Department of Signal Processing and Acoustics

Effects of spatial cue dynamics on the perceptual organization of sound

Henri Pöntynen

Aalto University

DOCTORAL
DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 85/2021

Effects of spatial cue dynamics on the perceptual organization of sound

Henri Pöntynen

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, on 05.08.2021 at 16:00.

Aalto University
School of Electrical Engineering
Department of Signal Processing and Acoustics
Communication Acoustics

Supervising professor
Professor Ville Pulkki, Aalto University, Finland

Thesis advisor
PhD Nelli Salminen, Aalto University, Finland

Preliminary examiners
Research Associate Professor Virginia Best, Boston University, Massachusetts, United States of America
DSc Piotr Majdak, Austrian Academy of Sciences, Austria

Opponent
Associate Professor Ewan Macpherson, Western University, Ontario, Canada

Aalto University publication series
DOCTORAL DISSERTATIONS 85/2021

ISBN 978-952-64-0423-3 (printed)
ISBN 978-952-64-0424-0 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-64-0424-0

Unigrafia Oy
Helsinki 2021

Finland

Printed matter
4041-0619

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Henri Pöntynen

Name of the doctoral dissertation
Effects of spatial cue dynamics on the perceptual organization of sound

Publisher School of Electrical Engineering

Unit Department of Signal Processing and Acoustics

Series Aalto University publication series DOCTORAL DISSERTATIONS 85/2021

Field of research Acoustics and Audio Signal Processing

Manuscript submitted 3 March 2021 **Date of the defence** 5 August 2021

Permission for public defence granted (date) 11 May 2021 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

The auditory system is constantly analyzing the mixture of sounds arriving at the ears to form mental representations of the sound sources present in the environment - a process known as the perceptual organization of sound. This process relies on heuristics derived from the statistical properties of sounds heard in natural environments, including those of their perceived directional properties. Auditory percepts have a salient spatial dimension that reveals the locations of sound sources with remarkable accuracy despite the fact that the sensory receptors of the auditory organs are not sensitive to sound direction. Rather, directional hearing is an inherently computational process wherein implicit spatial cues are extracted neurally from the acoustic waves arriving at the ears.

While the vast majority of spatial hearing research has focused on the perception of individual point-like sources under conditions where both the listener and the source remain static, natural listening scenarios are rarely this simplistic. Instead, when sounds are heard outside of laboratory conditions, the spatial cues available to listeners are constantly changing due to the combination of listener and source movements as well as acoustic interference between concurrently active sound sources. Yet, the role of spatial cue dynamics in the perceptual organization of sound remains an unexplored topic in many fields of auditory research.

The experiments included in this thesis address various auditory phenomena associated with dynamically varying spatial cues. Publications I, II, and IV document behavioral studies where the perceptual effects of spatial cue dynamics arising from the combination of listener and source motion (PI), listener motion alone (PII), or from acoustic-domain interference of multiple concurrently active sources (PIV) were assessed. The results of these studies show that cue dynamics can both enhance and degrade the accuracy of auditory perception.

Publication III documents a neuroscientific experiment where electroencephalography was used to assess the cortical responses evoked by random-chord stereograms — a type of auditory stimulus capable of evoking binaurally driven auditory illusions. The results show that these stimuli evoke robust cortical responses as indicated by various time-, frequency- and time-frequency-domain measures. Random-chord stereograms could therefore potentially provide a flexible research tool for neuroscientific experiments seeking to isolate binaurally driven processes in the perceptual organization of sound.

Overall, the results provide new insights into the role of spatial cue dynamics in auditory perceptual organization. The results are informative for the design of novel audio processing algorithms for binaural audio devices as well as for improving the ecological validity of auditory experiments across disciplines.

Keywords psychoacoustics, spatial hearing, auditory scene analysis, auditory neuroscience

ISBN (printed) 978-952-64-0423-3 **ISBN (pdf)** 978-952-64-0424-0

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2021

Pages 164 **urn** http://urn.fi/URN:ISBN:978-952-64-0424-0

Preface

The work for this doctoral thesis was conducted at the Department of Signal Processing and Acoustics at Aalto University, School of Electrical Engineering in Espoo, Finland. The work was funded by the Academy of Finland. Early stages of the work received financial support from HPY Research Foundation.

I want to thank Professor Ville Pulkki for supervising my doctoral thesis work and providing a research environment characterized by a great degree of freedom, trust and independence. Ville's broad-minded and undogmatic approach to doing science encouraged me to pursue research topics that would otherwise have felt inaccessible with my educational background. I want to express my sincere gratitude to PhD Nelli Salminen for being a fantastic thesis instructor and an excellent scientific mentor. Nelli's deep knowledge of neuroscientific and behavioral hearing research was instrumental in enabling me to see the big picture of the auditory field and inspired me to broaden my interests beyond audio-oriented psychoacoustics. Her seemingly endless patience and thoughtful sensitivity to the typical mental pitfalls of early-stage researchers played a crucial role in supporting this work to its completion. I would also like to thank the preliminary examiners, Research Associate Professor Virginia Best and DSc Piotr Majdak for their kind and helpful comments on the thesis manuscript.

During this work I've had the privilege of getting to work with many extraordinary people whose enthusiasm and talent for science and technology has been truly inspiring to witness. Therefore, I want to thank all of my colleagues at Aalto University and Facebook Reality Labs for the many intellectually stimulating conversations and continuously upholding an excellent work atmosphere. Special thanks to Fabián Esqueda, Geoffrey Gormond and Julian Parker for the numerous lengthy discussions about analog synthesizers that eventually materialized into scientific publications. Working on circuit modeling problems provided good balance to the perceptually oriented thesis work.

Finally, I want to thank my family and friends for their support and for reminding me of the important things in life. Most importantly, I am

deeply grateful to my partner Matilda for standing by my side through the ups and downs of the past years.

Helsinki, Finland, June 2021
Henri Pöntynen

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Henri Pöntynen, Olli Santala, Ville Pulkki. Conflicting dynamic and spectral directional cues form separate auditory images. In *Proceedings of the 140th Convention of the Audio Engineering Society*, Paris, France, Convention Paper 9582, June 4 - 7 2016.

II Henri Pöntynen, Nelli Salminen. Resolving front-back ambiguity with head rotation: The role of level dynamics. *Hearing Research*, Volume 377, pp. 196 – 207, June 2019.

III Henri Pöntynen, Nelli Salminen. Cortical processing of binaural cues as shown by EEG-responses to random-chord stereograms. Submitted to *Journal of the Association for Research in Otolaryngology*, December 2020.

IV Ville Pulkki, Henri Pöntynen, Olli Santala. Spatial perception of sound source distribution in the median plane. *Journal of the Audio Engineering Society*, Volume 67, Issue11, pp. 855 – 870, November 2019.

Author's Contribution

Publication I: "Conflicting dynamic and spectral directional cues form separate auditory images"

The present author designed and implemented the experiment, conducted the data analysis and wrote the manuscript. The co-authors contributed ideas to the experimental design and manuscript preparation.

Publication II: "Resolving front-back ambiguity with head rotation: The role of level dynamics"

The present author designed the experiments in collaboration with the second author, implemented and conducted the experiments and data analyses, and wrote the manuscript with the second author.

Publication III: "Cortical processing of binaural cues as shown by EEG-responses to random-chord stereograms"

The present author designed and implemented the experiment, gathered the EEG-data and performed the data analyses. The manuscript was prepared in collaboration with the second author.

Publication IV: "Spatial perception of sound source distribution in the median plane"

The present author contributed to the interpretation of the experimental results, data analyses and preparation of the manuscript.

List of Figures

Abbreviations

A1 Primary auditory cortex
A2 Belt area of auditory cortex
A3 Parabelt area of auditory cortex
CMAA Concurrent minimum audible angle
CN Cochlear nucleus
DCN Dorsal cochlear nucleus
EEG Electroencephalography
ERP Event-related potential
HRTF Head-related transfer function
IC Inferior colliculus
ICC Central nucleus of the inferior colliculus
IHC Inner hair cell
ILD Interaural level difference
IPD Interaural phase difference
ITP Interaural time difference
JND Just-noticeable difference
LL Lateral lemniscus
LSO Lateral superior olive
MSO Medial division of the superior olive
ORN Object-related negativity
RCS Random-chord stereogram
SC Superior colliculus
SOC Superior olivary complex
VAS Virtual auditory space

1. Introduction

The auditory system is faced with the task of transforming acoustic waves arriving at the ears into behaviorally relevant mental representations of sound sources in the surrounding environment — a process referred to as the perceptual organization of sound, or auditory scene analysis. At the level of sensory receptors, the organs of hearing are not sensitive to the directions of sounds nor do they explicitly encode which segments of sound were produced by any given sound source. Instead, spatial hearing — and perceptual organization of sound in general — are inherently computational processes, wherein auditory percepts are formed based on the implicit cues embedded into the acoustic waves arriving at the eardrums. Despite the lack of receptor-level encoding, more often than not the identity and spatial properties of auditory percepts match those of the physical sound sources to a remarkable degree of accuracy. Yet, many details of the perceptual organization process remain elusive and it is not fully understood which acoustic-domain parameters facilitate it or how it is implemented neurally by the auditory system.

One aspect of auditory perception that has often been overlooked in auditory studies across disciplines is that of dynamic spatial cues. Outside of laboratory conditions spatial cues rarely remain stable due to the combination of listener and source movements. Similarly, it is not uncommon for real soundscapes to contain more than one active sound source. When multiple sound sources are active concurrently, acoustic-domain interference between the sounds emitted by the individual sources introduce on-going, movement-independent dynamics to the spatial cues available to the listener. These source-interference-induced dynamics can result in severely distorted spatial perception.

Elucidating the perceptual effects associated with spatial cue dynamics would be advantageous for many fields of auditory research. First, it could serve to open up new avenues in scene analysis research — a field where self-motion cues have been largely ignored despite their behavioral relevance. Second, identifying perceptually salient auditory phenomena driven by spatial cue dynamics could aid in improving the ecological validity of

neuroscientific scene analysis studies by introducing novel stimulation paradigms that better approximate the complexity of natural soundscapes than the simplistic stimuli classically employed in auditory neuroscience. Finally, given that spatial cue dynamics can yield both enhancements as well as biases to auditory perception, an improved understanding of the role of spatial cue dynamics in auditory perception could prove invaluable in the development of various technological applications, including signal processing algorithms for bilateral hearing aids and virtual/augmented reality audio devices. To this end, the publications contained in this thesis sought to assess various aspects of auditory perceptual organization in listening scenarios involving dynamic spatial cues. The methods included behavioral experiments and electroencephalography (EEG).

2. Auditory spatial perception

In the case of vision and somatosensation — two sensory modalities with a salient spatial dimension — directional information is encoded already at the level of the sensory receptors by virtue of their topographical organization. However, in audition spatial features are not explicitly represented at the receptor level. Rather, they have to be extracted from implicit acoustic cues embedded into the ear canal signals (Blauert, 1997). Despite this apparent disadvantage, the spatial hearing system functions to a remarkable degree of accuracy, as shown by localization studies where errors as low as a few degrees are commonly reported (Mills, 1958; Wightman and Kistler, 1989b; Makous and Middlebrooks, 1990; Middlebrooks and Green, 1991; Blauert, 1997). To understand the processes underlying directional perception of sound, this chapter presents the anatomical and acoustic bases of human spatial hearing and reviews results from behavioral studies pertinent to the thesis work. Accordingly, the majority of the chapter is focused on describing spatial hearing phenomena during self-motion and in complex listening scenarios involving multiple concurrent sources.

2.1 Peripheral organs

The peripheral system transforms acoustic waves into neural impulses that project into the central auditory pathway via the cochlear nerve. Conventionally, peripheral organs are divided into the external-, middle- and inner ear (see Fig. 2.1). The external ear comprises the pinna and the ear canal, terminated at the eardum. The role of the external ear is twofold, it both amplifies the sounds impinging on the ear, as well as modifies the sound spectrum arriving at the eardrum in a direction-dependent manner; a crucial factor in enabling spatial hearing in three-dimensional space (Blauert, 1997).

The middle ear consists of the bones in the ossicular chain that transmit the vibrations of the eardrum to the high-impedance fluid of the inner ear through the oval window. By coupling the large surface area of the

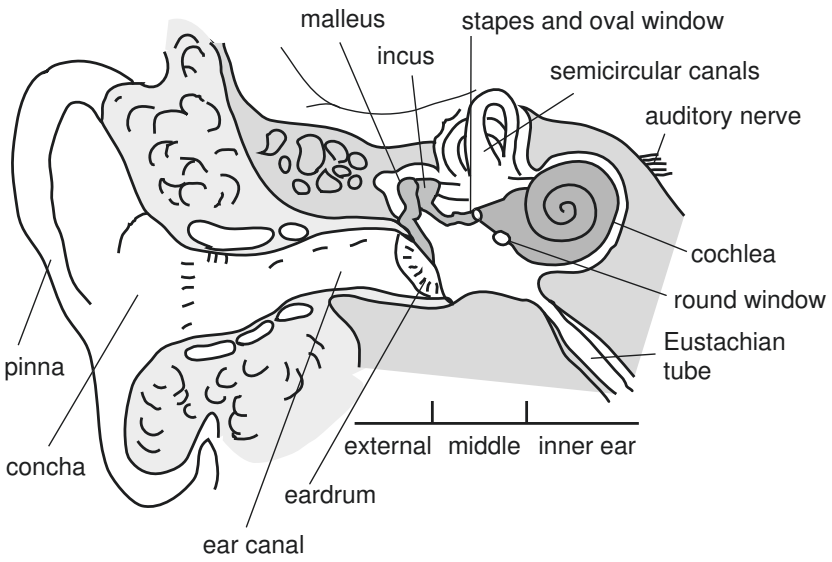

Figure 2.1. Peripheral hearing organs and the vestibular organs. Adopted from Pulkki and Karjalainen (2015).

eardrum to the much smaller surface area of the oval window, the ossicular chain achieves an impedance transformation that facilitates the efficient transfer of sound to the inner ear (Rossing, 2007).

The inner ear transforms the mechanical vibrations transmitted by the middle ear into electrical nerve impulses that are further transmitted to the nervous system via the auditory nerve. This transformation is achieved by the biomechanical properties of the inner ear in a frequency-dependent manner, so that different frequencies stimulate different sets of inner hair cells (IHC) — the sensory receptors of hearing — arranged tonotopically along the length of the cochlea (Plack, 2018). Tonotopy is a general property of the auditory system that is maintained throughout the auditory pathway (Saenz and Langers, 2014). Accordingly, the electrical impulse outputs of the IHCs are passed to the central auditory system along parallel tonotopically organized fibres in the auditory nerve.

Due to the anatomical arrangement of IHCs and the biomechanics of the cochlea, the frequency resolution of the inner ear transduction is non-uniform across frequency (Moore, 2012). At low center-frequencies, a larger quantity of unique IHCs is available per linear frequency interval than at high center-frequencies. Consequently, the frequency selectivity of the cochlea is best at low frequencies and degrades with increasing frequency. Bandwidths within which simultaneously presented frequencies stimulate overlapping sets of IHCs are referred to as critical- or auditory bands (Fletcher, 1938a,b, 1940; Moore, 2012). These bands represent the frequency resolution of the cochlea at different center-frequencies and

can be conceptualized as the minimum frequency separation required for two sounds to be accurately represented in the neural activity of separate auditory nerve fibres. The limited frequency resolution of the cochlea has implications for the acuity of spatial hearing as interaural spatial cues are extracted from auditory band-specific comparisons of cochlear outputs (e.g. Scharf et al., 1976). Therefore, sound energy within a given auditory band contributes to the spatial information extracted from the neural activity in the corresponding nerve fibres.

As in the case of frequency resolution, also the temporal resolution of the cochlea decreases with increasing frequency. At low frequencies, both the membrane potential of the IHCs (for instance, Palmer and Russell, 1986) as well as the spike rate activity at the auditory nerve (e.g. Rose et al., 1967; Anderson et al., 1971) reflect the temporal fine-structure of acoustic stimuli. At frequencies above a few kHz, temporal precision of the nerve fibre discharges deteriorates, so that the discharge times in individual fibres are no longer phase-locked to the stimulation.

2.2 Head-related acoustics and spatial hearing

The implicit acoustic cues that enable spatial hearing in humans arise from the systematic manner by which sounds arriving from different directions interact with the head of the listener. Sound is a wave phenomenon that follows the laws of diffraction and reflection; To a first approximation, when a sound wave encounters an obstacle with dimensions significantly smaller than its wavelength, the wave travels past the obstacle due to diffraction relatively unaffected (Kuttruff, 2007; Rossing, 2007). This results in comparable sound pressure levels at both sides of the obstacle but a difference in the time of arrival at either side of the obstacle. Conversely, when the dimensions of the obstacle are large in comparison to the sound's wavelength, an increasingly large proportion of the sound energy impinging on the surface of the object is reflected backwards (Rossing and Fletcher, 2012). Since the high-frequency wave is not able to travel past the obstacle unimpeded, a sound pressure level difference is introduced between the opposing sides of the obstacle. These basic properties of physics of sound form the acoustic-domain basis for human spatial hearing, as the human head presents an acoustic obstacle that perturbs the acoustic field formed around the head. These perturbations are direction- and frequency-dependent and serve as the acoustic basis for auditory spatial cues.

2.3 Binaural spatial cues

2.3.1 Temporal difference cues

When sounds impinge on the listener's head, an azimuth-dependent difference in time of arrival arises between the ear canal signals due to the path-length difference between the sound source and the positions of the two ears. For example, a sound emitted by a source on the right side of the listener will arrive earlier at the right ear than at the left ear and vice versa. This time-based spatial cue is referred to as an interaural time difference (ITD), or in the context of narrow-band sounds, an interaural phase difference (IPD). Temporal difference cues are the dominant localization cues that the auditory system exploits in the localization of low-frequency sounds (Wightman and Kistler, 1992).

The magnitude of ITD is a function of both the distance between the ears as well as the position of the sound source; ITDs are minimal for sources positioned along the vertical midline and grow larger at more lateral positions. In humans, the maximum naturally occurring ITDs at the extreme lateral angles of ± 90 degrees are as short as 600 - 750 μs, depending on the head-size of the subject (Feddersen et al., 1957; Kuhn, 1977). Moreover, humans are highly sensitive to changes in low-frequency ITD, with the smallest just-noticeable differences (JND) being as low as 10 μs (Klumpp and Eady, 1956). ITD-sensitivity varies with azimuth angle so that sensitivity is greatest for sources positioned at the midline and decreases at more lateral positions (Yost, 1974).

In addition to interaural differences in the temporal fine-structure of sounds, the auditory system derives spatial cues from the interaural delay of the amplitude envelopes of sounds (Henning, 1974; McFadden and Pasanen, 1976; Henning, 1980). These cues are however less salient than the phase-locked fine-structure IPD and are available only at relatively high-frequencies, above approximately 1.5 kHz. Moreover, the salience of envelope-ITD depends heavily on the sharpness of the envelope (Henning and Ashton, 1981; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1994, 2002, 2009; Laback et al., 2011), thus limiting its usefulness as a general spatial cue.

2.3.2 Level difference cues

In the case of high-frequency sounds, the wavelength of sound can be much smaller than the dimensions of the head. This has two important implications for binaural hearing. First, when the wavelength of sound is small compared to the distance between the ears, an unambiguous phase-relationship cannot be established for the interaural delay and it becomes

impossible to distinguish which of the two ears is leading and which is lagging based on on-going IPDs alone (Schnupp et al., 2011). This ambiguity makes temporal comparisons between the ears an unreliable localization cue for high-frequency sounds. Second, for high-frequency sounds, the head represents a large acoustic obstacle and diffraction effects become negligible. This imposes an acoustic shadowing effect at the contralateral side of the head, resulting in an interaural level difference (ILD) between the ears for laterally located sounds (Schnupp et al., 2011).

The magnitude of ILD depends both on frequency and the location of the sound source. For sound sources at the vertical midline, ILD is near zero due to the approximate symmetry of the human head (Blauert, 1969). In the case of wideband sounds that contain a wide range of frequencies, the magnitude of the overall ILD increases approximately monotonically with lateral angle (Blauert, 1997). In the case of high-frequency narrowband sounds however, the relationship between ILD and lateral angle is not strictly monotonic due to acoustic interaction between the impinging sound wave and the anthropometric details of the listener's head. As the frequency increases, these interactions become increasingly idiosyncratic and confound the nearly monotonic mapping between azimuth angle and ILD observed with wideband sounds (Schnupp et al., 2011). At the behavioral level, JNDs of ILDs can be as low as 0.5 - 1.0 dB under optimal conditions (Mills, 1960; Yost and Dye Jr, 1988). As in the case of ITD, sensitivity to changes in ILD is greater when initial ILDs are near-zero, corresponding to source positions along the vertical midline.

2.3.3 Ambiguity of binaural cues

Although binaural cues enable remarkably accurate localization performance in the left-right dimension, they provide no systematic cues to the front-back or up-down dimensions of sound source location. This spatial ambiguity arises from the fact that interaural difference cues are not unique to any given source position. Rather, the same magnitude of interaural differences in both timing and level can arise due to sound sources in a number of possible locations (Blauert, 1997; Schnupp et al., 2011; Moore, 2012). For instance, the geometric path-length difference from the source to the two ears is approximately zero for all sources positioned along the vertical midline. If the head is simplistically modelled as a sphere with the ears represented as two opposing points on its surface, the possible source locations that yield the same binaural differences corresponds approximately to a conical surface that is symmetric about the interaural axis and intersects with the actual location of the sound source (Blauert, 1969). This locus of possible source positions captures the spatial ambiguity inherent to binaural difference cues and is commonly referred to as the cone of confusion (Moore, 2012). The cone of confusion often manifests as

front-back confusions in spatial hearing tasks (see the left-hand side panel of Fig. 2.2). To overcome the spatial ambiguities inherent to binaural cues, the auditory system relies on spatial cues derived from monaural signal characteristics, as well as cross-modal cues derived from combining self-motion information with the spatial cue dynamics that co-occur naturally with listener movement.

