Chapter 1

c0005 A Brief Introduction to the Topic

st0020 1.1 TWO TASKS FOR THE AUDITORY SYSTEM

p0010 Whenever a detectable sound wave reaches our ears, the brain will try to assign meaning to the acoustic event. In a split second, the auditory system succeeds in *identifying* the nature of the sound source out of a virtually unlimited number of possibilities: is it noise (the wind, the rain, the sea, a sigh)? Was it perhaps a familiar or an unfamiliar human voice? Was it an animal vocalization? Maybe it was a car, some other man-made machine, a musical instrument, or an orchestra? Perhaps, the sound was caused by the ticking of an object (a fork?) against another object (a dinner plate?), etc.

p0015 At the same time, the auditory system *localizes* the sound source. But just like source identification, the seemingly simple localization task could refer to multiple possibilities: where is the sound source located in "external space," that is, in the world around us, through which we navigate? Or: where is the sound relative to the ears or head? Where is it relative to other landmarks in the environment? Surprisingly, as we will see later, the brain seems to be particularly interested in determining where the sound source is located *relative to your eyes*! Thus, the auditory system has evolved to perform the following major tasks on the acoustic input:

t0010

Auditory task	It answers
Identification	What?
Localization	Where?

p0020 Obviously, the ability to rapidly identify and localize sound sources is vital for survival. In any case, it was crucial when in a not too distant past, we as hominids, had to struggle fiercely to stay alive, as food sources (good for us) and predators (very bad for us) had to be identified and localized as fast as possible. It is therefore not surprising that throughout evolution, the auditory systems of virtually all the animal species have developed dedicated neural circuits to efficiently and accurately solve identification and localization tasks.

p0025 Although simply formulated, these tasks are in fact astonishingly ~~difficult~~ complex to perform. Current technological advances, despite the tremendous increase in computer speed and memory storage over the last decades, are still not able to execute these tasks with the same accuracy, speed, flexibility, and efficiency as biological

The Auditory System and Human Sound-Localization Behavior. http://dx.doi.org/10.1016/B978-0-12-801529-2.00001-5

ISBN: 978-0-12-801529-2; PII: B978-0-12-801529-2.00001-5; Author: VANOPSTALENGLISH; Document ID: 00001; Chapter ID: c0005

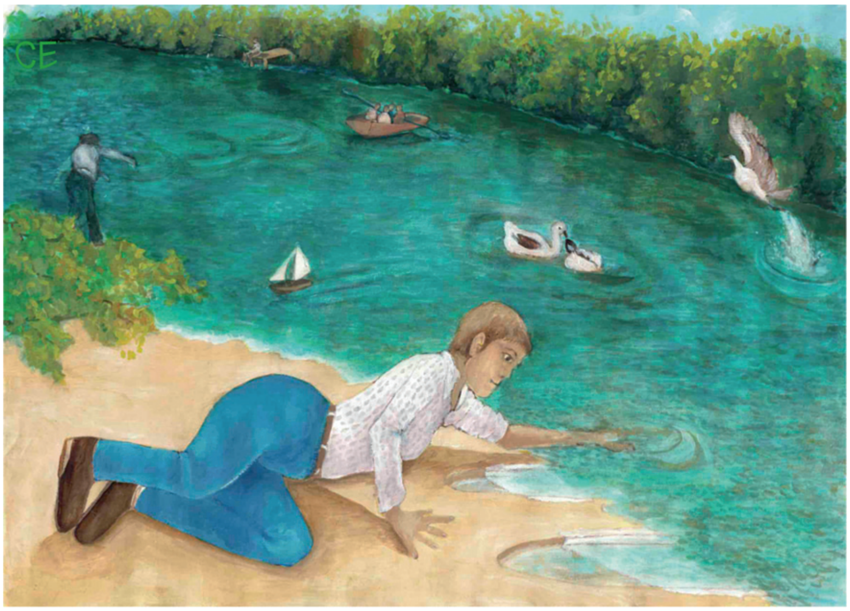
auditory systems. Indeed, the auditory system seems to carry a bag loaded with sophisticated tricks (neural algorithms) in order to do what it is supposed to.

p0030 How do we know all this, and how do we study, understand, and model the different aspects of sound processing in human and animal brains? Can we learn something essential from this, and perhaps implement this knowledge in future sound-recognition technologies, including healthcare applications, such as improved hearing aids and implants? This monograph forms my personal account of an exciting line of research on sound localization behavior in humans and nonhuman primates, which has kept me busy for well over 20 years, and is likely to keep me busy for the next decade as well. Some of the questions raised here form the central topic of this book.

p0035 To persuade the reader that sound processing in the brain is an interesting topic, wholly worthy of study, and above all, intellectually challenging and rewarding, here I would like to briefly illustrate, in a very general way, the sound–source identification problem as it presents itself to the auditory system.

st0025 1.2 AN ILL-POSED PROBLEM

p0040 The fundamental problem faced by audition has been particularly nicely formulated and illustrated by Albert Bregman (1990) in his "*man-at-the-lake*" analogy (Fig. 1.1). Imagine this guy, lying at the shore of a large lake. Although the story



f0010 FIGURE 1.1 **Our hero at the lake is allowed to only look at the movements of the two thin sheets to identify the sources that caused the water waves on the lake.** *(~~Courtesy:~~ Carmen* Artwork: courtesy of *Espadinha.)*

doesn't tell, we may assume that the man is either deaf, or deafened by earplugs, so he can't hear. He just dug two narrow, parallel channels that fill themselves with water from the lake, and has then draped and fastened two thin plastic sheets onto the water surface of each channel.

p0045 Then the game starts: the man is allowed to only look at the up and down movements of the two sheets. Meanwhile, the lake's water surface is continuously perturbed by all kinds of objects (toy boats, swimming and splattering kids, dropped stones, landing ducks and geese, wind, etc.) that each cause their own specific water waves to travel at some fixed speed, from some direction, along the surface of the lake. Of course, the man doesn't know all this, since he is not allowed to look at the lake, and he can't hear either.

p0050 However, a small part of these traveling waves will at some moment, enter each of his two channels. The challenge for our guy is to decide, only on the basis of the motion patterns of the two sheets (which he may assume to oscillate without any energy loss with the water motion, and which he may analyze in every possible way), what exactly is happening on the lake.

p0055 Clearly, the two channels symbolize our ear canals, while the lake is the air, set in vibration by different sound sources; the two plastic sheets represent our eardrums. The "guy" in this story represents the homunculus within our auditory system that can only "look" at the one-dimensional temporal vibrations of the two eardrums to analyze acoustic input.

p0060 One doesn't have to be very imaginative to recognize that this is a formidable, if not an *unsolvable*, challenge for the auditory system! Indeed, mathematics tells us that such a problem is in fact, *ill posed* (Kabanikhin, 2008). This means there is no unique solution to the problem as in reality there are infinitely many mathematically valid solutions! How can we appreciate the severity of this problem?

p0065 Recall the "Hitchhiker's Guide to the Galaxy" by Douglas Adams, in which the earthlings have been told the answer (which is "42"), but have to guess *the correct question* in order to save Planet Earth from total destruction. Was the question: how old is your wife's sister? How many hairs does your adolescent son have on his chest? How much is 6 × 7, or 43.5 × (42/43.5)? How many presidents had governed the USA when G.W. Bush took office in 2001? Clearly, infinitely many questions can be formulated, all with the same correct answer: "42."

