodal stimulus is aurally guided. This mechanism therefore predicts that both the SRT and accuracy of AV-saccades should equal A-saccades, which differs from the observed data. As with the race model, the aurally assisted visual search mechanism cannot explain the improved accuracy of AV-saccades beyond the level typical of A-saccades. Another prediction of this mechanism is that A100V-saccades should be the most accurate and V100A-saccades the most inaccurate. This also differed drastically from the observed data (Fig. 5.10).

A third explanation of the observed data could be that the auditory system provides a nonlocalized “warning cue” to the subject to initiate the saccade, irrespective of the spatial register of the stimuli (Kingstone and Klein 1993; Ross and Ross 1980, 1981). While this mechanism could explain a partial reduction of SRTs of bimodal saccades (Fig. 5.11; Table 5.5), the AV-control experiment demonstrated that spatial alignment of the visual and auditory stimuli was crucial in mediating further improvements in SRT and accuracy, counter to a warning-cue mechanism (Fig. 5.11; Table 5.5). It is also hard to imagine how a warning cue mechanism could explain the larger improvements in accuracy at lower S/N ratios (Fig. 5.9) or why saccade accuracy and SRT varied systematically with the different temporal asynchronies (Fig. 5.10). While we would have liked to have systematically altered the spatial congruity between the AV-stimuli in this experiment, such an experiment is a major undertaking and is the focus of a separate and ongoing series of experiments.

In conclusion, all three mechanisms assume that bimodal saccades are driven in response to one modality, and therefore predict that their timing and metrics should be identical to either V- or A-saccades. Yet, in none of the 11 different stimulus configurations tested did bimodal saccades have an accuracy-SRT profile identical to V- or A- saccades (Fig. 5.10). The observed properties of bimodal saccades combine aspects of both A-saccades and
Auditory-Visual Interactions

V-saccades to achieve the “best-of-both-worlds,” and accordingly the most parsimonious explanation is that auditory and visual stimuli are integrated in a way that depends on their spatial and temporal register.

Rules for multisensory integration of bimodal signals and comparison to previous work

In the intermediate layers of the mammalian SC, many neurons respond to multimodal stimuli (Stein and Meredith 1993). Studies in anesthetized preparations show that the form and magnitude of multisensory interactions in these neurons depend on the temporal and spatial alignment of the stimuli. Further, SC neurons obey the principle of inverse effectiveness, whereby the magnitude of multisensory interactions are largest when the multisensory stimuli are presented at near-threshold intensities (Meredith and Stein 1986). Studies of SC activity in awake preparations have confirmed these basic rules (Bell et al. 2001; Frens and Van Opstal 1998; Peck 1996; Peck et al. 1995; Wallace et al. 1998). However, linking these rules to behavior is not always straightforward. For example, mean SRTs in humans to high-intensity audio-visual stimuli are typically 10-50 ms shorter than SRTs to unimodal stimuli (Colonius and Arndt 2001; Engelken and Stevens 1989; Frens et al. 1995; Goldring et al. 1996; Harrington and Peck 1998; Hughes et al. 1994, 1998; Munoz and Corneil 1995; Nozawa et al. 1994). Studies with lower intensity stimuli have found that the SRT reductions to low-intensity bimodal stimuli are in the same range (Frens et al. 1995; Hughes et al. 1994; Nozawa et al. 1994), contrary to what would have been predicted given inverse effectiveness (Meredith and Stein 1986). Is it possible that inverse effectiveness is masked by other neural processes operating only in behavioral experiments? If so, could these processes also confound SRT and accuracy of bimodal saccades?

In light of these questions, we highlight several limitations or confounds of previous behavioral studies. First, auditory stimuli have been typically constrained to the horizontal plane, meaning that only sound-source azimuth needed to be extracted. Our extension to the 2-D oculomotor range, as well as the manipulation of the acoustic S/N ratio, provided the opportunity to observe major differences in the sensitivity of azimuth and elevation perception. The ability of the auditory system to extract stimulus elevation degraded at higher S/N ratios than stimulus azimuth (Fig. 5.4), consistent with recent studies (Good and Gilkey 1996; Zwiers et al. 2001b). This effect relates presumably to the different mechanisms the CNS uses to extract sound-source azimuth and elevation from the acoustic cues (see Blauert 1997 for review). Consequently, the accuracy improvements afforded by presenting bimodal targets at low S/N ratios were greater in elevation than in azimuth (Fig. 5.9, Table 5.4). Previous studies may have underestimated the contributions of multisensory integration to saccade accuracy by constraining targets to the horizontal plane.