2.4 Monaural spatial cues

Despite the fact that binaural cues are uninformative about spatial properties of sounds beyond the left-right dimension, behavioral studies show that listeners can indicate the elevation angle of wideband noise bursts with surprising accuracy. For instance, mean vertical localization errors can be as low as 3.5 degrees for sources positioned along the vertical midline (Makous and Middlebrooks, 1990). The spatial cues that enable this level of vertical localization performance arise from the linear spectral distortions imposed by the pinnae on the sound waves impinging on the head (Batteau, 1967; Middlebrooks and Green, 1991; Blauert, 1997). In specific, the direction-dependent acoustic filter formed by the pinna modifies the wideband sound spectrum formed at the eardrums in a manner that is unique to each angle of arrival and can therefore serve as a cue to sound source location also in the front-back and up-down dimensions (Shaw and Teranishi, 1968; Shaw, 1997; Carlile et al., 2005).

Due to their acoustic basis in wideband spectral distortions, the salience of monaural cues is heavily dependent on the source spectrum. For example, due to the acoustic shadow cast by the flanges of the pinnae, sounds arriving from directly behind the listener yield ear canal spectra with less high-frequency energy than identical sounds positioned directly in front of the listener (Asano et al., 1990). Accordingly, this direction-dependent difference in the magnitude of the high-frequency spectrum enables monaural disambiguation between source locations in front and behind the listener (Musicant and Butler, 1984; Oldfield and Parker, 1986), provided that the source spectrum contains energy in the high-frequency range (8 - 16 kHz) of the auditory spectrum (Langendijk and Bronkhorst, 2002). Similarly, idiosyncratic patterns of sound reflections in the face of the pinnae introduce elevation-dependent peaks and notches in the resultant ear canal spectra that serve as elevation cues (Middlebrooks and Green, 1991). These cues are predominantly imposed on the spectral envelope at frequencies above 4 kHz and accurate vertical localization therefore requires that the source spectrum contains energy in this frequency range (Hebrank and Wright, 1974; Asano et al., 1990; Noble et al., 1994; Langendijk and Bronkhorst, 2002). Accordingly, if the spectrum of the target sound is limited to low frequencies, both vertical and front-back localization are impaired.

Similarly, monaural cues cannot provide accurate spatial information for narrow-band sounds. Instead, results from spatial hearing experiments show that when narrow-spectrum sounds are presented from the vertical midline — where no binaural cues are available and localization has to rely on monaural cues — elevation percepts depend on the center-frequency of the sound, but not on the actual source elevation angle (Blauert, 1969; Butler and Helwig, 1983; Middlebrooks, 1992; Itoh et al., 2007). For example, when bursts of third-octave white noise centered around 8 kHz are presented over loudspeakers positioned directly in front, above, or behind listeners in an anechoic room, subjects consistently report that the sound is perceived above the head, regardless of which loudspeaker is actually used to present the stimulus (Blauert, 1969; Butler and Helwig, 1983; Middlebrooks, 1992; Itoh et al., 2007). Similar frequency-dependent distortions to spatial perception have been reported in the context of pure tones presented over headphones (Thakkar and Goupell, 2014).

2.5 Spatial cues derived from self-motion

Despite the limitations inherent to the mechanisms of human spatial hearing, the auditory system is able to overcome many of them by leveraging cross-modal information. Arguably, the cross-modal effects that have the greatest significance for spatial hearing in behaviorally relevant listening scenarios arise from the combination of the sense of self-motion and auditory information. The sense of self-motion itself is a multimodal percept derived by combining information from vision, muscle proprioception, efferent copies of motor commands and the vestibular sense (Wallach, 1939, 1940; Cullen, 2012; Greenlee et al., 2016). Here, the term is used generally to refer to the perceived state-of-motion (e.g. direction and velocity of head rotation) a listener experiences, regardless of the exact combination of sensory modalities that give rise to these percepts. While the significance of self-motion cues in auditory spatial perception was noted by many researchers already in the early 20th century (Young, 1928, 1931; Willey et al., 1937; Wilska, 1938; Wallach, 1939, 1940), they nevertheless remain a relatively unexplored topic in many fields of auditory research.

Self-motion cues arise from the fact that listeners are able to cohesively interpret changes in self-orientation with the accompanying spatial cue dynamics. For example, when a continuous sound devoid of spectral cues is presented to a listener, front-back confusions are common (e.g. Stevens and Newman, 1936) since binaural cues provide no information about the front-back dimension of the sound (see left-hand side panel of Fig. 2.2). Yet, when the head is rotated, the resultant changes in binaural cues are categorically different for sound sources in front of and behind the listener (see right-hand side panel of Fig. 2.2). In specific, the magnitude of the

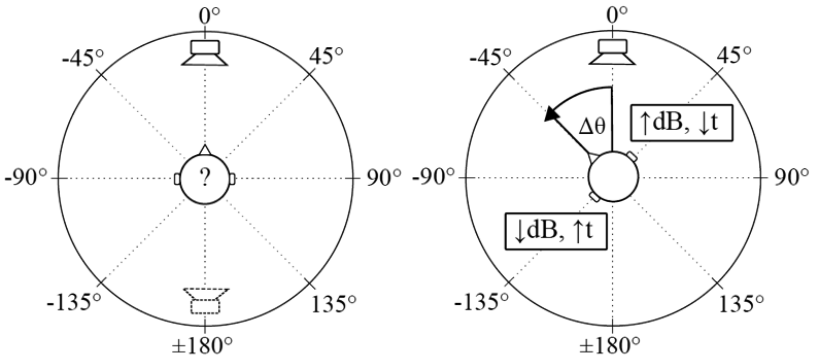

Figure 2.2. Left panel: Ambiguity of binaural difference cues. Binaural spatial cues stem from differences in the timing and level of the acoustic-domain ear canal signals between the two ears. Since the same magnitude of binaural differences in both level and timing can arise from a locus of possible source positions — the so-called "cone of confusion" — interaural difference cues contain an inherent spatial ambiguity. Consequently, when only binaural cues are available, estimates of lateral angle are often accurate but confusions about the front-back dimension are common. Right panel: acoustic basis of self-motion cues. When the head is turned, binaural differences increase or decrease depending on whether the source is in front of or behind the listener.

binaural disparities either increases or diminishes, depending on the front-back position of the source and the direction of head rotation. Thus, when the changing self-orientation information is successfully combined with the spatial cue dynamics that accompany self-motion, the front-back ambiguity inherent to binaural cues can be resolved without access to monaural spatial cues. Accordingly, behavioral studies show that head movements drastically reduce the frequency of front-back confusions that are otherwise common in spatial hearing experiments (see for instance Thurlow and Runge, 1967; Bronkhorst, 1995; Perrett and Noble, 1997a; Wightman and Kistler, 1999; Iwaya et al., 2003; Macpherson, 2011; Brimijoin and Akeroyd, 2012; Macpherson, 2013; Kim et al., 2013; McAnally and Martin, 2014; Brimijoin and Akeroyd, 2016; Pastore et al., 2018). In addition, localization errors in elevation angle are diminished if subjects are allowed to perform head movements (Thurlow and Runge, 1967; Perrett and Noble, 1997a,b; Iwaya et al., 2003; Kato et al., 2003; Vliegen et al., 2004), indicating that self-motion cues can enhance spatial perception beyond simple front-back disambiguation. Interestingly, these enhancements occur also for low-passed targets when movements are restricted to horizontal head rotations (Perrett and Noble, 1997b). This suggests that self-motion-driven enhancements to spatial perception are not directly attributable to re-orienting the head towards the source, so that the target is placed within the region of highest spatial acuity.

2.5.1 Dynamic front-back illusion

Wallach (1939, 1940) postulated an internalized association between spatial cue dynamics and perceived self-motion to be the basis for movement-induced enhancements to spatial hearing. In a series of classic experiments, Wallach demonstrated that the auditory system interprets changes in binaural cues in a context-dependent manner, so that when the listener perceives self-motion, concurrent changes in interaural disparities are implicitly attributed to a stationary sound source at a location that is consistent with the sequence of spatial cues; This became known as "the principle of least displacement" (Wallach, 1940). Interpreted geometrically, the least displacement principle predicts that the perceived source location corresponds to that spatial position that is common to all of the cones of confusion that arise during the motion. The spatial ambiguity arising from such a processing principle was demonstrated in a series of localization experiments where self-motion percepts induced via different sensory modalities (e.g. visually induced circular vection), were combined with source motion to bias spatial perception to illusory target locations (Wallach, 1939, 1940). One series of these experiments involved sounds that were moved within a circular loudspeaker array in tandem with the subjects' head rotations. Such a manipulation distorts the natural relationship between spatial cue dynamics and head rotation and can be leveraged to induce salient localization biases to self-motion-derived spatial percepts. Crucially, these experiments demonstrated that the principle of least displacement could be used to induce systematic hemiplane reversals by moving a sound source in azimuth by a factor of two relative to the angle of horizontal head rotation (see Fig. 2.3). The sequence of binaural cues arising from such a manipulation is approximately the same as the sequence that would arise naturally if the source was stationary in the opposing hemiplane. Accordingly, the source is perceived as a static sound source in the opposite hemiplane, following the least displacement principle. Here, this illusory hemiplane reversal effect is referred to as the "dynamic front-back illusion". The dynamic front-back illusion provides a useful methodological paradigm for identifying localization judgments derived from self-motion cues. Accordingly, in recent years, when camera-based real-time motion tracking apparatuses have become more widely available, many researchers have leveraged modern variants of the paradigm in localization studies focusing on self-motion cues (for instance, Macpherson, 2011; Brimijoin and Akeroyd, 2012; Macpherson, 2013; Yost et al., 2019). Some of these studies are discussed in more detail in section 4.3.2 and briefly in the section below.

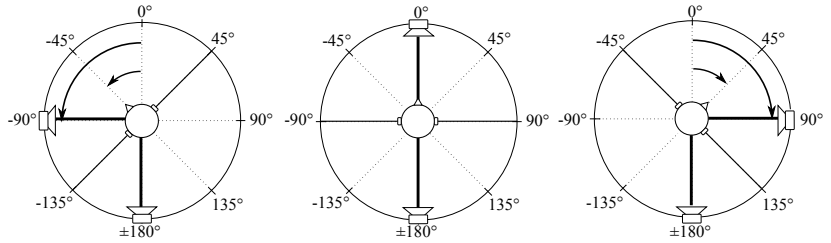

Figure 2.3. Two-dimensional illustration of the spatially ambiguous binaural dynamics underlying the dynamic front-back illusion. The thin solid line represents the interaural axis and the thick solid lines the projection of the cone of confusion onto the horizontal plane. The frontal hemiplane source represents the real source whose azimuthal position is determined by the instantaneous rotation angle of the listener's head, according to a 2:1-ratio. The rear hemiplane source represents the illusory spatial percept produced by the experimental manipulation. When the head is moved, it is implicitly assumed that the position of the sound in global coordinates is independent of the instantaneous position of the listener's head. Therefore, the ecologically valid interpretation of the sequence of binaural cues arising during self-motion is the one where the sound source remains stationary during the rotation; This corresponds to the spatial position that is common to the sequence of cones of confusion formed during the rotation. By displacing a sound source in azimuth by a factor of two for every degree of horizontal head rotation, the resultant sequence of binaural difference cues approximates those that the listener would experience if the sound source were stationary in the opposite hemifield.

2.5.2 Relative salience of dynamic spatial cue modalities

Although many behavioral studies have established self-motion-induced dynamic ITD to be a robust localization cue (e.g. Perrett and Noble, 1997a,b; Macpherson, 2011), the role of dynamic ILD cues is not as clearly established. Listeners are able to effectively use self-motion cues to resolve the front-back dimension of high-pass filtered target sounds that do not provide low-frequency ITD-cues and this ability appears to be retained even when monaural cues are confounded by experimental manipulations (e.g. Perrett and Noble, 1997b; Macpherson, 2011). This suggests that that dynamic ILD can provide salient spatial information. Yet, when listeners are tasked to use head rotations to localize narrow-band sounds presented from various azimuthal positions, the rate of front-back confusions increases for high-frequency stimuli and spatial perception appears to be driven by the directional biases induced by narrow-band stimulation rather than by information provided by dynamic ILD (Macpherson, 2011).

Similarly, if wideband stimuli are used in experiments leveraging the dynamic front-back illusion, monaural cues and self-motion-coupled dynamic binaural cues provide conflicting information about the position of the target and it is not clear which cue modality or combination of cues dominates spatial perception. For example, when listeners are tasked to perform head movements and localize spoken sentences presented according to the dynamic front-back illusion paradigm, directional percepts become unstable, resulting in the perception of a sound image that "flick-

ers" between the positions implied by the dynamic binaural information and the monaural cues (Brimijoin and Akeroyd, 2012) revealing no obvious spatial cue hierarchy.

2.6 Perceptual characteristics of headphone stimuli

Headphone stimulation is commonly used in auditory studies due to the precise control it provides over binaural signal characteristics. However, in its most basic form headphone stimulation evokes spatial percepts that are distinctly different from those evoked by sounds presented over loudspeakers or real sound sources. As such, spatial perception of headphone stimuli merits a brief discussion of its own.

When sounds are presented over headphones they are transmitted directly into the ear canal and the acoustic effects imposed by the head and pinnae are essentially bypassed. This results in binaural signals devoid of naturally occurring spatial cues. Accordingly, headphone stimuli generally evoke percepts of sound images inside the head, rather than out in the external environment. To reflect this difference, the term "lateralization" — rather than localization — is used to characterize the left-right dimension of sound sensations perceived inside the head (Durlach et al., 1992; Blauert, 1997). Similarly, the term "verticalization" is sometimes used to characterize the position of intracranial auditory images along the vertical dimension (Thakkar and Goupell, 2014).

When binaural cues are imposed artifically on headphone signals the resultant spatial percept can be characterized as a single "point-like" lateralized auditory image. However, this is not always the case. For instance, when noise is presented over headphones, the evoked spatial percept depend on the degree of interaural coherence between the left- and right channels. When the channels are fully coherent, the same noise sample is presented to both ears and the resultant spatial percept is point-like and positioned at the center of the head (for instance, Jeffress et al., 1962; Blauert, 1997). When the correlation between the binaural channels is decreased, the perceived width of the auditory image increases, until segregating into two separate images, lateralized to the two ears when correlation approaches zero (e.g. Blauert and Lindemann, 1986). Similar splitting of auditory images can be evoked with antiphasic binaural samples containing low-frequency energy (Wilska, 1938; Hirsh, 1948; Licklider, 1948; Blauert, 1997).

2.6.1 Factors influencing the externalization of sounds

While intracranial auditory images are most often associated with headphone stimulation, they are not an inherent characteristic of headphone-

based stimulation in itself. In fact, internalization can occur even for stimuli presented over loudspeakers under the appropriate combination of signal parameters (see for instance: Toole, 1970; Plenge, 1974; Levy and Butler, 1978; Brimijoin et al., 2013). Rather, internalization stems from the unnatural signal characteristics associated with inserting the stimuli directly into the ear canals and bypassing the acoustic environment. The missing signal characteristics can however be artificially imposed on the headphone signals to promote percepts of externalized sound images, located outside of the head. While many signal parameters contribute to externalization (recently reviewed in Best et al., 2020), here, the discussion is limited to the head-related transfer function (HRTF) and the use of head-tracking.

HRTFs represent the acoustic-domain effects associated with head-related acoustics. As such, they capture the direction- and frequency-dependent phase and magnitude distortions observed in the ear-canal signals when sounds interact with the head (Xie, 2013). HRTFs can be obtained via acoustical measurements of the sound pressure field formed at the eardrums by sound sources at different directions relative to the head. Processing headphone stimuli with digitally implemented HRTF-filters corresponding to a given source position, imposes the same spatial cues on the headphone signals that would emerge naturally if the sound was presented over a loudspeaker from that source position. Since HRTF-filters capture all of the acoustic-domain information that serve as the primary spatial cues (ITD, ILD and monaural cues), they can evoke salient percepts of externalized sound under headphone stimulation (Wightman and Kistler, 1989a,b; Kawaura et al., 1991; Xie, 2013).

Additionally, externalization can be greatly enhanced by introducing self-motion cues into the binaural signals with the aid of a real-time head-tracking system (for instance, Kawaura et al., 1991; Bronkhorst, 1995; Wightman and Kistler, 1999; Begault et al., 2001; Minnaar et al., 2001; Algazi et al., 2004; Kim and Choi, 2005; Brimijoin et al., 2013; Macpherson, 2013; Xie, 2013; Hendrickx et al., 2017). Importantly, head-orientation-coupled manipulation of binaural difference cues enables self-motion cues to evoke strong percepts of externalization in headphone listening, even for low-frequency- and narrow-band stimuli, that lack the spectral bandwidth required for the formation of accurate monaural cues (e.g. Loomis et al., 1990; Macpherson, 2011, 2013).

2.7 Spatial perception of multiple sound sources

While the vast majority of studies on spatial hearing have assessed auditory perception of single point-like sound sources, behaviorally relevant listening is rarely this simplistic. Natural soundscapes contain both unsyn-

chronized ensembles of individual sounds (e.g. a group of people talking) as well as volumetric sound sources whose spatial properties are not point-like (e.g. waves on a shore, waterfalls, orchestras, choirs, etc.). Since spatial cues are extracted computationally in auditory bands, the acuity of spatial hearing can be significantly worse in multi-source-scenes than with single sound sources if spectrally overlapping sources yield conflicting spatial information.

When two sounds are active at the same time, the acoustic field at the eardrums corresponds to the sum of the sounds arriving to the ears from both sources (Bauer, 1961). As a result, the instantaneous phase and level in each auditory band at both ears is described by the systematic phase and level distortions imposed by the acoustics of the head, as well as the confounding effects that arise from the summation of the energy emitted by the other source. This may result in binaural difference cues of much larger magnitude than what emerge in the case of single sources (Młynarski and Jost, 2014) as well as monaural cues that do not accurately represent the directional filtering characteristic associated with either source direction. Accordingly, the spatial information available to the auditory system can be severely degraded in multi-source scenes.

2.7.1 Azimuthally separated source pairs

Due to the tonotopical organization of the auditory system, the distortions to spatial cues imposed by concurrent activity of multiple sound sources depends on the degree of spectral overlap between the sources. If the sources stimulate separate auditory bands, frequency-dependent spatial cues can be extracted with relatively little detriment to the acuity of spatial perception. In contrast, if the sources overlap in the frequency-domain, spatial perception can be severely confounded; This is apparent in — for example — behavioral measurements of the concurrent minimum audible angle (CMAA, Perrott, 1984). CMAA describes the minimum angular separation required for subjects to be able to reliably characterize spatial deviations from a co-located presentation condition for synchronously presented stimulus pairs. CMAA-measurements conducted with tones show that while no CMAA could be established for tone pairs with small frequency disparities (i.e. the task was too difficult), CMAA decreases rapidly when the frequency-separation between concurrent tones increases, reaching a minimum of about 5 degrees for sources centered about the midline (Perrott, 1984). A similar, albeit less dramatic decrease in CMAA occurs when the sources are positioned more laterally, away from the vertical midline (Perrott, 1984). Similarly, experiments conducted in HRTF-based virtual auditory space (VAS) indicate that subjects become worse at assessing the relative positions of concurrently presented sounds of various bandwidths (e.g. amplitude- or frequency modulated tones, noise bursts) when spectral

overlap between the sounds increases and when the sources are positioned away from the midline (Divenyi and Oliver, 1989).

Besides frequency overlap, perception of concurrent sources is also strongly affected by the temporal correlation between the sources. When spectrally overlapping sources are temporally correlated — or coherent, the resultant spatial cues are stable across time and in general, result in the perception of a single fused auditory image at a spatial position determined by the aggregate of the spatial cues (Blauert, 1997). This phenomenon is referred to as summing localization and it forms the basis for stereophonic panning techniques that enable smooth manipulation of the apparent position of sound images between pairs of loudspeakers (Blauert, 1997; Pulkki, 1997; Pulkki and Karjalainen, 2015). In the context of the present work however, the more interesting scenarios are the ones where the sources share the same power spectrum, but are temporally uncorrelated.

In sound scenes consisting of uncorrelated sources, the moment-to-moment amplitude and phase of the sources are independent of one another. As a result, the summation of the source signals at the ear canal has the effect of introducing on-going random variations into the instantaneous interaural level- and phase differences within all of the auditory bands where the power spectra emitted by the sources overlap (Bauer, 1961; Takahashi and Keller, 1994; Blauert, 1997; Roman et al., 2003; Keller and Takahashi, 2005). Therefore, the binaural cues embedded into the acoustic-domain ear canal signals are formed as the combination of the deterministic interaural differences imposed by head-related acoustics and a stochastic component associated with the summation of the temporally incoherent sound energy arriving from the other source. Consequently, in the case of multi-source scenes, frequency-specfic binaural disparities are better characterized as probability distributions than as fixed values. This is illustrated in Fig. 2.4 where the IPD- and ILD-distributions associated with azimuthally separated pairs of uncorrelated broadband noises have been estimated computationally from the HRTFs of a binaural mannequin following the procedure outlined in Młynarski and Jost (2014). While all of the ILD-distributions in Fig. 2.4 are unimodal, the distributions of IPDs in many auditory bands become bimodal when the separation angle between the sources increases. Based on these acoustic-domain statistics, it appears that ITD/IPD-cues may be more salient in spatially driven perceptual segregation of concurrent sounds than ILD-cues.