p0070 A similar mind-boggling problem bugs the auditory system. Think about it: the sound wave that reaches the ears is a linear superposition of all the sound sources in the environment. Suppose there are N such sources, and that the time-varying sound pressure for each source is $s_n(t-\{\tau_n\})$, for $t \in \{\tau_n\}$, where $\{\tau_n\}$ are the on- and offset timings of sound n. Each sound occupies a certain location in space, which we here measure in relation to the ears. For simplicity, we denote a location by its two directional angles (ignoring distance), here indicated by H (for horizontal), and V (for vertical). As we will see in later chapters of this book, different locations/directions of given sound sources lead to systematically different acoustic patterns at the two ears. As a result, a sound source is uniquely described by its temporal variations (and hence, its spectral content, which is given by the signal's set of constituent frequencies, $\{\omega_n\}$) and by its directional information (the *what* and *where*): $s_n(t-\{\tau_n,\}, \{\omega_n\}, H_n, V_n)$. The resulting sound–pressure waves seen at the left (L) and right (R) eardrums are then given by:

$$S^{\mathrm{L,R}}(t) = \sum_{n=1}^{N} s_n(t-\{\tau_n\},\{\omega_n\}, H_n, V_n) \tag{1.1}$$

p0075 We will see later that the left- and right-ear sound patterns for a given sound, differ from each other in a systematic way, whenever the horizontal position of the sound source moves away from the midsagittal plane.[a] The acoustic problem for our auditory homunculus therefore boils down to:

b0010 *Find $\{\tau_n\}$, $\{\omega_n\}$, and H_n, V_n for all* n *sources in the acoustic scene.*

p0085 For natural acoustic environments we can, however, construct the following truth table for the general situation of the problem described by Eq. (1.1).

t0015

Acoustic feature	A priori knowledge
Number of sources, N	Unknown
Spectral content of individual sources, $\{\omega_n\}$	Unknown
On- and offset times of individual sources, $\{\tau_n\}$	Unknown
Locations of individual sources, H_n, V_n	Unknown

p0090 That is, when the auditory system has no knowledge about the constituent sources that make up the total sound wave, there is absolutely *no way* that it can uniquely segregate individual sources from the acoustic mixture described by Eq. (1.1). The problem is even more severe, since many sound sources in the environment will have considerable spectral- and temporal overlap. They may start and end in roughly similar time windows, and their spectral bandwidths may be quite similar too. The problem even remains ill posed if we would know for sure that there are only two sources in the scene (ie, if $N = 2$)! Indeed: 42 = 21 + 21 = 43.5–1.5 = 3 + 39 = … In other words, there is no hope for us, earthlings, to prevent ultimate disaster or is there?

fn0010 [a] Later we will see that a sound-source displaced in the horizontal plane yields binaural differences in the timings of sound on- and offsets: $\{\tau_n\}^{\mathrm{L,R}}$, and in the high-frequency content of the acoustic spectrum: $\{\omega_n\}^{\mathrm{L,R}}$. In this general introduction, we will not delve into this matter further.

st0030 1.3 DEALING WITH ILL-POSED PROBLEMS

p0095 It turns out that in fact the auditory system does a remarkably good job at segregating, identifying, and following particular sound sources in the environment, and localizing them with astonishing accuracy. Yet, the problem of Eq. (1.1) remains ill posed and doesn't just disappear. To understand its successful performance, the auditory system must somehow rely on more practical, useful strategies, than on trying to solve unsolvable problems.

b0015 *Strategy: Since the sound-identification problem cannot be solved, deal with it to the best of your abilities!*

p0105 A useful strategy to deal with ill-posed problems like Eq. (1.1) is to make clever assumptions regarding potential solutions, and to use these assumptions as efficiently as possible. In modern neuroscience theories (known under the collective name of Bayesian models) such assumptions are called "*priors*," as they refer to learned and stored probabilities of stimulus–response properties that the system may use as prior information, that is, advance knowledge, to deal (not "solve") with perceptual tasks. The idea behind these theories, which are essentially probabilistic, rather than deterministic in nature, is that the brain (ie, the auditory system) generates a response that can be considered its best *statistical estimate* for the current solution. In a (statistically) *optimal* system, such a response will, on average, have the smallest systematic error (highest accuracy) and variability (highest precision), given the uncertain and ambiguous sensory evidence of Eq. (1.1), and the potential advance prior knowledge (estimate, expectation) stored in the system. Now that is quite a mouth full, but we will come back to this topic in more detail later in this book. It here suffices to state that there is good evidence that our sensory and motor systems, the auditory system included, seem to operate according to such statistical (nearly) optimal principles.

p0110 In the case of sound sources, these assumptions could be based on certain relevant properties of the physical world, which have been learned through experience by interacting with the environment in many different ways: navigating through the environment, perceiving the same objects in the environment through our different sensory systems, orienting to objects in the environment, particular properties of familiar objects, etc. For example, our brains may have learned that in natural *physical* environments there can be only one object (read: sound source) at any given point in space at a certain time. Note that this statement excludes the possibility of artificial, technical devices, such as loudspeakers, or headphones, which can clearly violate this natural requirement! I nonetheless believe that all auditory systems have evolved to use this kind of natural logic in order to make sense of unknown acoustic environments.

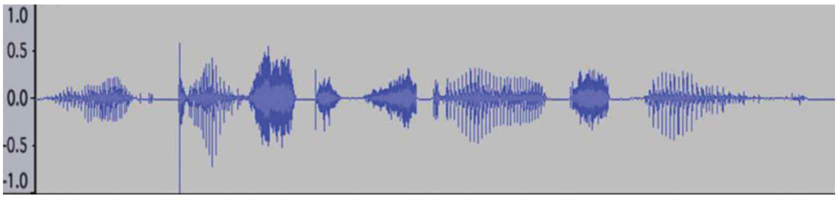
p0115 This is a very important point, which immediately touches on the central topic of this book, namely that sound processing and sound localization are *active* processes that involve planned behaviors! The idea is that by doing so, the system becomes better and better at dealing with ill-posed problems like formulated by Eq. (1.1).

st0035 1.4 SPECTRAL REPRESENTATION AND SOURCE PRIORS

p0120 The vibrations of the tympanic membrane, as described by Eq. (1.1), are transmitted via the middle-ear bones to the stapes at the entrance of the inner ear (the cochlea) (Purves and Augustine, 2012). Fig. 1.2 illustrates a typical, complex sound–source signal, produced by a male human voice, uttering the sentence "*Your test starts now*" (in phonetic script: "jʊə tɛst sta:rts naʊ"), which took about 2 s to complete.

p0125 In the cochlea, the organ of Corti performs the first nontrivial sensory transformation on this sound–pressure wave: from a pure *temporal* in- and outward movement of the oval window to a combined *spectral–temporal* representation along the length (coordinate x, in mm) of the cochlea. We will see in chapter: The Cochlea, that as a result of the intriguing micromechanics in the cochlea, specifically the position-dependent resonance properties of the basilar membrane (BM), in combination with local nonlinear feedback through the function of outer hair cells (OHC), the temporal sound–pressure signal $S(t)$, induces a transverse traveling wave along the BM, the amplitude of which reaches a sharply defined maximum at a frequency-specific location. Through this biomechanical mechanism, sound frequencies are topographically represented along the BM. In this *tonotopic representation*, frequency f, is mapped roughly logarithmically along the BM: $x = -4\log_2 (f/f_{\max})$ mm, with $f_{\max}$ the highest audible frequency at $x = 0$ (see chapter: The Cochlea). In cases where the sound is a single pure tone, that is, $S(t) = S_0\sin(2\pi f_0 t)$, high-frequency tones (in normal-hearing humans up to about 16–20 kHz) yield their maximum near the base ($x < 10$ mm) of the BM, low-frequency tones (between 40 and 100 Hz) vibrate maximally at the BM apex ($x \sim 30$–35 mm), while midrange frequencies (2–6 kHz) are encountered at more central locations.