A second limitation of previous studies is the use of a limited number of potential target positions. This could allow subjects to use prior knowledge about potential target positions to prepare movements before target presentation, which, if left unaccounted for, would also lead to underestimations of the contributions of multisensory integration to saccade accuracy. The present setup used 24 potential target locations, making such a strategy highly unlikely.

Third, subject instructions and experimental context affect the temporal expression of multisensory integration (i.e., SRT). For example, requiring a subject to orient specifically to one modality while ignoring the other yields SRT distributions that violate the race model when the instructed target is visual, but not when the instructed target is auditory (Corneil and Munoz 1996; Hughes et al. 1994). In general, requiring subjects to discriminate between modalities prolongs SRTs (Corneil and Munoz 1996) and could confound the estimation of the contributions of multisensory integration to SRT. In the present experiments, subjects could orient to both the auditory and visual stimulus, so this was not a concern. Overall, the
setup employed in our experiments allows for a behavioral assessment of the consequences of multisensory integration over both spatial and temporal domains, while being removed from confounds, such as the three discussed here, that affected the interpretation of previous studies.

A few behavioral studies have manipulated the temporal alignment between auditory and visual stimuli to address the temporal window over which stimuli may interact (Colonius and Arndt 2001; Corneil and Munoz 1996; Engelken and Stevens 1989; Frens et al. 1995). For the saccadic system, the temporal window is about ±100 ms, presumably allowing AV-integration in spite of differences in retinal versus cochlear transduction times (~50 ms and 2-10 ms, respectively) (Gouras 1967; Kraus and McGee 1992) and the speed of sound versus light over a large range of stimulus distances. In the SC of awake, behaving primates, auditory response latencies usually range around 30 ms (Bell et al. 2001; Jay and Sparks 1987) and visual response latencies around 60 ms (see Munoz et al. 2000 for review), suggesting that the more complex transformation of auditory responses into oculocentric coordinates does not greatly affect the relative arrival times at the SC. As shown in Fig. 5.10, the combination of SRT and accuracy of AV-saccades surpassed that for either unimodal A- and V-saccades, and we argued above that a race model could not account for these data. However, Fig. 5.10 also shows that the temporal window permitting excitatory interactions is not symmetrical around synchronously presented stimuli. V100A-saccades were initiated at SRTs that surpassed the race model prediction and were more accurate than any other saccade type. Conversely, A100V-saccades were initiated at SRTs that fell well short of the race model prediction (and were even slower than A-saccades) and were more inaccurate than AV-saccades and V-saccades at low S/N ratios (but were still more accurate than A-saccades). These findings suggest a nonlinearity in the interactions of delayed visual or auditory signals. Apparently, a delayed auditory signal facilitates saccade initiation and sharpens the accuracy of a developing visually guided saccade, but a delayed visual signal inhibits saccade initiation and worsens the accuracy of a developing aurally guided saccade. Although surprising, this nonlinearity bears some resemblance to multisensory recordings in the SC of anesthetized cats, wherein response enhancements are observed if the visual stimulus leads the auditory stimulus but response depressions are observed if the auditory stimulus leads the visual stimulus (Meredith et al. 1987). Understanding the neural mechanism(s) responsible for this nonlinearity requires neuronal recordings from awake, behaving preparations.