Some behavioral studies support the hypothesis that on-going temporal disparities are crucial for facilitating the spatial perception of concurrent sounds. For instance, Best et al. (2004) used individualized VAS-techniques to assess spatial perception of concurrent pairs of broadband noise bursts with various co-located and spatially separated source positions. The task of the subjects was to report whether they perceived the concurrent bursts

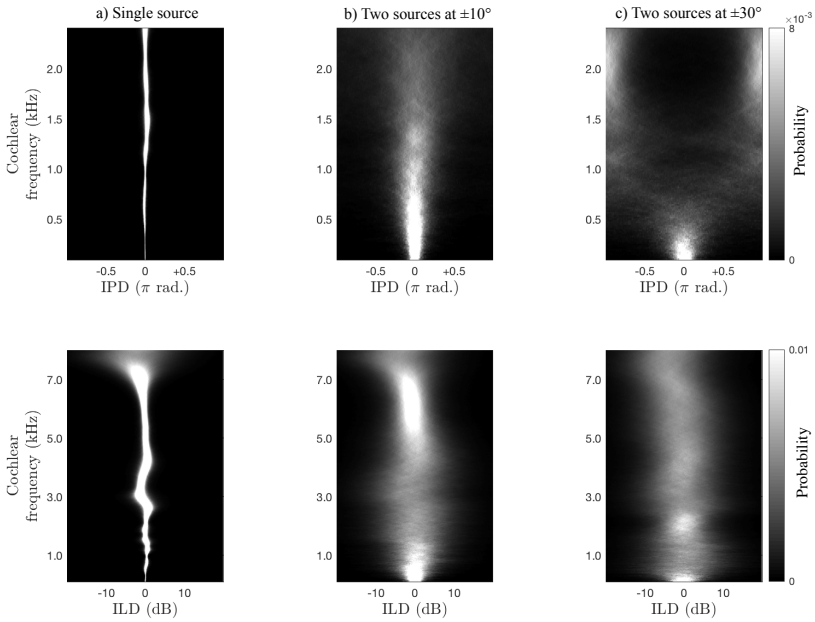

Figure 2.4. Probability distributions of the frequency-specific, instantaneous interaural difference cues arising from two concurrent white noise sources in the horizontal plane. Column a) - single source at 0 degs., column b) - two sources at ±10 degs., column c) - two sources at ±30 degs. For point-sources (column a), both IPD and ILD display narrow distributions centered around 0. When the separation increases to ±10 degs. (column b) both cue modalities are unimodal, but distributed across a wider range of values in all cochlear channels. For sources separated by ±30 degs. (column c) IPD-distributions display increasingly bimodal distributions in the cochlear channels centered above 500 Hz, but the ILD-distributions remain unimodal and centered on 0 dB. The images were derived by following the procedure outlined in Młynarski and Jost (2014). First, 5-s samples of uncorrelated white noise were filtered with the HRTFs of a binaural mannequin obtained from the HUTUBS HRTF-database (Brinkmann et al., 2019). The resultant ear canal signals were then divided into auditory bands with a gammatone filterbank (Slaney, 1998). A Hilbert transform was then applied to the narrow-band outputs of each of the cochlear channels to obtain an analytic signal for each auditory band. Interaural time- and level differences were then extracted from the differences in the phase (wrapped to π) and power of the analytic narrow-band signals.

of noise as emanating from a single direction or two distinct directions. The results showed, that in the case of azimuthally separated sources, subjects systematically reported percepts of two directions. Yet, when ITD-cues were removed from the stimulation, the rate of two-direction percepts diminished dramatically. This suggests that binaural cues — and temporal disparities in specific — facilitate the perceptual segregation of concurrently active uncorrelated sources and therefore appear to play an important role in the perceptual organization of complex soundscapes (see Chapter 4 for a discussion of auditory scene analysis and Section 4.3 for an extended discussion on the role of spatial cues in the perceptual organization of sound).

2.7.2 Vertically separated source pairs

Spatial perception of vertically separated sources is subject to similar severe acoustic-domain confounds as that of azimuthally separated sources. When two spectrally overlapping sources are active at the same time, the sound spectrum at the eardrum is formed as the superposition of the pinna-filtered spectra corresponding to the two source elevations. This spectral summation has the effect of obscuring the peaks and notches in the ear canal spectra that serve as elevation cues. In the case of correlated sources, the ear canal spectra are distorted, but stable across time. However, if the sources are uncorrelated the monaural spectra and the associated spatial cues are unstable across time as the sound pressure level in each auditory band may be dominated by the signal emitted by either one of the sources at any time instance. When pairs of vertically separated sources are placed along the auditory midline, binaural disparities diminish, and spatial perception has to rely on the confounded monaural cues only. Currently, not many behavioral studies have assessed spatial perception in such scenarios.

In the VAS-study of Best et al. (2004) subjects consistently reported perceiving sound from one direction only, when pairs of uncorrelated broadband noise bursts were presented with the HRTF-filters corresponding to two midline positions at different elevation angles. Based on the results of that study, it appears that monaural cues are not sufficient for perceptual segregation of concurrent sounds and that binaural cues mediate the organization of auditory scenes into separate sources in the case where spatial information is the only grouping cue available. Nevertheless, it is not clear if spatial perception of elevated broadband noise pairs is equally poor when stimuli are presented over loudspeakers and the perception of more complex median plane distributions remains unexplored.

2.7.3 Complex distributions

While spatial hearing studies involving pairs of concurrent sources are relatively rare, the literature is even more sparse in the case of experiments involving more complex source distributions consisting of three or more simultaneously active spectrally overlapping sources. At the level of the acoustic-domain signals arriving at the ears, scenarios involving more than two concurrent sources introduce similar confounds to spatial cues as in the case of two concurrent sources, with the exception that the contribution of the stochastic components arising from the interaction of the source signals upon summation at the eardrums becomes more significant as the number of sources increases. Accordingly, based solely on predictions derived from the characteristics of acoustic-domain phenomena, the acuity of spatial hearing is expected to decrease dramatically relative to the perception of

more simplistic source arrangements.

In the behavioral study by Santala and Pulkki (2011), subjects were presented with azimuthal distributions of uncorrelated pink noise (wideband noise characterized by a -3 dB per-octave power spectrum attenuation) and tasked to indicate the perceived directions of sound. Two types of distortions in spatial perception were consistent across subjects. First, continuous distributions of adjacent loudspeakers spanning various azimuthal widths of the frontal hemiplane were perceived to be narrower than the physical span of the source distribution, indicating that spatial perception of such sources is biased towards the centre of the source cluster (Santala and Pulkki, 2011). Second, spatial gaps in the distributions were not perceived until the separation between source clusters increased sufficiently, indicating that spatial discontinuities are masked by the surrounding sources when the separation angle is small; A result in line with the computational predictions shown in Fig. 2.4. In addition, it was observed that up to three individual sources in the distribution could be perceived with distinct directions in the frontal hemiplane if the source arrangement maximized the angular separation between the sources.

In Santala and Pulkki (2011), subjects were encouraged to move their heads to aid the assessment of the spatial properties of the source distributions. Since the study included no control condition where head movements were restricted, it is not clear to what extent self-motion cues contributed to the overall perception of the source distributions. Similarly, the study involved only azimuthal distributions where the statistical properties of the binaural cue distributions — rather than those of monaural cues — are expected to be the dominant source of perceptual variability between the source arrangements. Therefore, the contributions of monaural cue dynamics to the overall perception of such source distributions is difficult to evaluate based on the results of that study.

Chapter 2 summary

To conclude the chapter on auditory spatial perception, spatial hearing is an inherently computational process, as the sensory receptors in the peripheral hearing organs are not sensitive to the direction of arrival of sounds. Rather, the peripheral organs decompose sounds into narrow frequency bands and pass information to the central auditory system in parallel, tonotopically organized nerves. The auditory system deduces the spatial properties of sounds by extracting direction- and frequency-dependent signal features from the ear-canal signals; These include interaural disparities in timing (ITD) and level (ILD), as well as elevation-dependent spectral features (monaural cues). Due to the way in which acoustic waves interact with physical obstacles — such as the human head — ITD and ILD

are the salient spatial cues for low- and high-frequency sounds respectively. Interaural difference cues are limited to resolving the left-right dimension of the sound source position, but wideband sounds provide monaural cues that can reveal both the elevation angle as well as the front-back dimensions of sound sources.

In addition, spatial hearing is strongly affected by cross-modal interactions between audition and other sensory modalities. Proprioception interacts saliently with audition during self-motion, as movement induces systematic, source-position-dependent spatial cue dynamics that enable the directional ambiguities inherent to binaural cues to be resolved even if monaural cues are not available. Accordingly, artificially inserted binaural self-motion cues can evoke strong percepts of externalization for headphone stimuli that would otherwise be perceived to be located inside the head. Dynamic low-frequency ITD in specific has been shown to provide a robust self-motion cue that greatly enhances spatial perception but the usefulness of self-motion-induced dynamic ILD — or level dynamics in general — in active localization tasks is not clear and merits further investigation. While self-motion cues can enhance spatial perception, they can also introduce directional biases as demonstrated by the dynamic front-back illusion.

Finally, when multiple spectrally overlapping sounds are active at the same time, the signals at the eardrums consist of the sum of the signals emitted by each of the sources. When the sources are temporally incoherent, the resultant spatial cues are unstable from moment-to-moment; As such, they are better described statistically, than as a single fixed value. The lack of stable spatial cues distorts various aspects of spatial perception but many of these distortions remain uncharacterized. Median plane distributions represent an interesting but underexplored scenario in spatial hearing as the directional perception of such distributions has to rely on distorted monaural cues since ITD and ILD are uninformative of source elevation.

3. Neural processing of spatial cues

While the previous chapter focused on the relationship between acoustic phenomena in the physical domain and their corresponding spatial percepts in human listeners, this chapter elucidates how auditory information is processed along the neural pathways that eventually implement auditory perception. To this end, the neural computations performed by the relevant subcortical nuclei are described and the role of the auditory cortex is discussed in preparation for the next chapter on auditory scene analysis — the cognitive process of perceptual organization of sound.

3.1 Subcortical processing of binaural cues

The neural stations of the ascending auditory pathway pertinent to spatial hearing are shown schematically in Fig. 3.1. After sounds are decomposed into auditory bands in the cochlea, the tonotopically organized cochlear outputs project to the cochlear nucleus (CN) via the auditory nerve. Two major neural pathways ascend from the CN: The dorsal division of the CN (DCN) projects to the contralateral inferior colliculus (IC) via the lateral lemniscus (LL) and the ventral division of the CN (VCN) projects bilaterally to both the ipsi- and contralateral superior olivary complexes (SOC) (Grothe et al., 2010). The SOC represents the first stage of the auditory pathway where signals from the two ears are combined. It contains nuclei specialized in performing comparisons of interaural timing- (in the medial division of the superior olive, MSO) and level differences (in the lateral superior olive, LSO) with high temporal precision (Goldberg and Brown, 1969; Brand et al., 2002; Tollin and Yin, 2005; Grothe et al., 2010; Grothe and Pecka, 2014; Yin et al., 2019). As such, it serves as the main neural station for subcortical processing of binaural difference cues.

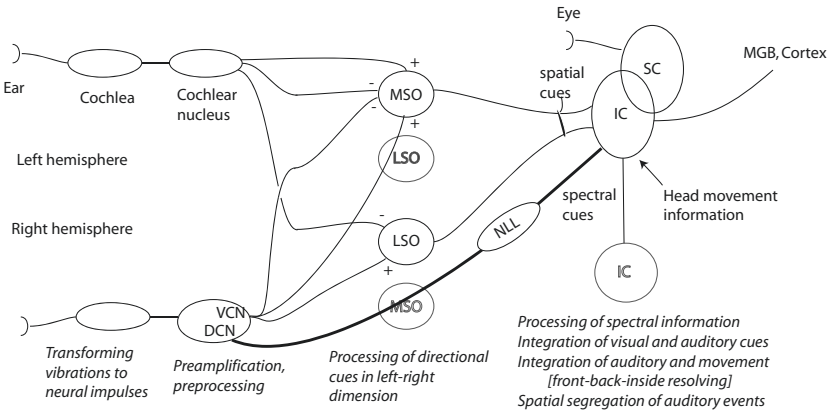

Figure 3.1. Schematic representation of the ascending auditory pathway. For clarity, stations of the pathway are shown for one side of the pathway only. Adopted from Pulkki and Karjalainen (2015).

3.1.1 Extraction of ITD in the MSO

The MSO is the main ITD-sensitive neural substrate in the SOC. Neurons in the MSO receive both excitatory and inhibitory inputs from both CNs; Whereas the excitatory inputs are direct projections from the CNs, the inhibitory connections are relayed through the trapezoid body (Cant and Casseday, 1986; Cant and Hyson, 1992; Smith et al., 1993; Kapfer et al., 2002; Yin et al., 2019). At the level of single-cell recordings, the outputs of low-frequency MSO-neurons have been shown to display tightly phase-locked activation in response to pure tone stimulation in a frequency-selective manner, as well as sensitivity to differences in interaural timing (for instance, Galambos et al., 1959; Yin and Chan, 1990; Batra et al., 1997; Pecka et al., 2008). Accordingly, MSO-neurons are characterized by a characteristic frequency and delay, that together define the combination of stimulus parameters that result in the strongest neuronal activation.

A large body of neurophysiological evidence gathered across mammalian species (for instance, gerbil: Brand et al., 2002; Pecka et al., 2008; Day and Semple, 2011, cat: Hancock and Delgutte, 2004, chinchilla: Bremen and Joris, 2013, see also Joris et al., 2006) shows that the range of ITDs represented by MSO-neurons varies as a function of the neuron's characteristic frequency. In specific, delays are represented only up to the so-called π-limit, the temporal delay corresponding to half of the period of the neuron's characteristic frequency (McAlpine et al., 2001; Franken et al., 2015). This frequency-dependent distribution of delay sensitivities suggests that the MSO processes interaural timing differences as auditory-band-specific IPDs, rather than as ITDs (McAlpine et al., 2001).

3.1.2 Extraction of ILD in the LSO

The LSO represents the main neural substrate responsible for extracting interaural level difference cues. LSO-neurons receive excitatory and inhibitory inputs from the ipsilateral and contralateral CNs, respectively (Galambos et al., 1959; Boudreau and Tsuchitani, 1968; Spangler et al., 1985; Cant and Casseday, 1986; Sanes, 1990). This complementary synapsing appears to make the LSO well-suited for evaluating disparities in interaural level. Accordingly, neurophysiological recordings have shown LSO-neurons to be sensitive to ILD (e.g. Tollin and Yin, 2002; Tollin et al., 2008). In addition, cells in the LSO display weak sensitivity to ITDs in the amplitude envelopes of binaural signals (for instance, Joris and Yin, 1995; Joris, 1996; Joris and Yin, 1998), implicating the LSO as a possible neural station for the extraction of envelope-delay cues.

3.2 Subcortical processing of monaural cues

Monaural spatial cues manifest as direction-dependent modifications to the spectral envelope formed at the eardrum. In specific, the sharp spectral notches formed by the out-of-phase summation of sound waves reflected within the pinna cavities are thought to represent the salient acoustic-domain signal features that the auditory system exploits in deducing sound source elevation (Xie, 2013). Accordingly, a neural substrate specialized in gathering spatial information from the sound spectrum should display sensitivity to narrow-band spectral minima within broadband sounds. Neurons that display such characteristics have been discovered in the DCN and the central nucleus of the inferior colliculus (ICC); The overall cue extraction process appears to take place via complementary activity in these two substrates (Grothe et al., 2010).

Physiological recordings in the cat auditory system have identified neurons in the DCN, whose response characteristics appear suitable for deriving directional information from the monaural spectrum (Young et al., 1992; Imig et al., 2000; May, 2000; Davis et al., 2003; Oertel and Young, 2004; Yin et al., 2019). These neurons show constant, frequency-independent inhibitory responses (relative to the spontaneous firing rate of the neuron) to pure tones, but display strong frequency-selectivity to wideband stimuli containing a spectral notch. Specifically, the neuron's output is excitatory when the spectral notch is not aligned with the characteristic frequency of the neuron, but when the two are aligned the activity of the neuron is inhibited and approaches its near-threshold activity (Davis et al., 2003). This inhibitory sensitivity to frequency-dependent spectral notches in wideband sounds appears to be the earliest stage of neural analysis that is well-suited for extracting spatial information from the monaural

spectrum.

The monaural projections from DCN-neurons connect with neurons in the ICC (Davis et al., 2003). The ICC-neurons targeted by the DCN-neurons display a similar, frequency-independent inhibitory response to pure tones as the DCN-neurons, but their response to wideband stimulation containing a spectral notch is reversed relative to the DCN-neurons. Namely, the neurons in the ICC are inhibited by spectral notches that are not aligned with the characteristic frequency of the neuron and conversely, exhibit excitatory responses to notches aligned with the characteristic frequency of the neuron (Aitkin and Martin, 1990; Delgutte et al., 1999; Davis et al., 2003). It has been suggested that the cascade of complementary activation of the DCN- and ICC-neurons enhances the frequency-selectivity of the neural notch identification process (Davis et al., 2003; Yin et al., 2019).

3.3 Convergence of neural pathways in the IC

The IC is a mandatory station along the auditory pathway that appears to facilitate many aspects of auditory perception. Overall, its role is not as clearly defined as those of e.g. the SOC-nuclei (for instance, see Casseday et al., 2002 for a review of the various functions associated with the IC). Importantly, the IC represents the first station along the ascending auditory pathway, where the neural streams associated with ITD, ILD and monaural cues converge on a common substrate for the first time. The IC receives inputs from both the ipsi- and contralateral SOCs, as well as a direct projection from the contralateral CN (Adams, 1979; Brunso-Bechtold et al., 1981; Grothe and Park, 2000; Davis et al., 2003; Grothe et al., 2010). The inputs from LSO- and DCN-neurons arrive to common neural segments in the IC (Oliver et al., 1997; Loftus et al., 2004), suggesting that the spatial information furnished by these neural streams (ILD and monaural cues) may be integrated upon arrival. Moreover, many cells in the IC have been shown to be sensitive to sound source direction in a cue-invariant manner (Chase and Young, 2008) supporting the cue-integration hypothesis. Nevertheless, some degree of cue modality separation appears to be retained, as some projections from the LSO and MSO associated with a common auditory band do not converge in the IC (Loftus et al., 2004). This suggests that information about conflicting cue modalities may be available beyond the level of the IC.

Cells in the IC across many mammalian species have been shown to be responsive to changes in various binaural signal features (for instance, puretone IPD and ILD in the cat: Rose et al., 1966, and gerbil: Spitzer and Semple, 1993, 1998, carrier- and envelope ITDs of wideband sounds in the guinea pig: Agapiou and McAlpine, 2008) at modulation rates exceeding hundreds of Hz (Joris, 1996; Joris et al., 2006). This indicates that the

auditory pathway maintains a high temporal resolution representation of binaural parameters at least until the level of the IC. However, this is not fully exploitable at the behavioral level as shown by the fact that performance in various listening tasks decreases rapidly already when binaural parameters are modulated at much slower rates (see for instance, Grantham and Wightman, 1978a,b; Grantham, 1982, 1984; Culling and Summerfield, 1995b, 1998; Akeroyd and Summerfield, 1999; Nassiri and Escabí, 2008; Kolarik and Culling, 2009). Thus, despite the high temporal precision of binaural processing at the brainstem level, perception of binaural parameters appears to unfold at a much slower rate.

3.4 Neural bases of cross-modal effects in spatial hearing

The IC is also connected with the superior colliculus (SC) — a brainstem nucleus that is not a part of the mandatory auditory pathway, but nevertheless plays an important role in the overall processing of auditory information. Neural projections associated with various sensory modalities (e.g. audition, vision, somatosensation) converge at the SC (Gordon, 1973; Palmer and King, 1982; Meredith and Stein, 1986), implicating it as an important subcortical nucleus in processing cross-modal sensory information. Accordingly, SC-neurons are organized topographically, so that both auditory and visual neurons tuned to a common spatial location are found close together in the layout of the SC (for a concise review of multimodal spatial maps in the SC, see e.g. King, 2004).

Some of the information furnished by the auditory and vestibular systems is combined already at the periphery (e.g. Burian and Gstoettner, 1988) and recent neurophysiological evidence suggests that vestibulo-auditory integration in the DCN disambiguates source-movement-induced changes in auditory spatial information from those induced by self-motion (Wigderson et al., 2016). Similarly, the IC receives projections from neural centers associated with motor function (Adams, 1980; Olaźbal and Moore, 1989; Moriizumi and Hattori, 1991), and at least some degree of auditory information processing appears to be modulated by somatosensory and motor inputs in the IC (Casseday et al., 2002). Overall, due to the multimodal nature of self-motion cues, the complete formation of auditory self-motion cues is likely to be achieved in a distributed manner in multiple processing stages along the auditory pathway.

3.5 Cortical processing

As mentioned previously, low-level binaural parameters are represented in the subcortical auditory pathway with high-temporal precision but percep-

tion appears to operate at a much slower time-scale. While this apparent inconsistency may make the role of the auditory cortex in spatial hearing seem ambiguous, sluggish responses at the perceptual level may nevertheless prove beneficial for facilitating successful directional perception in behaviorally relevant scenarios. In natural soundscapes, the instantaneous values of acoustic-domain auditory cues can vary unpredictably (due to e.g. reverberation or source interference in multi-source scenes), without providing meaningful information about the positions of sound sources (Młynarski and Jost, 2014; Pavão et al., 2020). Therefore, sluggish perceptual responses to sporadically varying binaural parameters may facilitate accurate spatial hearing by effectively down-weighting the instantaneous values of spatial cues that may provide unreliable information.

Although spatial cues are extracted already at subcortical levels, functional auditory cortices are nevertheless required for spatial hearing. This is demonstrated by studies where the deactivation of the auditory cortex in behaving animals (for instance, temporarily by cooling or permanently by ablation) results in the impairment or total loss of the ability to localize sounds (see e.g. Jenkins and Masterton, 1982; Thompson and Cortez, 1983; Jenkins and Merzenich, 1984; Heffner and Heffner, 1990; Heffner, 1997; Malhotra et al., 2004; Lomber and Malhotra, 2008 for studies conducted in various species). Similarly, human subjects with damaged auditory cortices display severely impaired performance in spatial hearing tasks (Zatorre and Penhune, 2001). Despite their importance, the cortical mechanisms that facilitate spatial hearing and auditory perception in general are not fully understood; As such, they represent an active field of on-going research (see King et al., 2018 and van der Heijden et al., 2019 for recent reviews).