p0130 Thus, evolution has figured out a nice way to extend the temporal representation of the sound mixture of Eq. (1.1) with an additional, now explicitly

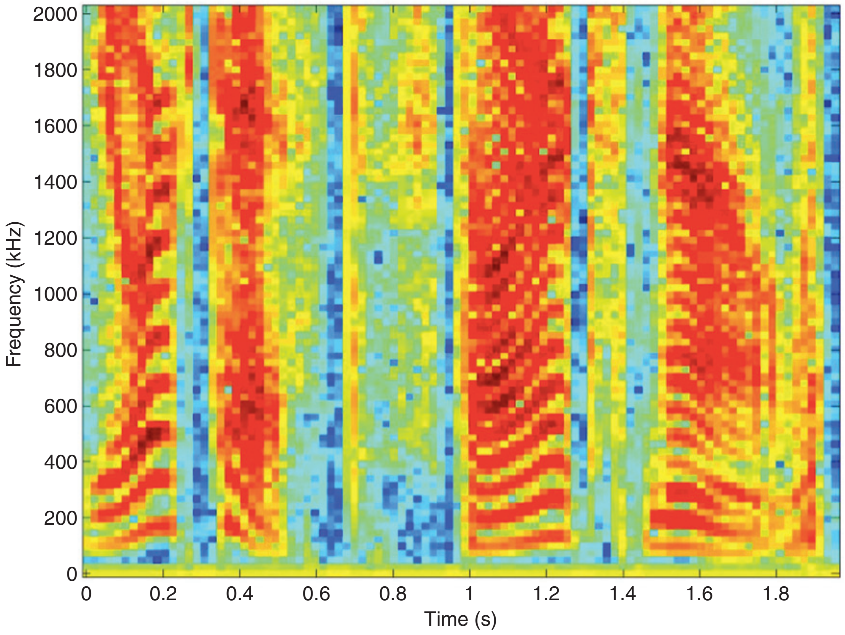


f0015 **FIGURE 1.2 Temporal vibration pattern of the complex sound–pressure wave, when a male human voice utters the sentence: "Your Test Starts Now."**

represented dimension (the sound's spectrum), which could further help the auditory system in its analysis of the acoustic scene.

p0135 Indeed, how the auditory system may be utilizing this additional representation is illustrated in the spectral–temporal representation (spectrogram) of the male utterance in Fig. 1.3. The features that make up the spectrogram suggest that the auditory system could make some useful prior assumptions that relate to the spectral–temporal structure of sound sources themselves. For instance, many potentially interesting sounds, like the vocalizations from fellow humans, or from a prey or a predator, are caused by mechanical vibrations of vocal chords or strings, and therefore unavoidably contain regular discrete harmonic complexes in their spectra. The frequencies in such vibrational spectra are thus related by: $\{\omega_n\} = \{n\omega_0\}$ with $n = 1, 2, 3, \cdots$, and where ω_0 is the fundamental (lowest) frequency in the sound spectrum.

p0140 Thus, the sonogram of Fig. 1.3 would allow rapid identification of some unique acoustic patterns, which are not obvious at all in the temporal signal of Fig. 1.2. Indeed, the distinct harmonic complexes in the voice (containing between 7 and 10 tones with a spectral interval of approximately 100 Hz, and a fundamental frequency of about 100 Hz) can be easily identified by visual



f0020 **FIGURE 1.3 Spectral–temporal representation (or sonogram) of the sound in Fig. 1.2.** Time runs along the abscissa (from 0 to 2.0 s), frequency along the ordinate (represented here on a linear scale from 0 to 2000 Hz to highlight the harmonic regularities better). Color represents amplitude (red: positive; blue: negative) of the frequency component. Any time point (a vertical line) relates to the instantaneous vibrational pattern of the BM.

ISBN: 978-0-12-801529-2; PII: B978-0-12-801529-2.00001-5; Author: VANOPSTALENGLISH; Document ID: 00001; Chapter ID: c0005

inspection in the time windows 0–0.2 s (“ʊə”), 1.0–1.25 s (“*a*:”) and 1.45–1.75 s (“aʊ”) in the 0–1000 Hz range. A second interesting feature to note is that all tones in the harmonic complexes start and end at the same time, and sweep synchronously upward and downward in frequency, as the voice pronounces the “ʊə” “*a*:” and “aʊ” vowels.

p0145 It is easy to imagine that joint synchronous movements in the spectral–temporal domain will provide strong cues to the auditory system for wanting to group them as a single acoustic object. Why? Obviously, because of mere statistics: it is extremely *unlikely* that multiple, independent acoustic sources will be precisely synchronized in their on- and offsets (time), and follow joint complex movements in the frequency domain at fixed spectral intervals. On the other hand, for a single sound source it is extremely *likely* (and even a prerequisite!) that tight synchrony and spectral harmony both occur together.

p0150 So, even though there is no unique exact solution for the general auditory problem of Eq. (1.1), sensible prior assumptions may readily discard many potential mathematical solutions, for the simple reason that they are nonphysical, extremely unlikely, or do not fit task-related expectations

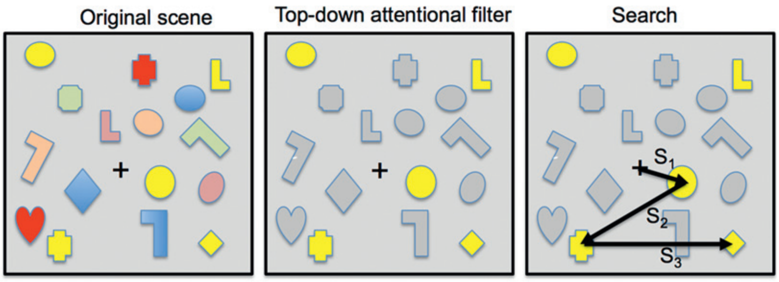
b0020 *The auditory system deals with ill-posed problems by making sensible statistical assumptions about the world, and about potential target sounds.*

st0040 1.5 TOP–DOWN SELECTIVE FILTERING

p0160 From the above, it could follow that the auditory system may not be interested at all in finding “the” solution of the *N*-sources problem, since it is impossible to identify them anyway. In fact, the truth table for the general auditory identification problem has in reality become irrelevant: all acoustic features that turn out to be “unknown” could possibly be replaced by: *Don’t know, and don’t care.*

p0165 Instead, we could assume that the auditory system would prefer to make educated prior guesses about the most important, task-relevant, regions within the acoustic scene. Such a mechanism would call for *selection*, *based on task-specific prior information*. It invokes top–down decision mechanisms that set priorities on the acoustic task at hand. Which sources are we interested in at this moment, and for this particular task? What do we know about their properties (prior information)? What do we consider as “target,” and what as “distracter,” or “background”? Once the system has defined and set its priority list (the “task”), it can efficiently start its probabilistic search in the acoustic input of Eq. (1.1). The task constraints may thus act as clever “filters” on the spectral–temporal input, which suppress those spectral–temporal regions in the input that are considered too remote from the target.

p0170 There may exist an interesting analogy to the way in which *vision* is thought to process information (Purves and Augustine, 2012). The visual system is thought to use an attentional “filter,” that in combination with a “covert”



f0025 **FIGURE 1.4** **Search in a cluttered scene with many competing objects.** The task is to find the yellow diamond. The fovea fixates the small cross. A top–down attentional filter enhances visual responses to yellow features in the scene, creating a "yellowness" saliency map. All different-colored objects become less salient. Saccadic eye movements are only directed to the conspicuous yellow targets in this saliency map. (A) Original scene; (B) top–down attentional filter; and (C) search.