**Conceptual model of auditory-visual interactions in a complex scene**

Figure 5.12 presents a conceptual model to explain how activity within the SC might evolve prior to A-, V-, and AV-saccades. We assume that visual and auditory signals are initially processed separately and converge on the SC, inducing modality-specific profiles of SC activity. At high intensities, aurally induced profiles arrive earlier than visually induced profiles, but with lower firing rates and a broader tuning (dashed lines and empty profiles in Fig. 5.12, B and C) (Bell et al. 2001; Frens and Van Opstal 1998; Jay and Sparks 1987; Peck et al. 1995; Wallace et al. 1996, 1998). These profiles continue to develop until a threshold is exceeded, here modeled by an integrated number of spikes. Achieving threshold silences the activity of omnipause neurons (OPNs), permitting the activation of the burst and pulse-step generators that results in saccade generation. Obviously, SRT relates to the time the SC threshold is surpassed (denoted by time in Fig. 5.12C). We assume that saccade accuracy is determined by the center of gravity of the SC activity profile at this time, so that, over multiple trials, a sharper profile leads to more accurate saccades. Although speculative, these assumptions are consistent with the premotor processing prior to visually or aurally guided saccades to single, high-intensity stimuli (see Findlay and Walker 1999; Munoz et al. 2000 for review).
Auditory-Visual Interactions

To our knowledge, there are no data from behaving animals on the SC activity patterns in the presence of the AV-background that address the effects of manipulations of S/N ratio and stimulus asynchronies. Our conceptual model makes predictions about these activity profiles that could be readily tested in future investigations. In the presence of the AV-background or low S/N ratios, unimodal V- or A-saccades have longer SRTs and/or are more inaccurate, respectively (Figs. 5.2-4; Tables 5.1 and 5.2), presumably related to the introduction of “noise” from competitive interactions within or prior to the SC. As a result, unimodal activity profiles within the SC are broader (shaded shapes in Fig. 5.12) and take longer to achieve threshold (Fig. 5.12 C). In such a “noisy” environment, the pairing of AV-stimuli permits nonlinear excitatory interactions that sharpen and increase the firing rate of the SC ac-

Figure 5.12. A: conceptual model of AV-interactions in the superior colliculus (SC). Auditory and visual information is initially processed separately. The signals are nonlinearly integrated in the SC (\( \Pi \)), and relayed to the downstream circuitry (OPn, on-mipause neurons; burst, burst generator; PSG, pulse step generator) when a saccade threshold is reached. B and C: sketch predicted spatial and temporal profiles of SC activity, respectively, prior to the generation of V-, A-, or AV-saccades, either in the presence of noise (indicated by +N; solid lines and shaded profiles) or not (dashed lines and empty profiles). In B, the width of the profiles determines the accuracy of the responses (e.g., note the ellipsoid shape for A-saccades, indicating more variable responses in elevation). In C, we assume that saccadic threshold is reached when an integrated number of spikes is exceeded, here denoted at time \( \tau \). Varying levels of noise (depending either on S/N ratio and/or the number of distractors) are assumed to reduce SC firing rates and smear the population activity in both the auditory and visual channels, resulting in prolonged SRTs and more inaccurate saccades. Provided SC activity induced by the auditory and visual channels overlap temporally and spatially, nonlinear interactions within the SC sharpen the population profile, leading to increased accuracy. See Discussion for further details.
tivity profile, culminating in more accurate saccades generated at shorter and more consistent SRTs. However, such excitatory interactions can only occur over restricted spatial and temporal windows. Inhibitory interactions are observed if the bimodal stimuli are not aligned in space (AV-control experiment, Fig. 5.11 and Table 5.5), or time (i.e., for A100V-saccades, Fig. 5.10). While such interactions could be mediated by an intrinsic inhibitory network within the SC (Kadunce et al. 1997; Meredith and Ramoa 1998; Munoz and Istvan 1998), determining the exact mechanism(s) requires recording in awake, behaving preparations.

Conclusions
The experimental architecture presented in this paper provides a novel and illuminating way to investigate the behavioral significance of multisensory integration in the saccadic system. The complexity of the background and the manipulations of the S/N ratio and temporal register of the stimuli present a formidable challenge to the saccadic system. Importantly, such features mimic those occurring in everyday life, wherein auditory-visual stimuli may be presented over a wide range of distances from the subject or embedded within a complex auditory-visual background. Our results demonstrate the importance of multisensory integration to facilitate orienting in a complex quasi-natural environment.

Acknowledgements
The authors gratefully acknowledge the technical support of T. Van Dreumel and H. Kleijnen. We thank R. Aalbers and P. Hofman for crucial contributions to the software, and M. Zwiers for repeatedly volunteering as a subject. We also thank Drs. I. Armstrong and Y. Kobayashi and C. Au, A. Bell, J. Gore, A. LeVasseur, and E. Marouka for input on an earlier version of the manuscript. These experiments were carried out in the Nijmegen laboratory as part of the Human Frontiers Science Program (Research Grant RG-0174/1998-B). This research was further supported by a doctoral travel award from Queen’s University to B. Corneil, the Canadian Institutes of Health Research to B. Corneil and D. P. Munoz, and the University of Nijmegen to A. J. Van Opstal and M. Van Wanrooij).
Chapter 2
Monaurally deaf people lack the binaural acoustic difference cues in sound level and timing that are needed to encode sound location in the horizontal plane (azimuth). It has been proposed that these people therefore rely on spectral pinna cues of their normal ear to localize sounds. However, the acoustic head-shadow effect (HSE) might also serve as an azimuth cue, despite its ambiguity when absolute sound levels are unknown. Here, we assess the contribution of either cue in the monaural deaf to two-dimensional (2D) sound localization. In a localization test with randomly interleaved sound levels, we show that all monaurally deaf listeners relied heavily on the HSE, whereas binaural control listeners ignore this cue. However, some monaural listeners responded partly to actual sound-source azimuth, regardless of sound level. We show that these listeners extracted azimuth information from their pinna cues. The better monaural listeners were able to localize azimuth on the basis of spectral cues, the better their ability to also localize the vertical direction (elevation) of the sound-source. In a subsequent localization experiment with one fixed sound level, monaural listeners rapidly adopted a strategy on the basis of the HSE. We conclude that monaural spectral cues are not sufficient for adequate 2D sound localization under unfamiliar acoustic conditions. Thus, monaural listeners strongly rely on the ambiguous HSE, which may help them to cope with familiar acoustic environments.

Chapter 3
This paper reports on the acute effects of a monaural plug on directional hearing in the horizontal and vertical planes of human listeners. Sound localization behavior was tested with rapid head-orienting responses toward brief high-pass filtered (> 3 kHz; HP) and broadband (0.5-20 kHz; BB) noises, with sound levels between 30-60 dBA. To deny listeners any consistent azimuth-related head-shadow cues, stimuli were randomly interleaved. A plug immediately degraded azimuth performance, as evidenced by a sound level-dependent shift ('bias') of responses contralateral to the plug, and a level-dependent change in the slope of the stimulus-response relation ('gain'). Although the azimuth bias and gain were highly correlated, they could not be predicted from the plug's acoustic attenuation. Interestingly, listeners performed best for low-intensity stimuli at their normal-hearing side. These data demonstrate that listeners rely on monaural spectral cues for sound-source azimuth localization as soon as the binaural difference cues break down. Also the elevation response components were affected by the plug: elevation gain depended on both stimulus azimuth and on sound level, and, like for azimuth, localization was best for low-intensity stimuli at the hearing side. Our results show that the neural computation of elevation incorporates a binaural weighting process that relies on the perceived, rather than the actual, sound-source azimuth. We conjecture that sound localization ensues from a weighting of all acoustic cues for both azimuth and elevation, in which the weights may be partially determined, and rapidly updated, by the reliability of the particular cue.

Chapter 4
Human sound localization results primarily from the processing of binaural differences in sound level and arrival time for locations in the horizontal plane and of spectral shape cues generated by the head and pinnae for positions in the vertical plane. The latter mechanism incorporates two processing stages: a spectral-to-spatial mapping stage and a binaural weighting stage that determines the contribution of each ear to perceived elevation as function of sound azimuth. We demonstrated recently that binaural pinna molds virtually abolish
Summary

the ability to localize sound-source elevation, but, after several weeks, subjects regained normal localization performance. It is not clear which processing stage underlies this remarkable plasticity, because the auditory system could have learned the new spectral cues separately for each ear (spatial-mapping adaptation) or for one ear only, while extending its contribution into the contralateral hemifield (binaural-weighting adaptation). To dissociate these possibilities, we applied a long-term monaural spectral perturbation in 13 subjects. Our results show that, in eight experiments, listeners learned to localize accurately with new spectral cues that differed substantially from those provided by their own ears. Interestingly, five subjects, whose spectral cues were not sufficiently perturbed, never yielded stable localization performance. Our findings indicate that the analysis of spectral cues may involve a correlation process between the sensory input and a stored spectral representation of the subject’s ears and that learning acts predominantly at a spectral-to-spatial mapping level rather than at the level of binaural weighting.

CHAPTER 5

This study addresses the integration of auditory and visual stimuli subserving the generation of saccades in a complex scene. Previous studies have shown that saccadic reaction times (SRTs) to combined auditory-visual stimuli are reduced when compared with SRTs to either stimulus alone. However, these results have been typically obtained with high-intensity stimuli distributed over a limited number of positions in the horizontal plane. It is less clear how auditory-visual interactions influence saccades under more complex but arguably more natural conditions, when low-intensity stimuli are embedded in complex backgrounds and distributed throughout 2D space. To study this problem, human subjects made saccades to visual-only (V-saccades), auditory-only (A-saccades), or spatially coincident auditory-visual (AV-saccades) targets. In each trial, the low-intensity target was embedded within a complex auditory-visual background, and subjects were allowed over 3 s to search for and foveate the target at 1 of 24 possible locations within the 2D oculomotor range. We varied systematically the onset times of the targets and the intensity of the auditory target relative to background [i.e., the signal-to-noise (S/N) ratio] to examine their effects on both SRT and saccadic accuracy. Subjects were often able to localize the target within one or two saccades, but in about 15% of the trials they generated scanning patterns that consisted of many saccades. The present study reports only the SRT and accuracy of the first saccade in each trial. In all subjects, A-saccades had shorter SRTs than V-saccades, but were more inaccurate than V-saccades when generated to auditory targets presented at low S/N ratios. AV-saccades were at least as accurate as V-saccades but were generated at SRTs typical of A-saccades. The properties of AV-saccades depended systematically on both stimulus timing and S/N ratio of the auditory target. Compared with unimodal A- and V-saccades, the improvements in SRT and accuracy of AV-saccades were greatest when the visual target was synchronous with or leading the auditory target, and when the S/N ratio of the auditory target was lowest. Further, the improvements in saccade accuracy were greater in elevation than in azimuth. A control experiment demonstrated that a portion of the improvements in SRT could be attributable to a warning-cue mechanism, but that the improvements in saccade accuracy depended on the spatial register of the stimuli. These results agree well with earlier electrophysiological results obtained from the midbrain superior colliculus (SC) of anesthetized preparations, and we argue that they demonstrate multisensory integration of auditory and visual signals in a complex, quasi-natural environment. A conceptual model incorporating the SC is presented to explain the observed data.
Samenvatting

Hoofdstuk 2
Eenzijdig doven hebben geen binaurale geluids niveau- en tijdsverschillen, die nodig zijn om geluidslokatie in het horizontale vlak (azimut) te coderen, ter beschikking. Men heeft voorgesteld dat deze mensen zich daarom op spectrale pinna kenmerken van hun horende oor baseren om geluiden te lokaliseren. Echter, het akoestische hoofdschaduweffect (HSE) zou ook als aanduiding voor azimut kunnen dienen, ondanks zijn ambiguïteit wanneer de absolute geluids niveaus onbekend zijn. In deze studie beoordeelden wij de bijdrage van één van beide richtingskenmerken aan tweedimensionale (2D) geluidslocalisatie van eenzijdig doven. In een localisatietest met willekeurig aangeboden geluids niveaus tonen wij aan dat alle eenzijdig dove luisteraars zich vooral op het HSE baseerden, terwijl de controle-groep van binauraal horende luisteraars dit richtingskenmerk negeerde. Echter, sommige monaurale luisteraars reageerden gedeeltelijk correct op de werkelijk aangeboden geluidsbron-azimut, ongeacht geluids niveau. Wij tonen aan dat deze luisteraars azimuthinformatie uit hun pinna-kenmerken extraheerden. Hoe beter de monaurale luisteraars in staat waren om azimut op basis van spectrale kenmerken te lokaliseren, hoe beter hun capaciteit was om de verticale richting (elevatie) van de geluidsbron te bepalen. In een volgend localisatieexperiment met slechts één bepaald geluids niveau maakten de monaurale luisteraars snel gebruik van een strategie op basis van het HSE. Wij concluderen dat de monaurale spectrale richtingskenmerken niet voldoende zijn voor adequate 2D geluidslocalisatie in onbekende akoestische omstandigheden. De eenzijdig dove luisteraars richten zich daarom vooral op het dubbelzinnige HSE, die hen kan helpen om aan vertrouwde akoestische omgevingen het hoofd te bieden.

Hoofdstuk 3
Dit hoofdstuk rapporteert over de acute gevolgen van een monaurale stop op het richtingshoren in de horizontale en verticale vlakken van menselijke luisteraars. Het geluidslocalisatiegedrag werd getest met snelle hoofd-oriënterende reacties naar kortdurende, gefilterde hoogfrequent-doorlatende (3 kHz; HP) en breedbandige (0,5-20 kHz; BB) ruis, met geluidsniveaus tussen 30-60 dBA. Om de luisteraars azimuth kenmerken op basis van het HSE te ontzeggen, werden de stimuli willekeurig aangeboden. Een stop verslechterde onmiddellijk de azimuthprestatie, wat gekenmerkt werd door een geluids niveau-afhankelijke verschuiving van reacties aan de contralaterale zijde van de stop, en een niveau-afhankelijke verandering in de helling (‘gain’) van de stimulus-respons relatie. Hoewel de azimuthverschuiving en de gain uiterst gecorreleerd waren, konden zij niet voorspeld worden op basis van de door de stop veroorzaakte geluids niveauverzwakking. Uiterst interessant was dat luisteraars het best presteerden voor stimuli van lage intensiteit aan hun normaal-horende zijde. Deze gegevens tonen aan dat de luisteraars zich baseren op monaurale spectrale richtingskenmerken voor geluidsbron-azimuth localisatie zodra de richtingskenmerken gebaseerd op binaurale verschillen van mindere kwaliteit worden. Ook werden de elevatie responsen beïnvloed door de stop: de elevatie-gain hing af van zowel stimulus-azimut als van geluids niveau, en, net zoals voor azimuth, was de localisatie het best voor stimuli van lage intensiteit aan de horende kant. Onze resultaten tonen aan dat bij de neurale berekening van elevatie een binauraal wegingsproces betrokken is, dat zich op de waargenomen, in plaats van de daadwerkelijke, geluidsbron azimuth baseert. Wij veronderstellen dat geluidslocalisatie tot stand komt als gevolg van een weging tussen alle akoestische richtingskenmerken voor zowel azimuth als elevatie, waarin de gewichten gedeeltelijk worden bepaald, en snel bijgewerkt, door de betrouwbaarheid van de verscheidene richtingskenmerken.
Samenvatting

HOOFDSTUK 4
Geluidslocalisatie bij mensen wordt voornamelijk bepaald uit de verwerking van binaurale tijds- en geluidsniveau-verschillen voor lokaties in het horizontale vlak en van spectrale richtingskenmerken, die door het hoofd en pinnae worden geproduceerd, voor posities in het verticale vlak. Het laatstgenoemde mechanisme is afhankelijk van twee verwerkingsstadia: een spectrum-naar-positie calibratiestadium en een binauraal wegingsstadium dat de bijdrage van elk oor tot waargenomen elevatie als functie van azimut bepaalt. Wij toonden onlangs aan dat binaurale pinnamallen localisatie van geluidsbronelevatie vrijwel onmogelijk maakten, maar na verscheidene weken, bereikten de proefpersonen hun normale localisatiestatus. Het is niet duidelijk welk verwerkingsstadium aan deze opmerkelijke plasticiteit ten grondslag ligt, omdat het auditieve systeem de nieuwe spectrale richtingskenmerken voor elk oor afzonderlijk (adaptatie van het ruimtelijke-calibratie stadium ) geleeerd kon hebben, of voor slechts één oor, terwijl de bijdrage van dat oor uitgebreid werd naar de contralaterale zijde (adaptatie van het binauraal-wegingsstadium). Om deze twee mogelijkheden van elkaar te onderscheiden, pasten wij een monaurale spectrale vervorming op lange termijn bij 13 proefpersonen toe. Onze resultaten tonen aan dat, in acht experimenten, de luisteraars leerden nauwkeurig te lokaliseren met behulp van spectrale richtingskenmerken die wezenlijk verschillen van de kenmerken die verstrekt werden door hun eigen oren. Opmerkelijk was dat vijf proefpersonen, van wie de spectrale richtingskenmerken niet voldoende verstoord waren, nooit een stabiel localisatiegedrag bereikten. Onze bevindingen wijzen erop dat bij de analyse van spectrale richtingskenmerken een correlatieproces tussen de sensorische input en een opgeslagen spectrale vertegenwoordiging van de oren van de proefpersoon betrokken zou zijn en dat het leren voornamelijk plaatsvindt in het spectrum-naar-positie stadium in plaats van op het binauraal-wegingsstadium.

HOOFDSTUK 5
Deze studie richt zich op de integratie van auditieve en visuele stimuli die ten grondslag ligt aan de generatie van saccades in een complexe scène. Vorige studies hebben aangetoond dat saccadische reactietijden (SRTs) naar gecombineerde auditief-visuele stimuli zijn gereduceerd vergeleken met SRTs naar één van beide stimuli alleen. Echter, deze resultaten zijn typisch verkregen met stimuli van hoge intensiteit, die verdeeld zijn over een beperkt aantal posities in het horizontale vlak. Het is minder duidelijk hoe de auditief-visuele interac- ties saccades in complexere, maar, betwistbaar, natuurlijke omstandigheden beïnvloeden, waarbij stimuli van lage intensiteit in complexe achtergronden worden aangeboden en verdeeld worden over de 2D ruimte. Om dit probleem te bestuderen, maakten menselijke proefpersonen saccades naar puur visuele (V-saccades), puur auditieve (A-saccades), of ruimtelijk samenvallende auditief-visuele doelen (AV-saccades). In elke proef werd een doel met lage intensiteit ingebed binnen een complexe auditief-visuele achtergrond, en de proefpersonen werd toegestaan om meer dan 3 s naar het doel te zoeken en hun fovea te richten op 1 van de 24 mogelijke posities binnen het 2D oculomotorbereik. Wij variëerden systematisch de begintijden van de doelen en de intensiteit van het auditieve doel in verhouding tot de achtergrond [d.w.z., de signaal-ruis verhouding (van S/R)] om hun gevolgen voor zowel SRT als saccadische nauwkeurigheid te onderzoeken. De proefpersonen konden vaak het doel binnen één of twee saccades lokaliseren, maar in ongeveer 15% van de proeven produceerden zij scanpatronen, die uit vele saccades bestonden. De huidige studie meldt slechts de SRT en de nauwkeurigheid van de eerste saccade in elke proef. Bij alle proefpersonen hadden A-saccades kortere SRTs dan V-saccades, maar ze waren onnauwkeuriger dan V-saccades wanneer ze gericht waren op auditieve doelen met een lage S/R verhouding. AV-saccades waren minstens zo nauwkeurig als V-saccades maar hadden SRTs die typisch waren voor A-saccades. De eigenschappen van AV-saccades hingen systematisch af van zowel stimulustiming als van de S/R verhouding van het auditieve
doel. Vergeleken met unimodale A- en V-saccades, waren de verbeteringen in SRT en nauwkeurigheid van AV-saccades het grootst wanneer het visuele doel synchroon was met, of voorafging aan het auditieve doel, en wanneer de S/R verhouding van het auditieve doel het laagst was. Verder waren de verbeteringen in saccadenauwkeurigheid groter in elevatie dan in azimut. Een controleeexperiment toonde aan dat een gedeelte van de verbetering in SRT aan een waarschuwingseffect toe te schrijven zou kunnen zijn, maar dat de verbetering in saccadenauwkeurigheid afhing van de ruimtelijke ordening van de stimuli. Deze resultaten komen goed overeen met eerdere elektrofysiologische resultaten die werden verkregen uit metingen in de Colliculus Superior in de middenhersenen van verdoofde preparaten, en wij argumenteren dat zij multisensorische integratie van auditieve en visuele signalen in een complexe, quasi-natuurlijke omgeving aantonen. Een conceptueel model dat de Colliculus Superior omvat wordt voorgesteld om de waargenomen gegevens te verklaren.
BIBLIOGRAPHY


Gouras P (1967) The effects of light-adaptation on rod and cone receptive field organization
of monkey ganglion cells. J Physiol (Lond) 192: 747-760.


Bibliography


Bibliography


DANKWOORD

Vele mensen hebben bijgedragen aan de totstandkoming van dit boekje. Bedankt:

John: Menigmaal heb je luidkeels niet alleen mij maar zelfs ook de data kunnen overtuigen van jouw gelijk. Jouw gedrevenheid heeft mijn enthousiasme in de wetenschap enorm versterkt.

Joyce: Jip&Janneke, Pluk, Nijn, Bugs, Zingende Vis. Desondanks werd ons kamertje een stuk gezelliger door jouw aanwezigheid.

Huib en Sigrid: bedankt, niet alleen voor de prettige samenwerking, maar ook voor het meerdere malen afleiden van Bram, Paco en Orc.

Andere (ex-)collega’s: jullie darts ontwijken op de gang was altijd een goede workout. Als de wetenschap het toeliet, waren een gezamenlijke lunch in de zomer op het helikopterplatform/parkeerplaats/bouwplek/grasveld, een filmpje, of een terrasje altijd zo welkom afwisseling van het werk.

Doug, Brian, Andrew: Blame Canada. Thanks for the hospitality.

Een heleboel radertjes zorgen voor een goed functionerend apparaat. Bedankt Ton, Hans (apparatuur), Co (electrodes), Alex, Fred, Frans (dierverzorging & operaties), Ger, Gunther (software & computers).

De vele proefpersonen: als jullie al geen hoorafwijking hadden, is met veel moeite getracht om jullie die toch te geven. En als zo’n hoorprobleem al niet erg genoeg was, waren de sociale consequenties van het rondlopen met een stuk roze kauwgum in je oor vaak niet te overzien. Bedankt voor jullie geduld en vertrouwen.

Patrick: Je hebt je studie opgegeven en bent vrachtwagenchauffeur geworden, zodat je mijn studie kon betalen. Hoe kan ik jou ooit terugbetalen?

Pap en Mam: zelfs in het getto hebben jullie mij liefdevol kunnen opvoeden. Ook al konden jullie niet altijd mijn vele vragen beantwoorden, jullie hebben mij wel altijd gestimuleerd om zoveel mogelijk te leren. Bedankt voor Het Grote Waarom-boek dat ik op mijn tiende verjaardag kreeg (het begin van mijn wetenschappelijke carrière).

Waarschijnlijk ben ik velen vergeten in dit dankwoord. In ieder geval wil ik de volgende mensen NIET bedanken voor het totstandkomen van dit proefschrift:

De Gemert-crew + Stefan: als jullie me niet meesleepten op jullie rescue-operations/raids/secret missions om onschuldigen leeg te zuigen, de wereld te verderven, macht te veroveren of een dictatuur te beginnen, dan leidde jullie feestgedrag in de echte wereld me wel af van de perikelen rondom het proefschrift.

Gang 66: vaak kon ik na een drukke werkdag niet genieten van een ontspannende aflevering ST:TNG, DSN, of Voyager; GTST, Get the Picture en Boulevard waren vaak de winnaar van de machtsstrijd in de keuken. En als ik toe was aan een goede nachtrust, werd ik
op sleeptouw meegenomen naar jazz-café’s, death-metal concerten, goth-feesten, dance-events, festivals, gangfeesten en Dio-uitstapjes. Het is bijna onmogelijk, maar gelukkig is mijn gehoor niet achteruitgegaan.