While subcortical recordings indicate rapid processing of auditory information, cortical responses to changes in auditory stimuli appear to operate at much more sluggish rates (for instance, Dajani and Picton, 2006; Picton, 2013). To account for the apparent inconsistency in temporal-responses between the brainstem and the cortex, a two-stage model of binaural processing has been proposed wherein brainstem nuclei process inputs at a high temporal resolution and project their outputs to sluggish cortical processes. According to this model, the task of the cortical processes is to integrate the low-level parameters provided by subcortical stations into coherent, higher-level, perceptual representations of sounds, i.e. auditory objects and streams (see Chapter 4). It has been suggested that the formation of these cognitive representations takes place at the level of the primary auditory cortex (A1) (Griffiths et al., 2000; Fishman et al., 2001; Nelken, 2004; Micheyl et al., 2005, 2007; King and Nelken, 2009; Shamma and Micheyl, 2010; Christison-Lagay et al., 2015; Alain et al., 2017; Lu et al., 2017; King et al., 2018) and that higher-level processes along the putative "where" and "what" pathways in the belt- (A2) and parabelt (A3)

areas of the auditory cortex (Rauschecker et al., 1995; Rauschecker, 1997, 1998; Kaas and Hackett, 1999; Romanski et al., 1999; Belin et al., 2000; Rauschecker and Tian, 2000; Zatorre et al., 2004; Ahveninen et al., 2006; Hackett, 2015; Perrodin et al., 2015; Puvvada and Simon, 2017; King et al., 2018; Retsa et al., 2018; Shiell et al., 2018) allocate perceptual properties, such as spatial position and sound category to these representations. This functional division resembles the neural processing streams found in the somatosensory and visual pathways, where low-level stimulus features are carried to the relevant primary sensory cortical areas through the thalamus and processed further along separate cortical pathways that appear to show distinct specialization to spatial- and stimulus identification tasks (Mishkin, 1979; Ungerleider, 1982; Murray and Mishkin, 1984; Morel and Bullier, 1990; Lomber and Malhotra, 2008).

Chapter 3 summary

In summary, the primary spatial cues (ITD, ILD and monaural cues) are extracted separately for the different auditory bands in distinct subcortical nuclei. Importantly, the different modalities of spatial cues are processed in parallel neural pathways up to the level of the IC, where the projections from the brainstem nuclei associated with spatial cue extraction converge on a common neural substrate for the first time. Auditory information is modulated by cross-modal information — including self-motion — at multiple subcortical levels. The brainstem nuclei project to the thalamus and finally to the auditory cortex, where the low-level information extracted at subcortical levels seem to be sluggishly integrated into coherent auditory percepts.

4. Perceptual organization of sound

The two previous chapters discussed behavioral and neural aspects of spatial hearing, underlining the fact that due to the lack of direction-sensitive receptors in the peripheral organs auditory spatial information has to be extracted computationally from implicit cues. However, the task faced by the auditory system —- constructing accurate mental representations of the sound sources in the environment — is ill-posed at a more fundamental level than just spatial perception. Namely, the tonotopic organization of the cochlear receptors does not provide a reliable basis for deriving explicit information about the identity or quantity of sound sources that produced the signals received at the ears. A comparison with vision elucidates the issue. In vision, information about objects and their positions in space is available already at the receptor level. Light reflected off the surfaces of physical objects arrives at the topographically arranged photosensitive cells covering the retina. This activates the retinal array in a way that carries information both about the direction-of-arrival of the light, as well as the shape and dimensions of the reflecting object. Moreover, individual objects in multi-object visual scenes are separable from each other at the receptor level, since each object tends to stimulate a distinct subset of receptors in the retinal array. In contrast, sounds arriving at the ears are combined in the ear canal and often stimulate overlapping receptors in the tonotopic cochlear array. Therefore, much like spatial properties, also the identities of sounds have to be deduced from implicit cues in the acoustic mixture received at the ears — a cognitive process known as auditory scene analysis, or perceptual organization of sound. The computational difficulties imposed by this task often remain underappreciated, perhaps due to the remarkably rare instances where the process fails in everyday listening scenarios.

The following chapter presents an overview of selected topics in behavioral and neuroscientific research of auditory scene analysis. Since scene analysis represents a relatively large sub-field of auditory research, the scope of the chapter is limited and the focus is on the role of spatial cues in the perceptual organization of sound. The interested reader is encouraged

to consult Bregman (1994); Darwin (1997); Moore and Gockel (2002); Carlyon (2004); McDermott (2009); Micheyl and Oxenham (2010b); Shamma et al. (2011); Moore and Gockel (2012) for reviews on human psychophysics, Feng and Ratnam (2000); Hulse (2002); Bee and Micheyl (2008); Fay (2008) for overviews of animal studies and Nelken (2008); Nelken and Bar-Yosef (2008); Bidet-Caulet and Bertrand (2009); Bizley and Cohen (2013); Simon (2015); Snyder and Elhilali (2017); King and Walker (2020) for reviews of neuroscientific scene analysis research.

4.1 Auditory objects and streaming

Psychophysics is the field of science concerned with characterizing the mappings between properties of physical stimuli and sensory percepts. In the case of vision and audition, the concept of "sensory objects" provides a useful conceptual framework for describing sensory experience, as many aspects of sensation emerge only as attributes of perceptual objects. For instance in audition, loudness, spatial position and timbre are not perceivable in isolation, but only as perceptual attributes characterizing a cognitive representation of a sound — or an auditory object. Although an exact definition of sensory objects might be elusive, Griffiths and Warren (2004) propose four principles that describe object analysis by any sensory modality:

I: " [. . .] *object analysis involves the analysis of information that corresponds to things in the sensory world.*"

II: " [. . .] *object analysis involves the separation of information related to the object and information related to the rest of the sensory world.* "

III: " [. . .] *object analysis involves the abstraction of sensory information so that information about an object can be generalized between particular sensory experiences in any one sensory domain,* [. . .] "

IV: " [. . .] *object analysis involves generalization between senses,* [. . .] "

In the case of audition, the above principles can be interpreted as follows. I: Auditory object analysis involves the analysis of information about sound sources in the environment. II: Auditory object analysis involves the perceptual separation of soundscapes into their constituent sources. III: Auditory object analysis involves the abstraction of acoustic information into higher-level cognitive representations, so that the identity assigned to sounds (e.g. human voice, siren, bird song, etc.) is invariant to changes in its low-level parameters. For instance, a sound perceived as a human voice

retains its identity as a voice despite modest variations in pitch, amplitude and timbre. IV: Auditory object analysis enables auditory percepts to be associated with the appropriate percepts in other sensory modalities, e.g. the sound of breaking glass is mentally associated to the sight of a shattered window.

In scene analysis literature the terms auditory object and auditory stream are sometimes used interchangeably. Here, the concept of an auditory stream is used to refer to a sequence of sounds that are grouped together perceptually across time and perceived to originate from the same physical source (Moore and Gockel, 2002; Shamma and Micheyl, 2010; Moore and Gockel, 2012). In contrast, the term "auditory object" refers to the individual perceptual tokens of audition that may together constitute a stream. For instance, a spoken sentence or a melody played on a musical instrument consists of a sequence of discrete sound events — syllables and musical notes, auditory objects — but perceptually, the entire sequence is associated with a single auditory stream.

4.2 Auditory grouping cues and natural sound statistics

Because receptor-level activity in the auditory periphery is uninformative about the identities of sound sources in the surrounding environment, the auditory system is faced with the task of deriving this information implicitly. Based on the ear canal signals, the task is mathematically ill-posed since an arbitrarily large number of source configurations can yield the same ear canal signals (Bregman, 1994). Yet, perception of complex soundscapes is highly correlated with the veridical arrangement of the physical sources in the scene, indicating that correct perceptual organization is achieved with surprising ease. Psychophysical studies have revealed that a large number of stimulation parameters contribute to the perceptual organization of sounds. These auditory grouping cues include for instance, onset- and offset times, harmonicity, fundamental frequency, frequency separation, power- and phase-spectra, spatial separation, amplitude- and pitch envelopes (see Bregman, 1994; Darwin, 1997; Moore and Gockel, 2002; Carlyon, 2004; McDermott, 2009; Micheyl and Oxenham, 2010a; Shamma et al., 2011; Moore and Gockel, 2012 for reviews). The relative salience of some grouping cues is heavily context-dependent, and their contribution to perceptual organization may depend on the combination of all cues available at any given moment as well as the behavioral goals of the listener (Moore and Gockel, 2012). Despite the complexity of the perceptual organization task, many aspects of the underlying processing principles can be elucidated by considering the statistical properties of natural sounds.

Empirical analyses of natural soundscapes show that statistical regu-

larities govern various aspects of the structure of sounds emerging in the natural world (for instance, ambient noise: Waser and Brown, 1986, amplitude envelopes: Attias and Schreiner, 1997, modulation spectra: Singh and Theunissen, 2003, timbral texture: McDermott and Simoncelli, 2011; McDermott et al., 2013; Mishra et al., 2020; Zhai et al., 2020, spectral decay of acoustic sounds: Voss and Clarke, 1975; De Coensel et al., 2003 and reverberation: Traer and McDermott, 2016a,b, time-frequency-domain activation patterns: Młynarski and McDermott, 2019, binaural difference cues: Młynarski and Jost, 2014; Pavão et al., 2020 and monaural spectra: Parise et al., 2014). Crucially, these analyses reveal that sounds encountered in natural listening scenarios represent a severely restricted subset of all possible sounds. The visual system appears to have developed adaptations that optimize the processing of statistical regularities of natural scenes and accommodate successful behavior in the visual world (see Kayser et al., 2004; Geisler, 2008; De Cesarei et al., 2017, for reviews). Similar adaptations to natural stimulus statistics would be beneficial also in the auditory domain, as heuristic processing principles derived from the statistical properties of natural sounds constrain the set of feasible solutions to the ill-posed scene analysis problem.

For example, it is statistically unlikely that independent sound sources would be activated in synchrony unless purposefully coordinated to do so, as in the case of e.g. string ensembles or other musical elements. Consequently, psychophysical tests show that coherent temporal activations in the form of common onset- and offset times and correlated amplitude envelope fluctuations represent salient perceptual grouping cues (Bregman, 1978; Darwin, 1984; Hall et al., 1984; Hall III and Grose, 1990; Roberts and Moore, 1991; Hukin and Darwin, 1995a; Elhilali et al., 2009). Similarly, the spectrum of many natural sounds (e.g. voiced speech and animal vocalizations) consists of harmonic overtones whose frequencies are integer multiples of the fundamental frequency. Accordingly, harmonicity (McAdams, 1982; Moore et al., 1986; Buell and Hafter, 1991; Darwin and Ciocca, 1992), pitch (Broadbent and Ladefoged, 1957; Assmann and Summerfield, 1990), timbre and spectral structure in general (Wessel, 1979; Assmann and Summerfield, 1989; Roberts and Brunstrom, 1998; Vliegen and Oxenham, 1999; Roberts and Brunstrom, 2001; Roberts et al., 2002), are strong cues for evaluating whether different frequency bands should be bound together into a single perceptual entity or not. Overall, it appears that the most salient cues to perceptual organization are derived from spectro-temporal information that can be detected and decoded under monaural listening conditions.

4.3 The role of spatial cues in perceptual organization

The role of spatial cues in scene analysis is not as clear as it may first appear. Intuitively, one could assume that spatial separation or co-location are effective cues for segregation or fusion of sound components; Afterall, spatially separated sounds are unlikely to originate from the same source under natural conditions. However, if the auditory system has developed adaptations to process natural scenes in specific, the reliability of spatial information in such scenes should be taken into consideration. Unlike in the case of laboratory experiments that are often conducted under anechoic or virtual acoustic conditions, most real hearing tasks take place in physical spaces bounded by reflective surfaces. These introduce delayed sound reflections into the ear canal signals that may have the effect of degrading the fidelity of the spatial information available to the listener (Shinn-Cunningham, 2005; Młynarski and Jost, 2014; Joris and van der Heijden, 2019; Pavão et al., 2020). Accordingly, binaural cues may not represent a robust enough signal parameter for reliable scene analysis judgments in natural acoustic environments (Darwin, 2005). Many behavioral studies have assessed the salience of spatial cues in scene analysis tasks with mixed results, depending on the combination of available cues and the requirements of the experimental task. Below, pertinent results from experiments involving both instantaneous- and sequential grouping tasks are reviewed.

4.3.1 Instantaneous grouping

Grouping by laterality

Several experiments have assessed the effect of lateralization in auditory grouping. While ITD, ILD and their combinations, can be used for lateralization, fully dichotic, “ear-of-entry” stimulus presentation is commonly used in perceptual grouping studies. In this stimulation paradigm, each ear receives a unique signal free of cross-talk from the signal presented over the other earphone. This produces interaural difference magnitudes that do not normally emerge under natural listening conditions; Namely, an infinite ILD — since the signal level at the contralateral ear is zero, and an undefined ITD — since the contralateral ear does not receive a delayed replica of the ipsilateral signal.

Some aspects of auditory perception appear to be unaffected by fully dichotic stimulation. For example, dichotically presented partials of a harmonic complex are fused together by the auditory system to form a coherent perception of pitch (Beerends and Houtsma, 1986, 1989; Darwin and Ciocca, 1992). Similarly, when the first two formant frequencies of synthetic vowels are presented dichotically, the resultant percept corre-

sponds to the vowel category implied by the formant combination, unless the signals presented to the two ears differ also in pitch (Broadbent and Ladefoged, 1957). These studies suggest that lateralization by ear-of-entry does not introduce obligatory perceptual segregation at least in the case of vowel and pitch perception.

Studies using double-vowel stimuli suggest that ear-of-entry can in some scenarios be leveraged as a grouping cue, if doing so facilitates the auditory task. Double-vowel stimuli are based on combinations of band-pass noise bursts whose center-frequencies correspond to the first two formant frequencies of vowels. Despite their abstract construction, such stimuli are perceived as belonging to the vowel category characterized by the formant combination. Ambiguous double-vowel stimuli can be devised by selecting four such noise bands, whose different pairings correspond to different vowel categories and presenting all four bands simultaneously (Culling and Summerfield, 1995a). Since any combination of the four formant bands yields a valid vowel category, the perceived identities of the concurrent noise-band vowels depends on which pairs of noise are grouped together perceptually. When bursts of double-vowel stimuli are presented so that the vowel category grouping is biased by ear-of-entry, subjects are able to consistently report the identities of the vowels according to laterality. In contrast, when ear-of-entry cues are not available and vowel grouping is instead biased by perceptually similar lateralization induced by large ITDs only, subjects are no longer able to perform the task (Culling and Summerfield, 1995a), unless extensive task-specific training is undertaken (Drennan et al., 2003). This suggests that ITDs represent a weak grouping cue that can not be effectively deployed even when doing so would facilitate the task. Similarly, when a single harmonic positioned along a phoneme boundary of a voiced vowel is presented to one ear, contralaterally to the rest of the partials, the contribution of the contralateral harmonic to the phoneme identity of the overall stimulus decreases relative to diotic conditions (Hukin and Darwin, 1995b; Darwin and Hukin, 1997). Yet, when similar lateralization percepts are imposed with large ITDs, the contribution of the contralateral harmonic to overall vowel identity is restored, implicating ITD as a weaker grouping cue than ear-of-entry (Hukin and Darwin, 1995b; Darwin and Hukin, 1997).

Overall, behavioral evidence suggests that ear-of-entry cues can be leveraged as an effective grouping cue to perceptual organization, if doing so facilitates the task. Yet, when spectro-temporal cues are absent and ITD is the only grouping cue available, spatial information is much more difficult to leverage and binaural cues appear to be weakly weighted in object formation.

IPD/ITD-driven perceptual segregation phenomena

When stimuli are presented over headphones, ITD and ILD can be assigned independently to yield combinations of binaural cues that imply source positions in opposite directions. When such a manipulation is imposed on a stimulus, the resultant auditory image is typically perceived at a position between the two lateral positions implied by the conflicting cues; A phenomenon known as time-intensity trading (see Deatherage and Hirsh 1959 for a review of early studies). Crucially, time-intensity trading demonstrates that even when ITD and ILD are incongruent — a feasible cue to the presence of two distinct sound sources — the majority of listeners perceive a single auditory image rather than two, supporting the idea that binaural disparities are weakly weighted in object formation. Yet, some studies have reported subsets of listeners who perceived stimuli with incongruent ITD and ILD as two distinct images (for instance, Banister 1926; Whitworth and Jeffress 1961; Hafter and Jeffress 1968), indicating that the weighting of spatial cues in perceptual organization may not be consistent across listeners.

Some psychoacoustic phenomena that break the trend of spatial cues being weakly weighted in instantaneous perceptual grouping seem to be specifically driven by temporal differences. As mentioned in Chapter 2, antiphasic stimuli presented over headphones form two separate point-like images with extreme lateralizations to the two ears if the stimulus provides low-frequency ITD-cues. This splitting of the auditory image into two separate percepts was reported already in the first half of the 20th century by several investigators (for instance, Wilska 1938; Hirsh 1948; Licklider 1948). IPD-cues are salient drivers of perceptual segregation also in Huggins- or dichotic pitch stimuli. Huggins pitch refers to the pitch sensation evoked by binaural noise samples that are otherwise diotic but contain an interaural phase disparity restricted to a narrow frequency band (Cramer and Huggins, 1958). These stimuli evoke a pitch sensation corresponding to the center-frequency of the phase disparity. The pitched component is perceptually segregated from the diotic noise component and emerges only under binaural listening (Culling et al., 1998). Since the effect is evoked by IPD, it is most salient in the low-end of the auditory range, where phase-locked IPD is available; Accordingly, the effect becomes gradually imperceivable at higher frequencies, where phase-locking degrades (Culling, 1999). Another line of evidence for the role of temporal differences in perceptual organization comes from behavioral studies assessing the perception of concurrent pairs of uncorrelated noise bursts. When listeners are presented with VAS-stimuli consisting of two spatially separated bursts of uncorrelated broadband noise yielding temporally unstable spatial cues, listeners systematically report perceiving two separate azimuthal directions (Best et al., 2004). Yet, when temporal difference cues are removed from the stimulation, listeners no longer report percepts

at two directions, suggesting that on-going ITD-dynamics are crucial to facilitating perceptual organization.

Overall, binaural cues within the naturally occurring range appear to be weakly weighted in many short-time-scale grouping tasks, but seem to be salient drivers of perceptual organization in some special instances involving unnatural IPD-values. Moreover, on-going ITD-dynamics appear to provide a salient perceptual cue to the perceptual organization of broadband noise bursts devoid of spectro-temporal grouping cues.

4.3.2 Sequential grouping

The role of spatial cues in perceptual organization becomes more pronounced when they co-occur with spectro-temporal grouping cues (Buell and Hafter, 1991; Shackleton and Meddis, 1992; Shackleton et al., 1994; Hukin and Darwin, 1995a; Darwin and Hukin, 1997; Darwin, 1997). This suggests that despite their relatively weak weighting in the short-time-scale process of object formation, spatial cues nevertheless play a crucial role in facilitating successful behavior in natural soundscapes (e.g. attending to a single speaker in a crowd), where spectro-temporal grouping cues are typically available and accompanied with spatial information. Accordingly, a large body of evidence from psychophysical studies shows that the advantage yielded by spatial information becomes especially meaningful in behaviorally relevant listening tasks, such as speech perception in complex "cocktail party" (Cherry, 1953) listening. For example, when speech is presented concurrently with another spectrally overlapping sound, the intelligibility of the speech target is significantly better when the two sounds are spatially separated than when they are co-located (Shinn-Cunningham, 2005; Litovsky, 2012). Moreover, spatial position represents a preferred grouping cue in multi-talker scenes. For instance, when presented with two concurrent spoken sentences differing in ITD, listeners tend to group words according to a common ITD, rather than according to pitch cues (Darwin and Hukin, 1999). While the benefits of static binaural cues in the spatial unmasking of speech are well established (see Bronkhorst, 2000; Freyman et al., 2001; Litovsky, 2012; Leibold et al., 2019 for overviews), here the focus is on studies leveraging other experimental paradigms.

Melody- and rhythm identification tasks

One class of studies that demonstrates the role of interaural disparities as an effective driver of sequential perceptual grouping employs melody identification tasks. When listeners are tasked to identify two temporally interleaved pure tone melodies, identification rates are drastically better when the two melodies are separated by IPD or by ear-of-entry, than when the melodies are presented diotically (Hartmann and Johnson, 1991). Similarly, when a tone complex consisting of sinusoids corresponding to the

frequencies of notes in a musical scale is presented binaurally, individual components of the complex can be biased to stand out perceptually from the on-going background complex by modulating the IPDs of individual sinusoids. By applying such modulations sequentially, binaurally encoded melodies that are not perceivable monaurally, can be embedded into tone complex stimuli (Kubovy et al., 1974; Kubovy and Daniel, 1983; Culling, 2000). Similar results have been reported with Huggins pitch sequences embedded into binaural noise stimuli (Dougherty et al., 1998).

Another line of evidence for the salient role of spatial cues in sequential organization comes from rhythmic masking release (RMR) experiments. In the RMR-paradigm, two rhythms formed from e.g. broadband noise burst sequences are presented simultaneously, so that their concurrent presentation obscures the temporal patterns of both sequences. RMR-experiments are often designed so that the experimental task is facilitated by perceptual segregation of the interleaved rhythms into separate streams. For instance, subjects may be tasked to discriminate between two known target rhythms in the presence of another random masking rhythm (Turgeon et al., 2002). Sach and Bailey (2004) showed that while subjects could not identify target rhythms consisting of tone bursts presented over headphones when the target was co-located with an arrythmic masker sequence, performance in the identification task improved significantly when the intracranial spatial separation of the target and masker sequences was increased by introducing binaural difference cues between the sequences. Similarly, in free-field stimulation using broadband noise bursts, spatial separation as small as 8 azimuthal degrees between the masker- and target streams can yield consistently successful performance in the RMR-task (Middlebrooks and Onsan, 2012). A larger spatial separation is needed between the target and the masker sequences for high-frequency broadband stimuli that do not provide fine-structure IPD-cues (Middlebrooks and Onsan, 2012), suggesting that low-frequency interaural timing cues facilitate the task more effectively than spatial cues derived from high-frequency channels.