(ie, attentional, mental), "spotlight," and "overt" (observable) fovea[b] efficiently deals with complex visual scenes. This strategy allows vision to quickly identify and localize a particular yellow object (suppose that this is the task) in a complex urban scene, like a busy street in which many advertisement boards, cars, and people all compete for attention.

p0175 Fig. 1.4 illustrates this idea for a simpler laboratory test, where the task is to find the yellow diamond among a set of different shapes with different colors. The attentional filter first "highlights/boosts" everything yellow in the scene (Fig. 1.4B), while the covert spotlight selects the next goal for foveation with the eyes (Fig. 1.4C). The sequence of fast (overt) saccadic eye movements across the scene (the arrows), will then quickly direct the fovea from one conspicuous yellow object to the other, until the visual system has identified the target (Koch and Ullman, 1985; Itty and Koch, 2000).

p0180 Saccades have to be executed one after the other, which is typically done with an intersaccadic interval of about 150–250 ms, so that our brains program and generate about three to four saccades per second. As a result, this visual search strategy may appear to be a time-consuming "serial" search, but because of the preprocessing by the attentional filter, which effectively acts as "parallel filtering," it actually becomes a "clever" serial search! The fovea will hardly ever land, for example, on irrelevant red or blue objects, so that the system avoids wasting valuable visual processing time. Moreover, use of additional forms of prior information (not only "yellowness," but perhaps other visual features, like orientation, relative size, shape, etc.) can in principle render the system even more efficient.

p0185 A second interesting property of saccadic eye movements, which could further increase the efficiency of goal-directed visual search, is the use of

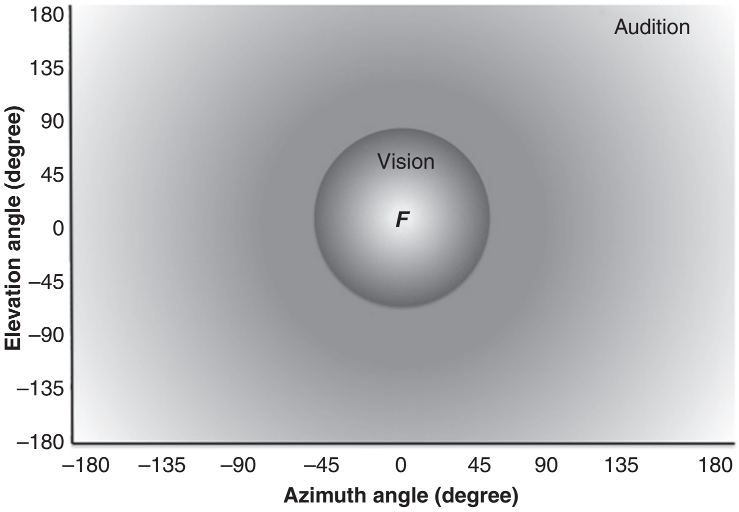
fn0015 [b] The fovea is the small area on the retina with a high spatial resolution (spanning less than a degree of visual angle), which is crucial for perceiving color and for analyzing fine details for visual object recognition. A saccade directs the fovea as fast as possible (at speeds up to 600–700 degrees/s) to a selected target in the periphery.

a mechanism called "*inhibition of return*" (IOR). This process prevents (or makes it more unlikely) that the next saccadic eye movement (eg, saccade S_3 in Fig. 1.4) revisits spatial locations that have just been explored (eg, the yellow oval, which was already foveated by saccade S_1) (Posner and Cohen, 1984; Klein, 2000; Wang and Klein, 2010).

p0190 The selection process just described illustrates the important idea that "seeing" (visual perception) and "looking" (exploring the visual field with eye movements) need to occur together, as two sides of the same coin. This idea entails that vision is an active process (*active vision*), which involves dynamic, adaptive filtering of the input, the use of prior, task-relevant information to select potential targets from a myriad of possibilities, and subsequent active exploration by the visuomotor system, which in turn involves clever decision making, and speed-accuracy trade off through goal-directed and fast orienting responses across the selected environment.

p0195 We will see that in the same realm the auditory system may employ a strategy of "*active hearing*" to perform the sound localization and identification task. By programming and generating combined eye–head movements to target sounds in the environment, results show that it can find the sound source within a few hundred milliseconds, with a reaction time, accuracy, and precision that often surpasses that of vision.

p0200 An important reason for the use of gaze control in both systems is that the spatial resolutions and spatial ranges for vision and audition are markedly



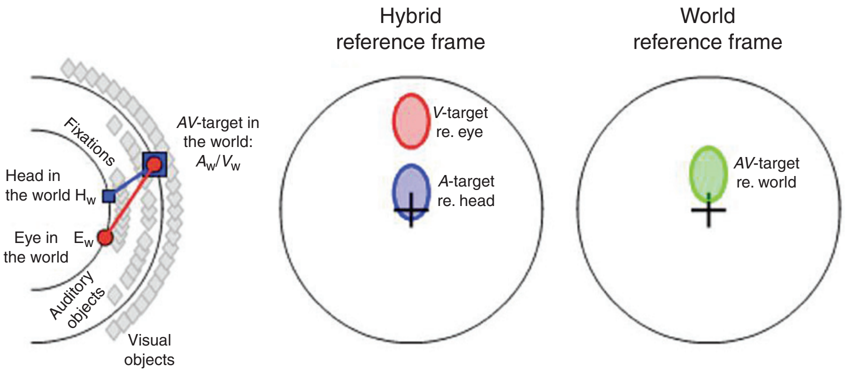
f0030 **FIGURE 1.5 Vision and audition have very different spatial resolutions and spatial perceptual ranges.** Whereas vision is restricted to a high resolution within a narrow (<10 degree) range around the fovea (*F*), audition has a lower, yet roughly constant spatial resolution across the entire directional space (−180 to +180 degree in both azimuth and elevation). Azimuth: angle in horizontal plane; elevation: angle in vertical plane. [*A*,*E*] = [0,0] = straight ahead; [180,180] = back; [0,90] = zenith.

different (Fig. 1.5). Whereas vision has a very high spatial resolution over a very limited viewing angle (ie, a parafoveal range of a few degrees), which decreases rapidly toward the visual periphery of about 45 degree and beyond, audition has a lower spatial resolution than vision, but an omnidirectional perceptual range (it covers the full sphere in azimuth and elevation). As a consequence, any peripheral target beyond the visual perceptual range, or in a range with very low spatial resolution, has to be acquired by the auditory system whenever possible! Fig. 1.5 suggests that the auditory system may in fact be the primary sensory system-driving eye–head orienting responses.