Although spatial cues can yield significant advantages in sequential grouping tasks, they appear to be deployable in a selective manner. This is demonstrated in experiments employing auditory tasks that are expected to become more difficult if perceptual organization follows the grouping implied by the spatial information; When engaged in such tasks, it is advantageous to ignore the spatially driven segregation cues. Studies assessing the role of ITD in obligatory stream segregation show that listeners can effectively ignore ITD-based grouping cues if attending to them hinders performance the psychophysical task (for instance, Boehnke and Phillips, 2005; Stainsby et al., 2011).

Motion-driven effects in auditory grouping

In the visual domain, movement represents a salient grouping cue and visual objects whose motion is not congruent with the background or surrounding objects tend to stand out perceptually (e.g. Abrams and Christ, 2003). Given the salience of motion-cues in visual grouping, the enhancements provided by spatial information in various auditory tasks involving sequential grouping and the significant role of self-motion cues in behaviorally relevant listening, it seems plausible that motion-driven grouping effects could manifest also in the auditory domain. Yet, only relatively few studies have addressed such effects in hearing; A selection of these studies is reviewed below.

One hypothetical auditory-motion-driven perceptual effect is related to spatial release from masking with moving target sounds. When a target sound moves against a background of spatially distributed masking sounds, the motion of the target could potentially serve as a grouping cue that makes the target sound stand out from the background sounds, analogously with the visual motion "pop-out" effect. To this end, Pastore and Yost (2017) measured the motion-induced improvements to word recognition rates for single-word speech stimuli presented against multiple concurrent masking talkers. However, no differences in recognition rates were found for static and moving target conditions, suggesting that the enhancements to perceptual organization by source motion are less salient than analogous effects in the visual domain.

Despite the apparent lack of source-motion-driven spatial masking release, it is possible that the spatial cue dynamics associated with self-motion could have a more pronounced effect on auditory perceptual organization than identical dynamics yielded by source motion. Since the work of Wallach (1939, 1940), it has been acknowledged that a meaningful interpretation of spatial cue dynamics requires information about how the dynamics emerged. Under self-motion, this information is readily furnished by non-auditory sensory systems (e.g. the vestibular system) but it is unavailable during source motion unless the movements of the source are controlled by the listener (e.g. Wightman and Kistler, 1999). Accordingly, self-motion cues have been shown to be more effective than source motion cues at providing information about sound scenes involving dynamic spatial cues. For instance, listeners are better at describing the relative positions of two concurrent sounds during self-motion than during source motion, even when the auditory inputs are identical under the two conditions (Brimijoin and Akeroyd, 2014). Therefore, it is possible that perceptual organization effects driven by motion-cues are more salient when they arise from self-motion than from source-motion.

One example of self-motion effects in auditory perceptual organization comes from behavioral studies using the ABAX-streaming paradigm (van Noorden, 1975). ABAX-stimuli consist of a pair of short tones or bursts

of narrow-band noise (A and B) followed by an silent period (X). Each element in the sequence (A, B and X) are of equal temporal duration. The perceptual organization of repeated ABAX-sequences is strongly affected by the frequency separation between the A- and B-segments of the stimulation. When A and B are close in frequency, the stimulation sequence is perceived as a single auditory stream forming a distinct triplet rhythm. In contrast, with large frequency intervals, the A- and B-segments form two separate auditory streams unfolding in isochronous rhythms. With intermediate frequency intervals, perception of ABAX-stimuli is bistable and a single stream percept may "build-up" to a segregated percept across a period of a few seconds; The segregated percept may then rapidly "reset" to a single stream percept if a sudden change in the stimulation parameters is introduced (see e.g. Carlyon, 2004; Moore and Gockel, 2012 and the references therein). Kondo et al. (2012) used ABAX-stimuli to measure the effects of motion-cues in scene analysis. By using a head-tracked VAS-system, it was possible to decouple listener movements from the spatial cue dynamics that would emerge under natural listening conditions. Listeners gave subjective reports of streaming during trials consisting of bistable ABAX-stimuli. In some trials, subjects were instructed to perform head movements while the stimuli were being presented. The results from these trials show that when self-motion was initiated during two-stream percepts, it was accompanied by a rapid resetting of auditory organization back to the single-stream triplet ABAX-rhythm (Kondo et al., 2012). Interestingly, this effect occurred regardless of whether or not self-motion was accompanied with spatial cue dynamics, suggesting that self-motion modulates auditory perceptual organization even in the absence of changes in auditory information (Kondo et al., 2012).

Another experimental paradigm that appears especially well-suited for revealing self-motion-driven effects in auditory perceptual organization is the dynamic front-back illusion discussed in Chapter 2. Although the paradigm has been leveraged in many behavioral studies in recent years (e.g. Macpherson, 2011; Brimijoin and Akeroyd, 2012; Macpherson, 2013; Brimijoin and Akeroyd, 2016; Yost et al., 2019), the focus of these studies has been on localization only and many of them have used stimuli that carry salient spectro-temporal grouping cues. It is therefore possible that self-motion-driven perceptual re-organization effects might not have emerged simply because such effects may have been masked by the salient monaural grouping cues. For instance, Brimijoin and Akeroyd (2012, 2016) implemented the dynamic front-back illusion with speech stimuli; A class of auditory stimuli that is strongly biased towards perceptual fusion of its constituent spectral components by a number of signal parameters (e.g. common envelope and pitch modulations, harmonicity, semantic content, etc.). Moreover, due to the temporal structure of speech, the short-term spectrum of speech sounds is modulated on a moment-to-moment basis

(Moore, 2012). These modulations have the secondary effect of varying the relative salience of the three major localization cues ITD, ILD and the monaural spectrum, making it unclear what spatial cues are available at any instance. Therefore, it is not clear to what extent monaurally perceivable, spectro-temporal grouping cues dominated perceptual organization in these experiments and whether or not the role of spatial cue dynamics associated with self-motion would be more pronounced for non-speech stimuli. As such, the potential of the dynamic-front back illusion in auditory perceptual organization studies appears to be underexploited. Identification of auditory re-organization effects driven by self-motion-induced spatial cue dynamics could be facilitated by selecting the experimental task so that the results are informative also about non-spatial aspects of perceptual organization and by using stimuli that are devoid of salient spectro-temporal grouping cues. Such considerations could potentially reveal self-motion-driven perceptual effects that might have been undetectable under the conditions of previous experiments.

4.4 Neural correlates of perceptual organization

In the field of auditory neuroscience, much work has been done to identify neural correlates of the cognitive processes associated with auditory perceptual organization (e.g. Alain et al., 2002; Dyson and Alain, 2004; Alain, 2007; Bendixen et al., 2010; Kocsis et al., 2016; Tóth et al., 2016a, see also Simon, 2015 as well as Snyder and Elhilali, 2017 for reviews). Identifying the neural markers associated with various aspects of auditory perception is not only important for advancing basic auditory research, but may also find valuable applications in the development of hearing diagnostics (e.g. Dougherty et al., 1998) and neural interfaces for hearing devices (for instance, Perron, 2017; Bech Christensen et al., 2018).

One line of this research is concerned with identifying the extracranial electromagnetic correlates of concurrent auditory object perception. Accordingly, it has been found that the neural responses associated with stimuli that evoke percepts of multiple auditory objects differ from those associated with single-object percepts (Snyder and Elhilali, 2017). For instance, in the time-domain representations of event-related responses obtained via EEG or magnetoencephalography, concurrent object perception is associated with an “object-related negativity” (ORN, Alain et al., 2001; Alain, 2007). This electromagnetic marker refers to the magnitude difference in the dominant waves (N1 and P2) of the event-related responses associated with stimuli that are acoustically similar, but nevertheless evoke different perceptual organizations. For example, tone complex stimuli can be biased to segregate into multiple auditory objects by slightly detuning or delaying individual partials in the complex (Snyder and Elhilali, 2017). ORN can

be evoked by a wide range of segregation-promoting cues (see for instance, Alain et al., 2001, 2002; Johnson et al., 2003; Hautus and Johnson, 2005; McDonald and Alain, 2005; Sanders et al., 2008; Bendixen et al., 2010; Kocsis et al., 2016; Tóth et al., 2016a for studies using a wide range of parameter manipulations), regardless of the listener's age or attentional state (Alain and Izenberg, 2003; Alain, 2007; Alain and McDonald, 2007; Folland et al., 2012; Bendixen et al., 2015). These consistencies suggest that ORN indexes a general neural process associated with auditory perceptual organization.

In addition to time-domain measures, time-frequency-domain markers of concurrent object perception have also been investigated. The notion that human brain function is mediated by oscillations of neuronal activity (Buzsáki and Draguhn, 2004; Buzsáki, 2006) suggests that electromagnetic correlates of perceptual organization could potentially be identified using time-frequency decomposition of neural responses. Accordingly, Tóth et al. (2016b) investigated whether neuronal oscillations captured by EEG could be used to index concurrent auditory object perception. When perceptual organization was manipulated by introducing segregation-promoting cues (mistuning, asynchronous onsets, spatial cues) between the partials of tone complex stimuli, bursts of increased θ-band (4 - 8 Hz) activity were found to be correlated with concurrent object perception (Tóth et al., 2016b). This was observed both under passive and active listening conditions, suggesting that they signal the activation of primitive scene analysis processes functioning across attentional states.

Aside from extracranial electromagnetic correlates of concurrent object perception in humans, intracellular recordings in a variety of animal species have demonstrated spatially driven enhancements in neural responses at different stages of the auditory pathway. For instance, recordings from cat primary auditory cortex show spatially selective responses to RMR-stimuli; When interleaved rhythmic sequences are presented from opposite hemifields, cortical neurons synchronize preferentially to one of the constituent rhythms according to its spatial location (Middlebrooks and Bremen, 2013). Crucially, the spatial tuning of these responses is sharper than what is commonly observed in cortical neurons under single-source stimulation, providing neural evidence for spatially driven enhancements to perceptual organization (King and Middlebrooks, 2011; Middlebrooks and Bremen, 2013). Sharpening of spatial selectivity in A1-neurons has been reported also in other species (see e.g. Yao et al., 2015 for recordings in the rat auditory system). Spatial cues have also been shown to increase the fidelity of neuronal representation of spectro-temporally overlapping sounds. For instance, recordings in the rat IC show that when two spatially separated narrow-band noises with the same center-frequency are presented simultaneously, the IC-responses synchronize preferentially to the temporal fine-structure of the ipsilateral noise (Luo et al., 2017). Simi-

larly, when a narrow-band noise target is presented simultaneously with a spectrally overlapping masker noise, the envelope of cortical responses displays better synchronization to the envelope of the target sound when the target and masker are spatially separated than when they are co-located (Xu et al., 2019; Luo et al., 2020). All in all, results from animal studies suggest that spatial cues facilitate perceptual organization by enhancing the fidelity of neural representations of the constituent sources at multiple stages of the auditory pathway.

4.4.1 Spectro-temporally complex stimulation paradigms

While the classical experimental paradigms of auditory scene analysis often involve simplistic stimuli that bear little resemblance to the complexity of natural soundscapes, recent years has seen the development of spectro-temporally complex stimuli that are suitable for assessing neural markers of perceptual organization in complex soundscapes (e.g. tone-cloud masker stimuli, Gutschalk et al., 2008, stochastic figure-ground stimuli, Teki et al., 2011; Tóth et al., 2016a, see also Snyder and Elhilali, 2017). These stimuli provide better approximations of the complexity of real auditory scenes than classical stimuli (e.g. ABAX-sequences). As such, they represent a compromise between the methodological constraints imposed by non-invasive neuroimaging methods and the on-going strive for improved ecological validity in auditory neuroscience (e.g. Griffiths et al., 2004; Nelken, 2004, 2008; King and Walker, 2020). However, all of these stimulation paradigms are driven by monaurally emergent grouping cues that necessarily co-exist with the spatial cues that may be embedded in the stimulation. As such, they are not ideal for neuroscientific experiments targeting binaural processes in specific as the contributions of monaural and binaural processes cannot be reliably detangled from the neural responses. Given the salient role that spatial cues play in the perceptual organization of complex soundscapes, it would be beneficial to establish a spectro-temporally complex stimulation paradigm for use in neuroscientific experiments that provides grouping cues specific to binaural hearing.

Such stimulation could potentially be based on Huggins pitch stimuli, as they have the appealing property of providing no monaural grouping cues and can therefore reliably isolate binaural contributions to auditory perceptual organization. Accordingly, these stimuli have been used in EEG-studies to index the electromagnetic correlates of binaurally driven concurrent auditory object perception (Johnson et al., 2003; Hautus and Johnson, 2005). However, since Huggins pitch is driven by IPD, it is necessarily restricted to the limited range of low frequencies where the auditory system can decode temporal fine structure (Culling, 1999). This is a major limitation to the ecological validity of the stimulation, as natural sounds encompass a much wider range of frequencies that need to be

correctly grouped together to identify and localize sound sources in the environment. To overcome the limitations imposed by IPD-based grouping cues, the novel binaural stimulation paradigm needs to leverage interaural information beyond the IPD. Recently, theoretical accounts of auditory perceptual organization have emphasized the role of temporal coherence across various stimulus features in the formation of auditory percepts (e.g. Elhilali et al. 2009; Shamma and Micheyl 2010; Shamma et al. 2013). As such, they offer a potentially useful conceptual framework for advancing the development an auditory stimulus capable of indexing binaurally driven perceptual organization processes. According to the predictions of the "temporal coherence theory" of auditory scene analysis, it seems plausible that time-frequency-specific modulations of interaural envelope correlation could potentially be leveraged to develop auditory stimuli with spectro-temporally complex grouping cues that are only defined binaurally.

Chapter 4 summary

This chapter has presented some of the problems faced by the auditory system in perceptual organization tasks. Receptor-level information is uninformative about the sources of sound present in the environment and the task of identifying the individual sources turns into an ill-posed computational problem. It appears that the way the auditory system solves this issue is by relying on heuristics derived from the statistical properties of natural sounds.

The implication of this is that due to reverberation and other acoustic-domain phenomena that degrade the reliability of spatial information in natural soundscapes, many aspects of short time scale perceptual organization are driven by non-spatial signal characteristics. Nevertheless, the ITD-dynamics arising from acoustic-domain interference of concurrent sound sources appear to provide an effective cue to perceptual organization. Moreover, spatial information becomes increasingly important in listening tasks that involve perceptual grouping of sound sequences across time (e.g. speech perception) and multi-source scenes. Therefore, spatial hearing provides a crucial advantage in behaviorally relevant listening tasks and it is plausible that auditory self-motion cues could further contribute to the perceptual organization of sound. Yet, relatively few studies have investigated such effects, revealing a gap in scene analysis research. To this end, the dynamic front-back illusion represents a promising experimental paradigm for probing self-motion-driven effects in perceptual organization.

Finally, despite the importance of spatial cues in scene analysis tasks, their role remains underexplored in auditory neuroscience experiments assessing the perceptual organization of complex soundscapes. A possible reason for this is the lack of an established stimulation paradigm that is

both specific to binaural hearing and provides binaural grouping cues of sufficient spectro-temporal complexity.

5. Summaries of the studies

The studies contained in this thesis investigated the effects of spatial cue dynamics on the perceptual organization of sound. In specific, **PI** explored the role of self-motion cues in the perceptual organization of noise stimuli devoid of non-spatial grouping cues. The experiments in **PII** were designed to answer open questions about the role of self-motion-induced level dynamics in active localization. **PIII** assessed the suitability of a previously unexplored class of binaurally driven stimuli for use in neuroscientific scene analysis experiments. **PIV** characterized the spatial perception of complex source distributions in the horizontal- and median planes. A summary of each study is provided below.

5.1 Study I - Conflicting dynamic and spectral directional cues form separate auditory images

The significant role that self-motion cues play in auditory spatial perception was acknowledged already in the early 20th century by several authors but their role in auditory perceptual organization remains largely unexplored. Here, we conducted a behavioral experiment employing a continuous psychoacoustics paradigm where subjects used horizontal head rotations to form impressions of simple sound scenes consisting of one or two sources presented over loudspeakers. We implemented the dynamic front-back illusion using head-tracking cameras and amplitude panning to move sound sources within a circular array of loudspeakers according to the subject's instantaneous head orientation. This enabled us to create self-motion-induced front-back confusions by distorting the relationship between head movements and the resulting spatial cue dynamics. In order to promote the emergence of perceptual re-organization effects associated with self-motion, we used steady-state noise stimuli devoid of salient spectro-temporal grouping cues, the presence of which could potentially obscure self-motion-induced effects.

The results show that low-pass filtered stimuli providing mainly dynamic

low-frequency ITD-cues were consistently localized to the hemiplane implied by the binaural dynamics arising from head rotations, regardless of the actual location of the sound source. Conversely, high-pass filtered stimuli providing dynamic ILD and monaural cues were localized to the correct hemiplane and were seemingly unaffected by the manipulations imposed on the spatial cue dynamics. In the case of wideband stimuli providing both dynamic ILD and ITD as well as high-frequency monaural cues, subjects systematically reported simultaneous percepts in both the front- and rear-hemiplanes, despite the fact that the stimulus was presented only from the frontal hemiplane.

While previous research on auditory self-motion cues has mainly focused on localization, the results of **PI** suggest that self-motion-driven spatial cue dynamics affect the interpretation of auditory scenes at a more fundamental level than previously considered. Scenarios where different modalities of spatial cue dynamics are in conflict with each other may result in categorically different perceptual organizations than scenarios where these cues are congruent. For instance, here conflicting self-motion cues and high-frequency monaural cues resulted in a single sound source being perceived as two separate sources in opposite hemiplanes, according to the directions implied by the conflicting cue modalities. This suggests that in addition to affecting spatial perception, self-motion-driven spatial cue dynamics contribute to the fundamental processes of object formation and stream segregation in a manner that has not been considered previously.

5.2 Study II - Resolving front-back ambiguity with head rotation: The role of level dynamics

Self-motion enhances spatial hearing by providing dynamic spatial cues that can resolve otherwise ambiguous spatial information, such as the front-back location of narrow-band sounds. Previous experiments on self-motion effects in spatial hearing have established dynamic ITD as salient cue modality but the role of ILD-dynamics in active localization remains unclear. The purpose of **PII** was to address the ambiguous role of level dynamics in spatial hearing. To this end, we conducted a set of four behavioral experiments, where we used head-tracking and real-time signal processing to restrict the head rotation range listeners could exploit in determining the front-back location of spatially ambiguous pure tone stimuli. Specifically, the range of head orientations where the target signal was audible was limited by complementarily ramping up a spatially diffuse noise masker and ramping down the target signal when head orientation exceeded experimentally imposed limits. Based on this paradigm, we assessed the usefulness of narrow-band level dynamics in resolving front-back ambiguity.

In experiment I, subjects indicated the front-back dimension of sinusoidal stimuli presented from the midline (0 or 180 degrees) under free-field conditions. Task performance was assessed as a function of the allowed head rotation range. In experiment II, subjects performed a similar task for free-field sources with various lateral offsets and a fixed head rotation range of ±20 degrees. Experiments III and IV were headphone-based replications of experiments I and II in simplistic virtual auditory space, where the binaural dynamics were derived from the acoustic transfer functions of a rigid sphere, approximately the size of an average human head.

The results of the free-field experiments show that listeners were capable of exploiting level dynamics to successfully resolve the front-back dimension of high-frequency sinusoids if a sufficiently wide movement range was available (±40 degrees). When the free-field conditions were replaced by simplistic headphone stimulation, front-back responses were in agreement with the simulated source directions even with relatively small movement ranges (±5 degrees) whenever monaural sound level and ILD changed monotonically in response to head rotation. In conclusion, level dynamics can be leveraged as a front-back cue when they are approximately monotonic and provide consistent spatial cue dynamics during self-motion. However, in free-field conditions and particularly for narrowband target signals, this is often not the case. All in all, **PII** suggests that the usefulness of self-motion-coupled level dynamics are limited by confounding acoustic-domain phenomena, rather than by the processing principles of the auditory system.

5.3 Study III - Cortical processing of binaural cues as shown by EEG-responses to random-chord stereograms

A large body of behavioral evidence shows that spatial hearing facilitates successful perceptual organization of complex soundscapes. Nevertheless, the role of binaural cues in auditory scene analysis has received relatively little attention in recent neuroscientific studies striving for improved ecological validity by employing spectro-temporally complex stimuli. A possible reason for this may be that an appropriate stimulation paradigm suitable for use in neuroscientific experiments has not yet been established. Ideally, such a paradigm would be specific to binaural hearing, so that the contributions of monaurally driven perceptual organization processes could be reliably detangled from the neural responses. Additionally, the paradigm should be flexible enough to provide binaurally derived grouping cues of arbitrary spectro-temporal complexity. Potentially, such a paradigm could be based on random-chord stereograms (RCS, Nassiri and Escabí, 2008); a class of auditory stimuli that leverages IPD and interaural en-

velope correlation to evoke salient auditory re-organization effects that emerge only under binaural listening. While RCS-stimuli appear promising, their usefulness in non-invasive measurements of cortical activity has not yet been evaluated.

Here, our aim was to assess the usability of the RCS-paradigm as a stimulus category for indexing cortical processing of binaural cues. We recorded EEG-responses to RCS-stimuli consisting of an initial 3-s noise segment, where the envelopes of the binaural channels were uncorrelated, followed by another 3-s segment, where interaural envelope correlation was modulated periodically as per the RCS-paradigm. Two types of modulation were used: In the case of "wideband stimuli", modulations encompassed the entire stimulus bandwidth. In the case of "ripple stimuli", modulations were applied to shifting frequency bands according to a spectro-temporal ripple. EEG-responses were recorded from a group of normal-hearing subjects to wideband stimuli at two modulation rates — 3 and 5 Hz — and to one ripple stimulus modulated at 3 Hz.

Event-related potentials and inter-trial phase coherence analyses show that EEG-responses at the fronto-central electrodes synchronized to wideband RCS-modulations. In the case of ripple stimuli, the transition from noise to the modulated segment induced a change-onset complex, but the steady-state response during the on-going modulations did not synchronize to the modulation rate of the stimulus. In the case of wideband modulations, frequency-domain analyses revealed spectral power increases at frequencies related to the RCS-modulations. Overall, **PIII** shows that the RCS-paradigm can yield robust cortical responses that reliably index binaurally driven effects in auditory perceptual organization.

5.4 Study IV - Spatial perception of sound source distribution in the median plane

Studies of auditory spatial perception have traditionally focused on simplistic scenarios involving point-like sources. Yet, real soundscapes often contain non-point-like volumetric sources of various spatial configurations extending both in horizontal and vertical dimensions. Such sound sources pose a challenge to spatial hearing since the directional cues they provide are not stable across time but instead, vary unpredictably due to the superposition of the sound energy arriving to the ears from the different sections of the source distribution. Consequently, spatially distributed sound sources yield spatial cue dynamics that do not arise from listener or source movement but rather from acoustic-domain interference between multiple temporally uncorrelated radiators of sound. Currently, spatial perception of such sources is poorly understood and remains largely uncharacterized especially in the case of vertical source distributions.

Here, we investigated the accuracy of human directional hearing in the case of spatially distributed sources in two behavioral experiments. To this end, we used horizontal and vertical loudspeaker arrays to create various source configurations emitting uncorrelated pink noise. The task of the subjects was to describe the resultant spatial percepts by indicating the directions of loudspeakers that corresponded to the perceived directions of sound. Listeners could not perform head movements to facilitate the task. We were especially interested in exploring spatial perception of vertically extended source distributions positioned along the auditory midline, where the contributions of binaural cues are negligible and assessment of spatial properties relies predominantly on monaural signal characteristics.

In the case of horizontally distributed sources, the findings of **PIV** are similar to previous results reported in Santala and Pulkki (2011), where a similar experimental task was employed but head movements were not prohibited. This suggests that self-motion cues do not provide significant enhancements to the perception of azimuthal distributions in the frontal hemiplane. In **PIV**, subjects did not perceive small gaps in horizontal source distributions, but when the azimuthal separation between source pairs or clusters was sufficiently wide, subjects reported separate distinct directions, suggesting perceptual segregation. In addition, horizontal distributions tended to be perceived as narrower than the span of the physical distributions. Overall, the results from horizontal distributions show that listeners are able to make use of unstable spatial cues in auditory perceptual organization, even if the short-term cue dynamics are unreliable. In the case of vertical distributions, the results show that spatial perception of median plane source distributions is qualitatively different from that of horizontal distributions. For distributions consisting of multiple sources, subjects tended to severely underestimate the number of active sources in the distributions, reporting only two or three directions for source configurations spanning angular areas as large as ±45 degrees in elevation. However, the results demonstrate that subjects were nevertheless able to reliably identify pairs of median plane sources if the angular separation between the sources was sufficiently large. This suggests that the temporally unstable spatial cues arising from the acoustic-domain interaction of the concurrent sources facilitated successful perceptual organization despite the fact that the resultant ear canal spectra did not represent the directional spectrum associated with either source elevation.

6. Discussion

Overall, the results from the studies included in this thesis show that spatial cue dynamics contribute significantly to the perceptual organization of sound. Below, the implications of the results are discussed in the context of spatial audio technologies and basic auditory research.

6.1 Technological applications

The goal of many spatial sound technologies is to evoke spatial impressions that are indistinguishable from reality. Arguably, there are two main categories of spatial synthesis technologies with fundamentally different approaches to achieving this goal; Namely, physically driven methods and perceptually driven methods. While physically driven methods, such as wavefield synthesis (Snow, 1955; Berkhout et al., 1993; Ahrens, 2012) and ambisonics, (Gerzon, 1973; Zotter and Frank, 2019) aim to replicate the physical properties of the target sound field as accurately as possible, perceptually driven methods (e.g. DirAC, Pulkki, 2007) are motivated by psychophysics. These methods capitalize on the perceptual limitations of the auditory system in the design of digital signal processing algorithms that do not seek to reproduce the sound field perfectly, but rather to replicate only its perceptually relevant spatial features. Accordingly, perceptually driven synthesis techniques are designed according to targets that are often more feasible to achieve than a perfect reconstruction of the physical sound field.

While perceptually indistinguishable reproduction of reality has been the main driver for much of the work in spatial sound technology for the past decades, the emergence of ubiquitous virtual reality and augmented hearing technologies introduces new targets both for spatial sound technologies and basic research on spatial hearing. An ambitious future goal for binaural reproduction techniques is to achieve a level of sophistication that enables the evocation of spatial percepts in complex auditory scenes that would not necessarily ever emerge naturally under the constraints

imposed by the physics of sound in the real world. Development of such *hyper-real* spatial audio techniques requires that the systematic perceptual biases associated with spatial hearing and related cross-modal effects can be identified, quantified and compensated for. Crucially, successful development of such techniques requires a paradigm shift wherein technological goals are defined in terms of arbitrary auditory percepts, rather than benchmarked against the percepts experienced in real soundscapes.

Hyper-real spatial synthesis algorithms have already emerged in the context of spatial cue dynamics in head-tracked, headphone-based spatial synthesis. For example, recent work by Brimijoin (2018) and Brimijoin et al. (2020) showed systematic angle-dependent distortions in the perceived angular displacement of real sound sources rotated in a circular loudspeaker array. By quantifying these spatial distortions across azimuth, the systematic warping of auditory space could be compensated for, enabling the synthesis of hyperstable auditory stimuli; Such sounds are perceived to be more stable during self-motion than real sound sources. Some aspects of the experimental results documented in the current work merit a discussion in the hyper-reality context.

The results of **PII** suggest that the auditory system does not maintain frequency dependent azimuth-to-ILD mappings to facilitate the processing of narrow-band ILD dynamics emerging during head movements. Rather, the results imply that spatial decoding of dynamic ILD information relies on the assumption that the overall ILD computed across the auditory bands associated with any given auditory object varies monotonically with lateral position. While this is generally not valid for narrow-band sounds, such a processing principle would still enable successful behavior in the majority of ecologically relevant listening scenarios, since narrow-band sounds represent a rather limited subset of sounds encountered in natural environments. Further, the results of the headphone experiments in **PII** show that in the case of narrow-band sounds, simplistic ILD dynamics appear to enhance, rather than degrade auditory spatial perception. This could potentially be exploited in hyper-real spatial synthesis of narrow-band sounds in augmented- or virtual reality devices. For instance, if it is detected that the localization of a given target sound — be it virtual or a binaurally delivered, augmented version of a real sound in the environment — is mostly driven by high-frequency narrow-band ILD, spatial perception of that sound can likely be enhanced by applying simplistic azimuth-to-ILD mappings, rather than a set of real HRTFs. Such simplified mappings could be derived from e.g. the azimuth-dependent wideband ILD of the listeners true HRTFs.

While the results of **PII** suggest that the spatial cue dynamics delivered by augmented hearing devices do not have to fully correspond to those arising naturally during listener movement, the results of **PI** show that this aspect cannot be completely ignored either. In **PI**, the relationship between

spatial cues and head movements was manipulated artificially by using a head-tracking apparatus to control a panning algorithm. The manipulation revealed that conflicting spatial cue modalities arising during an active localization task consistently resulted in a fundamentally erroneous perceptual organization of the auditory scene. In specific, low-frequency ITD-sequences manipulated according to the dynamic front-back illusion resulted in systematic hemiplane reversals of low-frequency sounds. When the same manipulations were applied to wideband sources providing both dynamic binaural cues as well as spectral cues, the interpretation of the auditory scene was erroneous at a more fundamental level, as subjects consistently reported percepts of two simultaneous auditory images occupying opposite hemiplanes even though the stimulation consisted of a single wideband source. These results have immediate implications for spatial synthesis algorithms in virtual- and augmented reality applications. Perhaps more importantly, they — together with the results of **PII** — also underline the importance of preserving sufficiently accurate spatial cue dynamics in bilateral hearing aids. The distortions imposed on the binaural signals by hearing aids (e.g. compression, delay, spectral distortion) should be evaluated by considering their perceptual implications in behaviorally relevant listening scenarios, where the listener is in motion and experiences dynamic spatial cues at least as often as static ones. Failure to do so may lead to situations where the auditory percepts experienced by the users of such devices are severely distorted.

Finally, the results of **PIV** show that the spatial percepts evoked by various sound source distributions deviated from the spatial properties of the actual source distributions, especially when the sources were positioned in the median plane. Since the acuity of spatial hearing was demonstrated to be rather poor in these scenarios, their reproduction in the perceptually motivated spatial synthesis paradigm is relatively straight-forward. In the context of **PIV**, hyper-real synthesis of complex distributions would compensate for the spatial distortions associated with distributed sources and evoke auditory spatial percepts that closely match the spatial properties of the source arrangement even if such percepts would not emerge under real-world conditions. While the results of **PIV** do not make it obvious how this could be achieved, it is clear that obtaining a deeper understanding of the salient signal features that drive perceptual organization of complex soundscapes is part of the solution.

6.2 Towards improved ecological validity in auditory perceptual organization studies

Although scene analysis is a relatively mature field of auditory research, it has mainly focused on revealing the dominant low-level processing

principles deployed in simplistic listening scenarios involving stationary listeners and point-like sources. Because successful detection of many salient auditory grouping cues does not require binaural hearing, spatial cues or their dynamic variations have rarely been the focus of perceptual organization experiments. As such, the experimental scenarios of many scene analysis studies are far removed from the listening tasks faced in real life. In an attempt to open up new avenues in behavioral and neuroscientific research on auditory perceptual organization, the studies contained in the present work incorporated aspects of real-world listening that have previously received little attention in scene analysis research.

The results from the experiments documented in **PI** & **PII** show that self-motion cues yield a strong influence on how auditory scenes are perceived both spatially and at the more fundamental level of object- and stream formation. Given that listeners are typically in motion when engaged in behaviorally relevant listening, the incorporation of self-motion-driven listening tasks to behavioral experiments presents a fruitful direction for future studies on auditory perceptual organization. Such an approach would take scene analysis studies a step closer to ecological validity by acknowledging the fundamentally cross-modal nature of human audition.

Similarly, **PIV** showed that spatial cue dynamics arising from the acoustic-domain interference associated with concurrent sound sources plays a role in the perceptual organization of spatially complex soundscapes even when no other grouping cues are available. Crucially, **PIV** revealed that both horizontal and vertical spatial separation of spectrally overlapping sources contributed to how many distinct directions — and implicitly, how many separate auditory objects — the scene was divided into in the perceptual domain. This suggests that the role of spatial cues in the perceptual organization of ecologically relevant complex soundscapes merits further investigation.

Finally, due to a number of methodological constraints, auditory neuroscience has traditionally employed simplistic stimuli that bear little resemblance to real soundscapes. In recent years however, spectro-temporally complex synthetic stimuli that serve to improve the ecological validity of neuroscientific scene analysis experiments have become more common. Despite this development, no such stimulation paradigm has emerged that is able to index binaurally driven perceptual organization processes in specific. To this end, **PIII** evaluated the usefulness of random-chord stereograms as a potential stimulus category to fill this methodological gap, revealing that these stimuli evoke robust cortical responses that are attributable to binaural processing only. As such, the RCS-paradigm appears to present a promising means of introducing spectro-temporal complexity to neuroscientific scene analysis studies focused on binaurally driven perceptual organization processes.

7. Conclusions

The work documented in this thesis elucidates the role of spatial cue dynamics in experimental scenarios where salient monaural auditory grouping cues were not available. Spatial cue dynamics emerged in each experiment in one of three ways: 1) as a result of listener movement (**PI** & **PII**), 2) by synthetic binaural stimulation (**PIII**), or 3) by acoustic-domain interaction between spectrally overlapping, incoherent sound sources (**PIV**). Overall, spatial cue dynamics were found to have a salient effect on perceptual organization at the level of object-formation and spatial perception. The results have implications for various fields of auditory research and the development of spatial synthesis technologies.

PI investigated how a simple auditory scene is perceived when the spatial information delivered by monaural cues and dynamic binaural cues arising from listener self-motion are in conflict with one another. By leveraging the dynamic front-back illusion, it was found that the conflicting spatial cues resulted in the perception of two separate auditory images at directions implied by the two cue modalities. The results of **PI** indicate that cross-modal effects between the spatial auditory system and the sense of self-motion can have salient effects on the perceptual organization of sounds during listener movements. Accordingly, these effects should be accounted for in the design of audio processing algorithms for hearing devices and augmented reality systems with dynamic spatial synthesis engines.

PII explored the role of self-motion-induced level dynamics as a localization cue. It was found that front-back localization of high-frequency narrow-band targets is confounded by the naturally occurring head-related acoustics. Yet, when simplistic headphone stimulation was used to impose simplistic level dynamics, front-back discrimination performance increased substantially despite the unnaturalness of the stimulation. Therefore, spatial perception of high-frequency narrow-band targets is not limited by the processing principles of the auditory system itself, but rather, by the idiosyncratic spatial cue dynamics emerging in the acoustic domain. This suggests that in the case of augmented reality devices, and other electroacoustic systems where dynamically spatialized sound content is delivered

binaurally directly to the ear canals, the spatial impression evoked by high-frequency narrow-band targets can be made more robust by using simplified azimuth-to-ILD mappings, rather than the naturally occurring, band-specific ILD-dynamics embedded into human HRTFs.

PIII assessed the usefulness of random-chord stereograms in neuroscientific auditory scene analysis experiments. These stimuli are an attractive option for assessing cortical processing of binaural cues, as the auditory re-organization effects that they evoke are driven by binaural rather than monaural signal features. The results of **PIII** showed that simple variants of RCS-stimuli, where binaural envelope correlation is modulated periodically across the entire stimulus bandwidth evoked robust cortical responses. As such, the RCS-paradigm appears to provide a promising basis for the design of neuroscientific experiments seeking to elucidate the role of binaural processing in auditory scene analysis without introducing monaural confounds into the neural response data. Moreover, the RCS-paradigm could potentially be suitable for developing novel EEG-based diagnostics of binaural processing and facilitating the fitting of bilateral hearing devices.

PIV evaluated the accuracy of directional perception in spatially complex auditory scenes involving distributed sources along the median- and horizontal planes. Results from horizontal distributions were largely in line with those from a previous study employing the same stimulation and response paradigms despite the fact that in the previous study subjects were encouraged to move their heads and here static listening conditions were imposed. The similarity of the results suggests that in the case of spatially distributed sources in the frontal hemiplane, head movement cues do not provide salient enhancements to spatial perception. The median plane results revealed that spatial perception of multiple vertically separated sound sources is remarkably poor as indicated by e.g. the consistent underestimation of the number of active sources in multi-source vertical distributions. Nevertheless, listeners were able to indicate the vertical positions of two concurrent sources when the angular separation between them was sufficiently large, suggesting that at least two simultaneous directions can be decoded from temporally unstable monaural spectral cues. In short, the results from **PIV** show that spatial perception of vertically distributed median plane sources is considerably worse than in the case of horizontal distributions and point-like sources. These limitations can be capitalized on in spatial synthesis applications that seek to reproduce the spatial percepts evoked by real source distributions.

All in all, the results of the present work provide new insights into how dynamic spatial cues are integrated in the perceptual organization of sound and open up new lines of basic behavioral and neuroscientific research on auditory scene analysis. In addition, the results of the behavioral studies find practical applications in informing the design of spatial audio

algorithms in user-worn binaural devices and aiding the perceptually informed positioning of elevated audio channels in multichannel loudspeaker systems.

Bibliography

RA Abrams and SE Christ. Motion onset captures attention. *Psychol Sci*, 14(5): 427–432, 2003.

JC Adams. Ascending projections to the inferior colliculus. *J Comp Neurol*, 183 (3):519–538, 1979.

JC Adams. Crossed and descending projections to the inferior colliculus. *Neurosci Lett*, 19(1):1–5, 1980.

JP Agapiou and D McAlpine. Low-frequency envelope sensitivity produces asymmetric binaural tuning curves. *J Neurophysiol*, 100(4):2381–2396, 2008.

J Ahrens. *Analytic methods of sound field synthesis*. Springer Science & Business Media, 2012.

J Ahveninen, IP Jääskeläinen, T Raij, G Bonmassar, S Devore, M Hämäläinen, S Levänen, FH Lin, M Sams, BG Shinn-Cunningham, et al. Task-modulated "what" and "where" pathways in human auditory cortex. *Proc Natl Acad Sci USA*, 103(39):14608–14613, 2006.

L Aitkin and R Martin. Neurons in the inferior colliculus of cats sensitive to sound-source elevation. *Hear Res*, 50(1-2):97–105, 1990.

MA Akeroyd and Q Summerfield. A binaural analog of gap detection. *J Acoust Soc Am*, 105(5):2807–2820, 1999.

C Alain. Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear Res*, 229(1-2):225–236, 2007.

C Alain and A Izenberg. Effects of attentional load on auditory scene analysis. *J Cogn Neurosci*, 15(7):1063–1073, 2003.

C Alain and KL McDonald. Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. *J Neurosci*, 27(6):1308–1314, 2007.

C Alain, SR Arnott, and TW Picton. Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. *J Exp Psychol Hum Percept Perform*, 27(5):1072, 2001.

C Alain, BM Schuler, and KL McDonald. Neural activity associated with distinguishing concurrent auditory objects. *J Acoust Soc Am*, 111(2):990–995, 2002.

C Alain, JS Arsenault, L Garami, GM Bidelman, and JS Snyder. Neural correlates of speech segregation based on formant frequencies of adjacent vowels. *Sci Rep*, 7(1):1–11, 2017.

VR Algazi, RO Duda, and DM Thompson. Motion-tracked binaural sound. *J Audio Eng Soc*, 52(11):1142–1156, 2004.

DJ Anderson, JE Rose, JE Hind, and JF Brugge. Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects. *J Acoust Soc Am*, 49(4B):1131–1139, 1971.

F Asano, Y Suzuki, and T Sone. Role of spectral cues in median plane localization. *J Acoust Soc Am*, 88(1):159–168, 1990.

PF Assmann and Q Summerfield. Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency. *J Acoust Soc Am*, 85(1):327–338, 1989.

PF Assmann and Q Summerfield. Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J Acoust Soc Am*, 88(2):680–697, 1990.

H Attias and CE Schreiner. Temporal low-order statistics of natural sounds. In *Adv Neural Inf Process Syst*, pages 27–33, 1997.

H Banister. Three experiments on the localization of tones. *Br J Psychol*, 16(4): 265, 1926.

R Batra, S Kuwada, and DC Fitzpatrick. Sensitivity to interaural temporal disparities of low-and high-frequency neurons in the superior olivary complex. i. heterogeneity of responses. *J Neurophysiol*, 78(3):1222–1236, 1997.

DW Batteau. The role of the pinna in human localization. *Proc Royal Soc B*, 168 (1011):158–180, 1967.

BB Bauer. Phasor analysis of some stereophonic phenomena. *J Acoust Soc Am*, 33(11):1536–1539, 1961.

C Bech Christensen, RK Hietkamp, JM Harte, T Lunner, and P Kidmose. Toward EEG-assisted hearing aids: Objective threshold estimation based on ear-EEG in subjects with sensorineural hearing loss. *Trends Hear*, 22:2331216518816203, 2018.

MA Bee and C Micheyl. The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol*, 122(3): 235, 2008.

JG Beerends and AJM Houtsma. Pitch identification of simultaneous dichotic two-tone complexes. *J Acoust Soc Am*, 80(4):1048–1056, 1986.

JG Beerends and AJM Houtsma. Pitch identification of simultaneous diotic and dichotic two-tone complexes. *J Acoust Soc Am*, 85(2):813–819, 1989.

DR Begault, EM Wenzel, and MR Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J Audio Eng Soc*, 49 (10):904–916, 2001.

P Belin, RJ Zatorre, P Lafaille, P Ahad, and B Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–312, 2000.

A Bendixen, SJ Jones, G Klump, and I Winkler. Probability dependence and functional separation of the object-related and mismatch negativity event-related potential components. *Neuroimage*, 50(1):285–290, 2010.

A Bendixen, GP Háden, R Németh, D Farkas, M Török, and I Winkler. Newborn infants detect cues of concurrent sound segregation. *Dev Neurosci*, 37(2):172–181, 2015.

AJ Berkhout, D de Vries, and P Vogel. Acoustic control by wave field synthesis. *J Acoust Soc Am*, 93(5):2764–2778, 1993.

LR Bernstein and C Trahiotis. Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise. *J Acoust Soc Am*, 95(6):3561–3567, 1994.

LR Bernstein and C Trahiotis. Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli". *J Acoust Soc Am*, 112(3):1026–1036, 2002.

LR Bernstein and C Trahiotis. How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli. *J Acoust Soc Am*, 125(5):3234–3242, 2009.

V Best, A Van Schaik, and S Carlile. Separation of concurrent broadband sound sources by human listeners. *J Acoust Soc Am*, 115(1):324–336, 2004.

V Best, R Baumgartner, M Lavandier, P Majdak, and N Kopčo. Sound externalization: A review of recent research. *Trends Hear*, 24:2331216520948390, 2020.

A Bidet-Caulet and O Bertrand. Neurophysiological mechanisms involved in auditory perceptual organization. *Front Neurosci*, 3:25, 2009.

JK Bizley and YE Cohen. The what, where and how of auditory-object perception. *Nat Rev Neurosci*, 14(10):693–707, 2013.

J Blauert. Sound localization in the median plane. *Acustica*, 22:205–213, 1969.

J Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

J Blauert and W Lindemann. Spatial mapping of intracranial auditory events for various degrees of interaural coherence. *J Acoust Soc Am*, 79(3):806–813, 1986.

SE Boehnke and DP Phillips. The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences. *Percept Psychophys*, 67(6):1088–1101, 2005.