st0045 1.6 AUDIOVISUAL INTEGRATION

p0205 Vision and audition both use gaze shifts (gaze is the direction in which the fovea points in external space) as a natural response to rapidly look at (ie, localize) targets that were perceived in the periphery of sensory space. However, as in natural environments, objects tend to emit both visual and auditory signals it is of crucial importance that the two sensory systems collaborate intensively to perform the localization task as fast and as accurately as possible. There are many interesting and nontrivial issues involved in this integration, all of which will be dealt with in this book. Here, I will only briefly illustrate one of them, which is the need to bring the two sensory systems into "*register*." As the eyes can rotate freely within the head, and the head can independently rotate on the neck and trunk, the typical situation will be that the head (your nose) will point in a different direction in space than the fovea (gaze line) when a sudden audiovisual target appears in the periphery.

p0210 Fig. 1.6 describes the situation of a single target, perceived by the auditory and visual systems, when the nose is pointing upward, while the eyes are



f0035 **FIGURE 1.6 A single audiovisual object is represented by different coordinates for the auditory and visual signals, whenever the eyes and head are not in alignment.** V_E, visual target relative to the eye; and A_H, auditory target relative to the head. To bring the disjoint sensory representations into a common reference frame, the brain should perform a sensorimotor coordinate transformation that uses motor signals about the misalignment of eyes and head.

looking straight ahead. The sound-localization system perceives the acoustic input from the target slightly upward with respect to the nose, while the visual location of the same target is projected more upward on the retina. Clearly, the two sensory locations are not in register, and could in principle have arisen from two different objects. So how can the brain decide whether or not the two sensory inputs came from the same location in space, and hence, according to the prior about natural spaces, from a single object? To that end, it should use the difference in alignment between the two sensory systems (ie, either the head orientation on the trunk, the orientation of the eyes in the head, or both) to recalibrate the sensory signals. For example, if the target is to be referenced with respect to the head, the visual retinal signal should be combined with the motor signal of the eye-in-head orientation, E_H: $V_H = V_E + E_H$. Alternatively, the auditory head-referenced signal could be combined with the motor signal about eye-in-head orientation, like: $A_E = A_H - E_H$. Note that both sensory signals could also be represented in a more general spatial reference frame, for example, relative to the body, which would require the use of motor information from both the head and the eyes. In that case:

$$A_S = A_H + aE_H + bH_S \text{ and } V_S = V_E + cE_H + dH_S. \qquad (1.2)$$

p0215 As an exercise, the reader may verify that in the case of a single audiovisual target, the coefficients [a,b] and [c,d] of Eq. (1.2) are different for the visual and auditory modalities. In other words, these coordinate transformations are highly context and modality specific. Yet, audiovisual integration is crucial for adequate identification and localization of stimuli in the environment, and hence they need to be carried out fast with accuracy and precision.

st0050 1.7 HOW TO USE THIS BOOK

p0220 This monograph may serve as a full-semester course on the application of experimental and theoretical psychophysics to the human auditory and audiomotor systems. Much of the material covered in this book has been used in my own university courses, which were aimed at neurobiology, cognitive neuroscience, and physics bachelor students with a keen interest in quantitative understanding of the mechanisms of perception and sensorimotor integration. We believe this book will also appeal to biomedical engineering students, and to any natural scientist or engineer (biomedical, robotics, virtual reality, etc.) interested in the workings of the human brain. A basic understanding of elementary mathematics (calculus, standard differential equations, basic integration, linear algebra, and some vector calculus) and physics (mechanics) is needed, but sections that would require more advanced technical skills will be marked (*) as such. These sections can be skipped without losing touch with the narrative of the book.

p0225 Divided over 14 chapters, the reader will encounter a variety of related topics that range from the physical acoustic input stimulus of the system (see chapter: The Nature of Sound), to top–down neural mechanisms that may underlie audiovisual integration in goal directed–gaze orienting behavior (see chapter: Multisensory Integration), and in hearing impairments (see chapter: Impaired Hearing and Sound Localization).

p0230 The book covers topics ranging from the auditory periphery to central mechanisms of sound encoding and decoding, as well as mechanisms that underlie sound-localization encoding and the planning, and generation of coordinated eye- and head movements to sound sources. The book thus consists of roughly two equal-sized parts that could in principle be taught and studied to a large extent independently as two half-semester or trimester courses. In Chapters 2–7 (on *acoustics, theoretical concepts, and the neurobiology of audition*), the reader encounters relevant topics in acoustics, from the physical principles underlying the propagation of sound waves, to the physical–mathematical modeling of cochlear hydrodynamics, including active and nonlinear properties of cochlear OHC (see chapters: The Nature of Sound; The Cochlea). As a general background, the first part also provides some in-depth coverage of systems theory, linear as well as nonlinear, and Fourier analysis (including Laplace transforms, see chapters Linear Systems Analysis and Nonlinear Systems). The concepts from these two chapters are also needed in case one decides to study only the second half of the book. The first seven chapters also present neurobiological underpinnings of the auditory system, by describing some important neural encoding principles (and their analyses) observed in the activity of neurons at different stages in the ascending auditory pathway of mammals, like spike timing, rate coding, phase locking, spectral tuning, nonlinear interactions, monaural and binaural interactions, spatial sensitivity, and spectral–temporal receptive fields (see chapters: The Auditory Nerve; Sound-Localization Cues; and Assessing Auditory Spatial Performance).

p0235 The second half of the book (see chapters 8–14, on "*gaze control*, *sound-localization behavior, spatial updating and plasticity, and audiovisual integration*") deals with the behavioral aspects of sound-evoked orienting. These chapters are built on the idea that sound localization is an active process, in which the generation of rapid orienting responses of the eyes and head are needed to close the action-perception cycle. Additional theoretical background is given on psychometric methods and signal-detection theory (see chapter: The Gaze Orienting System), including the use of eye and eye–head movements as accurate, absolute, and fast sound-localization probes. We will also describe the neurophysiological underpinnings of eye–head gaze control in more detail by focusing on the involvement and modeling of the midbrain (see chapter: The Midbrain Colliculus). From the basic (static) sound-localization responses to single sound sources in the typical silent laboratory environment, we then proceed to the need for dynamic sensorimotor feedback control under more realistic auditory localization tasks, and how the required coordinate transformations may be represented in the activity patterns of neural populations (see chapter: Coordinate Transformations). In Chapter: Sound-Localization Plasticity, we describe the experimental evidence for some remarkable plasticity in the human auditory localization system, which extends well into adulthood. To some extent this topic returns in the chapter on impaired hearing (see chapter: Impaired Hearing and Sound Localization). In Chapter: Multisensory Integration, we will deal with the interesting interactions between the visual and auditory systems, and highlight some of the modern theoretical concepts of Bayesian inference that could underlie many of the phenomena observed in sound-localization behavior, and in audiovisual integration.

p0240 *To the student*: depending on your mathematics and physics background, this book can be studied either in its entirety, or by skipping the sections marked (*), which require advanced understanding of physics and/or mathematics. The same holds for the exercises that are given at the end of each chapter. Additional material, like computer scripts needed to run some of the simulations (in Matlab), or some extra hints that may help solve some of these exercises, are provided on the website, http://www.mbfys.ru.nl/~johnvo/LocalizationBook.html

p0245 *To the lecturer*: as the book is roughly divided into two equal parts that cover complementary fields of study (the auditory system and sound-localization behavior, respectively), it can be used as a full-semester (14 weeks) textbook for a course on human auditory psychophysics, in which each chapter can be covered in a typical week of 2 h of lectures, supplemented by two to three practical hours for the exercises and computer simulations. Alternatively, the book can be used in two separate semisemester (7–8-week quarters) or two trimester courses, which deal with the first eight chapters and with Chapters 9–14, respectively. Different combinations may be considered too. For example, the chapters on sound-localization plasticity (see chapter: Sound Localization Behavior and Plasticity) and on impaired hearing (see chapter: The Auditory System and Human Sound-Localization Behavior) could be included in the first course.

p0250 Physics students at the bachelor level (ie, in the second or third year of their curriculum), or engineering and informatics students with an equivalent math background, should be able to study all the material in this book. Students from the life sciences (eg, neurobiology, medical biology), or cognitive neuroscience and functional psychology, may want to skip the sections (*) that require some more advanced technical knowledge of math and/or physics.

p0255 Students are particularly encouraged to make the exercises and run the computer simulations (in Matlab) that are provided at the end of each chapter. Lecturers can obtain the fully worked-out problems of the exercises from the author upon request (j.vanopstal@donders.ru.nl).

p0260 All figures in the book are available for your PowerPoint presentations on the book's website, at http://www.mbfys.ru.nl/~johnvo/LocalizationBook.html

p0265 As a final remark, the author will be extremely grateful for any constructive feedback, interesting additional exercises, errors, etc. that could help to improve future editions of this book. Your contributions will of course be fully acknowledged.

st0055 1.8 OVERVIEW

p0270 *Chapter*: *The Nature of Sound* provides the physical basis of the sensory input stimulus to the auditory system: the sound wave, which is a longitudinal mechanical perturbation of the vibration of molecules that propagates at high velocity through the medium. From first principles of thermodynamics we deduce the propagation speed of monochromatic sound waves through air. We then look at the mechanical wave equation (both in homogeneous and in inhomogeneous media), the dispersion relation between the sound frequency and its wavelength, and at the transmission of acoustic power through the medium. We discuss reflection and transmission at the transition boundary of different media, which leads to the concept of acoustic impedance. We then introduce Fourier analysis of periodic signals, and finally define the phase velocity and group velocity of modulated acoustic signals.

p0275 *Chapter*: *Linear Systems Analysis* gives a thorough introduction into the field of linear black box analysis, a mathematical modeling technique that has become quite popular in the engineering sciences, but has also seen many useful applications in computational neuroscience and psychophysics, and in auditory science in particular. Starting from the superposition principle it is shown that this simple idea leads to a series of profound consequences: the concept of the system's impulse response as the system's memory, and the convolution integral to predict a system's response to arbitrary inputs. We apply Fourier analysis to extend the time-domain analysis of linear systems to the frequency domain, meanwhile introducing the system's transfer characteristic, and the Bode plot as a useful graphical tool to analyze its properties. We then introduce the Laplace transform as a convenient general method for dealing with a broad range of linear systems and signals in a semiintuitive way. We end the chapter with the introduction of the Gaussian white noise signal as a stimulus, and the auto- and cross-correlation functions of signals.

p0280 *Chapter*: *Nonlinear Systems* extends the linear black box theory of Chapter 3 to smooth nonlinear systems by introducing the Volterra and Wiener functional approaches to systems identification. In these nonlinear descriptions the system's response is described as a series representation of higher-order (ie, nonlinear) contributions. The core descriptors in these approaches are the system *kernels*, which are a natural extension of the impulse–response kernel for linear systems. Although the Volterra approach is a general systems analysis technique that may be applied to arbitrary input stimuli, it is not mathematically feasible to identify the underlying system kernels in a black box input–output measurement approach, because they are mutually dependent. This property has prompted researchers to develop a method that allows one to independently extract the nonlinear system kernels. The Wiener series, which is based on the autocorrelation properties of Gaussian White Noise as the system's input, is the most commonly used technique. Interestingly, a three-layered feed-forward artificial neural network with an appropriate input signal representation turns out to be mathematically equivalent to the complete Volterra series. It thus provides an alternative model description for the nonlinear system at hand. As a consequence, it is possible to independently extract the corresponding Volterra kernels from the trained synaptic weights in the network.

p0285 *Chapter*: *The Cochlea* applies several of the concepts described in the previous chapters to build and discuss a physical and physiologically realistic model of the cochlea. The chapter starts with the idea of tonotopy in the cochlea, and elaborates on the interesting analogy with the physics of water waves to the traveling wave along the cochlea's basilar membrane. It then describes the linear model based on the Nobel prize-winning work of Géorg von Békésy, as further elaborated by Joseph Zwislocki, and proceeds with the modern work on modeling the role of the electromotile response of OHC in combination with the tectorial membrane to explain the high sensitivity and huge dynamic range of the auditory system.

p0290 *Chapter*: *The Auditory Nerve* discusses the first neural stages in the auditory processing chain. Auditory nerve recordings indicate that responses faithfully reflect the motion patterns of the basilar membrane. Reverse correlation analysis links the tuning of low-frequency auditory nerve fibers to sharply tuned gammatone filters. Using a clever stimulation method the nerve recordings of single units can be used to fully reconstruct the amplitude and phase characteristics of the BM motion. The chapter concludes with a pragmatic model of the auditory periphery.

p0295 *Chapter*: *Sound-Localization Cues* describes the implicit localization cues that are available to the human auditory system in the horizontal plane (azimuth angle), the medial plane (elevation angle, and front/back), as well as sound–source distance. Binaural differences in time and intensity encode the sound–source azimuth angle for low- and high frequency sounds, respectively. The elevation angle is determined on the basis of complex spectral-shape cues that arise by acoustic reflections and diffraction within the pinna. Models that explain these computational processes are discussed. The chapter also describes the neural mechanisms for these cues that have been identified in animal studies.

p0300 *Chapter*: *Assessing Auditory Spatial Performance* discusses different methods that are used to study spatial hearing. The methods range from the mere detection of auditory stimuli, to left/right lateralization or discrimination of stimuli with respect to the center of the head to the localization of sounds in absolute external space. As the latter is measured with continuous pointing methods, the former measurements typically yield binary responses, and rely on the concepts of signal detection theory. We argue that the visual system is a natural ally of the sound-localization system, and therefore eye movements (gaze shifts) provide a natural, fast, and accurate pointer to study sound-localization behavior. We finally dwell on the method to virtual acoustics to study naturalistic sound-localization behavior with headphones, and which gives the flexibility to selectively manipulate the acoustic localization cues.

p0305 *Chapter*: *The Gaze Orienting System* describes the saccadic eye-movement system and the control of combined eye–head gaze-orienting movements. The latter is implicated in natural sound-localization behavior. Because of their inhomogeneous retina, foveate animals optimize their orienting behavior through rapid, accurate, and precise saccadic eye movements. The saccadic system is characterized by a prominent kinematic nonlinearity, which is implemented as a central vectorial pulse generator that provides a common eye-velocity drive to the horizontal and vertical motor plants. We argue that this same vectorial motor command drives the eye–head motor systems, albeit that it is modified by appropriate coordinate transformations that let the eyes and head move in the direction of auditory and visual targets. We discuss several quantitative models that explain the underlying circuitry of eye- and eye–head gaze shifts.

p0310 *Chapter: The Midbrain Colliculus* discusses how a localized population of saccade-related cells in the superior colliculus (SC) encodes the vector for the upcoming gaze shift. We model the mechanism underlying the afferent complex-logarithmic SC motor map, and how the efferent projections of its cells to the brainstem contribute to encode the saccade, and the tuning characteristics of SC movement fields. We then present evidence that the firing rates of SC cells determine the saccade kinematics, such that the population effectively acts as the nonlinear vectorial pulse generator, hypothesized for the common source control of eye–head gaze shifts. We show that a spatial gradient of peak firing rates along the rostral–caudal axis of the motor map, together with a fixed number of spikes in the burst, provide the nonlinear mechanism of the main sequence. Finally, we discuss the potential role of the midbrain inferior colliculus (IC) in spatial hearing, and how the tonotopic to spatial transformation in the auditory system might be understood.

p0315 *Chapter*: *Coordinate Transformations* describes the different egocentric reference frames that are relevant for the control of orienting gaze shifts to brief sounds and lights: oculocentric, craniocentric, and world-centered coordinates. We then discuss how experiments and quantitative models of eye–head gaze orienting can dissociate the different coordinate systems. Neurophysiological experiments have identified two different mechanisms that could potentially cope with coordinate transformations: gain fields and predictive remapping. We describe the static and dynamic double-step paradigms as prime examples of studying the nature of spatial target updating, and discuss the limits of spatial updating performance in situations where the system lacks crucial information about stimulus motion. Finally, we show how eye- and head-position signals interact within the auditory system, and postulate that the sound-localization system represents acoustic targets in a world-centered reference frame.

p0320 *Chapter*: *Sound-Localization Plasticity* discusses plasticity of the human sound-localization system in response to a variety of acoustic manipulations. We distinguish explicit perceptual learning from implicit sensory–motor feedback learning, and describe experiments that illustrate both types of learning. We introduce the phenomenal plasticity of the barn owl's auditory system, and describe the immediate localization effects of unilateral ear–canal plugging of human listeners. We discuss why spectral cues are essential to cope with this binaural perturbation. We then describe the adaptive response of humans to long-term unilateral plugging. When the pinna cues are perturbed by molds that change the pinna geometry, subjects relearn to map the new spectral cues to target elevation within a few weeks. Interestingly, these manipulations do not interfere with the original spectral cues, and there are no aftereffects. We argue that pattern-recognition learning (no aftereffects) differs in essential ways from parametric learning (with aftereffects). We demonstrate that the visual system is needed to fine tune the spatial mapping of spectral cues, by comparing sound localization of congenital blind listeners with normal sighted listeners. Finally, we demonstrate that perturbed spatial vision will perturb the sound-localization mappings in a similar way.

p0325 *Chapter: Multisensory Integration* discusses the mechanisms that guide multisensory integration. In particular, we focus on audiovisual and audiovestibular interactions that influence spatial perception. We describe the spatial–temporal factors that modulate the strength of multisensory integration, and introduce the phenomenon of inverse effectiveness. We forward three types of models to account for multisensory interactions on reaction times and response accuracy: race models, interactions within the superior collicular motor map, and Bayesian inference. We discuss the latter framework as a mathematical–statistical model to understand two prominent effects on response trajectories to multisensory stimuli: response averaging and decreased response variability. We show that audiovisual integration in cluttered audiovisual environments follows the spatial–temporal rules of multisensory integration, and provides strong support for inverse effectiveness at the behavioral level. We show that the sensory–motor system integrates stimuli only when they are spatially and temporally congruent, and keeps track of the stimulus statistics in the environment to assess the probability of audiovisual alignment. Finally, we apply the Bayesian framework to explain the considerable mislocalization of sounds during head tilts with respect to gravity.

p0330 *Chapter: Impaired Hearing and Sound Localization* describes the sound-localization abilities of hearing-impaired listeners. We distinguish listeners with a conductive hearing loss from sensory–neural impairments, and describe current technologies to restore impaired hearing: hearing aids, bone-conduction devices, middle-ear implants, and cochlear implants. Single-sided deaf listeners lack binaural hearing and may only use spectral cues from their hearing ear, and the head–shadow effect. Despite the fact that the latter cue is ambiguous, all SSD patients rely heavily on this cue. We argue that this behavior reflects a Bayesian strategy in which prior information is weighted more heavily than less reliable spectral cues. Listeners with a unilateral conductive hearing loss adopt a flexible localization strategy in which they weigh potentially remnant binaural cues, spectral cues, and the head–shadow effect to estimate the azimuth of a sound source under varying acoustic conditions. Finally, we propose that modest age-related, high-frequency hearing loss may be partially compensated by the acoustic effects of pinna growth. For certain sounds the elderly may thus outperform young listeners in localizing source elevation.

st0065 1.9 EXERCISES

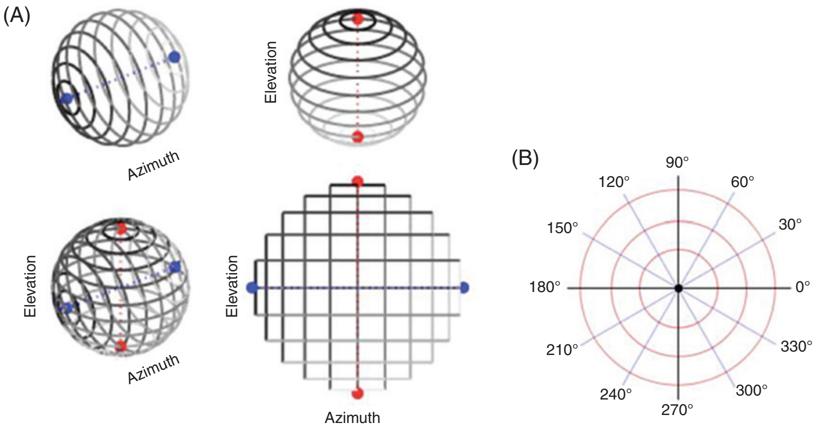
p0340 *Problem 1.1*: Coordinates and coordinate transformations of sound locations.

p0345 The azimuth (A)–elevation (E) system is a *double-pole* coordinate system, in which the azimuth angle in the horizontal plane is specified by a rotation about a head-vertical rotation axis from the midsagittal plane to the source position, while the elevation angle in the vertical plane is described by a rotation about the inter-aural axis from the horizontal plane to the source (Fig. 1.7A).

o0010 **(a)** Suppose a target sound is described by azimuth–elevation angles (A_S,E_S). You wish to foveate the sound with your eyes, which initially fixate at straight-ahead, that is, (A,E) = (0,0) degree. The sound location has to be transformed into oculocentric polar coordinates, expressing the rotation amplitude, given by eccentricity angle, R, and direction, Φ (Fig. 1.7B). Do the transformation, that is, calculate $R(A_S,E_S)$ and $\Phi(A_S,E_S)$.

o0015 **(b)** Show that for all sound locations in the frontal hemifield: $A + E \leq \pi/2$.

p0360 *Problem 1.2*: Audiovisual integration is useful only when sound S and visual target V are both at the *same* spatial location (Fig. 1.6)! Often, sensory coordinates may differ substantially, but when they emerge from a single object they



f0040 **FIGURE 1.7** (A) Double pole–azimuth elevation coordinate system to specify sound locations with respect to the head of the subject. Top plots show the isoazimuth (left) and isoelevation (right) contour lines. Bottom plots show the full sphere (left) and the projection of (A,E) contour lines as seen from straightahead. (B) polar coordinates, showing iso eccentricity lines (circles) and iso direction lines (spokes).

should be integrated. From Eq. (1.2), determine the coefficients [a,b,c,d] for adequate audiovisual integration.

p0365 Often, however, the *sensory* coordinates of S and V may be identical $(A_S,E_S) = (R_V,\Phi_V)$, yet originate from *different* objects. In such cases the stimuli should not be integrated.

p0370 Draw that situation. What happens if the coefficients that you just determined for integration are applied by the sensorimotor system? Give an argument as to why this may or may not help in the identification and localization tasks.

p0375 *Problem 1.3*: Inverse problems are often ill posed. A good example is given by the problem of perception: on the basis of a limited number of measurements (sensory observations, eg, foveation of points in the visual scene through saccades), the brain has to make an estimate (inference) about the environment and the stimuli that caused the percept. Mathematically, a problem is well posed if there exists a unique and stable solution to the problem. A solution is stable if it resists (small) perturbations of the starting values. If solutions are not stable, or unique, the problem is ill posed.

p0380 The latter may even happen for seemingly trivial calculations such as taking a derivative. As a numerical exercise we look at the following example: suppose that we have to determine the derivative of a function, $f(x)$:

$$q(x) = \frac{df}{dx}$$

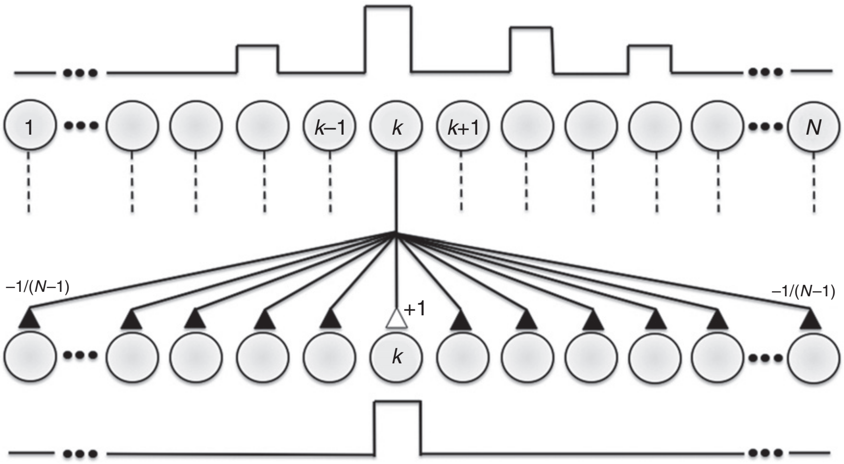
p0385 However, it could occur that instead of $f(x)$, we have to deal with a slightly perturbed measurement of this function, say:

$$f_n(x) = f(x) + \frac{\sin(nx)}{\sqrt{n}}$$

p0390 Clearly, for $n \to \infty$ the difference between perturbed and original function, $\|f - f_n\|$, approaches zero. Show that for the derivative, however, this difference becomes arbitrarily high. This means that the operation is unstable, and hence calculating the derivative is ill posed.

p0395 *Problem 1.4*: Fig. 1.8 shows a simple two-layered (input–output) neural network consisting of N linear neurons in each layer. The activity of the input layer is drawn above the neurons. Only neurons $k-2$, k, $k+2$, and $k+4$ receive input strengths of 1, 3, 2, and 1 units, respectively.

p0400 The neurons of the network have repetitive connection patterns; only the connections of input neuron k are drawn for clarity. Each neuron in the input layer excites its corresponding output neuron with synaptic strength +1, and inhibits all the other neurons of the network with synaptic strengths $-1/(N-1)$.



f0045 **FIGURE 1.8 A WTA two layer–feed forward network of linear neurons.** The connection scheme (highlighted here for input neuron *k* only) is identical for all neurons.

In this way, the total synaptic weight from each input neuron sums to zero. The activity of an output neuron is determined by:

$$y_k = \sum_{n=1}^{N} w_{kn} x_n$$

with w_{kn} the synaptic connection from input neuron n to output neuron k, and x_n the activity of input neuron n. Take $N = 11$ and $k = 5$. Show that the network indeed operates as a WTA network by calculating the activities of all N output neurons. In modeling saliency maps, WTA networks play an important role, as they weed out the contributions from all competing active neurons except from the one neuron with the strongest activation.

st0060 ACKNOWLEDGMENT

p0335 ~~This book is dedicated to Dr Dick Donker (Dec 21, 1934–Oct 1, 2014): neurologist, life artist, a lover of music, and auditory science.~~ I thank my colleagues, technicians, PhD students, post-docs, and graduate students, who greatly enriched my lab for more than 25 years. In particular, I am indebted ~~to Maarten~~ Frens, Jeroen Goossens, Marcel Zwiers, Paul Hofman, Denise van Barneveld, Tom van Grootel, Joyce Vliegen, Peter Bremen, Rob van der Willigen, Marc van Wanrooij, and Martijn Agterberg, who designed and completed the many tedious experiments and sophisticated data analyses that are presented in Chapters 7–14 of this monograph.

to my mentor Jan van Gisbergen, who
teached me everything about neuroscience,
and whose influence on my thinking is so
clearly visible in this book. I am grateful
Maarten

bi0010 REFERENCES

General Readings

bib0010 Bregman, A.S., 1990. Auditory Scene Analysis. MIT Press, Cambridge, MA.

bib0015 Purves, D., Augustine, G.J. (Eds.), 2012. Vision: the eye, Central visual pathways, and The auditory system. Neuroscience, fifth ed. Sinauer Associates Inc., Sunderland, MA (Chapters 10, 11, and 12), pp. 229–314.

ISBN: 978-0-12-801529-2; PII: B978-0-12-801529-2.00001-5; Author: VANOPSTALENGLISH; Document ID: 00001; Chapter ID: c0005

bib0020 Purves, D., Augustine, G.J. (Eds.), 2012. Eye movements and sensory motor integration. Neuroscience, fifth ed. Sinauer Associates Inc., Sunderland, MA, pp. 453–468.

On Ill-Posed Problems

bib0025 Kabanikhin, S.I., 2008. Definitions and examples of inverse and ill-posed problems. J. Inv. and Ill. Probs. 16, 317–357.

On Saliency Maps

bib0030 Koch, C., Ullman, S., 1985. Shifts in selective visual attention. Hum. Neurobiol. 4, 219–227.

bib0035 Itty, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40, 1489–1506.

On Inhibition of Return

bib0040 Posner, M.I., Cohen, Y., 1984. Components of visual orienting. In: Bouma, H., Bouwhuis, D.G. (Eds.), Attention and Performance. Hillsdale, NJ, Erlbaum (Chapter 32), pp. 531–556.

bib0045 Klein, R.M., 2000. Inhibition of return. Trends Cogn. Sci. 4, 138–147.

bib0050 Wang, Z., Klein, R.M., 2010. Searching for inhibition of return in visual search: a review. Vis. Res. 50, 220–228.