JC Boudreau and C Tsuchitani. Binaural interaction in the cat superior olive S segment. *J Neurophysiol*, 31(3):442–454, 1968.

A Brand, O Behrend, T Marquardt, D McAlpine, and B Grothe. Precise inhibition is essential for microsecond interaural time difference coding. *Nature*, 417 (6888):543–547, 2002.

AS Bregman. Auditory streaming: Competition among alternative organizations. *Percept Psychophys*, 23(5):391–398, 1978.

AS Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

P Bremen and PX Joris. Axonal recordings from medial superior olive neurons obtained from the lateral lemniscus of the chinchilla (chinchilla laniger). *J Neurosci*, 33(44):17506–17518, 2013.

WO Brimijoin. Angle-dependent distortions in the perceptual topology of acoustic space. *Trends Hear*, 22:2331216518775568, 2018.

WO Brimijoin and MA Akeroyd. The role of head movements and signal spectrum in an auditory front/back illusion. *i-Perception*, 3:179–181, 2012.

WO Brimijoin and MA Akeroyd. The moving minimum audible angle is smaller during self motion than during source motion. *Front Neurosci*, 8:273, 2014.

WO Brimijoin and MA Akeroyd. The effects of hearing impairment, age, and hearing aids on the use of self-motion for determining front/back location. *J Am Acad Audiol*, 27:588–600, 2016.

WO Brimijoin, WA Boyd, and MA Akeroyd. The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE*, 8:e83068, dec 2013.

WO Brimijoin, S Featherly, and P Robinson. Mapping the perceptual topology of auditory space permits the creation of hyperstable virtual acoustic environments. *Acoust Sci Technol*, 41(1):245–248, 2020.

F Brinkmann, M Dinakaran, R Pelzer, P Grosche, D Voss, and S Weinzierl. A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses. *J Audio Eng Soc*, 67(9):705–718, 2019.

DE Broadbent and P Ladefoged. On the fusion of sounds reaching different sense organs. *J Acoust Soc Am*, 29(6):708–710, 1957.

AW Bronkhorst. Localization of real and virtual sound sources. *J Acoust Soc Am*, 98(5):2542–2553, 1995.

AW Bronkhorst. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acust united Ac*, 86(1):117–128, 2000.

JK Brunso-Bechtold, GC Thompson, and RB Masterton. HRP study of the organization of auditory afferents ascending to central nucleus of inferior colliculus in cat. *J Comp Neurol*, 197(4):705–722, 1981.

TN Buell and ER Hafter. Combination of binaural information across frequency bands. *J Acoust Soc Am*, 90(4):1894–1900, 1991.

M Burian and W Gstoettner. Projection of primary vestibular afferent fibres to the cochlear nucleus in the guinea pig. *Neurosci Lett*, 84(1):13–17, 1988.

RA Butler and CC Helwig. The spatial attributes of stimulus frequency in the median sagittal plane and their role in sound localization. *Am J Otolaryngol*, 4 (3):165–173, 1983.

G Buzsáki. *Rhythms of the Brain*. Oxford University Press, 2006.

G Buzsáki and A Draguhn. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929, 2004.

NB Cant and JH Casseday. Projections from the anteroventral cochlear nucleus to the lateral and medial superior olivary nuclei. *J Comp Neurol*, 247(4):457–476, 1986.

NB Cant and RL Hyson. Projections from the lateral nucleus of the trapezoid body to the medial superior olivary nucleus in the gerbil. *Hear Res*, 58(1):26–34, 1992.

S Carlile, R Martin, and K McAnally. Spectral information in sound localization. *Int Rev Neurobiol*, 70:399–434, 2005.

RP Carlyon. How the brain separates sounds. *Trends Cogn Sci*, 8(10):465–471, 2004.

JH Casseday, T Fremouw, and E Covey. The inferior colliculus: A hub for the central auditory system. In *Integrative functions in the mammalian auditory pathway*, pages 238–318. Springer, 2002.

SM Chase and ED Young. Cues for sound localization are encoded in multiple aspects of spike trains in the inferior colliculus. *J Neurophysiol*, 99(4):1672–1682, 2008.

EC Cherry. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am*, 25(5):975–979, 1953.

KL Christison-Lagay, AM Gifford, and YE Cohen. Neural correlates of auditory scene analysis and perception. *Int J Psychophysiol*, 95(2):238–245, 2015.

EM Cramer and WH Huggins. Creation of pitch through binaural interaction. *J Acoust Soc Am*, 30(5):413–417, 1958.

KE Cullen. The vestibular system: multimodal integration and encoding of self-motion for motor control. *Trends Neurosci*, 35(3):185–196, 2012.

JF Culling. The existence region of Huggins' pitch. *Hear Res*, 127(1-2):143–148, 1999.

JF Culling. Auditory motion segregation: A limited analogy with vision. *J Exp Psychol Hum Percept Perform*, 26(6):1760, 2000.

JF Culling and Q Summerfield. Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *J Acoust Soc Am*, 98(2):785–797, 1995a.

JF Culling and Q Summerfield. The binaural temporal window. *Br J Audiol*, 29: 74–75, 1995b.

JF Culling and Q Summerfield. Measurements of the binaural temporal window using a detection task. *J Acoust Soc Am*, 103(6):3540–3553, 1998.

JF Culling, Q Summerfield, and DH Marshall. Dichotic pitches as illusions of binaural unmasking. i. Huggins' pitch and the "binaural edge pitch". *J Acoust Soc Am*, 103(6):3509–3526, 1998.

HR Dajani and TW Picton. Human auditory steady-state responses to changes in interaural correlation. *Hear Res*, 219(1-2):85–100, 2006.

CJ Darwin. Perceiving vowels in the presence of another sound: Constraints on formant perception. *J Acoust Soc Am*, 76(6):1636–1647, 1984.

CJ Darwin. Auditory grouping. *Trends Cogn Sci*, 1(9):327–333, 1997.

CJ Darwin. Simultaneous grouping and auditory continuity. *Percept Psychophys*, 67(8):1384–1390, 2005.

CJ Darwin and V Ciocca. Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *J Acoust Soc Am*, 91(6):3381–3390, 1992.

CJ Darwin and RW Hukin. Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *J Acoust Soc Am*, 102(4): 2316–2324, 1997.

CJ Darwin and RW Hukin. Auditory objects of attention: the role of interaural time differences. *J Exp Psychol Hum Percept Perform*, 25(3):617, 1999.

KA Davis, R Ramachandran, and BJ May. Auditory processing of spectral cues for sound localization in the inferior colliculus. *J Assoc Res Otolaryngol*, 4(2): 148–163, 2003.

ML Day and MN Semple. Frequency-dependent interaural delays in the medial superior olive: implications for interaural cochlear delays. *J Neurophysiol*, 106 (4):1985–1999, 2011.

A De Cesarei, GR Loftus, S Mastria, and M Codispoti. Understanding natural scenes: Contributions of image statistics. *Neurosci Biobehav Rev*, 74:44–57, 2017.

B De Coensel, D Botteldooren, and T De Muer. 1/f noise in rural and urban soundscapes. *Acta Acust united Ac*, 89(2):287–295, 2003.

BH Deatherage and IJ Hirsh. Auditory localization of clicks. *J Acoust Soc Am*, 31 (4):486–492, 1959.

B Delgutte, PX Joris, RY Litovsky, and TCT Yin. Receptive fields and binaural interactions for virtual-space stimuli in the cat inferior colliculus. *J Neurophysiol*, 81(6):2833–2851, 1999.

PL Divenyi and SK Oliver. Resolution of steady-state sounds in simulated auditory space. *J Acoust Soc Am*, 85(5):2042–2052, 1989.

RF Dougherty, MS Cynader, BH Bjornson, D Edgell, and DE Giaschi. Dichotic pitch: A new stimulus distinguishes normal and dyslexic auditory function. *Neuroreport*, 9(13):3001–3005, 1998.

WR Drennan, S Gatehouse, and C Lever. Perceptual segregation of competing speech sounds: The role of spatial location. *J Acoust Soc Am*, 114(4):2178–2189, 2003.

NI Durlach, A Rigopulos, XD Pang, WS Woods, A Kulkarni, HS Colburn, and EM Wenzel. On the externalization of auditory images. *Presence-Teleop Virt*, 1 (2):251–257, 1992.

BJ Dyson and C Alain. Representation of concurrent acoustic objects in primary auditory cortex. *J Acoust Soc Am*, 115(1):280–288, 2004.

M Elhilali, L Ma, C Micheyl, AJ Oxenham, and SA Shamma. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*, 61(2):317–329, 2009.

RR Fay. Sound source perception and stream segregation in nonhuman vertebrate animals. In *Auditory perception of sound sources*, pages 307–323. Springer, 2008.

WE Feddersen, TT Sandel, DC Teas, and LA Jeffress. Localization of high-frequency tones. *J Acoust Soc Am*, 29(9):988–991, 1957.

AS Feng and R Ratnam. Neural basis of hearing in real-world situations. *Annu Rev Psychol*, 51(1):699–725, 2000.

YI Fishman, DH Reser, JC Arezzo, and M Steinschneider. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res*, 151(1-2):167–187, 2001.

H Fletcher. Loudness, masking and their relation to the hearing process and the problem of noise measurement. *J Acoust Soc Am*, 9(4):275–293, 1938a.

H Fletcher. The mechanism of hearing as revealed through experiment on the masking effect of thermal noise. *Proc Natl Acad Sci USA*, 24(7):265, 1938b.

H Fletcher. Auditory patterns. *Rev Mod Phys*, 12(1):47, 1940.

NA Folland, BE Butler, NA Smith, and LJ Trainor. Processing simultaneous auditory objects: Infants' ability to detect mistuning in harmonic complexes. *J Acoust Soc Am*, 131(1):993–997, 2012.

TP Franken, MT Roberts, L Wei, NL Golding, and PX Joris. In vivo coincidence detection in mammalian sound localization generates phase delays. *Nat Neurosci*, 18(3):444–452, 2015.

RL Freyman, U Balakrishnan, and KS Helfer. Spatial release from informational masking in speech recognition. *J Acoust Soc Am*, 109(5):2112–2122, 2001.

R Galambos, J Schwartzkopff, and A Rupert. Microelectrode study of superior olivary nuclei. *Am J Physiol*, 197(3):527–536, 1959.

WS Geisler. Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol*, 59:167–192, 2008.

MA Gerzon. Periphony: With-height sound reproduction. *J Audio Eng Soc*, 21(1): 2–10, 1973.

JM Goldberg and PB Brown. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization. *J Neurophysiol*, 32(4):613–636, 1969.

B Gordon. Receptive fields in deep layers of cat superior colliculus. *J Neurophysiol*, 36(2):157–178, 1973.

DW Grantham. Detectability of time-varying interaural correlation in narrowband noise stimuli. *J Acoust Soc Am*, 72(4):1178–1184, 1982.

DW Grantham. Discrimination of dynamic interaural intensity differences. *J Acoust Soc Am*, 76(1):71–76, 1984.

DW Grantham and FL Wightman. Detectability of varying interaural temporal differences. *J Acoust Soc Am*, 63(2):511–523, 1978a.

DW Grantham and FL Wightman. Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. *J Acoust Soc Am*, 63(S1): S31–S31, 1978b.

MW Greenlee, SM Frank, M Kaliuzhna, O Blanke, F Bremmer, J Churan, LF Cuturi, PR MacNeilage, and AT Smith. Multisensory integration in self motion perception. *Multisens Res*, 29(6-7):525–556, 2016.

TD Griffiths and JD Warren. What is an auditory object? *Nat Rev Neurosci*, 5 (11):887–892, 2004.

TD Griffiths, GGR Green, A Rees, and G Rees. Human brain areas involved in the analysis of auditory movement. *Hum Brain Mapp*, 9(2):72–80, 2000.

TD Griffiths, JD Warren, SK Scott, I Nelken, and AJ King. Cortical processing of complex sound: a way forward? *Trends Neurosci*, 27(4):181–185, 2004.

B Grothe and Thomas J Park. Structure and function of the bat superior olivary complex. *Microsc Res Tech*, 51(4):382–402, 2000.

B Grothe and M Pecka. The natural history of sound localization in mammals – a story of neuronal inhibition. *Front Neural Circuits*, 8:116, 2014.

B Grothe, M Pecka, and D McAlpine. Mechanisms of sound localization in mammals. *Physiol Rev*, 90(3):983–1012, 2010.

A Gutschalk, C Micheyl, and AJ Oxenham. Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol*, 6(6):e138, 2008.

TA Hackett. Anatomic organization of the auditory cortex. In *Handbook of clinical neurology*, volume 129, pages 27–53. Elsevier, 2015.

ER Hafter and LA Jeffress. Two-image lateralization of tones and clicks. *J Acoust Soc Am*, 44(2):563–569, 1968.

JW Hall, MP Haggard, and MA Fernandes. Detection in noise by spectro-temporal pattern analysis. *J Acoust Soc Am*, 76(1):50–56, 1984.

JW Hall III and JH Grose. Comodulation masking release and auditory grouping. *J Acoust Soc Am*, 88(1):119–125, 1990.

KE Hancock and B Delgutte. A physiologically based model of interaural time difference discrimination. *J Neurosci*, 24(32):7110–7117, 2004.

WM Hartmann and D Johnson. Stream segregation and peripheral channeling. *Music Percept*, 9(2):155–183, 1991.

MJ Hautus and BW Johnson. Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *J Acoust Soc Am*, 117 (1):275–280, 2005.

J Hebrank and D Wright. Spectral cues used in the localization of sound sources on the median plane. *J Acoust Soc Am*, 56(6):1829–1834, 1974.

HE Heffner. The role of macaque auditory cortex in sound localization. *Acta Otolaryngol*, 117(sup532):22–27, 1997.

HE Heffner and RS Heffner. Effect of bilateral auditory cortex lesions on sound localization in Japanese macaques. *J Neurophysiol*, 64(3):915–931, 1990.

E Hendrickx, P Stitt, JC Messonnier, JM Lyzwa, BFG Katz, and C De Boishéraud. Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *J Acoust Soc Am*, 141(3):2011–2023, 2017.

GB Henning. Detectability of interaural delay in high-frequency complex waveforms. *J Acoust Soc Am*, 55(1):84–90, 1974.

GB Henning. Some observations on the lateralization of complex waveforms. *J Acoust Soc Am*, 68(2):446–454, 1980.

GB Henning and J Ashton. The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay. *Hear Res*, 4(2):185–194, 1981.

IJ Hirsh. The influence of interaural phase on interaural summation and inhibition. *J Acoust Soc Am*, 20(4):536–544, 1948.

RW Hukin and CJ Darwin. Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Percept Psychophys*, 57(2):191–196, 1995a.

RW Hukin and CJ Darwin. Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *J Acoust Soc Am*, 98 (3):1380–1387, 1995b.

SH Hulse. Auditory scene analysis in animal communication. *Advances in the Study of Behavior*, 31:163–200, 2002.

TJ Imig, NG Bibikov, P Poirier, and FK Samson. Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus. *J Neurophysiol*, 83 (2):907–925, 2000.

M Itoh, K Iida, and M Morimoto. Individual differences in directional bands in median plane localization. *Appl Acoust*, 68:909–915, 2007.

Y Iwaya, Y Suzuki, and D Kimura. Effects of head movement on front-back error in sound localization. *Acoust Sci Technol*, 24(5):322–324, 2003.

LA Jeffress, HC Blodgett, and BH Deatherage. Effect of interaural correlation on the precision of centering a noise. *J Acoust Soc Am*, 34(8):1122–1123, 1962.

WM Jenkins and RB Masterton. Sound localization: effects of unilateral lesions in central auditory system. *J Neurophysiol*, 47(6):987–1016, 1982.

WM Jenkins and MM Merzenich. Role of cat primary auditory cortex for sound-localization behavior. *J Neurophysiol*, 52(5):819–847, 1984.

BW Johnson, MJ Hautus, and WC Clapp. Neural activity associated with binaural processes for the perceptual segregation of pitch. *Clin Neurophysiol*, 114(12): 2245–2250, 2003.

PX Joris. Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive. *J Neurophysiol*, 76(4):2137–2156, 1996.

PX Joris and M van der Heijden. Early binaural hearing: the comparison of temporal differences at the two ears. *Annu Rev Neurosci*, 42:433–457, 2019.

PX Joris and TC Yin. Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. *J Neurophysiol*, 73(3):1043–1062, 1995.

PX Joris and TCT Yin. Envelope coding in the lateral superior olive. III. Comparison with afferent pathways. *J Neurophysiol*, 79(1):253–269, 1998.

PX Joris, B van de Sande, A Recio-Spinoso, and M van der Heijden. Auditory midbrain and nerve responses to sinusoidal variations in interaural correlation. *J Neurosci*, 26(1):279–289, 2006.

JH Kaas and TA Hackett. 'What' and 'where' processing in auditory cortex. *Nat Neurosci*, 2(12):1045–1047, 1999.

C Kapfer, AH Seidl, H Schweizer, and B Grothe. Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons. *Nat Neurosci*, 5 (3):247–253, 2002.

M Kato, H Uematsu, M Kashino, and T Hirahara. The effect of head motion on the accuracy of sound localization. *Acoust Sci Technol*, 24(5):315–317, 2003.

J Kawaura, Y Suzuki, F Asano, and T Sone. Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *J Acoust Soc Jpn*, 12(5):203–216, 1991.

C Kayser, KP Körding, and P König. Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol*, 14(4):468–473, 2004.

CH Keller and TT Takahashi. Localization and identification of concurrent sounds in the owl's auditory space map. *J Neurosci*, 25(45):10446–10461, 2005.

J Kim, M Barnett-Cowan, and EA Macpherson. Integration of auditory input with vestibular and neck proprioceptive information in the interpretation of dynamic sound localization cues. In *Proc Meet Acoust ICA 2013*, volume 19, page 050142. ASA, 2013.

SM Kim and W Choi. On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. *J Acoust Soc Am*, 117(6):3657–3665, 2005.

AJ King. The superior colliculus. *Curr Biol*, 14(9):R335–R338, 2004.

AJ King and JC Middlebrooks. Cortical representation of auditory space. In *The auditory cortex*, pages 329–341. Springer, 2011.

AJ King and I Nelken. Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat Neurosci*, 12(6):698, 2009.

AJ King and Kerry MM Walker. Listening in complex acoustic scenes. *Curr Opin Physiol*, 2020.

AJ King, S Teki, and BDB Willmore. Recent advances in understanding the auditory cortex. *F1000Res*, 7, 2018.

RG Klumpp and HR Eady. Some measurements of interaural time difference thresholds. *J Acoust Soc Am*, 28:859–860, 1956.

Z Kocsis, I Winkler, A Bendixen, and C Alain. Promoting the perception of two and three concurrent sound objects: An event-related potential study. *Int J Psychophysiol*, 107:16–28, 2016.

AJ Kolarik and JF Culling. Measurement of the binaural temporal window using a lateralisation task. *Hear Res*, 248(1-2):60–68, 2009.

HM Kondo, D Pressnitzer, I Toshima, and M Kashino. Effects of self-motion on auditory scene analysis. *Proc Natl Acad Sci USA*, 109(17):6775–6780, 2012.

M Kubovy and JE Daniel. Pitch segregation by interaural phase, by momentary amplitude disparity, and by monaural phase. *J Audio Eng Soc*, 31(9):630–638, 1983.

M Kubovy, JE Cutting, and R McGuire. Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues. *Science*, 186(4160): 272–274, 1974.

GF Kuhn. Model for the interaural time differences in the azimuthal plane. *J Acoust Soc Am*, 62(1):157–167, 1977.

H Kuttruff. *Acoustics: an introduction*. CRC Press, 2007.

B Laback, I Zimmermann, P Majdak, WD Baumgartner, and SM Pok. Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearing. *J Acoust Soc Am*, 130(3):1515–1529, 2011.

EHA Langendijk and AW Bronkhorst. Contribution of spectral cues to human sound localization. *J Acoust Soc Am*, 112(4):1583–1596, 2002.

LJ Leibold, E Buss, and L Calandruccio. Too young for the cocktail party? *Acoust Today*, 12(1), 2019.

ET Levy and RA Butler. Stimulus factors which influence the perceived externalization of sound presented through headphones. *J Aud Res*, 1978.

JCR Licklider. The influence of interaural phase relations upon the masking of speech by white noise. *J Acoust Soc Am*, 20(2):150–159, 1948.

RY Litovsky. Spatial release from masking. *Acoust Today*, 8(2):18–25, 2012.

WC Loftus, DC Bishop, RL Saint Marie, and DL Oliver. Organization of binaural excitatory and inhibitory inputs to the inferior colliculus from the superior olive. *J Comp Neurol*, 472(3):330–344, 2004.

SG Lomber and S Malhotra. Double dissociation of 'what' and 'where' processing in auditory cortex. *Nat Neurosci*, 11(5):609–616, 2008.

JM Loomis, C Hebert, and JG Cicinelli. Active localization of virtual sounds. *J Acoust Soc Am*, 88(4):1757–1764, 1990.

K Lu, Y Xu, P Yin, AJ Oxenham, JB Fritz, and SA Shamma. Temporal coherence structure rapidly shapes neuronal interactions. *Nat Commun*, 8(1):1–12, 2017.

L Luo, Q Wang, and L Li. Neural representations of concurrent sounds with overlapping spectra in rat inferior colliculus: Comparisons between temporal-fine structure and envelope. *Hear Res*, 353:87–96, 2017.

L Luo, N Xu, Q Wang, and L Li. Disparity in interaural time difference improves the accuracy of neural representations of individual concurrent narrowband sounds in rat inferior colliculus and auditory cortex. *J Neurophysiol*, 123(2): 695–706, 2020.

EA Macpherson. Head motion, spectral cues, and Wallach's principle of least displacement in sound localization. In *Principles and applications of spatial hearing*, pages 103–120. World Scientific, 2011.

EA Macpherson. Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation. In *Proc Meet Acoust ICA 2013*, volume 19, page 050131. ASA, 2013.

JC Makous and JC Middlebrooks. Two-dimensional sound localization by human listeners. *J Acoust Soc Am*, 87(5):2188–2200, 1990.

S Malhotra, AJ Hall, and SG Lomber. Cortical control of sound localization in the cat: unilateral cooling deactivation of 19 cerebral areas. *J Neurophysiol*, 92(3): 1625–1643, 2004.

BJ May. Role of the dorsal cochlear nucleus in the sound localization behavior of cats. *Hear Res*, 148(1-2):74–87, 2000.

S McAdams. Contributions of sub-audio frequency modulation and spectral envelope constancy to spectral fusion in complex harmonic tones. *J Acoust Soc Am*, 72(S1):S11–S11, 1982.

D McAlpine, D Jiang, and AR Palmer. A neural code for low-frequency sound localization in mammals. *Nat Neurosci*, 4(4):396–401, 2001.

K I McAnally and R L Martin. Sound localization with head movement: implications for 3-d audio displays. *Front Neurosci*, 8:210, 2014.

JH McDermott. The cocktail party problem. *Curr Biol*, 19(22):R1024–R1027, 2009.

JH McDermott and EP Simoncelli. Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, 71(5):926–940, 2011.

JH McDermott, M Schemitsch, and EP Simoncelli. Summary statistics in auditory perception. *Nat Neurosci*, 16(4):493–498, 2013.

KL McDonald and C Alain. Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *J Acoust Soc Am*, 118(3):1593–1604, 2005.

D McFadden and EG Pasanen. Lateralization at high frequencies based on interaural time differences. *J Acoust Soc Am*, 59(3):634–639, 1976.

MA Meredith and BE Stein. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*, 56(3):640–662, 1986.

C Micheyl and AJ Oxenham. Objective and subjective psychophysical measures of auditory stream integration and segregation. *J Assoc Res Otolaryngol*, 11(4): 709–724, 2010a.

C Micheyl and AJ Oxenham. Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. *Hear Res*, 266(1-2):36–51, 2010b.

C Micheyl, B Tian, RP Carlyon, and JP Rauschecker. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48(1): 139–148, 2005.

C Micheyl, RP Carlyon, A Gutschalk, JR Melcher, AJ Oxenham, JP Rauschecker, B Tian, and EC Wilson. The role of auditory cortex in the formation of auditory streams. *Hear Res*, 229(1-2):116–131, 2007.

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

JC Middlebrooks and P Bremen. Spatial stream segregation by auditory cortical neurons. *J Neurosci*, 33(27):10986–11001, 2013.

JC Middlebrooks and DM Green. Sound localization by human listeners. *Annu Rev Psychol*, 42:135–159, 1991.

JC Middlebrooks and ZA Onsan. Stream segregation with high spatial acuity. *J Acoust Soc Am*, 132(6):3896–3911, 2012.

AW Mills. On the minimum audible angle. *J Acoust Soc Am*, 30(4):237–246, 1958.

AW Mills. Lateralization of high-frequency tones. *J Acoust Soc Am*, 32(1):132–134, 1960.

P Minnaar, SK Olesen, F Christensen, and H Moller. The importance of head movements for binaural room synthesis. In *Proc Int Conf Aud Disp ICAD 2001*, 2001.

M Mishkin. Analogous neural models for tactual and visual learning. *Neuropsychologia*, 17(2):139–151, 1979.

AP Mishra, NS Harper, and JWH Schnupp. Exploring the distribution of statistical feature parameters for natural sound textures. *bioRxiv*, 2020. doi: 10.1101/2020.08.28.271528. URL https://www.biorxiv.org/content/early/2020/08/28/2020.08.28.271528.

W Młynarski and J Jost. Statistics of natural binaural sounds. *PLoS ONE*, 9(10): e108968, 2014.

W Młynarski and JH McDermott. Ecological origins of perceptual grouping principles in the auditory system. *Proc Natl Acad Sci USA*, 116(50):25355–25364, 2019.

BCJ Moore. *An introduction to the psychology of hearing*. Brill, 2012.

BCJ Moore and HE Gockel. Factors influencing sequential stream segregation. *Acta Acust united Ac*, 88(3):320–333, 2002.

BCJ Moore and HE Gockel. Properties of auditory stream formation. *Philos Trans R Soc Lond, B, Biol Sci*, 367(1591):919–931, 2012.

BCJ Moore, BR Glasberg, and RW Peters. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J Acoust Soc Am*, 80(2): 479–483, 1986.

A Morel and J Bullier. Anatomical segregation of two cortical visual pathways in the macaque monkey. *Vis Neurosci*, 4(6):555–578, 1990.

T Moriizumi and T Hattori. Pallidotectal projection to the inferior colliculus of the rat. *Exp Brain Res*, 87(1):223–226, 1991.

EA Murray and M Mishkin. Severe tactual as well as visual memory deficits follow combined removal of the amygdala and hippocampus in monkeys. *J Neurosci*, 4(10):2565–2580, 1984.

AD Musicant and RA Butler. The influence of pinnae-based spectral cues on sound localization. *J Acoust Soc Am*, 75(4):1195–1200, 1984.

R Nassiri and MA Escabí. Illusory spectrotemporal ripples created with binaurally correlated noise. *J Acoust Soc Am*, 123(4):EL92–EL98, 2008.

I Nelken. Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol*, 14(4):474–480, 2004.

I Nelken. Processing of complex sounds in the auditory system. *Curr Opin Neurobiol*, 18(4):413–417, 2008.

I Nelken and O Bar-Yosef. Neurons and objects: the case of auditory cortex. *Front Neurosci*, 2:9, 2008.

W Noble, D Byrne, and B Lepage. Effects on sound localization of configuration and type of hearing impairment. *J Acoust Soc Am*, 95(2):992–1005, 1994.

JM Nuetzel and ER Hafter. Discrimination of interaural delays in complex waveforms: Spectral effects. *J Acoust Soc Am*, 69(4):1112–1118, 1981.

D Oertel and ED Young. What's a cerebellar circuit doing in the auditory system? *Trends Neurosci*, 27(2):104–110, 2004.

UE Olaźbal and JK Moore. Nigrotectal projection to the inferior colliculus: horseradish peroxidase transport and tyrosine hydroxylase immunohistochemical studies in rats, cats, and bats. *J Comp Neurol*, 282(1):98–118, 1989.

SR Oldfield and SP Parker. Acuity of sound localisation: A topography of auditory space. III. Monaural hearing conditions. *Perception*, 15(1):67–81, 1986.

DL Oliver, GE Beckius, DC Bishop, and S Kuwada. Simultaneous anterograde labeling of axonal layers from lateral superior olive and dorsal cochlear nucleus in the inferior colliculus of cat. *J Comp Neurol*, 382(2):215–229, 1997.

AR Palmer and AJ King. The representation of auditory space in the mammalian superior colliculus. *Nature*, 299(5880):248–249, 1982.

AR Palmer and IJ Russell. Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hear Res*, 24(1): 1–15, 1986.

CV Parise, K Knorre, and MO Ernst. Natural auditory scene statistics shapes human spatial hearing. *Proc Natl Acad Sci USA*, 111(16):6104–6108, 2014.

MT Pastore and WA Yost. Spatial release from masking with a moving target. *Front Psychol*, 8:2238, 2017.

MT Pastore, S Natale, WA Yost, and MF Dorman. Head movements allow listeners bilaterally implanted with cochlear implants to resolve front-back confusions. *Ear Hear*, 39(6):1224, 2018.

R Pavão, ES Sussman, BJ Fischer, and JL Peña. Natural ITD statistics predict human auditory spatial perception. *eLife*, 9:e51927, 2020.

M Pecka, A Brand, O Behrend, and B Grothe. Interaural time difference processing in the mammalian medial superior olive: The role of glycinergic inhibition. *J Neurosci*, 28(27):6914–6925, 2008.

S Perrett and W Noble. The contribution of head motion cues to localization of low-pass noise. *Percept Psychophys*, 59:1018–1026, 1997a.

S Perrett and W Noble. The effect of head rotations on vertical plane sound localization. *J Acoust Soc Am*, 102:2325–2332, 1997b.

C Perrodin, C Kayser, TJ Abel, NK Logothetis, and CI Petkov. Who is that? Brain networks and mechanisms for identifying individuals. *Trends Cogn Sci*, 19(12): 783–796, 2015.

M Perron. Hearing aids of tomorrow: Cognitive control toward individualized experience. *Hear J*, 70(11):22–23, 2017.

DR Perrott. Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity. *J Acoust Soc Am*, 75(4):1201–1206, 1984.

TW Picton. Hearing in time: evoked potential studies of temporal processing. *Ear Hear*, 34(4):385–401, 2013.

CJ Plack. *The sense of hearing*. Routledge, 2018.

G Plenge. On the differences between localization and lateralization. *J Acoust Soc Am*, 56(3):944–951, 1974.

V Pulkki. Virtual sound source positioning using vector base amplitude panning. *J Audio Eng Soc*, 45(6):456–466, 1997.

V Pulkki. Spatial sound reproduction with directional audio coding. *J Audio Eng Soc*, 55(6):503–516, 2007.

V Pulkki and M Karjalainen. *Communication acoustics: an introduction to speech, audio and psychoacoustics*. John Wiley & Sons, 2015.

KC Puvvada and JZ Simon. Cortical representations of speech in a multitalker auditory scene. *J Neurosci*, 37(38):9189–9196, 2017.

JP Rauschecker. Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol*, 117(sup532):34–38, 1997.

JP Rauschecker. Cortical processing of complex sounds. *Curr Opin Neurobiol*, 8 (4):516–521, 1998.

JP Rauschecker and B Tian. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci USA*, 97(22):11800–11806, 2000.

JP Rauschecker, B Tian, and M Hauser. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207):111–114, 1995.

C Retsa, PJ Matusz, JWH Schnupp, and MM Murray. What's what in auditory cortices? *NeuroImage*, 176:29–40, 2018.

B Roberts and JM Brunstrom. Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *J Acoust Soc Am*, 104(4):2326–2338, 1998.

B Roberts and JM Brunstrom. Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. *J Acoust Soc Am*, 110(5):2479–2490, 2001.

B Roberts and BCJ Moore. The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony. *J Acoust Soc Am*, 89(6):2922–2932, 1991.

B Roberts, BR Glasberg, and BCJ Moore. Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J Acoust Soc Am*, 112(5):2074–2085, 2002.

N Roman, D Wang, and GJ Brown. Speech segregation based on sound localization. *J Acoust Soc Am*, 114(4):2236–2252, 2003.

LM Romanski, B Tian, J Fritz, M Mishkin, PS Goldman-Rakic, and JP Rauschecker. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci*, 2(12):1131–1136, 1999.

JE Rose, NB Gross, CD Geisler, and JE Hind. Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source. *J Neurophysiol*, 29(2):288–314, 1966.

JE Rose, JF Brugge, DJ Anderson, and JE Hind. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J Neurophysiol*, 30(4):769–793, 1967.

TD Rossing. *Springer handbook of acoustics*. Springer Science & Business Media, 2007.

TD Rossing and NH Fletcher. *Principles of vibration and sound*. Springer Science & Business Media, 2012.

AJ Sach and PJ Bailey. Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Percept Psychophys*, 66(8):1379–1387, 2004.

M Saenz and Dave RM Langers. Tonotopic mapping of human auditory cortex. *Hear Res*, 307:42–52, 2014.

LD Sanders, AS Joh, RE Keen, and RL Freyman. One sound or two? Object-related negativity indexes echo perception. *Percept Psychophys*, 70(8):1558–1570, 2008.

DH Sanes. An in vitro analysis of sound localization mechanisms in the gerbil lateral superior olive. *J Neurosci*, 10(11):3494–3506, 1990.

O Santala and V Pulkki. Directional perception of distributed sound sources. *J Acoust Soc Am*, 129(3):1522–1530, 2011.

B Scharf, M Florentine, and CH Meiselman. Critical band in auditory lateralization. *Sens Processes*, 1(2):109–126, 1976.

JWH Schnupp, I Nelken, and AJ King. *Auditory neuroscience: Making sense of sound*. MIT press, 2011.

TM Shackleton and R Meddis. The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs. *J Acoust Soc Am*, 91(6):3579–3581, 1992.

TM Shackleton, R Meddis, and MJ Hewitt. The role of binaural and fundamental frequency difference cues in the identification of concurrently presented vowels. *Q J Exp Psychol Sec A*, 47(3):545–563, 1994.

SA Shamma and C Micheyl. Behind the scenes of auditory perception. *Curr Opin Neurobiol*, 20(3):361–366, 2010.

SA Shamma, M Elhilali, and C Micheyl. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci*, 34(3):114–123, 2011.

SA Shamma, M Elhilali, L Ma, C Micheyl, AJ Oxenham, D Pressnitzer, P Yin, and Y Xu. Temporal coherence and the streaming of complex sounds. In *Basic Aspects of Hearing*, pages 535–543. Springer, 2013.

EAG Shaw. Acoustical features of the human external ear. In *Binaural and spatial hearing in real and virtual environments*. Mahwah, NJ: Lawrence Erlbaum, 1997.

EAG Shaw and R Teranishi. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J Acoust Soc Am*, 44(1):240–249, 1968.

MM Shiell, L Hausfeld, and E Formisano. Activity in human auditory cortex represents spatial separation between concurrent sounds. *J Neurosci*, 38(21): 4977–4984, 2018.

BG Shinn-Cunningham. Influences of spatial cues on grouping and understanding sound. In *Proc Forum Acust*, volume 29. Citeseer, 2005.

JZ Simon. The encoding of auditory objects in auditory cortex: Insights from magnetoencephalography. *Int J Psychophysiol*, 95(2):184–190, 2015.

NC Singh and FE Theunissen. Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am*, 114(6):3394–3411, 2003.

M Slaney. Auditory toolbox. *Interval Research Corporation, Tech. Rep*, 10(1998): 1194, 1998.

PH Smith, PX Joris, and TCT Yin. Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive. *J Comp Neurol*, 331(2):245–260, 1993.

W Snow. Basic principles of stereophonic sound. *IRE Trans Audio*, 2:42–53, 1955.

JS Snyder and M Elhilali. Recent advances in exploring the neural underpinnings of auditory scene perception. *Ann NY Acad Sci*, 1396(1):39–55, 2017.

KM Spangler, WB Warr, and CK Henkel. The projections of principal cells of the medial nucleus of the trapezoid body in the cat. *J Comp Neurol*, 238(3): 249–262, 1985.

MW Spitzer and MN Semple. Responses of inferior colliculus neurons to time-varying interaural phase disparity: effects of shifting the locus of virtual motion. *J Neurophysiol*, 69(4):1245–1263, 1993.

MW Spitzer and MN Semple. Transformation of binaural response properties in the ascending auditory pathway: influence of time-varying interaural phase disparity. *J Neurophysiol*, 80(6):3062–3076, 1998.

TH Stainsby, C Füllgrabe, HJ Flanagan, SK Waldman, and BCJ Moore. Sequential streaming due to manipulation of interaural time differences. *J Acoust Soc Am*, 130(2):904–914, 2011.

SS Stevens and EB Newman. The localization of actual sources of sound. *Am J Psychol*, 48(2):297–306, 1936.

TT Takahashi and CH Keller. Representation of multiple sound sources in the owl's auditory space map. *J Neurosci*, 14(8):4780–4793, 1994.

S Teki, M Chait, S Kumar, K von Kriegstein, and TD Griffiths. Brain bases for auditory stimulus-driven figure–ground segregation. *J Neurosci*, 31(1):164–171, 2011.

T Thakkar and MJ Goupell. Internalized elevation perception of simple stimuli in cochlear-implant and normal-hearing listeners. *J Acoust Soc Am*, 136(2): 841–852, 2014.

GC Thompson and AM Cortez. The inability of squirrel monkeys to localize sound after unilateral ablation of auditory cortex. *Behav Brain Res*, 8(2):211–216, 1983.

W R Thurlow and PS Runge. Effect of induced head movements on localization of direction of sounds. *J Acoust Soc Am*, 42(2):480–488, 1967.

DJ Tollin and TCT Yin. The coding of spatial location by single units in the lateral superior olive of the cat. I. Spatial receptive fields in azimuth. *J Neurosci*, 22 (4):1454–1467, 2002.

DJ Tollin and TCT Yin. Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive. *J Neurosci*, 25(46):10648–10657, 2005.

DJ Tollin, K Koka, and JJ Tsai. Interaural level difference discrimination thresholds for single neurons in the lateral superior olive. *J Neurosci*, 28(19):4848–4860, 2008.

FE Toole. In-head localization of acoustic images. *J Acoust Soc Am*, 48(4B): 943–949, 1970.

B Tóth, Z Kocsis, GP Háden, Á Szerafin, BG Shinn-Cunningham, and I Winkler. EEG signatures accompanying auditory figure-ground segregation. *Neuroimage*, 141:108–119, 2016a.

B Tóth, Z Kocsis, G Urbán, and I Winkler. Theta oscillations accompanying concurrent auditory stream segregation. *Int J Psychophysiol*, 106:141–151, 2016b.

J Traer and JH McDermott. Statistics of natural reverberation enable perceptual separation of sound and space. *Proc Natl Acad Sci USA*, 113(48):E7856–E7865, 2016a.

J Traer and JH McDermott. The perception of reverberation is constrained by environmental statistics. *J Acoust Soc Am*, 139(4):2210–2210, 2016b.

M Turgeon, AS Bregman, and PA Ahad. Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. *J Acoust Soc Am*, 111(4):1819–1831, 2002.

LG Ungerleider. Two cortical visual systems. *Analysis of visual behavior*, pages 549–586, 1982.

K van der Heijden, JP Rauschecker, B de Gelder, and E Formisano. Cortical mechanisms of spatial hearing. *Nat Rev Neurosci*, 20(10):609–623, 2019.

LPAS van Noorden. *Temporal coherence in the perception of tone sequences*. Institute for Perceptual Research Eindhoven, the Netherlands, 1975.

J Vliegen and AJ Oxenham. Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am*, 105(1):339–346, 1999.

J Vliegen, TJ Van Grootel, and AJ Van Opstal. Dynamic sound localization during rapid eye-head gaze shifts. *J Neurosci*, 24(42):9291–9302, 2004.

RF Voss and J Clarke. '1/f' noise in music and speech. *Nature*, 258(5533):317–318, 1975.

H Wallach. On sound localization. *J Acoust Soc Am*, 10:270–274, 1939.

H Wallach. The role of head movements and vestibular and visual cues in sound localization. *J Exp Psychol*, 27:339–368, 1940.

PM Waser and CH Brown. Habitat acoustics and primate communication. *Am J Primatol*, 10(2):135–154, 1986.

DL Wessel. Timbre space as a musical control structure. *Comput Music J*, pages 45–52, 1979.

RH Whitworth and LA Jeffress. Time vs intensity in the localization of tones. *J Acoust Soc Am*, 33(7):925–929, 1961.

E Wigderson, I Nelken, and Y Yarom. Early multisensory integration of self and source motion in the auditory system. *Proc Natl Acad Sci USA*, 113(29): 8308–8313, 2016.

FL Wightman and DJ Kistler. Headphone simulation of free-field listening. I: Stimulus synthesis. *J Acoust Soc Am*, 85(2):858–867, 1989a.

FL Wightman and DJ Kistler. Headphone simulation of free-field listening. II: Psychophysical validation. *J Acoust Soc Am*, 85(2):868–878, 1989b.

FL Wightman and DJ Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am*, 91(3):1648–1661, 1992.

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

CF Willey, E Inglis, and CH Pearce. Reversal of auditory localization. *J Exp Psychol*, 20(2):114, 1937.

A Wilska. *Studies on Directional Hearing. English translation*. Aalto University School of Science and Technology, Department of Signal Processing and Acoustics, 2010. PhD thesis originally published in German as 'Untersuchungen über das Richtungshören', University of Helsinki, 1938.

B Xie. *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

N Xu, L Luo, Q Wang, and L Li. Binaural unmasking of the accuracy of envelope-signal representation in rat auditory cortex but not auditory midbrain. *Hear Res*, 377:224–233, 2019.

JD Yao, P Bremen, and JC Middlebrooks. Emergence of spatial stream segregation in the ascending auditory pathway. *J Neurosci*, 35(49):16199–16212, 2015.

TC Yin and JC Chan. Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol*, 64(2):465–488, 1990.

TCT Yin, PH Smith, and PX Joris. Neural mechanisms of binaural processing in the auditory brainstem. *Compr Physiol*, 9(4):1503–1575, 2019.

WA Yost. Discriminations of interaural phase differences. *J Acoust Soc Am*, 55(6): 1299–1303, 1974.

WA Yost and RH Dye Jr. Discrimination of interaural differences of level as a function of frequency. *J Acoust Soc Am*, 83(5):1846–1851, 1988.

WA Yost, MT Pastore, and KR Pulling. Sound-source localization as a multisystem process: The Wallach azimuth illusion. *J Acoust Soc Am*, 146(1):382–398, 2019.

ED Young, GA Spirou, JJ Rice, and HF Voigt. Neural organization and responses to complex stimuli in the dorsal cochlear nucleus. *Philos Trans R Soc Lond, B, Biol Sci*, 336(1278):407–413, 1992.

PT Young. Auditory localization with acoustical transposition of the ears. *J Exp Psychol*, 11(6):399, 1928.

PT Young. The role of head movements in auditory localization. *J Exp Psychol*, 14:95–124, 1931.

RJ Zatorre and VB Penhune. Spatial localization after excision of human auditory cortex. *J Neurosci*, 21(16):6321–6328, 2001.

RJ Zatorre, M Bouffard, and P Belin. Sensitivity to auditory object features in human temporal neocortex. *J Neurosci*, 24(14):3637–3642, 2004.

X Zhai, F Khatami, M Sadeghi, F He, HL Read, IH Stevenson, and M Escabi. Distinct neural ensemble response statistics are associated with recognition and discrimination of natural sound textures. *Proc Natl Acad Sci USA*, 2020.

F Zotter and M Frank. *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

Aalto-DD 85/2021

ISBN 978-952-64-0423-3 (printed)
ISBN 978-952-64-0424-0 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Signal Processing and Acoustics
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS