

Psychoacoustic and neurophysiologic investigations of auditory continuity

Citation for published version (APA):

Riecke, L. (2009). *Psychoacoustic and neurophysiologic investigations of auditory continuity*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20090401lr>

Document status and date:

Published: 01/01/2009

DOI:

[10.26481/dis.20090401lr](https://doi.org/10.26481/dis.20090401lr)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Psychoacoustic and neurophysiologic investigations of auditory continuity

Lars Riecke

Cover design & layout: L. Riecke
Production: Datawyse / Universitaire Pers Maastricht

© 2009 L. Riecke, Maastricht

ISBN: 978 90 5278 816 6

The work presented in this thesis was funded by the Cognition program of the Netherlands Organization for Scientific Research (NWO) and was conducted at Maastricht University.

Psychoacoustic and neurophysiologic investigations of auditory continuity

Dissertation

To obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. G.P.M.F. Mols
in accordance with the decision of the Board of Deans,
to be defended in public on Wednesday 1st of April 2009, at 12.00 hours

by

Lars Riecke

Supervisor:

Prof. dr. R. Goebel

Co-supervisor:

Dr. E. Formisano

Assessment Committee:

Prof. dr. P. De Weerd (Chairman)

Prof. dr. P. van Dijk (University of Groningen, The Netherlands)

Dr. A. Gutschalk (Ruprecht-Karls-Universität Heidelberg, Germany)

Prof. dr. B. Jansma

Contents

Chapter 1	General introduction	7
Chapter 2	Quantification of acoustic influences on illusory continuity and proposal of a neural filter model	17
Chapter 3	FMRI-based localization of neural representations of acoustic and illusory continuity in auditory cortex	39
Chapter 4	EEG-based chronometry of cortical processing of acoustic and illusory continuity	57
Chapter 5	FMRI-based localization of neural representations of illusory voice continuity in auditory cortex	79
Chapter 6	Identification of contextual determinants of illusory continuity	91
Chapter 7	General discussion	107
	Summary	115
	Samenvatting	117
	Acknowledgements	121
	Curriculum vitae	123

Chapter 1

General introduction

Hearing serves many ecological purposes such as communication or identification of events. In natural environments, meaningful auditory perception requires the tracking of sounds in the presence of other concurrent sounds. Sounds are pressure waves generated by mechanical vibrations. When multiple sound sources are present, the emitted waves combine so as to produce a composite wave in which the acoustic characteristics of the individual waves, such as frequency and amplitude, are distorted. Nevertheless, the auditory system is reliable in extracting sounds of interest even under such challenging acoustic conditions. This remarkable aspect of hearing lies at the heart of our ability to form meaningful perceptual entities and has been described in an influential theoretical framework called Auditory Scene Analysis (Bregman, 1990).

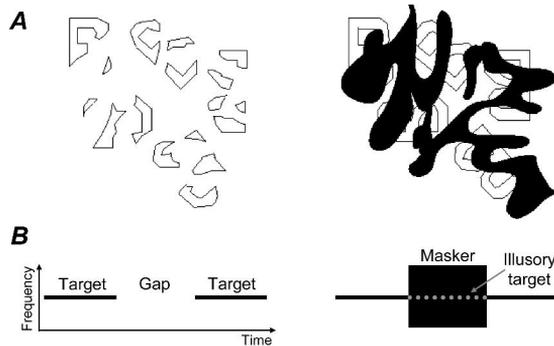


Figure 1. The continuity illusion. **A.** Visual continuity illusion (adapted from Bregman, 1990). Random visual shapes (left) become recognizable as fragments of the letter “B” when a visual mask fills the gaps between the fragments (right). **B.** Auditory continuity illusion. When a silent gap in an interrupted target sound (left) is filled with a masking sound, the target can be perceived as continuing through the masker (right). This may be accompanied by the perceptual restoration of the target during the masker (gray dotted line).

A striking illustration of scene analysis and of the robustness of the auditory system against interfering sounds in particular is the auditory continuity illusion (Fig. 1B). The continuity illusion describes the phenomenon of perceiving a sound as apparently continuous, although the sound is physically interrupted by other louder sounds. This phenomenon may be accompanied by the perceptual restoration of the missing sound during the interruption. Continuity illusions apply to sounds of different complexity, such as noise, tones, sweeps, melodies, and speech (for reviews, see Bregman, 1990; Warren, 1999), and they have been reported for several animals such as birds (Braaten & Leary, 1999), cats (Sugita, 1997), and monkeys (Miller *et al.*, 2001; Petkov *et al.*, 2003). The existing evidence suggests that the phenomenon may emerge from fundamental mechanisms that facilitate smooth perception of fragmented sounds in noisy environments. However, these mechanisms are poorly understood: It is yet unclear where and how in the brain the continuity illusion may arise.

Understanding these mechanisms is crucial not only for advancing the knowledge of the auditory system in general, but also for improving hearing aids and artificial

speech recognition devices that perform sub optimally in noisy environments. The studies presented in this thesis aim at broadening this knowledge. Before tackling the underlying mechanisms in the following chapters, the present chapter first introduces important concepts for studying the continuity illusion. Specifically, the notion of sensory-perceptual transformations of sounds is discussed, and general principles of sound processing in the brain are outlined. Furthermore, an overview of the methods for investigating sensory-perceptual phenomena is given, followed by a review of previous research on the continuity illusion. Finally, an outline of the research aims and content of the thesis is given.

1 Sensory-perceptual transformations: A theoretical framework for studying the neural basis of the continuity illusion

The continuity illusion implies the subjective experience of a sound that is not present in the actual sound wave that entered the ear. Therefore, the illusory sound is constructed from the sensory input by the auditory system. This constructive process can be understood as a sensory-perceptual transformation of the sound wave. Sound waves produced in natural environments generally comprise a wide variety of acoustic features, such as frequency-modulations or amplitude-modulations. These features depend on the sound sources that generate these waves, and thus they are critical for sound source identification. However, sound object perception may not depend on the rapidly varying acoustic features alone but also on the perceptual interpretation of these features on a larger time scale. Especially in noisy scenes, the acoustic input may be highly redundant and therefore irrelevant features can be ignored. This is illustrated, for example, by the classical cocktail party effect (Cherry, 1953), in which a relevant sound is experienced as more salient than other ambient sounds.

In terms of information processing, subjective hearing experiences can be understood as arising from ‘transformed’ representations of sound waves (Smith & Lewicki, 2006; Wang *et al.*, 2008). This emergent property of sound may be achieved in the auditory system by redundancy compression (Barlow, 1961; Chechik *et al.*, 2006) and abstraction (Griffiths & Warren, 2004). More specifically, basic acoustic features can be extracted from the auditory stimulus and may be analyzed for their non-randomness (regularities, or patterns) (Näätänen *et al.*, 2001). The extracted regularities may be further retained as more abstract stimulus representations that are then analyzed for higher-order regularities, and so on. In these theoretical terms, meaningful auditory perception emerges gradually from interpreting non-random acoustic features of sensory stimuli.

The transformation of a sound wave into an auditory experience may not be fixed: Sound percepts generally depend not only on the actual sound wave, but also on other acoustic and cognitive factors. Experimental manipulation of these factors thus can be exploited for investigating how the auditory brain ‘makes sense’ of the auditory world, which is of major interest for the studies presented in this thesis. Important factors are, for example, the acoustic context and prior auditory

experience: Exposure to loud sounds, for instance, may render subsequent sounds softer (e.g. Durlach & Braida, 1969). Selective attention also plays a role (Broadbent, 1958; for reviews see Carlyon, 2004; Alain & Bernstein, 2008), as exemplified by bistable perceptual phenomena and the cocktail party problem. As mentioned before, auditory illusions are especially useful in studying sensory-perceptual transformations as they involve constructed percepts of sounds which are not present in the actual sensory input.

2 Basic principles of the neural processing of sound

Sound processing in the brain involves several stages of stimulus analysis that are defined by the functional characteristics of auditory neurons. The neural processing of sound begins in receptor cells in the cochlea of the inner ear. These so-called hair cells translate sound waves that entered the ear into neural signals. These signals are transmitted via the auditory nerve to a series of interconnected neuronal structures, collectively referred to as the auditory pathway. The most relevant structures include the cochlear nuclei and the superior olivary nuclei in the brainstem, the inferior colliculi in the midbrain, the medial geniculate nuclei in the thalamus, and the primary and secondary auditory cortices in the temporal lobes of the cerebrum.

A characteristic feature of neurons along the auditory pathway is their tuning to particular sound frequencies: Individual auditory neurons are excited selectively by sounds that contain frequencies within particular bandwidths. These neurons thus 'decompose' auditory stimuli into their constituting spectral components, analogous to a mathematical procedure called Fourier analysis. Given their broad frequency tuning, auditory neurons can be conceptualized as bandpass filters that transmit restricted portions of the stimulus spectrum (Fletcher, 1940). The filtering characteristics of auditory neurons are flexible: Frequency tuning becomes more complex along the ascending auditory pathway (Miller *et al.*, 2002). Different regions in auditory cortex, for example, may be specifically tuned to sounds of different complexity, such as melodies (Patterson *et al.*, 2002) or voices (Belin *et al.*, 2000). Frequency tuning may be further modulated by 'bottom-up' factors such as the preceding auditory input, and by 'top-down' factors like attention (Fritz *et al.*, 2003). These bidirectional influences are facilitated by afferent and efferent connections throughout the auditory pathway.

Another characteristic feature of the auditory pathway is that neurons tuned to neighboring frequencies are situated adjacently on the epithelium. Arrays of these neurons constitute topographic representations of the sound spectrum. These so-called tonotopic maps can be conceptualized as banks of bandpass filters which represent auditory stimuli in orderly arranged and partially overlapping frequency channels. In terms of information processing, the auditory pathway may thus implement a hierarchy of serial and parallel stimulus analysis stages in which the neural representation of the sound spectrum undergoes several transformations (Rauschecker *et al.*, 1997; Lee *et al.*, 2004). The functional architecture of the auditory pathway may enable sound processing as depicted above (see

paragraph 1): While sensory representations in the cochlea closely resemble the acoustic properties of auditory stimuli, representations in the auditory cortex may reflect more the perceptual properties of the sensory input (Nelken, 2004; Las *et al.*, 2005; Chechik *et al.*, 2006).

3 Methods for investigating sensory-perceptual transformations

Sensory-perceptual phenomena can be studied at the behavioral level by using psychoacoustic methods or at a neurophysiologic level by using functional magnetic resonance imaging (fMRI) or electroencephalography (EEG), for instance. These methods were also employed in the studies presented in this thesis.

3.1 Psychoacoustic methods

Psychoacoustic methods involve systematic varying one or more acoustic dimensions of an auditory stimulus, while measuring the effects on the subject's auditory experience (Fechner, 1860). These methods are useful, for example, for quantifying the sensitivity of the auditory system to changes in particular sound properties. Perceptual limits can be assessed on basis of thresholds, which are specific values along an acoustic dimension at which subjects can just detect the presence of a particular sound (Gescheider, 1997). However, detection is not an all-or-none phenomenon: Subjects may or may not detect a target sound across presentations of stimuli close to threshold. This implies that identical stimuli may evoke different perceptual conditions. Psychoacoustic designs may generally benefit from prior knowledge of neural principles of sound processing (see paragraph 2) in the sense that stimuli can be designed more specifically to tap particular neural mechanisms. For example, notched-noise stimuli can be used to estimate the filter characteristics of cochlear neurons (Patterson, 1976). Psychoacoustic methods were applied in the studies presented in chapter two and six of this thesis and guided the design of all the stimuli employed in the neurophysiologic investigations described in the other chapters.

3.2 Functional magnetic resonance imaging

Neurons in the brain increase their oxygen consumption when they are activated by sensory stimuli and/or cognitive processes. This increased oxygen demand is compensated for by increased local inflow of oxygenated blood. This hemodynamic response is slightly delayed and induces local variations in blood volume and in the relative concentration of oxygenated and deoxygenated blood. Based on the different magnetic resonance properties of oxygenated and deoxygenated blood, fMRI can be used to localize brain regions of increased blood-oxygenation level (Ogawa *et al.*, 1992).

fMRI provides a non-invasive and indirect measure of localized neural activity on basis of the hemodynamic response, at a spatial scale of a few millimeters. It should be noted that the gradient coils of the MR scanner produce loud acoustic artifacts

during the acquisition of MR images. To avoid masking of auditory stimuli by scanner noise, auditory fMRI experiments require the presentation of stimuli in silence between successive image acquisitions. With such 'sparse' sampling methods, uncontaminated stimulus-evoked BOLD signals can be measured after each stimulus presentation, facilitated by the sluggishness of the hemodynamic response (e.g. Belin *et al.*, 1999; Hall *et al.*, 1999; Talavage & Edmister, 2004). The temporal resolution of sparsely sampled fMRI is inherently low, typically in the range of several seconds. A major application for auditory fMRI is the localization of brain regions in which neural activities correlate with particular cognitive aspects of a given auditory task. fMRI was applied in the studies presented in chapter three and five.

3.3 Electroencephalography

When sensory stimuli and/or cognitive processes trigger neuronal action potentials, post-synaptic dendritic currents are induced, accompanied by electrical field potentials. When several parallel-oriented neurons in a cortical patch are synchronously active, their field potentials sum up. Using EEG, the resulting net cortical field potentials can be measured at scalp sites that are perpendicular to the evoking neurons. Scalp EEG activities oscillate at multiple frequencies, even in the absence of sensory input (Berger, 1930). To extract time-locked changes in the ongoing EEG, as evoked by experimental stimuli for example, EEG data need to be averaged across several stimulus presentations. Two different features of the EEG signal are relevant in this regard: First, phase-locked EEG changes can be isolated by averaging in the time domain. These event-related potentials (ERPs) typically show prominent peaks at several latencies with respect to the evoking event. Second, time-locked, frequency-specific EEG changes can be extracted by averaging in the frequency domain. These signals display event-related modulations in oscillations, which may be related to changes in the synchronized interactions of different neuronal populations (Lopes da Silva, 1991; Singer, 1993).

EEG provides a non-invasive measure of neural activity on basis of electrical field potentials, at a spatial scale of several centimeters (Epstein, 1983). This low spatial resolution is related to the fact that EEG is insensitive to activity in neurons that are oriented tangentially to the scalp. Furthermore, field potentials are distorted during their propagation through brain tissue, resulting in loss of spatial resolution at the scalp. Nevertheless, EEG provides high temporal resolution in the range of a few milliseconds. A major application for EEG is thus the chronometry of neural activities: Potentials or oscillations that are time-locked to particular cognitive aspects of a given experimental task may be used as indicators for the timing of particular neural processes. EEG was applied in the study presented in chapter four of this thesis.

3.4 Statistical considerations: Controlling for confounding variables

While psychoacoustic methods allow only limited insights into neural mechanisms, their combination with neurophysiologic methods, such as fMRI or EEG, allows directly comparing subjective auditory measures with objective measures of neural

activity. Specific insights can be gained, for example, by experimental manipulation of the acoustic stimulus (independent variable), while measuring simultaneously the consequences on neural activity and auditory perception (dependent variables). A problem with interpreting the measured neural activities may arise from multicollinearity among the different variables. Specifically, neural activity may correlate not only with perception, but also with the stimulus manipulation, thus limiting conclusions about sensory-perceptual transformations in the brain. Fortunately, there are ways of disentangling acoustic influences and perceptual influences (e.g. Logothetis & Schall, 1989). First, partial correlations may be extracted by decorrelating the two collinear factors. This permits investigating the contribution of one factor that is uncorrelated to the contribution of another factor. Second, switches in perception may be extracted and analyzed across presentations of the same bistable stimulus. This facilitates the investigation of perceptual influences independent of the influences of the evoking stimulus. Finally, illusory stimuli may be used as they can be compared to stimuli that evoke closely matched non-illusory percepts. This allows isolating acoustic influences.

Controlling for confounding variables is also achieved by using acoustically simple stimuli, such as synthetic tones. A problem with natural stimuli like speech, for example, is that subjects may spontaneously switch their attention to different sound features, such as the identity or the gender of a speaker. Compared to natural sounds, the acoustic properties of tones can be controlled more precisely and they are less likely to trigger undesired confounding mental processes. Since the auditory system decomposes sounds into their frequency components (see paragraph 2), tones can be considered as schematic 'building blocks' of more complex sounds and hence they are useful to study more general auditory mechanisms. These methodological considerations are crucial for all the studies presented in this thesis.

4 Previous research on the auditory continuity illusion

Since its discovery by Miller and Licklider (1950), the acoustic conditions that determine the continuity illusion have been investigated extensively (for reviews, see Bregman, 1990; Warren, 1999). One prerequisite for the phenomenon is the masking of the silent gap in the interrupted sound (target) by the louder sound (Houtgast, 1972; Warren *et al.*, 1972). Masking may occur when the louder sound (masker) contains similar frequencies as the target, thus rendering the gap inaudible. Another prerequisite is that the target fragments that surround the gap contain similar frequencies. Furthermore, the fragments need to constitute meaningful extensions of each other (Ciocca & Bregman, 1987). Under these conditions, the fragments may be grouped and produce a continuity illusion of the target. The continuity illusion is not an all-or-none phenomenon, but a graded perceptual state which may fade out or fade in during long interruptions (Wrightson & Warren, 1981; Warren *et al.*, 1994; Bregman, 1990).

Despite these and several other psychoacoustic findings, the neural mechanisms underlying the continuity illusion remain poorly understood. At the beginning of this

project, only two neurophysiologic studies had addressed this issue. The first study identified neurons in cat primary auditory cortex presumably involved in the processing of the spectrotemporal properties of the target and the masker (Sugita, 1997). The only study in humans (Micheyl *et al.*, 2003) demonstrated that changes in the masking of an interrupted target sound are accompanied by changes in the mismatch-negativity. The mismatch-negativity is a short-latency ERP thought to reflect the pre-attentive detection of unexpected acoustic events in auditory cortex (Näätänen *et al.*, 2001). Therefore, these results suggested that continuity illusions may recruit auditory cortical regions involved in pre-attentive processing of masking changes. The problem with this interpretation is that neither of the two studies could investigate influences of actual continuity illusions since behavioral data were not obtained during the physiological measurements. Neurophysiologic evidence regarding the mechanisms underlying the continuity illusion is thus scarce. This issue is addressed in the studies presented in chapter three to five.

5 Research aims and outline of the thesis

The general purpose of the present thesis is to broaden our understanding of the ability to track sounds in noisy environments. The main research question addressed by the different studies is: How does the auditory system construct smooth and meaningful sound percepts from noisy acoustic stimuli? The mechanisms underlying this phenomenon are investigated by means of the continuity illusion, and the findings are interpreted in the proposed framework for sensory-perceptual transformations (see paragraph 1).

The psychoacoustic study presented in chapter two aims at fine-tuning different acoustic parameters of masking noise for neurophysiologic investigations of the continuity illusion of tones. Another purpose of this study is the modelling of interactions between the acoustic parameters that induce masking, and of their effects on the continuity illusion. The stimulus parameters derived from that study are used in a subsequent fMRI study presented in chapter three. This study attempts to localize auditory cortical regions that are involved in processing of acoustic properties of masking noise and in perceiving continuity illusions. The same stimuli and paradigm are applied in the EEG study discussed in chapter four. This study investigates the relative timing of the previously identified processes. Chapter five presents an fMRI study that is identical to the previous fMRI study, except that voice sounds are employed as stimuli. This study asks whether cortical processes involved in continuity illusions of complex natural sounds differ from those involved in illusions of simple synthetic tones. Chapter six describes a psychoacoustic study that addresses the question whether the continuity illusion adapts to different acoustic contexts. Finally, chapter seven summarizes the findings of these studies and proposes a neural model for the continuity illusion.

References

- Alain C, Bernstein LJ (2008). From sounds to meaning: the role of attention during auditory scene analysis. *Curr Opin Otolaryngol Head Neck Surg*, 16, 485–489.
- Barlow HB. (1961). Possible principles underlying the transformation of sensory messages. In *Sensory Communication*, IW Rosenblith, ed. (Cambridge, MA: MIT Press), pp. 217–234.
- Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, 10, 417–429.
- Belin P, Zatorre JR, Lafaille P, Ahad P, Pike B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312.
- Berger H. (1930). Über das Elektrenkephalogramm des Menschen II. *J Psychol Neurol*, 40, 160–179.
- Braaten RF, Leary JC. (1999). Temporal induction of missing birdsong segments in European starlings. *Psychol Sci*, 10, 162–166.
- Bregman AS. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Broadbent D. (1958). *Perception and Communication*. London: Pergamon Press.
- Carlyon RP. (2004). How the brain separates sounds. *Trends Cogn Sci*, 8, 465–471.
- Chechik G, Anderson MJ, Bar-Yosef O, Young ED, Tishby N, Nelken I. (2006). Reduction of information redundancy in the ascending auditory pathway. *Neuron*, 51, 359–368.
- Cherry EC. (1953). Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am*, 25, 975–979.
- Ciocca V, Bregman AS. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Percept Psychophys*, 42, 476–484.
- Durlach NI, Braida LD. (1969). Intensity perception: I. A theory of intensity resolution. *J Acoust Soc Am*, 46, 372–383.
- Epstein CM. (1983). *Introduction to EEG and evoked potentials*. Philadelphia: JB Lippincott Co.
- Fechner G. (1860). *Elements of Psychophysics*. New York: Holt, Rinehart, and Winston.
- Fletcher H. (1940). Auditory patterns. *Rev Modern Physics*, 12, 47–65.
- Fritz J, Shamma SA, Elhilali M, Klein DJ. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*, 6, 1216–1223.
- Gescheider G. (1997). *Psychophysics: the fundamentals*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Griffiths TD, Warren JD. (2004). What is an auditory object? *Nat Rev Neurosci*, 5, 887–892.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. (1999). 'Sparse' temporal sampling in auditory fMRI. *Hum Brain Mapp*, 7, 213–223.
- Houtgast T. (1972). Psychophysical evidence for lateral inhibition in hearing. *J Acoust Soc Am*, 51, 1885–1894.
- Las L, Stern EA, Nelken I. (2005). Representation of tone in fluctuating maskers in the ascending auditory system. *J Neurosci*, 25, 1503–1513.
- Lee CC, Imaizumi K, Schreiner CE, Winer JA. (2004). Concurrent tonotopic processing streams in auditory cortex. *Cereb Cortex*, 14, 441–451.
- Logothetis N, Schall D. (1989). Neuronal correlates of subjective visual perception. *Science*, 245, 761–763.
- Lopes da Silva F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr Clin Neurophysiol*, 79, 81–93.
- Micheyl C, Carlyon RP, Shtyrov Y, Hauk O, Dodson T, Pullvermüller F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *J Cogn Neurosci*, 15, 747–758.
- Miller CT, Dibble E, Hauser MD. (2001). Amodal completion of acoustic signals by a nonhuman primate. *Nat Neurosci*, 4, 783–784.
- Miller GA, Licklider JCR. (1950). The intelligibility of interrupted speech. *J Acoust Soc Am*, 22, 167–173.
- Miller LM, Escabi MA, Read HL, Schreiner CE. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol*, 87, 516–527.

Chapter 1

- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends Neurosci*, 24, 283–288.
- Nelken I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol*, 14, 474–480.
- Ogawa S, Tank D, Menon R, Ellerman JM, Kim SG, Merkle H, Ugurbil K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89, 5951–5955.
- Patterson RD. (1976). Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am*, 48, 640–654.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.
- Petkov CI, O'Connor KN, Sutter ML. (2003). Illusory sound perception in macaque monkeys. *J Neurosci*, 23, 9155–9161.
- Rauschecker JP, Tian B, Pons T, Mishkin M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol*, 382, 89–103.
- Singer W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Ann Rev Physiol*, 55, 349–374.
- Smith EC, Lewicki MS. (2006). Efficient auditory encoding. *Nature*, 439, 978–982.
- Sugita Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *Neuroreport*, 8, 1155–1159.
- Talavage TM, Edmister WB. (2004). Nonlinearity of fMRI responses in human auditory cortex. *Hum Brain Mapp*, 22, 216–228.
- Wang X, Lu T, Bendor D, Bartlett E. (2008). Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience*, 154, 294–303.
- Warren RM, Obusek CJ, Ackroff JM. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149–1151.
- Warren RM, Bashford JA, Healy EW, Brubaker BS. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Percept Psychophys*, 55, 313–322.
- Warren RM. (1999). *Auditory Perception: A new Analysis and Synthesis*. Cambridge, UK: Cambridge University Press.
- Wrightson JM, Warren RM. (1981). Incomplete auditory induction of tones alternated with noise: Effects occurring below the pulsation threshold. *J Acoust Soc Am*, 69, S105–106.

Quantification of acoustic influences on illusory continuity and proposal of a neural filter model

Abstract

A sound that is briefly interrupted by a silent gap is perceived as discontinuous. However, when the gap is filled with noise, the sound may be perceived as continuing through the noise. It has been shown that this continuity illusion depends on the masking of the omitted target sound, but the underlying mechanisms have yet to be quantified thoroughly. In this paper we systematically quantify the relation between perceived continuity and the duration, relative power, or notch width of the interrupting broadband noise for interrupted and non-interrupted amplitude-modulated tones at different frequencies. We fitted the psychometric results to estimate the range of the noise parameters that induce auditory grouping. To explain our results within a common theoretical framework, we applied a power spectrum model to the different masking results and estimated the critical bandwidth of the auditory filter that may be responsible for the continuity illusion. Our results set constraints on the spectral resolution of the mechanisms underlying the continuity illusion, and provide a stimulus set that can be readily applied for neurophysiologic studies of its neural correlates.

1 Introduction

The human auditory system can identify and select relevant sounds in loud and noisy environments in which the sensory input is a mixture of acoustic signals. One illustration of this phenomenon is the auditory continuity illusion (Miller & Licklider, 1950) in which a sound is perceived as continuous even though parts of it have been replaced by another sound, e.g. a noise burst (Fig. 1). The illusion has been observed for steady-state tones, frequency-modulated (FM) sweeps, familiar melodies, and even for complex speech signals. In all cases, the addition of an interrupting sound makes the overall stimulus more perceptible: Thus the phenomenon has been labeled perceptual restoration, synthesis, fusion, or induction. The illusion has been demonstrated in humans as well as in non-human species (monkeys [Miller *et al.* 2001; Petkov *et al.*, 2003], cats [Sugita, 1997], and birds [Braaten & Leary, 1999]), suggesting a general constructive mechanism of auditory perception that operates at multiple levels of abstraction. The illusion seems to reflect perceptual sensitivity to faint, but expected sounds and robustness against contaminating noise, resulting in an enhanced perceptual signal-to-noise ratio (SNR).

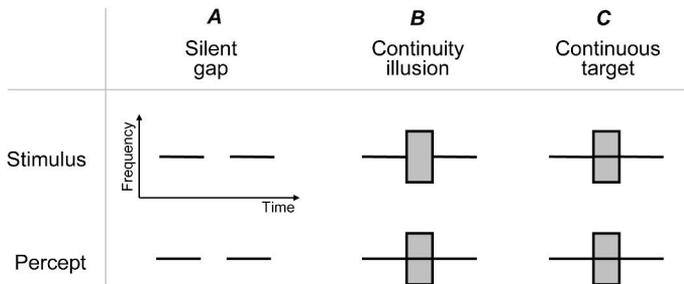


Figure 1. The auditory continuity illusion, exemplified for steady-state tones. **A**, Two tones interrupted by a silent gap are perceived as two independent entities. **B**, Illusion of a single entity when a broadband noise masker is added to the gap, a percept that is similar to a physically continuous tone (**C**).

But how can the continuity illusion of the target sound arise from a stimulus in which that sound is in fact discontinuous? This question is of particular interest as it may shed light into more general mechanisms of auditory perception. It is known that the omitted target can be perceptually restored only when it has been replaced by another sound that would be able to mask the target if it were actually present (Warren *et al.*, 1988; for reviews, see Bregman, 1990; Warren, 1999). Auditory masking refers to a perceptual phenomenon in which one sound, the masker, renders another sound, the target, inaudible. Masking occurs when the spectrogram of the masker obliterates the spectrogram of the target. This effect can be attenuated by removal of the overlapping frequency band from the masker (*i.e.* insertion of a spectral notch), and it can be enhanced by increasing the masker's intensity relative to that of the target (*i.e.* decreasing the SNR) (Bregman, 1990). Masking may result

in the continuity illusion if the target precedes and continues after the masker (Warren, 1999), which reflects the tight coupling between the two phenomena (Houtgast, 1972; Warren *et al.*, 1972). The salience of the continuity illusion thus depends on the degree of masking which, in turn, depends on the degree of spectrotemporal concealment of the target by the masker (Bregman, 1990; Warren, 1999). The salience of the continuity illusion is further influenced by the duration of the masker (Kluender & Jenison, 1992). Earlier studies with pure tones have demonstrated prominent continuity illusions with masker durations of 10–300 ms (Warren, 1999). For longer-duration maskers, however, the illusion tends to fade, evoking percepts of partially continuous targets. Thus, the illusion is not an all-or-none phenomenon, but reflects a gradual extrapolation of the preceding target that extends perceptually through (parts of) the masker (Bregman, 1990; Warren *et al.*, 1994).

According to the Gestalt principles of auditory grouping (Wertheimer, 1938; Bregman, 1990), spectral and temporal proximity of the interrupted tone and masking noise burst are required to induce continuity illusions. However, these Gestalt principles ignore the effect of the relative spectral power of masker and target (see above). Furthermore, despite extensive psychophysical evidence for the proximity requirement, the underlying grouping mechanisms remain poorly quantified. For example, it is unknown whether the proximity principles vary with center frequency. A conceptual problem in the study of the continuity illusion arises with the use of very brief maskers (100 ms or less) (e.g. Petkov *et al.*, 2003; Darwin, 2005) and of maskers that contain power in the target's frequency (*i.e.* the on-frequency) band (e.g. Darwin, 2005; Lyzenga *et al.*, 2005). Without comparison of continuity illusions to non-illusory continuity percepts of appropriate control stimuli, it remains unclear whether listeners could hear the actual targets in very brief maskers. Strictly speaking, if the masker contains the on-frequency band the frequency channel of the target sound is physically continuous and no illusory filling occurs in that channel. Furthermore, previous studies have used different target sounds to investigate how individual stimulus parameters may influence the illusion (Bregman, 1990; Warren, 1999), but more abstract parameters (e.g. interactions between the spectrotemporal concealment and the relative spectral power) were not investigated and thus cannot be derived *ad hoc* from the individual earlier studies.

According to a current view the continuity illusion can be considered to result from neural mechanisms that extract abstract stimulus properties in order to form perceptual objects (Griffiths & Warren, 2004; Darwin, 2005). In the present study we will apply the concept of the auditory filter to provide a constraint on such neural mechanisms. The auditory filter model considers the auditory system as a bank of overlapping linear band-pass filters, and accounts for the fact that masking of a tone is restricted to a narrow spectral band, the so-called critical band (CB) around the tone's frequency (Fletcher, 1940). According to the power spectrum model (Patterson & Moore, 1986), the detection of a tone in noise occurs in the auditory filter closest to the tone's frequency, and the total noise power passing through that filter determines the amount of masking. The CB is often expressed in terms of the equivalent

rectangular bandwidth (ERB), which is considered as a measure of the spectral resolution of the auditory system around the tone's frequency (Moore, 2003).

Typically, auditory filters and ERBs have been used to interpret masking threshold data from rippled- or notched-noise experiments (Patterson, 1976). Here, we consider as masking thresholds a range of limiting stimulus parameters that lead to continuity illusions (Houtgast, 1974) and apply the concept of an auditory filter to the continuity illusion. We propose that the "illusory" filter properties need not be identical to those of the classical filter. According to this idea, the auditory filter contributes to the formation of a continuity illusion by grouping all frequencies within its CB. For example, if the sensory input to the filter comprises a spectral notch smaller than the CB, and if masking conditions are met, then the frequencies within the critical band, whether they are physically present or not, are merged to fill in the notch at the output of the filter. Such a filter might thus function as an integrator whose ERB represents only the frequencies that are capable to yield the continuity illusion.

The present study aims at the quantification of the proposed mechanisms involved in the continuity illusion. We designed a series of psychophysical experiments in which we systematically assessed the perceptual consequences of varying the masking noise across large parameter ranges. Noise burst-interrupted amplitude-modulated (AM) tones of different frequencies were used to generate spectrally balanced stimulus sets comprising a physically discontinuous and a physically continuous tone, respectively. Using the method of constant stimuli and a scaling procedure in a repeatedly presented paradigm, we analyzed the perceptual responses from a large pool of listeners, and quantified the relation between the salience of the perceived continuity of the target tone and the duration, relative power, or notch width of the noise masker while controlling for potential frequency effects. These parameters are known to control the masking potential of the noise and, therefore, the salience of the continuity illusion. We analyzed the salience of our illusions by comparing illusory to non-illusory continuity evoked by physically discontinuous and continuous targets, respectively. By fitting the psychometric results we predicted the range over which each noise parameter could establish the illusion. Adopting the power spectrum model, we determined the ERB of the auditory filters as a function of the tone's frequency for a range of detection thresholds. Our results indicate that strong illusions involved filters similar to those reported by classical masking studies, suggesting that common masking mechanisms may account for these illusions. For partial illusions the filters' width increased, suggesting a neural mechanism that may be different from that involved in masking.

2 Materials and methods

2.1 Participants and stimuli

29 volunteers (age: 22 ± 2 years, mean \pm standard deviation [SD], 25 females), mainly undergraduate students from the University of Maastricht, were paid for their participation and gave their informed consent. All participants had normal hearing

abilities as assessed with an initial hearing test before every experiment. Out of the 29 listeners, 16, 12, and 12 listeners participated in the noise duration, SNR, and notch width experiment, respectively, with nine listeners participating in two experiments, and one listener (one of the authors) participating in all three experiments. Except this last listener, all listeners were uninformed about the experimental background.

Tones interrupted by a noise burst were used to generate two different stimulus types. The experimental stimulus comprised three temporally non-overlapping sequential segments. The target tone was removed from the center segment and replaced by a noise masker, resulting in a discontinuous target. The control stimulus was identical to the experimental one in all other respects, except that the target remained in the center segment while the temporally overlapping masker was superimposed, resulting in a continuous target. The non-illusory continuity percepts of targets in control stimuli served to assess the salience of illusory continuity percepts evoked by experimental stimuli. Moreover, a correct continuity percept of the control stimuli provided listeners with a reference for perceptual ratings of the respective experimental stimulus.

Stimulus duration was set to 2800 ms. To equate for overall sound level, stimuli were equated on the root mean square (RMS) amplitude. This might have resulted in slight intensity differences between targets. According to informal observations and listeners' reports, the putative differences were virtually inaudible. Tones of five different frequencies (500, 930, 1732, 3223, 6000 Hz) were used as targets to test for the generality of the continuity illusion across different frequencies. Carrier tones were amplitude-modulated with a sinusoidal modulator (3-Hz frequency, 100% depth). All tones had linear rise and fall times of 3 ms. Broadband white Gaussian noise bursts were temporally and spectrally centered on the stimulus at linear and logarithmic scales, respectively. The non-modulated noise was filtered with a two-octave band-pass finite impulse response (FIR) filter centered on the target frequency. To create a spectral notch, frequency bands around the target frequency were removed using a FIR band-reject filter. Noise had linear rise and fall times of 3 ms. Ramp centers were synchronized with those of the respective tone off- and onsets. Stimuli were digitally generated in Matlab 7.0.1 (The MathWorks, Inc., Natick, MA, USA) using a 44.1-kHz sampling rate and 16 bits per sample. Non-spatial mono stimuli were delivered binaurally via headphones using a Creative Sound Blaster Audigy 2ZS sound card (Creative Technology, Ltd., Singapore) and Presentation 9.30 software (Neurobehavioral Systems, Inc., Albany, CA, USA).

Three experiments were conducted (Fig. 2). The noise in the stimulus was characterized by three parameters: duration (in ms), relative amplitude (in dB) given by

$$\text{SNR(dB)} = 20 \times \log_{10} \left(\frac{\text{Tone amplitude}}{\text{Noise amplitude}} \right) \quad (1)$$

and notch width (in octaves). In each of the experiments, one parameter was manipulated while the other two were set to their near-limiting values still associated with continuity illusions.

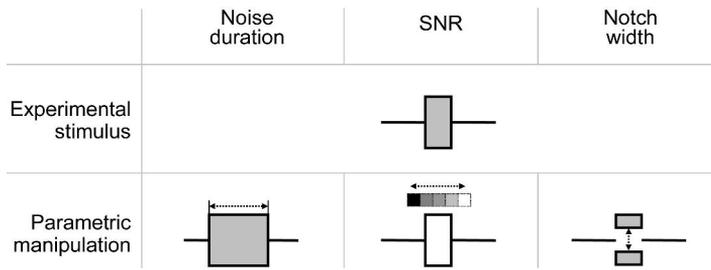


Figure 2. Experimental manipulations. Per experiment, a different noise parameter was varied in six steps across trials. Parametric manipulations were applied similarly to both experimental and control stimuli. The gray scale represents different levels of masker intensity.

In the first experiment, noise duration was varied in six steps (200, 400, 600, 900, 1400, 2000 ms), while noise amplitude was set to equal tone amplitude (SNR: 0 dB) and the notch width was set to zero. In the second experiment, noise duration was set to 600 ms (a near-limiting value in the preceding experiment, see Results), whereas SNR was varied in six steps (-8, -6, -3.5, 0, +6, +12 dB) and notch width was set to zero. In the third experiment, noise duration was set to 600 ms, and SNR to -3.5 dB (near-limiting values in the preceding experiments, see Results), whereas notch width was varied (0, 0.25, 0.5, 0.75, 1, 1.25 octaves). The stimulus parameters are summarized in Table 1.

	Noise duration experiment	SNR experiment	Notch width experiment
Noise duration	200–2000 ms	600 ms	600 ms
SNR	0 dB	-8 to +12 dB	-3.5 dB
Notch width	0 octaves	0 octaves	0–1.25 octaves

Table 1. Stimulus parameter settings in the experiments. The following parameters were fixed across experiments: Stimulus duration (2800 ms), tone frequency (500, 930, 1732, 6000 Hz), tone AM (3 Hz), noise bandwidth (2 octaves).

2.2 Task and design

Listeners set the overall sound intensity to an individual hearing level (~82 dB SPL on average) that remained fixed throughout the experiment. The current task was a modified version of the standard continuity illusion paradigm (Bregman, 1990) that we designed specifically for potential future applications in neuroimaging environments. Listeners were instructed on a screen to attend to targets in stimuli and to rate how likely the tone appeared to continue during the entire noise burst on each trial. Ratings were entered through a button press on a four-point scale (1: “most likely continuous”, 2: “probably continuous”, 3: “probably discontinuous”, 4: “most likely discontinuous”) while the rating-button key remained visible on the screen throughout

the task. Within each trial, a 2800-ms stimulation period was visually indicated by a green cross that turned red during the following response period. The upper response limit was 5000 ms; earlier responses terminated response periods earlier and initiated the next trial. After reading the instructions, listeners practiced on 18 test trials. The test stimuli also included the minima and maxima of the parametric stimulus range so that listeners could adjust their rating scale to the stimulus scale available in the subsequent experiment.

In each experiment, three factors or independent variables (indicated in italics in the following) were systematically varied: stimulus (two types: experimental, control), frequency (five values, see Stimulus generation), and noise parameter (duration, SNR, or notch width with six levels each: see Tab. 1 for details). This resulted in a $2 \times 5 \times 6$ within-subject design with 60 conditions per experiment. One stimulus was presented and one response was measured per trial. Trial duration varied from 3300 to 8300 ms (500-ms inter-trial interval + 2800-ms stimulus + 0–5000 ms response time). In total, 360 trials were presented per experiment resulting in an average experiment duration of approximately 30 minutes. Trials were organized in three types of blocks each of which included two adjacent levels of the noise parameter. More precisely, the first, second, and third block type comprised stimuli at noise parameter levels one and two, three and four, and five and six, respectively. Each block type was presented three times resulting in a total number of nine blocks presented per experiment. Block order was balanced and pseudo-randomized so that successive blocks were never of the same type. Successive blocks were always separated by a task break. Listeners were free to terminate breaks once they felt confident to resume the experiment. Within each block, $2 \times 5 \times 2 = 20$ different conditions (stimulus \times frequency \times noise per block) were presented twice resulting in a total number of 40 trials per block. The two noise levels and the five frequencies were both balanced and randomized within blocks. Out of all trials, $\frac{2}{3}$ was experimental trials and $\frac{1}{3}$ was control trials; the stimulus types were randomized within and between blocks. We quantified the mean response on each condition, averaged across nine (experimental stimulus) and three (control stimulus) repetitions, ranging from value one (“most likely continuous”) to value four (“most likely discontinuous”).

2.3 Statistical analysis and calculation of ERBs

Group data were statistically analyzed in SPSS 12.0.1 (SPSS inc., Chicago, ILL, USA) using general linear models (GLMs), and univariate as well as multivariate tests for repeated measures. Stimulus, frequency, and noise were included as three within-subject factors in the model with two, five, and six levels, respectively. For all three experiments, the three-way interaction among these factors turned out to be non-explanatory regarding the variance in the perceptual response. The non-significant interaction term was thus removed and the reduced GLM was re-analyzed. Two-way interactions were treated step-wise according to the same schema. Two independent factors were analyzed separately for main effects, while two dependent factors were analyzed for simple effects (*i.e.* one factor was investigated separately per level of

the other factor and vice versa). Pair-wise comparisons between conditions within experiments were performed using Paired Samples *t*-tests, whereas pair-wise comparisons between conditions between experiments were performed using Independent Samples *t*-tests. Inflated type-I error probabilities caused by multiple comparisons were corrected for using Bonferroni's method.

The power spectrum model (Patterson & Moore, 1986) and the notched-noise method (Patterson, 1976) were adopted to estimate the auditory filter bandwidth from detection thresholds of the illusory tone in the notched-noise masker at two different notch widths. Specifically, a fixed-tone paradigm was used in which the expected level P_s (in dB SPL) of an illusory tone at frequency f_0 (in Hz) was kept constant while the noise spectrum level N_0 (in dB SPL/Hz) varied with the notch width Δf (in Hz) at detection threshold. P_s was defined as the level of the tones surrounding the noise. Thresholds of the continuity illusion in the SNR and notch width (NW) experiments were defined by an average rating score of 2.5. Since the continuity illusion is a gradual phenomenon (see Introduction) the choice for a subjective criterion may have affected ERB estimation. To take this into account a range of additional thresholds was also examined. The corresponding noise spectrum levels were predicted from fitting a third-order polynomial to the psychometric results, using the least-squares error criterion. Assuming a perfectly rectangular filter centered on f_0 with constant weighting function in its pass-bands, the efficiency of the target detection process at filter output is

$$K = \frac{P_s}{W \times N_0} \quad (2)$$

where W is the masker bandwidth within the filter bandwidth ERB at detection threshold. Note that for notched noise, W is smaller than ERB, whereas for non-notched noise W , is smaller than or equals ERB. Assuming that masker bandwidth (BW: two octaves) exceeded ERB (Fig. 3), Equation (2) becomes

$$K_{SNR} = \frac{P_{s,SNR}}{ERB \times N_{0,SNR}} \quad (3)$$

for the SNR experiment and

$$K_{NW} = \frac{P_{s,NW}}{(ERB - \Delta f) \times N_{0,NW}} \quad (4)$$

for the notch width experiment. According to the power spectrum model (Patterson & Moore, 1986) the threshold for detecting a tone corresponds to a constant ratio of tone power to total masker power at the output of the filter. Thus,

$$K_{SNR} = K_{NW} \quad (5)$$

at detection threshold, and Equations (3) and (4) can be solved for ERB:

$$ERB = \frac{\Delta f}{1 - \frac{P_{s,NW} \times N_{0,SNR}}{P_{s,SNR} \times N_{0,NW}}} \quad (6)$$

where P_s and N_0 are expressed on a linear scale.

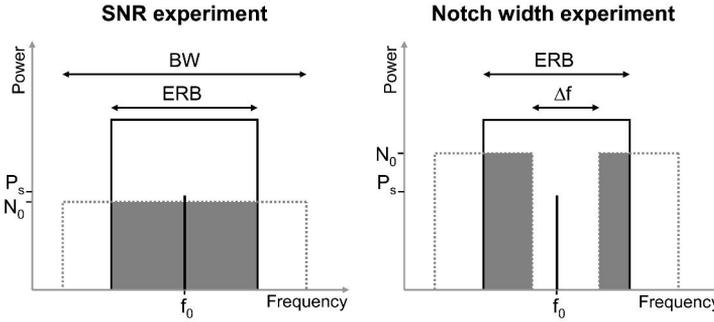


Figure 3. Auditory filter and power spectrum model. Spectra of tone targets and noise maskers at detection threshold in the SNR (left) and notch width (right) experiment are outlined in relation to the auditory filter on linear-linear scales. The black rectangles delineate the assumed shape of the auditory filter with an equivalent rectangular bandwidth ERB. It is assumed that for detection of a tone with frequency f_0 and fixed level P_s (represented by the black vertical lines) in noise a filter is recruited that is linearly centered on f_0 . The dotted gray lines delineate the white noise masker with bandwidth BW and constant spectrum level N_0 . In the notch width experiment, the masker comprised a spectral notch of width Δf placed non-symmetrically around f_0 on a linear scale. The shaded area represents the total noise power transmitted by the filter at detection threshold which is assumed to be constant across the two different conditions.

The ERBs were calculated according to Equation (6) from group-averaged data for each tone frequency, which implies that listeners used a constant rating scale between experiments. To test this assumption, ratings were first analyzed across experiments considering only conditions including identical stimuli (*i.e.* noise duration: 600 ms versus SNR: 0 dB, and SNR: -3.5 dB versus notch width: 0 octaves). Comparison between identical conditions in different experiments revealed that rating differences between experiments did not reach significance neither for control nor for experimental stimuli at any frequency. The identical conditions were either ranked closely (noise duration: 600 ms and SNR: 0 dB) or more remotely (SNR: -3.5 dB and notch width: 0 octaves) to each other on the respective stimulus scale. Thus, identical stimuli were rated similarly and independently of the stimulus scale available in the various experiments, supporting the assumption of constant rating scales and justifying the application of the power spectrum model to our data.

3 Results

3.1 Effects of noise duration, SNR, and notch width on perceived continuity

Figure 4A–C shows the psychometric curves in the noise duration, SNR, and notch width experiment, respectively, averaged across listeners and frequencies, whereas Figure 5 shows the corresponding curves for each frequency. The effects of noise duration, SNR, and notch width on ratings differed significantly for experimental and control stimuli (stimulus \times duration: $p_{5,11} < .00001$; stimulus \times SNR: $p_{5,7} < .00005$; stimulus \times notch: $p_{5,7} < .0005$): Experimental stimuli were rated as progressively more continuous when noise duration, SNR, and notch width decreased (Fig. 4A–C, solid

circles); these effects were significant at all frequencies (duration: $p_{5,11} < .01$; SNR: $p_{5,7} < .0005$; notch: $p_{5,7} = .0005$), except for 1732-Hz and 6000-Hz stimuli in the SNR experiment ($p_{5,7} > .05$). Control stimuli were rated as progressively less continuous when noise duration, SNR, and notch width decreased (Fig. 4A–C, open circles), but these trends did not reach significance at any frequency (duration: $p_{5,11} > .1$; SNR: $p_{5,7} > .05$; notch: $p_{5,7} > .1$). Thus, the perceived continuity of physically discontinuous targets, but not that of physically continuous targets, depended on the noise masker properties.

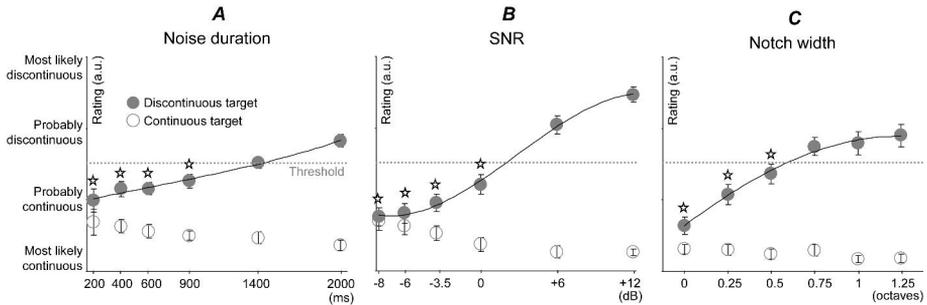


Figure 4. Mean percept (\pm SE) of experimental stimuli (solid circles) and control stimuli (open circles) in the noise duration (A), SNR (B), and notch width (C) experiment averaged across listeners and frequencies. The solid lines represent the fitted polynomial functions and the dotted horizontal lines represent perceptual ambiguity between discontinuity and continuity. Illusory stimuli are indicated by asterisks. a.u., arbitrary units.

Regarding differences between illusory and non-illusory continuity, listeners perceived physically discontinuous targets as continuity illusions (*i.e.* rating scores of 2.5 or less) in 73.3, 63.3, and 53.3 % of the experimental conditions of noise duration, SNR, and notch width experiment, respectively (Fig. 4A–C, Fig. 5, see asterisks). The control stimuli were generally rated as continuous, irrespective of noise duration, SNR, or notch width. Occasional incorrect discontinuity ratings in individual listeners were observed in 6.7, 6.1, and 2.5 % of the control conditions of noise duration, SNR, and notch width experiment, respectively, and occurred mainly at short noise durations and low SNRs. Thus, the short noise durations and low SNRs evoked the most continuity illusions as well as the most incorrect discontinuity ratings. The perceptual differences between experimental and control stimuli (*i.e.* the vertical distances between solid and open circles in Fig. 4A–C) decreased with decreases in noise duration, SNR, and notch width, suggesting that listeners had illusions that gradually assimilated the ‘real’ continuity percepts of the control stimuli. Control stimuli were generally rated as more continuous than experimental stimuli, a difference that was evident for 81.8 % of the conditions associated with continuity illusions (duration: $p_{15} < .001$; SNR: $p_{11} < .0005$; notch: $p_{11} < .001$). Only for low SNRs of -8 and -6 dB these stimulus type differences did not reach significance ($p_{11} > .1$). Thus, except for very intense noise maskers, listeners could implicitly differentiate between illusory and non-illusory continuity.

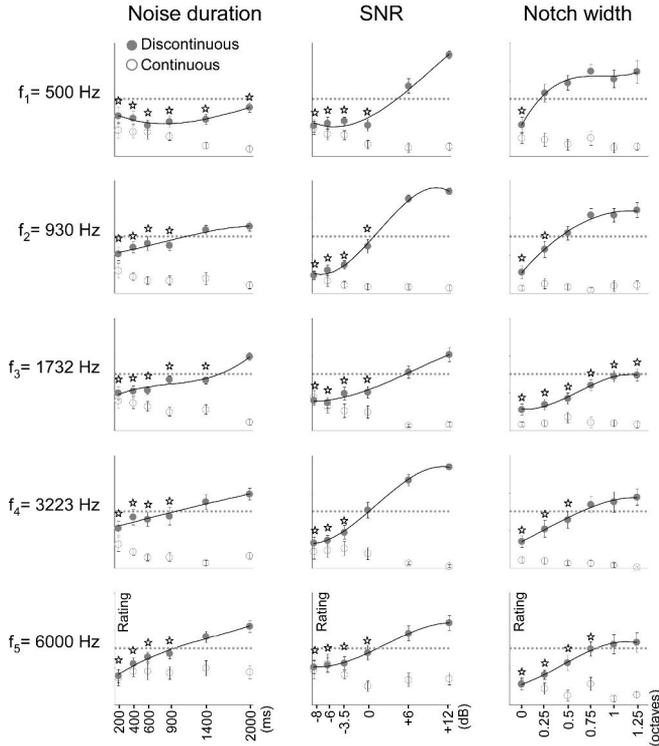


Figure 5. Mean percept (\pm SE) of experimental stimuli (solid circles) and control stimuli (open circles) per frequency (rows) across listeners. The solid lines represent the fitted polynomial functions and the dotted horizontal lines represent perceptual ambiguity between discontinuity and continuity. Illusory stimuli are indicated by asterisks. The plots have the same scales as those in Figure 4A–C.

Regarding limiting stimulus parameter values, continuity illusions were reported for noise durations up to about 900 ms, for noise intensities at least as high as tone intensity (*i.e.* $\text{SNR} \leq 0$ dB), and for notch widths up to half an octave, pooled across frequencies (Fig. 4A–C, see asterisks). The limiting noise parameter values at which percepts of experimental stimuli would shift between non-illusory and illusory continuity were predicted by the fitted functions and are shown in Table 2. Fitting of the polynomials to the experimental psychometric results yielded a coefficient of determination R^2 of 0.99 ± 0.01 (mean \pm SD) for the frequency-pooled data and an R^2 of 0.96 ± 0.03 for the non-pooled data.

	Noise duration [ms]	SNR [dB]	Notch width [octaves]	ERB [octaves]
f₁ = 500 Hz	> 2000	1.7	0.19	0.23
f₂ = 930 Hz	1100	1.2	0.41	0.56
f₃ = 1732 Hz	1568	1.9	> 1.25	1.4
f₄ = 3223 Hz	956	1.0	0.64	0.96
f₅ = 6000 Hz	938	1.2	0.79	1.02
Pooled f	1406	1.3	0.58	–

Table 2. Predicted limiting noise parameter values inducing shifts between non-illusory and illusory continuity percepts (*i.e.* detection threshold 2.5) of physically discontinuous targets and estimated auditory filter bandwidths in octaves, displayed for each frequency and pooled across frequencies.

3.2 Critical bandwidth of the auditory filters

The predicted limiting SNR and notch width values (*i.e.* rating score of 2.5) were considered as detection thresholds (Tab. 2). For these conditions, the average testing level P_s was estimated as 81.7 dB SPL. Small testing level differences between experiments ΔP_s might have been induced by stimulus level normalization (average ΔP_s : 1.4 dB SPL) and could be neglected (Moore & Glasberg, 1981; Glasberg & Moore, 1982). The average noise levels were estimated as 79.7 and 84.5 dB SPL in the SNR and notch width experiment, respectively. These values corresponded to noise spectrum levels (power densities) N_0 of 45.6 and 52.1 dB SPL/Hz, respectively. The ERB values at the individual frequencies were estimated as 78, 356, 1571, 2072, and 4084 Hz, and are plotted in Figure 6 (circles). The corresponding ERB values in units of octaves were 0.23, 0.56, 1.4, 0.96, and 1.02 octaves (Tab. 2).

Fitting of a linear function revealed that the obtained ERB values could be approximated by the equation

$$\text{ERB}(f_0) = 0.7 \times f_0 - 124.3 \quad (7)$$

with an accuracy of $R^2 = 0.97$, where ERB and f_0 are expressed in Hz. The obtained ERB values exhibited a monotonic increase with f_0 , consistent with filter widths obtained from classical masking experiments (crosses; Glasberg & Moore, 1990). Thus, the bandwidth of the masker frequencies that were required for tonal continuity illusions increased with the center frequency of the illusion. Application of additional detection thresholds was constrained to those rating scores that were common to SNR and NW experiment for each center frequency. Figure 7 shows that when threshold was increased (*i.e.* the continuity illusion faded) the estimated ERBs values increased consistently across center frequencies.

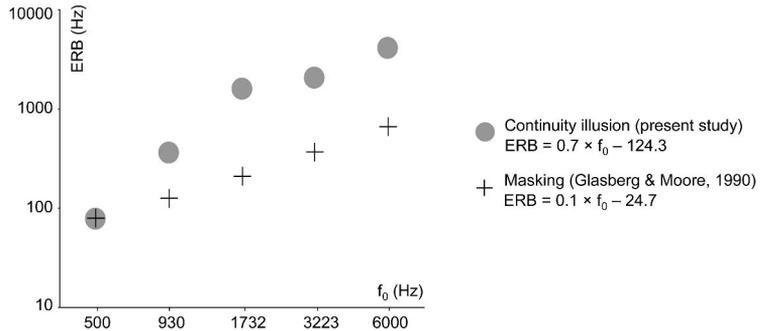


Figure 6. Equivalent rectangular bandwidth ERB of the auditory filters underlying continuity illusions (circles) as a function of tone frequency f_0 , plotted on logarithmic-logarithmic scales. The fitted linear function (see upper equation) approximates the obtained ERB values, exhibiting a monotonic increase with f_0 . A similar psychometric function was suggested by Glasberg and Moore (1990, see lower equation) on basis of classical notched-noise masking studies. The ERB values predicted by this function (crosses) exhibit a similar trend but the increases are smaller at high center frequencies.

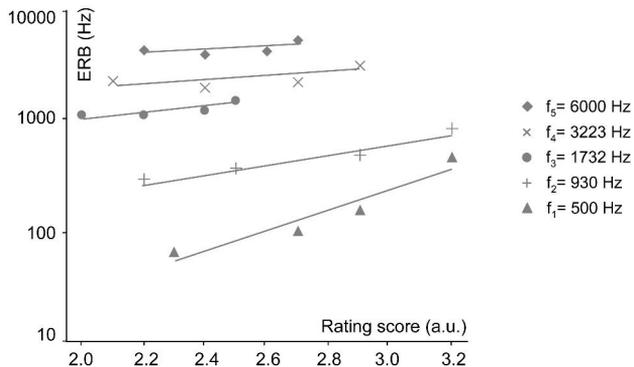


Figure 7. Equivalent rectangular bandwidth ERB of auditory filters underlying continuity illusions of varying strength as a function of detection threshold, plotted on logarithmic-linear scales for each center frequency. The fitted linear functions (gray lines) approximate the obtained ERB values (symbols), exhibiting an increase at higher detection thresholds (*i.e.* at weaker continuity illusions) consistently across center frequencies. Detection thresholds of 2.0 and 3.0 correspond to the ratings 'probably continuous' and 'probably discontinuous', respectively. The ERB values predicted for the intermediate threshold (rating score 2.5) correspond to the circles in Figure 6. a.u., arbitrary units

3.3 Frequency effects

As suggested by Figure 5, the center frequency had no effect on the effects of noise duration, SNR, and notch width, except for experimental conditions in the noise duration and SNR experiment (frequency \times duration: $p_{1,15} < .05$ [$> .1$]; frequency \times SNR: $p_{1,11} < .01$ [$> .1$]; frequency \times notch: $p_{1,11} > .1$ [$> .1$]; control stimuli in

parentheses). The absence of a general frequency \times noise interaction supported the pooling of the results across frequencies (Fig. 4A–C). As indicated by the asterisks in Figure 5, stimuli of 500, 930, 1732, 3223, and 6000 Hz were rated as continuity illusions in 61.1, 50, 83.3, 55.6, and 66.7 % of the experimental conditions across experiments. Regarding the respective control conditions, 3.8, 1.7, 5.8, 2.5, and 11.25 % were rated incorrectly as discontinuous. There was no significant effect of frequency on the number of illusory continuity or incorrect discontinuity percepts, but a high correlation (Spearman's $R = 0.9$) between the two percept types across frequencies. Listeners perceived the longest continuity illusions mainly at 500 Hz, whereas the most filled-in illusions (*i.e.* illusions at the largest notch width) were perceived mainly at 1732 Hz (Fig. 5). Remarkably, these illusions lasted up to 2000 ms or involved noise maskers with spectral gaps of up to 1.25 octaves.

4 Discussion

The aim of our research was to quantify the mechanisms that may underlie the continuity illusion. Based on our psychometric results, we estimated detection thresholds for continuity illusions for different noise maskers at different center frequencies and calculated the critical bandwidth of the proposed auditory filters that may underlie continuity illusions. By including continuous control stimuli we assessed the salience of illusory continuity percepts relative to that of 'real' (non-illusory) continuity percepts.

4.1 Masker duration, level, and bandwidth influence illusory but not non-illusory continuity

Our results show that the duration, level, and bandwidth of a noise masker influence the perceived continuity of physically discontinuous tones (Fig. 4A–C). As expected, the most salient continuity illusions were observed for short gaps that were strongly masked. For longer gaps or weaker maskers, the illusions gradually shifted to non-illusory discontinuity percepts. The physically continuous tones were always perceived as continuous independent of the noise maskers. The absence of a significant masker effect for the control stimuli suggests that listeners could not differentiate between masked and non-masked continuous tones. The auditory system might thus smoothly embed continuous sounds of interest in the ongoing context irrespective of the background noise level.

The result that the continuity illusion becomes increasingly prominent with decreases in SNR and notch width is consistent with previous results (Houtgast, 1972; Warren *et al.*, 1972; Warren *et al.*, 1988; Kluender & Jenison, 1992; Sugita, 1997; Petkov *et al.*, 2003) and supports the spectral proximity principle. In the SNR experiments, the masker contained the on-frequency band and thus had spectrotemporal overlap with the target. The results indicate that the continuity illusion requires an average SNR of 1.3 dB or less, depending on the frequency. Thus, mere acoustic power within the on-frequency channel is not sufficient for the illusion to arise. Unless masked by a sound of similar or greater intensity, the target's

offset might induce an unexpected spectral edge that might be interpreted as an object offset, thereby yielding a discontinuity percept (Bregman, 1990).

In the notch-width experiments, the masker was sufficiently intense to mask the target (*i.e.* SNR: -3.5 dB), although the spectral notch did not conceal the spectrogram of the interrupted target. Interestingly, despite the physical discontinuity in the on-frequency band, our data from the notched-noise experiment predicted continuity illusions for average notch widths of up to 0.58 octaves, which supports that physical continuity within the target frequency band is not required for the illusion to arise (Houtgast, 1972). We observed the strongest effects at 1732 Hz, for which a 1.25-octave notch still produced continuity illusions. The perceived continuity and masking can thus not be explained solely on the basis of concealment among two spectrograms. Rather, the missing on-frequency band must have been filled-in by a mechanism that may be related to masking.

4.2 A proposed role of the auditory filter in the continuity illusion

Our notched-noise experiments show that the continuity illusion of a tone requires only the frequencies that match and/or surround the tone's frequency, whereas more remote frequencies do not seem to have an influence. In addition, the SNR experiments indicate a minimal noise power below which the continuity illusion disappears. The results of these two experiments can be jointly understood through the concept of the auditory filter.

The auditory filter is a physiologically plausible model as the frequency tuning curves of auditory neurons have a certain bandwidth around the cell's best frequency. This bandwidth appears to increase along the ascending auditory pathway. Given this neuronal property, a notched-noise stimulus may be represented as a neurally filtered version of its spectrogram (Shamma, 2001) in which sharp edges and narrow notches have been smoothed and partially restored. Depending on the size of the neurons' spectrotemporal receptive fields, the processing of notched noise during an interruption of a less intense target sound might induce spatial-temporal overlap of activations in representations of proximate frequency bands. Given the continuity in the target frequency band before and after the noise, this may account for the perceptual filling-in of the preserved continuous activation in the target frequency representation (Warren *et al.*, 1972; Beauvois & Meddis, 1996). Electrophysiological research recently reported a possible neural correlate of such a mechanism in monkeys' primary auditory cortex during the integration of auditory streams (Fishman *et al.*, 2004).

We found that the reconstructed critical bandwidth of spectral integration (*i.e.* ERB) increases monotonically with the filter's center frequency (Fig. 6). Comparison of the ERBs at different center frequencies in octave units (Tab. 2) showed that the highest relative ERB occurred at 1732 Hz. This result is also reflected by listeners' ratings in the notch width experiments, showing the largest ratio of illusions as well as the largest range of filled-in notches for these stimuli. Presumably, frequencies around the 1732-Hz band (which is at the logarithmic center of the human hearing range) are more likely to be perceptually grouped than more peripheral frequencies. Stated

differently, the spectral resolution of the mechanisms underlying continuity illusions may be low when they operate on bands that are well-resolved otherwise.

Our results are qualitatively consistent with classical masking studies (for review, see Glasberg & Moore, 1990) and with an earlier pulsation threshold study (Houtgast, 1974), both of which also reported monotonic increases of ERB with center frequency. Our results show that this increase is more pronounced for continuity illusions than for masking (slope 0.7 versus 0.1, Fig. 6). Note that Houtgast (1974) reported ERB values that were two times smaller for pulsation thresholds than for simultaneous masking thresholds, irrespective of the center frequency. Several causes may underlie this apparent difference. First, Houtgast investigated the perceived temporal character of pure tones that alternated repeatedly with rippled-noise bursts of very short duration (125 ms). Second, the testing levels in Houtgast's study were about 40 dB lower than in our study. Previous research has shown that the ERB increases with the testing level (e.g. Weber, 1977; Lufti & Patterson, 1984; Moore & Glasberg, 1987) especially at high center frequencies (Rosen & Stock, 1989; Baker & Rosen, 2006). However, the reported effects of level were smaller than those observed in our study. Third, listeners in Houtgast's study were aware of the tone's absence and received feedback about their performance by adjusting the noise level on each trial; these two factors may have lowered the detection threshold.

Differences in detection threshold between pulsation and masking were already reported by Warren and colleagues (1988) who also alternated maskers and targets of short (300 ms) duration and provided listeners with feedback during threshold estimation. The estimated detection thresholds for illusions of infratone sounds in noise were lower than those for masking and were about 8 dB SPL lower than those reported in our study. Since the continuity illusion is a gradual phenomenon (see Introduction) and masking filters are typically estimated from more objective criteria, the reported difference in pulsation and masking threshold possibly reflects a difference between estimation techniques (i.e. detection of a perceptually restored sound versus a physically present sound). Interestingly, the detection threshold for the illusion was reported to increase when the target's amplitude fluctuated below 10 Hz, which may also explain the observed increases in filter width for our illusory tones which were amplitude-modulated at 3 Hz (for a potential neural account see Warren *et al.*, 1988). Investigating whether the choice for a particular detection threshold affected our estimation of filter widths, we found that increases in detection threshold (i.e. focusing on less complete illusions like those obtained at long noise durations, Fig. 4) were indeed associated with qualitative increases in ERB estimation, whereas at stricter criteria for continuity the width of the illusory filter may assimilate that of the masking filter (Fig. 7).

In sum our results suggest that decreases in the strength of the continuity illusion can be accounted for by filters of increasing bandwidth. Such wide-band filters may 'smooth' the spectrogram of tone and noise more and, therefore, facilitate their integration across larger bands while weakening the noise's local masking effect. This may lead to drops in the perceptual SNR, i.e. to fuzzier and less certain continuity percepts. For very strong continuity illusions, narrow-band filters seem to

be involved. Such filters may integrate the stimulus within narrower bands while preserving the noise's masking power, supporting the idea that masking-related mechanisms may underlie salient illusions. More research is required to investigate whether illusory perceptual grouping and tonal masking obey different principles at different stages of processing along the auditory pathway.

4.3 A proposed role of an acoustic short-term buffer in the continuity illusion

Our result that the illusion becomes increasingly prominent for decreases in masker duration is consistent with previous results (Kluender & Jenison, 1992; Warren, 1999) and supports the temporal proximity principle. While earlier pure tone studies have reported continuity illusions of up to 300 ms, our results show that appropriate stimuli can evoke illusions of up to 900 ms on average, even extending up to 1400 ms. One potential reason for these long illusions may be the amplitude modulation of our target sounds. A potential amplitude modulation effect on the duration of the illusion has so far not been investigated, but the modulation rate and depth of the continuity illusion are preserved with non-modulated maskers (Lyzenga *et al.*, 2005) and may influence the strength of masking (Warren *et al.*, 1988). A second contributing factor to long illusions might be the explicit instruction to listeners to attend to targets. Attention has also been shown to facilitate perceptual grouping of auditory events in streaming tasks (Alain & Arnott, 2000; Carlyon *et al.*, 2001; Cusack *et al.*, 2004).

We observed that the upper noise duration limit and thus the temporal proximity principle depend on the center frequency of the continuity illusion (Fig. 5). In particular, we obtained the longest continuity illusions of 2000 ms for 500-Hz tones. Since the illusion arises *post hoc*, *i.e.* based on the continuity of the target after the masker (Warren, 1999), an acoustic short-term buffer (Cowan, 1984) is likely involved. Specifically, the proposed buffer may retain the pre-masker target and match it with the post-masker target within a critical window of temporal integration. Our results put constraints on the proposed buffer's latency and suggest that within the preferred human hearing range the buffer performs best for low frequency bands.

Interestingly, we observed that the ratio of continuity illusions correlated with the ratio of incorrect discontinuity percepts of physically continuous tones. Presumably, listeners were more prone to interpret discontinuous targets as continuity illusions when the tone, if present during the same masker, was difficult to identify. This was indeed the case for the shortest noise duration and the lowest SNR. Consistent with the proposed acoustic short-term buffer (Cowan, 1984), the perceptual difficulties with 200-ms noise bursts might be ascribed to the time scales at which unified identifiable auditory objects are represented: The time scales might be too large to perceptually resolve partial sensory information occurring within such a short time window. The proposed mechanism might sample the continuity illusion stimuli more globally as integrated abstract objects, rather than as redundant continuous signals (*e.g.* Darwin, 2005).

4.4 Potentially confounding ranging effects

A limitation of the method of constant stimuli is that the observed effects of the stimulus parameterization might be contaminated by listeners' response bias. Moreover, the blocked design of our experiments might have encouraged listeners to adjust their rating scales to the different stimulus scales available within different blocks, rather than to the entire experiment, which would have confounded between-condition comparisons. However, there are two reasons why listeners were more likely to adjust their rating scales to the entire experiment. First, our results show a significant monotonic trend across conditions in all three experiments (Fig. 4A–C), which would not be expected for ratings based on within-block scaling. Second, before each experiment listeners were trained on 18 non-blocked trials including the most extreme conditions on the stimulus scales, familiarizing them with the entire stimulus scale presented in the subsequent experiment. Similarly, our result that ratings of identical conditions did not differ significantly between experiments further suggests that listeners used a relatively stable rating scale independent of the available blocked stimulus scales. Thus, the assumptions underlying the between-condition and between-experiment comparisons were supported by our data.

4.5 A well-adapted stimulus set for neurophysiologic studies on the continuity illusion

So far, human neurophysiological data on the neural mechanisms underlying the continuity illusion are scarce (Micheyl *et al.*, 2003) and there is a strong need for further research (Husain *et al.*, 2005). Our results provide a well-adapted stimulus set for future neurophysiologic studies of the continuity illusion. Specifically, the center frequencies are logarithmically centered on the human's best hearing range and have representations roughly equidistant along the tonotopic gradient on the cortical surface (Merzenich *et al.*, 1976), rendering them ideal for experiments on tonotopy (e.g. Formisano *et al.*, 2003). Moreover, the stimuli evoke AM tone illusions of several hundreds of milliseconds, *i.e.* the putative neural correlates may be sampled even with low temporal resolution techniques. AM stimuli were previously reported to enhance the SNR of auditory stimulus-evoked functional magnetic resonance imaging (fMRI) signals (Hart *et al.*, 2003). Furthermore, the salience of the illusions can be parametrically altered by notch width changes in the noise, providing a paradigm that can be readily applied. A similar paradigm has already been applied successfully by Davis & Johnsrude (2003) who investigated the neural correlates of speech intelligibility across different maskers and masking levels with fMRI.

A potential problem is that brain signals related to percept changes are difficult to dissociate from those related to stimulus changes. However, one could exploit our paradigm to define perceptually ambiguous stimuli that were rated equally often as continuous and discontinuous (Tab. 2). Such a repeatedly presented ambiguous stimulus would allow to define the experimental protocol *post hoc* according to the different perceptual responses, which is an efficient strategy to eliminate confounds from physical stimulus changes (e.g. Cusack, 2005).

Acknowledgements

We would like to thank Peter de Weerd and Noël Staeren for comments on the experimental design, Gerard van Breukelen for support with the statistical analysis, and Hanna Renvall, Noël Staeren, Nicolette Siep, and Jasper van den Bosch for constructive comments on the manuscript.

References

- Alain C, Arnott SR. (2000). Selectively attending to auditory objects. *Front Biosci*, 5, 202–212.
- Baker RJ, Rosen S. (2006). Auditory filter nonlinearity across frequency using simultaneous notched-noise masking. *J Acoust Soc Am*, 119, 454–462.
- Beauvois MW, Meddis R. (1996). Computer simulation of auditory stream segregation in alternating tone-sequences. *J Acoust Soc Am*, 99, 2270–2280.
- Braaten RF, Leary JC. (1999). Temporal induction of missing birdsong segments in European starlings. *Psychol Sci*, 10, 162–166.
- Bregman AS. (1990). *Auditory scene analysis: The perceptual organization of sound* (4th Ed.). Cambridge, MA: The MIT Press, Bradford Books.
- Carlyon RP, Cusack R, Foxtan R, Robertson IH. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perform*, 27, 115–127.
- Cowan N. (1984). On short and long auditory stores. *Psychol Bull*, 96, 341–370.
- Cusack R, Deeks J, Aikman G, Carlyon RP. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J Exp Psychol Hum Percept Perform*, 30, 643–656.
- Cusack R. (2005). The intraparietal sulcus and perceptual organization. *J Cogn Neurosci*, 17, 641–651.
- Darwin CJ. (2005). Simultaneous grouping and auditory continuity. *Percept Psychophys*, 67, 1384–1390.
- Davis MH, Johnsrude IS. (2003). Hierarchical processing in spoken language comprehension. *J Neurosci*, 23, 3423–3431.
- Fishman YI, Arezzo JC, Steinschneider M. (2004). Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am*, 116, 1656–1670.
- Fletcher H. (1940). Auditory patterns. *Rev Modern Physics*, 12, 47–65.
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40, 859–869.
- Glasberg BR, Moore BCJ. (1982). Auditory filter shapes in forward masking as a function of level. *J Acoust Soc Am*, 71, 946–949.
- Glasberg BR, Moore BCJ. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear Res*, 47, 103–138.
- Griffiths TD, Warren JD. (2004). What is an auditory object? *Nat Rev Neurosci*, 5, 887–892.
- Hart HC, Palmer AR, Hall DA. (2003). Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cereb Cortex*, 13, 773–781.
- Houtgast T. (1972). Psychophysical evidence for lateral inhibition in hearing. *J Acoust Soc Am*, 51, 1885–1894.
- Houtgast T. (1974). *Lateral Suppression in Hearing*. Doctoral dissertation, Vrije Universiteit, Amsterdam.
- Husain FT, Lozito TP, Ulloa A, Horwitz B. (2005). Investigating the neural basis of the auditory continuity illusion. *J Cogn Neurosci*, 17, 1275–1292.
- Kluender KR, Jenison RL. (1992). Effects of glide slope, noise intensity, and noise duration on the extrapolation of FM glides through noise. *Percept Psychophys*, 51, 231–238.
- Lufti RA, Patterson RD. (1984). On the growth of masking asymmetry with stimulus intensity. *J Acoust Soc Am*, 76, 739–745.
- Lyzenga J, Carlyon RP, Moore BC. (2005). Dynamic aspects of the continuity illusion: Perception of level and of the depth, rate, and phase of modulation. *Hear Res*, 210, 30–41.
- Merzenich MM, Kaas JH, Roth GL. (1976). Auditory cortex in the grey squirrel: tonotopic organization and architectonic fields. *J Comp Neurol*, 166, 387–402.
- Micheyl C, Carlyon RP, Shtyrov Y, Hauk O, Dodson T, Pullvermüller F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *J Cogn Neurosci*, 15, 747–758.
- Miller CT, Dibble E, Hauser MD. (2001). Amodal completion of acoustic signals by a nonhuman primate. *Nat Neurosci*, 4, 783–784.
- Miller GA, Licklider JCR. (1950). The intelligibility of interrupted speech. *J Acoust Soc Am*, 22, 167–173.

- Moore BCJ. (2003). *An Introduction to the Psychology of Hearing* (5th Ed.). San Diego, CA: Academic Press.
- Moore BCJ, Glasberg BR. (1981). Auditory filter shapes derived in simultaneous and forward masking. *J Acoust Soc Am*, 70, 1003–1014.
- Moore BCJ, Glasberg BR. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear Res*, 28, 209–225.
- Patterson RD. (1976). Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am*, 48, 640–654.
- Patterson RD, Moore BCJ. (1986). Auditory filters and excitation patterns as representations of frequency resolution, in *Frequency Selectivity in Hearing*, edited by B.C.J. Moore (Academic, London).
- Petkov CI, O'Connor KN, Sutter ML. (2003). Illusory sound perception in macaque monkeys. *J Neurosci*, 23, 9155–9161.
- Rosen S, Stock D. (1989). Auditory filter bandwidths as a function of level at low frequencies (125 Hz– 1 kHz). *J Acoust Soc Am*, 92, 773–781.
- Shamma S. (2001). On the role of space and time in auditory processing. *Trends Cogn Sci*, 5, 340–348.
- Sugita Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *Neuroreport*, 8, 1155–1159.
- Warren RM. (1999). *Auditory Perception: A new Analysis and Synthesis*. Cambridge, U.K.: Cambridge University Press.
- Warren RM, Bashford JA, Healy EW, Brubaker BS. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Percept Psychophys*, 55, 313–322.
- Warren RM, Obusek CJ, Ackroff JM. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149–1151.
- Warren RM, Wrightson JM, Puretz J. (1988). Illusory continuity of tonal and infratonal periodic sounds. *J Acoust Soc Am*, 84, 1338–1342.
- Weber DL. (1977). Growth of masking and the auditory filter. *J Acoust Soc Am*, 62, 424–429.
- Wertheimer M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed.). *A source book of Gestalt psychology* (pp. 71–88). London: Routledge & Kegan Paul. (Original work published 1923).

FMRI-based localization of neural representations of acoustic and illusory continuity in auditory cortex

Abstract

A sound that is interrupted by silence is perceived as discontinuous. However, when the silence is replaced by noise, the target sound may be heard as uninterrupted. Understanding the neural basis of this continuity illusion may elucidate the ability to track sounds of interest in noisy auditory scenes, but yet little is known. In the present functional magnetic resonance imaging study in humans we report that activity in primary auditory cortex reflects perceived continuity of illusory tones in noise. Exploiting a parametric manipulation of the illusory stimuli, we show that stimulus-evoked activity does not correlate with the basic acoustic properties of tones or noises, but rather with the abstract dependencies among them. Importantly, changes of neural responses to acoustically identical stimuli parallel changes of listeners' report of perceived continuity of these same stimuli, thus confirming the perceptual nature of these responses. Our findings show that, beyond the sensory representation of an auditory scene, primary auditory areas play a constructive role in the grouping of scene segments into unified auditory percepts.

Reprinted from: Riecke L, Van Opstal J, Goebel R, Formisano E. (2007). Hearing illusory sounds in noise: Sensory-perceptual transformations in primary auditory cortex. *J Neurosci*, 27, 12684–12689., with permission from The Society of Neuroscience.

1 Introduction

We can recognize sounds of interest remarkably well even when these are masked by background noise. For example, when a tone is interrupted by noise, the interrupted sound may be heard as continuing through the noise. This continuity illusion illustrates the constructive nature of perception that may serve to enhance sensitivity to expected signals and to ensure robustness against background noise. Psychoacoustic research investigated the continuity illusion for various sounds (tones, sweeps, melodies, vocalizations, and speech), and in several species (birds [Braaten & Leary, 1999], cats [Sugita, 1997], monkeys [Miller *et al.*, 2001; Petkov *et al.*, 2003], humans [*e.g.* Warren, 1999]), suggesting that common perceptual mechanisms may operate at multiple levels of abstraction. The current view holds that the gap induces sudden energy changes in the frequency channel of the interrupted sound (target) which are therefore perceptually interpreted as target off- and onsets. When such changes in the so-called on-frequency band are masked by another sound (masker), they become less salient and the masker may be interpreted as containing the target. The continuity of the target after the gap facilitates the emergence of the illusory target during the interruption (Bregman, 1990).

The neural correlates of the continuity illusion are poorly understood. Previous research identified neurons in cats' primary auditory cortex (AC) presumably involved in the processing of the spectrotemporal properties of targets and maskers (Sugita, 1997). A recent electrophysiology study (Petkov *et al.*, 2007) reported that neurons in monkey core area AI respond to illusory tones interrupted by noise as if these tones were continuous. These results thus suggest that activity in primary AC reflects perceptual rather than acoustic stimulus properties. In humans, the only neurophysiological study (Micheyl *et al.*, 2003) so far demonstrated that changes in the on-frequency band of the target tone are accompanied by changes in the level of the mismatch-negativity. The mismatch-negativity is a short-latency event-related scalp potential thought to originate from AC; it may reflect the processing of an acoustic short-term buffer that detects unexpected acoustic deviances in predictable stimulus sequences (Näätänen *et al.*, 2001). Therefore, these results suggest that stimuli evoking continuity illusions recruit areas in AC involved in pre-attentive detection of unexpected on-frequency band changes. Unfortunately, neither study could relate brain activities to listeners' actual percepts as behavioral data were not obtained during the physiological measurements.

In the present human functional magnetic resonance imaging (fMRI) study we measure blood oxygenation level dependent (BOLD) responses in AC to illusory and non-illusory tones (targets) with different levels of masking and we identify the regions in which activity levels co-vary with the subjective report of perceived continuity. The masking of the omitted on-frequency bands was varied across four parametric levels, and listeners rated the targets' overall perceived continuity on a four-point scale while functional images were collected. To dissociate brain regions related to continuity illusions from those related to non-illusory continuity percepts we

presented interrupted and uninterrupted targets. We optimized stimuli and task for fMRI demands in a preceding psychoacoustics study (Riecke *et al.*, 2008).

2 Materials and methods

2.1 Participants

Eleven human volunteers (mean age: 26 years, seven women) with normal hearing abilities participated after providing informed consent. The local ethics committee approved the experimental protocol.

2.2 Stimuli, design, and task

Tones of varying frequency (500, 930, 1732, 3223 Hz) with sine-modulated amplitude (3 Hz) were used as targets. Broadband Gaussian-noise bursts were used as maskers. Noise bursts were band-passed and band-stopped (notched) with both filters centered on the on-frequency band. To vary the masking level, the width of the spectral notch was parameterized across 0, 0.25, 1.25, and 2 octaves. Maskers were superimposed on the spectrotemporal center of targets on a logarithmic-linear scale. In interrupted target stimuli, onsets and offsets of gaps and maskers were synchronized. All onsets and offsets were linearly ramped with 3-ms rise–fall times. Uninterrupted target stimuli were spectrotemporally matched, except that no gaps were inserted in the target. All stimuli were matched for overall spectral power and sampled with 16 bits at 44.1 kHz in Matlab 7.0.1 (The MathWorks Inc., Natick, MA, USA). Stimulus parameter settings are summarized in Figure 1A.

Before imaging, listeners were trained to attend to targets and to rate their overall perceived continuity on a four-point scale (labeled with “most likely continuous”, “probably continuous”, “probably discontinuous”, “most likely discontinuous”). Experimental conditions (Fig. 1B) were presented during blocks of stimulation, each of which comprised four targets of either low (500, 930 Hz) or high (1732, 3223 Hz) frequencies in randomized order delivered diotically via Commander XG headphones (Resonance Technology Inc., Northridge, CA, USA) at 70 dB sound pressure level (SPL). Different stimulation blocks were matched and randomized with respect to their number and order so that successive blocks always differed. Stimulation blocks alternated with stimulation-free baseline blocks of same duration (20 s) during which listeners rated the preceding stimuli.

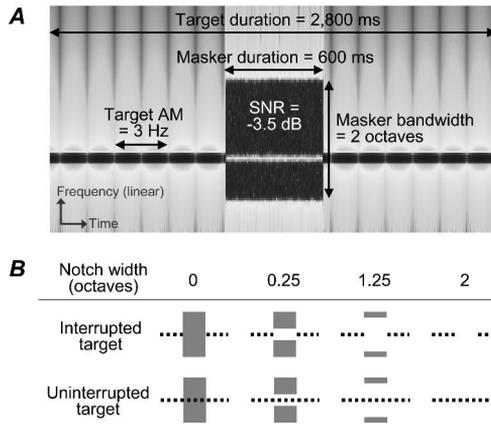


Figure 1. Stimulus and experimental design. **A**, Spectrogram and parameter settings, exemplified for a target stimulus of 1732 Hz, interrupted for 600 ms by a noise burst containing a 0.25-octave notch. The noise and the notch were logarithmically centered on the target frequency. AM, amplitude modulation; SNR, signal-to-noise ratio. **B**, Schematic stimulus spectrograms and labels of all eight stimulus conditions. Dots represent AM of tone targets which contain a temporal gap (upper) or not (lower). Gray rectangles represent noise maskers comprising a spectral notch of varying width.

2.3 Imaging and data preprocessing

Images were collected with a Siemens Allegra 3-Tesla magnetic resonance imaging system (Siemens Medical Systems, Erlangen, Germany). BOLD signal changes were measured with a head coil using a gradient echo planar imaging (EPI) sequence (time to echo: 30 ms, acquisition time: 2200 ms, repetition time [TR]: 5000 ms, field of view: $256 \times 256 \text{ mm}^2$, matrix size: 128×128 , slice thickness: 2 mm). During each experiment, 134 volumes were collected, each comprising 28 axial contiguous slices centered on the Sylvian fissure and covering the entire AC. In total, 40 functional runs were obtained (seven listeners \times four runs, four listeners \times three runs). Structural T1-weighted volumes optimized for gray-white matter contrast were obtained using three-dimensional modified-driven equilibrium with Fourier transform pulse sequences (voxel resolution: $1 \times 1 \times 1 \text{ mm}^3$).

The clustered volume EPI technique that was applied allowed for presentation of auditory stimuli in silence between subsequent volume acquisitions (e.g. Jäncke *et al.*, 2002; Van Atteveldt *et al.*, 2004). To increase statistical detection power and support listeners' attentiveness (Sha *et al.*, 2000; Gaab *et al.*, 2007) we used a relatively short TR, compared to that of other sparse temporal sampling techniques (e.g. Hall *et al.*, 1999). To minimize the resulting risk of the overlapping between stimulus-related BOLD responses and residual BOLD responses to preceding acoustic scanner noise that may arise from usage of such short TR (e.g. Belin *et al.*, 1999; Hall *et al.*, 1999; Talavage & Edmister, 2004) we took the following steps. First, we used carrier frequencies that avoided the peaks of the EPI noise spectrum (fundamental frequency at $\sim 1250 \text{ Hz}$) and presented the stimuli in silence, which

resulted in clear spectrotemporal separation between the acoustic stimuli and the scanner noise. Second, our listeners wore double ear protection (ear plugs and ear muffs) which attenuated the EPI noise level by ~50 dB SPL and allowed for stimulus presentation at a level perceived as much higher than that of the preceding EPI noise. Finally, listeners were instructed to attend to carrier frequencies, but not to the ambient EPI noise, which may have further enhanced the contrast-to-noise (CNR) ratio of the measured BOLD signals (Jäncke *et al.*, 1999; Hall *et al.*, 2000). The results (see supplemental Fig. S1A) indicate appropriate CNR in our BOLD data.

The collected fMRI data were analyzed using Brain Voyager QX 1.7 (Brain Innovation, Maastricht, The Netherlands) and Matlab. Pre-processing included head motion correction (no head motion exceeded 1.5 mm), inter-slice acquisition time correction, temporal high pass filtering (cutoff: 7 cycles per time course), functional-anatomical image co-registration, normalization to Talairach space, and cortical surface reconstruction. To maximally preserve anatomical specificity, group analysis was performed using cortical alignment methods (Goebel *et al.*, 2006) and non-smoothed functional data. Analysis was also performed in conventional Talairach space using functional data spatially smoothed with a 4-mm full-width half-maximum Gaussian kernel.

2.4 Statistical analysis

2.4.1 General linear model analysis

Listeners' perceived continuity rating data were averaged across block repetitions and analyzed in SPSS 12.0.1 (SPSS inc., Chicago, ILL, USA) using a general linear model (GLM) and an analysis of variance for repeated measures. For fMRI data, group statistical maps were obtained with a vertex-by-vertex (or voxel-by-voxel, for Talairach space-based analysis) two-level random-effect analysis of the BOLD signal time series. At the first-level, a GLM was computed for each experiment using separate predictors for each listener. The predicted time courses were adjusted for the hemodynamic response delay by convolution with a double-gamma hemodynamic response function. At the second level, group contrast *t*-maps were obtained based on the parameter estimates (beta values) derived from the first level analysis.

In the stimulus-based analysis of the fMRI data, the GLM included two predictors for interrupted and uninterrupted targets, respectively: A binary predictor coded for target presentation, and a linearly scaled predictor coded for the notch width parameterization, respectively. The contrast between the notch width parameterization for interrupted versus uninterrupted targets was computed to identify voxels exhibiting significant masker \times gap interactions. The subsequent analyses in regions of interest (ROIs) involved assessment of BOLD response differences between individual conditions, based on a GLM that included a binary predictor for each condition (Fig. 1B).

In the subsequent percept-based analysis of the fMRI data, two linearly scaled predictors that coded for listeners' ratings of perceived continuity of interrupted and

uninterrupted targets (respectively) were added to the GLM. To allow for detection of perceptual changes that were unrelated to stimulus changes, the two predictors were orthogonalized with respect to all other stimulus-related predictors in the model as described in the following.

2.4.2 Orthogonalization

A general problem in physiological studies of perception stems from the fact that percept changes are typically induced by concomitant stimulus changes (stimulus-related changes). Given this collinearity, correlated BOLD signal changes can be attributed to changes in the stimulus changes and in the concomitant percept. However, percept changes may also occur in the absence of concomitant stimulus changes (stimulus-unrelated changes); e.g. the same stimulus may be perceived differently, depending on subjective factors like the amount of attention that the listener paid to the stimulus. BOLD signal changes that correlate with such stimulus-unrelated changes can be attributed exclusively to percept changes but not to stimulus changes.

In the current study, to test whether there was such additional stimulus-unrelated variance that could be explained exclusively by subjective changes (*i.e.* by perceived continuity changes), stimulus-related variance was first removed from the measured BOLD responses. To this end, the stimulus-unrelated variance from listeners' perceived continuity data was extracted, using a linear algebra procedure called Gram-Schmidt orthogonalization (Wilf, 1962). Since this procedure operates in vector space, the variance in each predictor (see GLM analysis) was first transformed into its vector representation (Bandettini *et al.*, 1993). The Gram-Schmidt process then extracted the component of a given vector that was orthogonal to all the components of another given vector. Specifically, the procedure extracted the part of variance in one predictor that was linearly uncorrelated to the variance in another predictor. This allowed for computation of two perceived continuity predictors (one for interrupted and one for uninterrupted targets) that were unrelated to the respective notch width predictors. Each of these linearly scaled predictors thus allowed identifying voxels with significant rating-related BOLD changes that occurred without concomitant changes in the respective notch width predictor. In other words, these predictors allowed for an analysis of partial correlations, *i.e.* the calculation of the relative amount of variance in the BOLD response that was explained exclusively by changes in the perceived continuity.

2.4.3 Thresholding

Random-effect maps were thresholded based on a three-dimensional extension of the randomization procedure described by Forman *et al.* (1995). First, a voxel-wise threshold was set to $t_{10} = 3.5$ (uncorrected $p < 0.005$). Thresholded maps were then submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on Monte Carlo simulations for estimating cluster-level false-positive rates after 1000 iterations. Maps were applied the minimum cluster size

threshold which yielded a corrected cluster-level false-positive rate of $p < 0.05$. Multiple comparison-corrected maps were superimposed on the cortical surface that was obtained from alignment of individual cortices (cortex-based analysis), or on the average volume (Talairach space-based analysis). The most significant clusters were defined as ROIs.

2.4.4 ROI time course analysis

ROI analysis of BOLD signal time courses was based on mean amplitudes defined by a time window around the peak of the condition-related average time courses. Mean amplitudes associated with the different notch width conditions for interrupted and uninterrupted target stimuli were normalized with respect to the mean amplitude in the respective zero-octave notch condition.

2.4.5 Best-frequency mapping

To identify the locations of auditory cortical fields (ACFs), a best-frequency mapping technique (Formisano *et al.*, 2003) was applied to the cortically aligned functional data. The analysis considered only vertices in which activity differed significantly between targets versus baseline (supplemental Fig. S1A). The two lower and upper carrier frequencies were binned into a low-frequency and high-frequency condition, respectively. Using GLM analysis, a frequency-selectivity index FS of the two associated predictors was computed as the ratio of their subtracted beta values to their summed beta values. Ratios were color-coded at each vertex with hue and saturation coding for the field sign and value of FS, respectively.

3 Results

3.1 Behavioral results

Consistent with our recent results obtained outside the scanner (Riecke *et al.*, 2008), behavioral data analysis revealed that perceived continuity increased significantly with masking levels for only interrupted targets but not for uninterrupted targets (masker \times gap interaction: $F_{3,79} = 164.4$, $p < 10^{-34}$; Fig. 2B). fMRI data analyses on the cortical sphere and in Talairach space yielded consistent results (see below).

3.2 Stimulus-based imaging results

Interrupted and uninterrupted targets evoked widespread activities in bilateral AC relative to baseline ($t_{10} = 4.6$, cluster size > 100 , corrected $p < 0.05$; supplemental Fig. S1A). The analysis of the parametric manipulation of the masker in these regions (based on the contrast between the parametric notch width predictors for interrupted versus uninterrupted targets) revealed that activity levels in the right lateral transverse gyrus (Heschl's gyrus [HG]) and in adjacent superior temporal gyrus (STG) correlated significantly differently with the masking levels of interrupted versus that of uninterrupted targets (masker \times gap interaction: $t_{10} = 2.6$, cluster size > 50 ,

corrected $p < 0.05$; Fig. 2A, supplemental Figs. S1B, S2, blue regions). Comparable effects were also observed in the left hemisphere ($t_{10} = 3.5$, cluster size > 100 , corrected $p < 0.05$; supplemental Fig. S1B). The most significant clusters in the right hemisphere were defined as ROIs and the patterns of their masker \times gap interactions were further specified (Fig. 2A, supplemental Fig. S1B, blue arrows). For interrupted targets, increases in masking were associated with significant increases in the strength of the continuity illusion ($F_{3,79} = 98.4$, $p < 10^{-26}$; Fig. 2B, supplemental Fig. S3) and significant decreases in brain activity in lateral HG and STG (Fig. 2B,C, supplemental Fig. S3; ROI analysis: $t_{10} = 2.6$ and 4.6 , $p = 0.02$ and 0.0009), whereas for uninterrupted targets no such parametric effects were observed (ROI analysis: $t_{10} = -1.1$ and -0.2 , $p = 0.3$ and 0.8). At the highest masking level (no-notch condition), interrupted targets were perceived as continuity illusions, and were not associated with significant activity differences in STG compared to the non-illusory continuity percepts of uninterrupted targets (ROI analysis: $t_{10} = -1.9$, $p = 0.08$).

3.3 Percept-based imaging results

The analysis so far suggested that lateral HG and STG play a role in the detection of changes in the on-frequency band as well as in the concomitant change in the perception of the overall continuity of the stimulus. To investigate whether the same regions were also involved in changes in listeners' actual percepts alone (*i.e.* without concomitant on-frequency band changes) another analysis was performed based on the reports of listeners' perceived continuity that had been collected during the functional measurements (see Materials and Methods). Strikingly, this analysis revealed that the changes in perceived continuity that were unrelated to acoustic stimulus changes correlated significantly with activity in a lateral portion of the right HG for only interrupted targets (rating \times gap interaction: $t_{10} = 2.6$, cluster size > 50 , corrected $p < 0.05$, Fig. 2A, green regions; ROI analysis for interrupted and uninterrupted targets: $t_{10} = 4.9$ and -0.6 , $p = 0.0005$ and 0.5 respectively, Fig. 2C; ROI locations indicated in Fig. 2A, supplemental Fig. S1B, green arrows). Comparable effects were also observed in the left hemisphere ($t_{10} = 3.5$, cluster size > 100 , corrected $p < 0.05$; supplemental Fig. S1B). As these changes in the BOLD responses were not related to stimulus-driven changes, they likely reflected spontaneous changes in stimulus interpretation. Note that these BOLD response changes were largest for the most illusory percepts (interrupted targets, no-notch condition, supplemental Fig. S3).

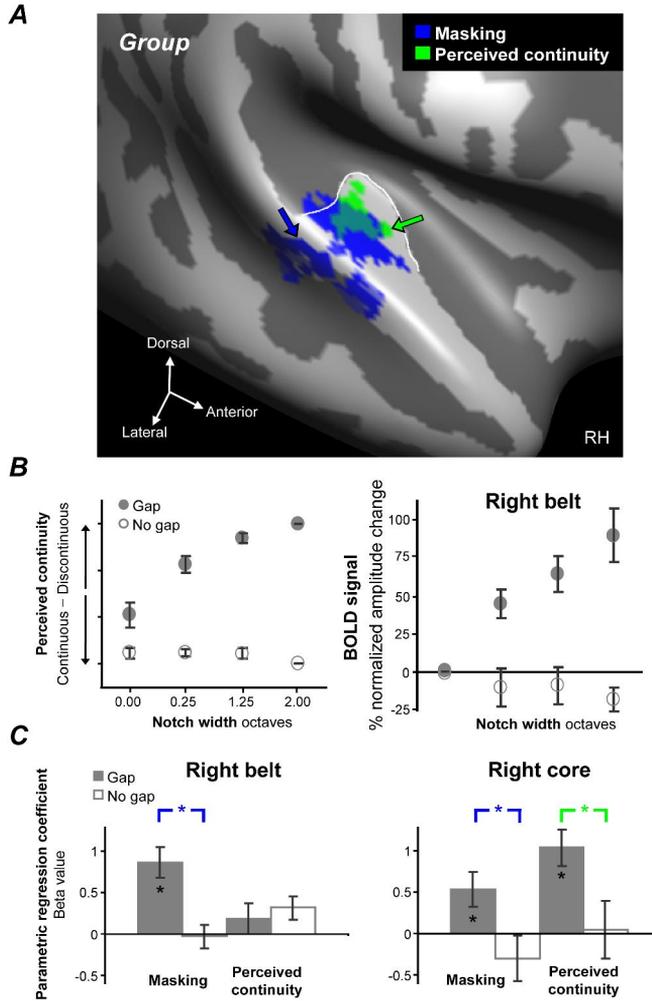


Figure 2. Neuroimaging and behavioral group results. **A**, Parametric activation maps projected onto the inflated average cortical surface of the right hemisphere (RH) after cortical alignment across listeners. Activity in core and adjacent belt areas correlated with masking (blue regions) and perceived continuity (green regions) for only interrupted targets (random-effect analysis: corrected $p < 0.05$). The white line delineates the location of Heschl's gyrus. **B**, Perceived continuity (left) and BOLD signal changes (right) in belt areas (panel A, blue arrow) showed parametric effects of masking for interrupted (solid circles) but not for uninterrupted (open circles) targets across listeners. Circles and bars represent mean value and standard error (SE) across listeners. To allow for comparison between interrupted and uninterrupted targets, mean BOLD signals were normalized to the respective no-notch conditions. **C**, Parametric changes in masking had significant effects on BOLD responses to interrupted (solid rectangles) but not to uninterrupted (open rectangles) targets in core and belt areas. In the core areas (panel A, green arrow), perceived continuity of only interrupted targets significantly explained additional variance in BOLD responses. Rectangles and bars represent mean value and SE of the parametric regression coefficients across listeners. Asterisks indicate significance at $p < 0.05$. The differential effects of masking (blue asterisks) and perceived continuity (green asterisk) defined the activated brain regions shown in panel A.

3.4 Best-frequency mapping results

Best-frequency mapping identified a low frequency-selective cluster on HG that was surrounded by high frequency-selective clusters in more medial, rostral, and caudal regions (Fig. 3). This result was highly consistent across individual listeners (supplemental Fig. S4) and is compatible with results from previous human tonotopy studies that support the notion of multiple, mirror-symmetric primary ACFs along HG (e.g. Formisano *et al.*, 2003; Talavage *et al.*, 2004). Projection of the cluster that we found to co-vary with listeners' perceived continuity (Fig. 2A, green region) onto the best-frequency map (Fig. 3, green outline) revealed that this region overlapped with the low-frequency selective region in HG, suggesting its inclusion within the ACFs of the primary core. The other cluster that we found to co-vary with the broadband masker for only interrupted targets was located in a more lateral region in STG (Fig. 2A, lateral blue region) which may correspond to the ACFs of the secondary belt (Hackett *et al.*, 2001, Sweet *et al.*, 2005).

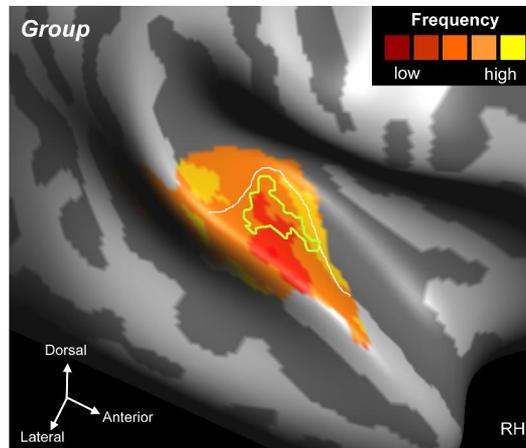


Figure 3. Best-frequency mapping group results. Best-frequency group maps projected onto the inflated average cortical surface of the right hemisphere (RH) after cortical alignment across listeners, showing cortical clusters of different frequency-selectivity (corrected $p < 0.05$; $FS > 0.29$; red, low frequencies; yellow, high frequencies). The region that was found to co-vary with listeners' perceived continuity (Fig. 2A, green region) was projected (green outline) onto the frequency selectivity map, revealing partial overlap with a low-frequency cluster in lateral Heschl's gyrus that may be part of the human primary core (Hackett *et al.*, 2001; Morosan *et al.*, 2001; Rademacher *et al.*, 2001; Sweet *et al.*, 2005; see Results for details). The white line delineates the location of Heschl's gyrus.

4 Discussion

Our finding that masking parametrically affects BOLD responses in the putative human core and belt areas for only interrupted targets suggests that neural activities in these areas do not reflect masking *per se* but rather the salience of on-frequency band changes, which depends on the level of masking. Thus, these modulations

point to perceptual differences between the two stimuli (*i.e.* perceived continuity differences, which varied across masking levels) rather than to spectrotemporal differences between interrupted and uninterrupted targets (*i.e.* differences in the gap, which were constant across masking levels). In line with current views on auditory object formation (Griffiths & Warren, 2004), we interpret these effects as reflecting the extraction of abstract relations between the basic spectrotemporal properties of targets and maskers and their transformation into perceptual and task-related representations. This interpretation is consistent with a recent electrophysiological study of the continuity illusion which suggested that neurons in the monkey core represent tones and interrupting noise bursts as if these were integrated (Petkov *et al.*, 2007). The result that discontinuity percepts evoke stronger BOLD responses than continuity illusions suggests that the observed activities primarily reflect the perception of onsets of new auditory objects. This interpretation is consistent with recent evidence from human neuroimaging studies (Mustovic *et al.*, 2003; Warren *et al.*, 2005; Herdener *et al.*, 2007; Wilson *et al.*, 2007) that employed non-illusory stimuli.

The finding that percept changes without concomitant stimulus changes also correlate with BOLD responses in lateral HG is consistent with recent evidence from studies on auditory streaming, a form of perceptual grouping that may be related to the continuity illusion (Bregman *et al.*, 1999). Perceived changes in streaming in the absence of stimulus changes were reported to correlate with neuromagnetic scalp signals that likely arise from lateral HG (Gutschalk *et al.*, 2005), and with single-cell responses in monkey core (Micheyl *et al.*, 2005). During perceptual switches these and other (Cusack, 2005) brain regions may be modulated by top-down processes. Similar regions have been shown to be involved in the processing of illusory pitch (Patel & Balaban, 2001; Bendor & Wang, 2005), especially in the right hemisphere (Zatorre, 1988; Zatorre *et al.*, 2002).

Regarding the precise functional localization of our effects, the results from the best-frequency mapping suggest that listeners' ratings of perceived continuity are included in a region that is most likely part of the human auditory core. However, the impossibility to compute the tonotopic gradient and the exact borders between the ACFs in the present data (related to the small number of frequencies and the group analysis) prevents specifying of particular core field(s). Similar conclusions can be drawn by comparing our results with previous cytoarchitecture studies that reported that the human homologue of the primary core is likely located in HG (Hackett *et al.*, 2001; Morosan *et al.*, 2001; Sweet *et al.*, 2005), even though the exact locations of primary area borders may vary in or around the gyrus (Rademacher *et al.*, 2001). In fact, the location of the region that we found to co-vary with listeners' perceived continuity appears in good agreement with the cytoarchitecturally defined human primary core area Te1.0 in HG (Morosan *et al.*, 2001; see supplemental Fig. S1B). Similarly, the regions that we localized in more lateral regions on STG probably coincide with the human secondary belt which may surround the core (Hackett *et al.*, 2001; Sweet *et al.*, 2005). Taken together our physiological and anatomical

considerations suggest that the detected regions in HG and in STG (Fig. 2A) are likely located in primary core and secondary belt, respectively.

It should be noted that our analyses complemented each other in differentiating two different types of cortical processing associated with our auditory task: The stimulus-based analysis identified cortical regions involved in the perceptual processing of abstract spectrotemporal stimulus properties. The percept-based analysis, on the other hand, revealed regions involved in the perceptual processing of cognitive and stimulus-unrelated factors. This differentiation was enabled by using the orthogonalization procedure (see Materials and Methods). Taken together our results indicate that: 1) the activity of early auditory cortical regions in humans does not correlate with the basic acoustic properties of tones or noise bursts alone, but also with the complex dependencies among them; and that 2) changes of neural responses in a core region (lateral HG) to acoustically identical sounds parallel listeners' reports of perceived continuity to these same sounds.

Together with the consideration that illusory continuity percepts are disrupted by the removal of the targets after the noise (Bregman, 1990), our findings suggest a two-stage model for cortical processing of stimuli evoking the continuity illusion. At a first stage, neuronal populations in the core and belt areas entail a representation of the auditory scene (tone plus noise) based on abstract dependencies between the individual scene segments, such as spectral edges between tone and noise or changes in the on-frequency band of the target. The result of this stage may be regarded as the formulation of a sensory-constrained perceptual hypothesis (Bregman, 1990), with the two possible outcomes of segregation of the scene into two (one continuous tone and noise) or three (two discontinuous tones and noise) perceptual objects. This hypothesis is resolved at a second stage in the core, in which the representation of the scene is re-evaluated based on the sensory evidence provided by the post-noise interval. Whereas the first processing stage is possibly fast and automatic, the second may be influenced by contextual, attentional and task-related processes, thus explaining the co-variance of neural and behavioral responses to the same illusory stimuli. The idea of a duplex analysis of the auditory scene in early auditory areas is in line with recent electrophysiological evidence in cats showing that the same primary AC neurons can represent stimuli at multiple time scales (Nelken, 2004; Ulanovsky *et al.*, 2004). The latter may provide the neural basis for an acoustic short-term buffer that would be required for retention of the perceptual hypotheses in our model. Consistently, the mismatch-negativity results from a previous study (Micheyl *et al.*, 2003) suggest that the analysis of changes in the target on-frequency band recruits a buffer in auditory cortical areas.

In conclusion, our findings provide strong evidence that beyond sensory representation of an auditory scene, primary auditory cortical areas also play a constructive role in the grouping of scene segments into unified percepts. Furthermore, they put forward a model for neural computation of auditory scenes that may inform analyses of physiological data obtained at high temporal resolution, such as with invasive or non-invasive electrophysiology and magnetoencephalography.

Acknowledgements

We thank Armin Heinecke and Fabrizio Esposito for help with the data analysis, Niclas Kilian-Huetten for support with the MR image segmentation, and Daniel Mendelsohn for constructive comments on the script.

References

- Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magn Reson Med*, 30, 161–173.
- Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, 10, 417–429.
- Bendor D, Wang X. (2005). The neural representation of pitch in primate auditory cortex. *Nature*, 436, 1161–1165.
- Braaten RF, Leary JC. (1999). Temporal induction of missing birdsong segments in European starlings. *Psychol Sci*, 10, 162–166.
- Bregman AS. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bregman AS, Colantonio C, Ahad PA. (1999). Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? *Percept Psychophys*, 61, 195–205.
- Cusack R. (2005). The intraparietal sulcus and perceptual organization. *J Cogn Neurosci*, 17, 641–651.
- Fischl B, Sereno MI, Tootell RBH, Dale AM. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*, 8, 272–284.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magn Reson Med*, 33, 636–647.
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40, 859–869.
- Gaab N, Gabrieli JDE, Glover GH. (2007). Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. *Hum Brain Mapp*, 28, 703–720.
- Goebel R, Esposito F, Formisano E. (2006). Analysis of functional image analysis contest (FIAC) data with Brain Voyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp*, 27, 392–401.
- Griffiths TD, Warren JD. (2004). What is an auditory object? *Nat Rev Neurosci*, 5, 887–892.
- Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci*, 25, 5382–5388.
- Hackett TA, Preuss TM, Kaas JH. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol*, 411, 197–221.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. (1999). “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp*, 7, 213–223.
- Hall DA, Haggard MP, Akeroyd MA, Summerfield AQ, Palmer AR, Elliott MR, Bowtell RW. (2000). Modulation and task effects in auditory processing measured using fMRI. *Hum Brain Mapp*, 10, 107–119.
- Herdener M, Esposito F, Di Salle F, Lehmann C, Bach DR, Scheffler K, Seifritz E. (2007). BOLD correlates of edge detection in human auditory cortex. *Neuroimage*, 36, 194–201.
- Howard MA III, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res*, 724, 260–264.
- Jäncke L, Mirzazade S, Sha NJ. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci Lett*, 266, 125–128.
- Jäncke L, Wüstenberg T, Scheich H, Heinze HJ. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, 15, 733–746.
- Micheyl C, Carlyon RP, Shtyrov Y, Hauk O, Dodson T, Pullvermüller F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *J Cogn Neurosci*, 15, 747–758.
- Micheyl C, Tian B, Carlyon RP, Rauschecker JP. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48, 139–148.

- Miller CT, Dibble E, Hauser MD. (2001). Amodal completion of acoustic signals by a nonhuman primate. *Nat Neurosci*, 4, 783–784.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage*, 13, 684–701.
- Mustovic H, Scheffler K, Di Salle F, Esposito F, Neuhoff JG, Hennig J, Seifritz E. (2003). Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage*, 20, 429–434.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends Neurosci*, 24, 283–288.
- Nelken I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol*, 14, 474–480.
- Patel AD, Balaban E. (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nat Neurosci*, 4, 839–844.
- Petkov CI, O'Connor KN, Sutter ML. (2003). Illusory sound perception in macaque monkeys. *J Neurosci*, 23, 9155–9161.
- Petkov CI, O'Connor KN, Sutter ML. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54, 153–165.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage*, 13, 669–683.
- Riecke L, Van Opstal AJ, Formisano E. (2008). The auditory continuity illusion: A parametric investigation and filter model. *Percept Psychophys*, 70, 1–12.
- Shah NJ, Steinhoff S, Mirzazade S, Zafiris O, Grosse-Ruyken ML, Jäncke L, Zilles K. (2000). The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination. *Neuroimage*, 12, 100–108.
- Sugita Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *Neuroreport*, 8, 1155–1159.
- Sweet RA, Dorph-Petersen KA, Lewis DA. (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *J Comp Neurol*, 491, 270–289.
- Talavage TM, Edmister WB. (2004). Nonlinearity of fMRI responses in human auditory cortex. *Hum Brain Mapp*, 22, 216–228.
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol*, 91, 1282–1296.
- Ulanovsky N, Las L, Farkas D, Nelken I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci*, 24, 10440–10453.
- Van Atteveldt N, Formisano E, Goebel R, Blomert L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43, 271–282.
- Warren JD, Jennings AR, Griffiths TD. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage*, 24, 1052–1057.
- Warren RM. (1999). *Auditory Perception: A new Analysis and Synthesis*. Cambridge, UK: Cambridge University Press.
- Wiif HS. (1962). *Mathematics for the Physical Sciences*. New York, NY: Dover Publications Inc.
- Wilson EC, Melcher JR, Micheyl C, Gutschalk A, Oxenham AJ. (2007). Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J Neurophysiol*, 97, 2230–2238.
- Zatorre RJ. (1988). Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am*, 84, 566–572.
- Zatorre RJ, Belin P, Penhune VB. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn Sci*, 6, 37–46.

Supplemental results

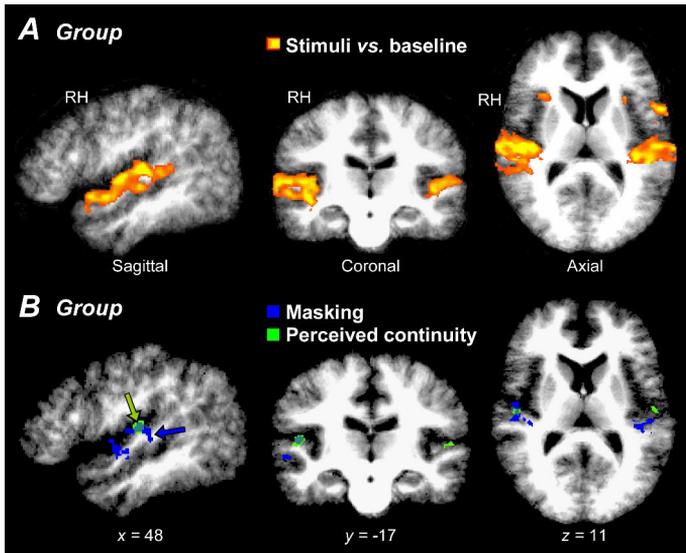


Figure S1. Parametric group activation maps projected onto listeners' average anatomical slices in Talairach space in neurological orientation. **A**, Task-related activations defined by the contrast stimuli versus baseline (corrected $p < 0.05$). **B**, The regions and peak regions (arrows) in Talairach space in which activity correlated with masking (blue) and perceived continuity (green) of only interrupted targets (corrected $p < 0.05$) were consistent with those obtained from cortex-based group analysis (Fig. 2A). Slice positions are indexed by Talairach coordinates (x , y , z) in units of millimeters. RH, right hemisphere.

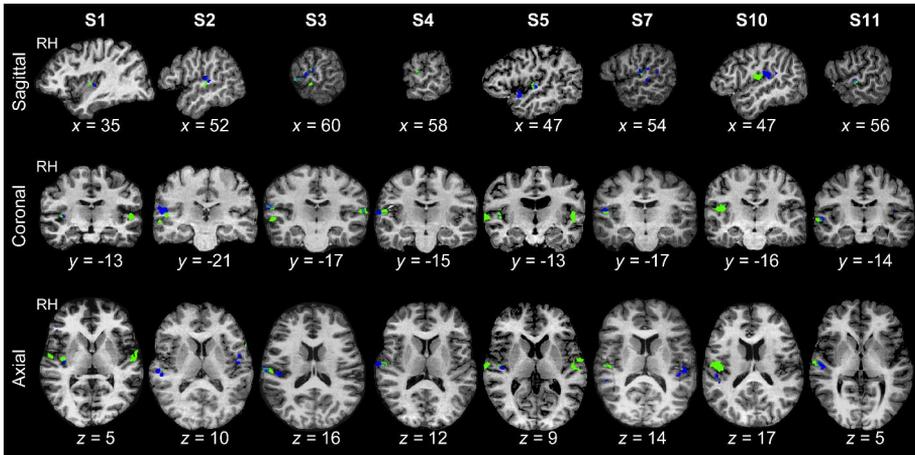


Figure S2. Parametric single-subject activation maps from eight representative listeners (columns) projected onto anatomical slices (rows) of the respective listener in neurological orientation. Activity in right core areas correlated with masking (blue regions) and perceived continuity (green regions) of only interrupted targets ($t_{385-516} = 1.0-2.7$, cluster size > 30). Slice positions are indexed by Talairach coordinates (x, y, z) in units of millimeters. RH, right hemisphere.

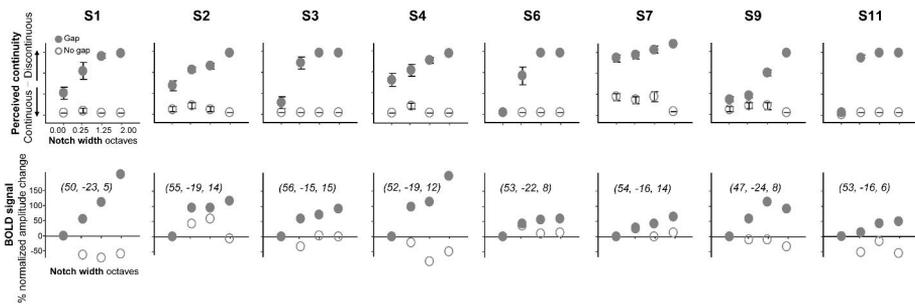


Figure S3. Behavioral and region of interest analysis results from eight representative listeners (columns), averaged across trial repetitions. Mean \pm SE of perceived continuity (upper row) and mean of BOLD signal change (lower row) show parametric trends depending on masking for interrupted (solid circles) but not for uninterrupted (open circles) targets. BOLD signals were obtained from peak activations in right auditory cortex indexed by Talairach coordinates (x, y, z) in units of millimeters (supplemental Fig. S1B, blue regions). To allow for comparison between interrupted and uninterrupted targets, mean BOLD signals were normalized to the respective no-notch conditions.

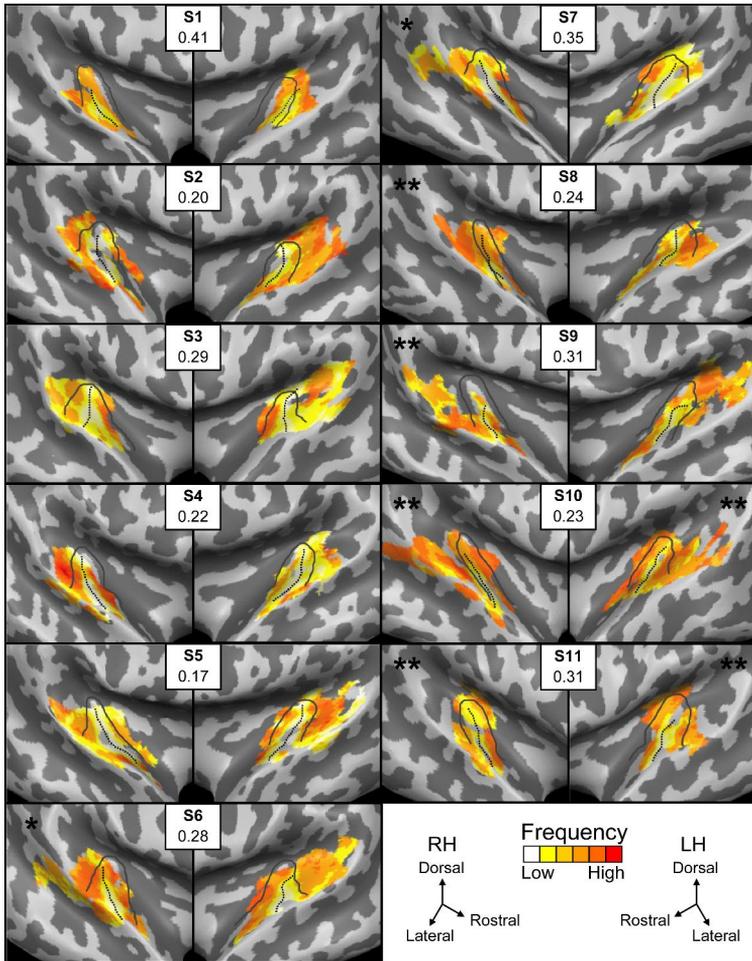


Figure S4. Best-frequency mapping single-subject results. Best-frequency maps from all listeners (S1–S11) projected onto the inflated cortical surface of the left and right hemisphere (LH, RH) of the respective listener, showing cortical clusters of different frequency selectivity (light colors, low frequencies; dark colors, high frequencies; decimal numbers, frequency-selectivity index FS). Solid lines delineate the locations of first Heschl's gyrus. Hemispheres comprising bifurcated or duplicated Heschl's gyri are indicated by single or double asterisks, respectively. Dashed lines approximate estimated low-frequency borders.

EEG-based chronometry of cortical processing of acoustic and illusory continuity

Abstract

Stable hearing in noisy environments is facilitated by restoration mechanisms in the auditory system. These mechanisms may illusorily fill interruptions in a sound when the gaps contain acoustic energy of a concurrent noise, yielding a continuity illusion of the interrupted sound. While the acoustic energy in filled sound gaps is represented already in peripheral neurons, the neural processes underlying the perception of illusory fillings are poorly understood. In this study, we used electroencephalography (EEG) to determine the timing of neural activations associated with continuity illusions. We parameterized the energetic filling of gaps in noise-interrupted tones, and asked listeners to rate their illusory fillings while scalp potentials were measured. Comparing the power spectra of EEG components induced by continuity illusions to those induced by true discontinuity percepts of the same interrupted stimuli revealed significant modulations in theta- and beta-oscillations, during and after the gap respectively, indicating effects of illusory fillings. Further comparisons of EEG component activities evoked by interrupted tones versus physically uninterrupted tones indicated effects of energetic fillings around 130 ms after gap onset and gap offset, respectively. We argue that top-down factors such as attention may modulate the salience of sensory gaps and thus determine the continuity illusion. The results set constraints on the timing of cortical processes assumed to be involved in the acoustical analysis and the perceptual interpretation of illusorily continuous sounds.

Based on: Riecke L, Esposito F, Bonte M, Formisano E. Sound restoration suppresses early auditory cortical processing of acoustic gaps. *Submitted*.

1 Introduction

The auditory continuity illusion (Miller & Licklider, 1950) is a striking illustration of the constructive nature of hearing. In this phenomenon an interrupted sound is illusorily heard as continuing through a louder noise. The continuity illusion has been observed for several sounds such as tones, sweeps, melodies, voices, or speech (for reviews, see Bregman, 1990; Warren, 1999) and in several species including birds, cats, monkeys, and humans (Sugita, 1997; Braaten & Leary, 1999; Miller *et al.*, 2001; Petkov *et al.*, 2003). These findings suggested that the illusory completion of the fragmented sound is based on a generic auditory mechanism which allows stable tracking of sounds in noisy environments.

The continuity illusion depends on the masking of the gap in the interrupted sound (target) (Houtgast, 1972; Warren *et al.*, 1972). Masking may occur when the interrupting noise (masker) contains acoustic energy in the frequency range of the target, thus rendering the gap inaudible. Masking may even occur when the masker spectrum is slightly segregated from the target band as peripheral auditory neurons may 'fill' the omitted energy in the critical band due to their band-pass filtering properties (Fletcher, 1940). In either case, the target and the masker may stimulate the same fibers in the auditory nerve (Delgutte, 1990) and evoke uninterrupted excitations that 'smear' the sensory evidence for the gap (Warren, 1970; Warren *et al.*, 1972; for review, see Warren, 1999), which we will refer to as 'energetic filling'.

Another requirement for the continuity illusion is that the fragment of the target that follows the gap continues the spectral trajectory of the fragment that preceded the gap (Ciocca & Bregman, 1987; see Fig. 1A). This implies that the illusion may lag behind the actual occurrence of the gap, suggesting 'retroactive' perceptual processes that differ fundamentally from the more instantaneous energetic processes. When the continuation of the target after the gap is given, the sensory energy that is filled in during the gap by the masker may be attributed to the surrounding fragments, and the target may be interpreted as illusorily continuous. This holistic representation may be derived from matching the sensory stimulus representation with perceptual schemas (Repp, 1992; for review, see Bregman, 1990) and may be accompanied by perceptual restoration of the target, which we will refer to as 'illusory filling'. The continuity illusion reflects a subjective perceptual state, which may fade out or fade in during long gaps that are filled with continuous energy (Wrightson & Warren, 1981; Warren *et al.*, 1994).

While the neural mechanisms underlying masking have been studied for decades, the neural basis of the continuity illusion has only recently received attention. Electrophysiological measurements in primary auditory cortex (AC) of monkeys have shown that excitations in neurons which respond to gaps in tones may be suppressed when these gaps are energetically filled by noise (Petkov *et al.*, 2007; for review, see Recanzone & Sutter, 2008). A subset of neurons which responded more to uninterrupted tones showed sustained excitations during the filled gaps, consistent with earlier observations in cats (Sugita, 1997). These findings demonstrate that neurons in monkey primary AC respond to interrupted sounds as if they were

continuous, provided that these interruptions are energetically filled. In humans, functional magnetic resonance imaging (fMRI) revealed that hemodynamic activities in primary AC scale with energetic fillings of gaps in tones (Riecke *et al.*, 2007). The activities were more widespread when the gaps were filled with less energy, suggesting neural processes related to the detection of sensory gaps. Another recent fMRI study in humans reported consistent results, showing enhanced activities in the primary AC to vowel sounds with spectral gaps, as opposed to energetically filled vowels (Heinrich *et al.*, 2008). Previous electroencephalography (EEG) results further indicated that energetic fillings of short tone gaps are indexed by the mismatch-negativity (MMN) (Micheyl *et al.*, 2003). The MMN is a short-latency event-related scalp potential (ERP) that may reflect auditory cortical processes related to the detection of deviances in predictable sound sequences (Näätänen *et al.*, 2001). While listeners did not attend to the interrupted sounds, MMN levels were found elevated within 200 ms after the onset of deviant energetic fillings.

A problem with these studies is that subjects' reports of continuity illusions during the physiologic measurements were either not obtained or not incorporated in the analyses, so a direct link between the continuity illusion and physiology could not be established. We addressed this point by investigating continuity illusions reported during fMRI, which showed that listeners' subjective experience of continuity illusions is paralleled by hemodynamic activities in primary AC (Riecke *et al.*, 2007).

A key question which has not been fully answered yet is when exactly the continuity illusion may arise. In the present study we investigate the timing of neural processes that may be related to the illusory filling and energetic filling of sound gaps, by combining behavioral measures of continuity illusions with EEG measures. Event-related EEG oscillations in the theta frequency range and beta frequency range have been proposed as neurophysiologic indicators for auditory attention (Cacace & McFarland, 2003; McFarland & Cacace, 2004), and the N1 ERP component has been associated with sound gap detection (Michalewski *et al.*, 2005; Lister *et al.*, 2007). Since the continuity illusion may rely on similar auditory processes (Bregman, 1990) these event-related EEG measures may provide useful insights into the timing of the neural mechanisms underlying continuity illusions. We presented listeners with interrupted and uninterrupted target tones which overlapped with a noise masker. We parameterized the strength of the energetic filling of the 600-ms gap by varying the spectral segregation between the target band and the noise. We also included ambiguous stimulus conditions individually designed to evoke bistable illusory fillings. Listeners rated the subjective strength of continuity illusions, while scalp potentials were measured simultaneously. Stimuli and task were fine-tuned based on a behavioral study (Riecke *et al.*, 2008) and similar to those used in our previous fMRI investigation. We assessed the effects of switches in continuity illusions and of changes in acoustic energy on event-related theta and beta oscillations, and on N1. We hypothesized that the perceptual and acoustical properties of the continuity illusion would modulate these EEG measures after the gap and during the gap, respectively.

2 Materials and methods

2.1 Participants

14 volunteers (age: 26 ± 4 years, mean \pm standard deviation [SD], mainly students from Maastricht University) who reported normal hearing and motor abilities, participated in the study after providing informed consent. All participants could hear the tone illusion as determined in a training session. Five participants also participated in the previous fMRI study. The local ethical committee approved the procedure.

2.2 Stimuli

Stimuli were assembled of tones (duration: 2800 ms, carrier frequency: 930 Hz) with sine-modulated amplitude (modulation rate: 3 Hz), and band-passed white noise (duration: 600 ms, bandwidth: 2 octaves) superimposed on the spectrotemporal center of the tones on a logarithmic-linear scale (Fig. 1A). For interrupted tones, the offsets and onsets of gaps and noise were synchronized. Uninterrupted tones were spectrotemporally matched, except that no gaps were inserted. The noise was band-stopped (notched) at the tone's frequency. To parameterize the spectral segregation of the tone and the noise, the width of this spectral notch was varied across three levels (Fig. 1B, left), including an individually determined intermediate level (see Design and Task). Spectral segregation was defined as half the notch width on a logarithmic scale. The effects of these acoustic manipulations on the sensory stimulus representation in the auditory nerve (Fig. 1B, right) were estimated based on a model by Patterson, Allerhand, and Giguere (1995). The different levels of the evoked energetic fillings were expected to induce illusory fillings of different strengths. Noise and stimuli were equated on their root mean square amplitudes, respectively. Onsets and offsets of tones and noise were linearly ramped with 3-ms rise-fall times. Stimuli were sampled at 44.1 kHz with 16 bit resolution using Matlab 7.0.1 (The MathWorks Inc., Natick, MA, USA). Stimuli were presented diotically at 70 dB sound pressure level (SPL) using Presentation 9.30 software (Neurobehavioral Systems, Inc., Albany, CA, USA), a Creative Sound Blaster Audigy 2ZS sound card (Creative Technology, Ltd., Singapore), and Control 25 speakers (JB professional, Rochester, NY, USA).

2.3 Design and task

Prior to the main EEG experiments, participants' ability to hear continuity illusions and to perform the task (see next paragraph) was assessed in 75 training trials. Individual continuity illusion thresholds were estimated based on the method of limits (Fechner, 1860): Multiple series of interrupted stimuli were presented in which the segregation between target and noise was either gradually decreased from 0.6 octaves, or increased from zero octaves, in average steps of 0.05 octaves. Thresholds were estimated as the average segregation at which listeners' ratings switched from non-illusory discontinuity to illusory continuity, or vice versa. The

illusion thresholds were implemented as intermediate levels of the segregation parameter (Fig. 1B) to define individual conditions that were expected to evoke bistable illusory fillings (see Results).

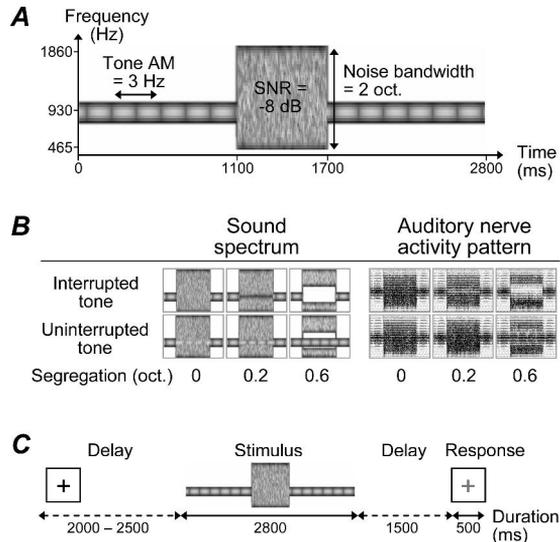


Figure 1. Stimulus and experimental design. **A**, Spectrogram and acoustic parameter settings for an amplitude-modulated target tone interrupted by a noise masker (noise duration: 600 ms). AM, amplitude modulation; SNR, signal-to-noise ratio; oct., octaves. **B**, Truncated stimulus spectrograms (left) and estimated auditory nerve excitation patterns (right) for all stimulus conditions. The energy in the gap was parameterized by varying the spectral segregation between the target and the noise, using a spectral notch of varying width. An individually-defined, intermediate masking condition (average target–masker segregation: 0.2 octaves) was used to evoke ambiguous fillings. **C**, Schematized trial of the auditory task. Listeners attended to targets and rated their continuity on a four-point scale after a delay. Stimulus and response intervals were indexed by visual cues and interleaved with silent intervals.

In the EEG experiments, participants were seated in a comfortable chair in an electrically shielded and sound-attenuated room and performed a forced-choice, delayed-response task. Trials comprised stimulation and response intervals which were indexed by a fixation cross changing in color (Fig. 1C). Response intervals were dissociated from stimulus intervals by silent intervals of 1.5 s so that motor responses did not interfere with auditory-evoked EEG signals. Participants were instructed on an LCD screen to attend to targets during the stimulus intervals, and to rate their continuity on a four-point scale (labeled with “most likely continuous”, “probably continuous”, “probably discontinuous”, “most likely discontinuous”). Ratings were performed by a button press with the index or middle finger of the left or right hand during the delayed response intervals. Participants were further instructed to fixate, and to delay eye blinks or other movements to the response intervals.

The six experimental conditions were presented in randomized order in blocks of 35-s duration, using an interstimulus interval (ISI) of 2–2.5 s that was jittered across trials in order to reduce expectations of listeners. In total 80 differently randomized

blocks were presented so that successive trials always differed. Subjects participated in the experiment five times and were allowed to take breaks in between. After the session, they estimated in a questionnaire how much time they needed for deciding for each of the different ratings.

2.4 EEG recordings and data preprocessing

2.4.1 Electrodes

EEG was measured with 61 equidistant Ag/AgCl scalp electrodes that were mounted to a standard electrode cap (Easycap GmbH, Herrsching-Breitbrunn, Germany) and referenced to the left mastoid. Inter-electrode impedances were kept below 5 k Ω by abrading the scalp. Electrooculography (EOG) was monitored using an electrode placed below the left eye.

2.4.2 Amplification, A/D conversion, and signal analysis

EEG was recorded with a Neuroscan SynAmps² system (Neuroscan Labs, El Paso, TX, USA), using an analog pass band of 0.05–70 Hz and a 50-Hz notch (filter slopes -12 dB/octave) to remove slow linear drifts and line noise. After digitization (sampling rate 250 Hz, resolution 0.168 μ V/bit), EEG data were processed in Matlab and EEGLAB (Delorme & Makeig, 2004). The data were high-pass filtered (cutoff: 1 Hz), and re-referenced to an average reference based on the mean activity of all channels. 4.5-s epochs locked to stimulus onsets were extracted, including a 1-s pre-stimulus interval.

2.4.3 Artifact removal

Data channels containing partially sustained artifacts, probably related to poor skin contacts, were excluded from further analysis (two channels per dataset on average). Epochs containing non-repetitive artifacts related to movements or single-channel noise were identified semi-automatically using a ± 75 - μ V criterion and visual inspection. Epochs during which listeners were not actively engaged in the task were identified based on lapsed button responses. Both types of trial were discarded. Epochs affected by blinks, eye movements, or other repetitive artifacts were not discarded as the underlying spatially fixed EOG processes could be separated from brain-related EEG processes using independent component analysis (ICA, see below).

2.4.4 Independent component analysis

Scalp potentials generally reflect a mixture of localized activities in synchronized neuronal ensembles and other non-cortical sources, conducted through the tissue. The contributions of functionally distinct sources can be partially isolated based on their temporal independence. To determine maximally temporally independent EEG and EOG processes, we applied an ICA algorithm to the EEG time series (runica,

see Bell & Sejnowski, 1995; Makeig *et al.*, 1996; Makeig *et al.*, 1997; Lee *et al.*, 2000). The algorithm computes a 'un-mixing' matrix that decomposes multi-channel data into a linear sum of spatially fixed components with maximal joint entropy (independent components [ICs]). ICA was applied to the individual artifact-cleaned datasets comprising on average 386 ± 26 trials (about 750 samples per trial per channel). After decomposition, ICs primarily accounting for blinks or eye movements were identified based on their frontal scalp distributions and their irregular occurrences across trials. Remaining artifacts were identified based on their non-dipolar scalp maps, flat activity spectra, and irregular occurrences (Makeig *et al.*, 1996; Jung *et al.*, 2000a, 2000b; Delorme *et al.*, 2007). Removal of these ICs resulted in individually pruned datasets comprising twelve ICs per dataset on average.

2.4.5 Computation of EEG measures

EEG power spectra and ERPs were computed for each IC in the individual datasets. Component power spectra were computed from a wavelet analysis by sliding an 1100-ms window across the single-trial activity time series. The number of cycles per Morlet wavelet was linearly increased from 3 to 20 across the investigated frequency range (3–40 Hz). Mean event-related spectral power changes were estimated relative to the mean baseline spectrum on a logarithmic scale for each frequency band, using time-frequency bins of 17-ms duration and 0.1-Hz bandwidth. Statistical analyses were based on averaging the power estimates across all bins within selected time-frequency windows. These windows were defined from the most prominent peaks in the average component cluster spectrum (see below) and were fixed across conditions and subjects.

Component ERPs were determined by averaging the single-trial source activity time series. N1 latencies were defined from the most negative peak in the average component cluster waveform in a 90–180-ms interval after the evoking event. N1 amplitudes were defined for each stimulus condition by averaging across a 50-ms window centered on the average peak. The average waveforms were low-passed filtered (cutoff: 20 Hz) for display purposes only.

2.4.6 Component clustering

Functional-anatomically similar EEG processes were identified based on similarities in their average component power spectra, waveforms, and scalp topographies (Jung *et al.*, 2001; Makeig *et al.*, 2002, 2004; Contreras-Vidal & Kerick, 2004; Onton *et al.*, 2005, 2006). To reduce their number of features, spectral power and ERP measures were decomposed into five principal components using principal component analysis (PCA). Similarity among individual ICs was then estimated from the 'distance' of their principal components in the Euclidean space that was defined by the joint feature-reduced measures (*i.e.* a 10-dimensional source space). ICs from all subjects were partitioned into three main IC clusters using an algorithm that minimizes the summed squared distance of the principal components from the estimated mean IC cluster (*k*-

means, Matlab Signal Processing Toolbox). The clustering was further refined by visual inspection to improve uniformity in scalp topography among IC within each cluster, resulting in 63 ICs per cluster on average. The polarity of each IC was adjusted so that its scalp topography correlated positively with the respective cluster topography.

The resulting clusters showed clear differences in their average scalp topography, indicating source activities projected to anterior, central, and posterior scalp regions, respectively. In the present study, we focused on the cluster associated with a clear central topography (supplemental Fig. S1) because the component power spectra and ERPs of this cluster were most strongly time-locked to the auditory task and most consistent across subjects.

2.5 Statistical analysis

Listeners' continuity rating data were averaged across the 80 trials presented in each condition. Statistical analysis of behavioral data involved paired t-tests and a two-way analysis of variance (ANOVA) for repeated measures, including segregation and gap as fixed factors, and listeners as a random factor ($N = 14$).

EEG data were analyzed using two complementary procedures. First, the effects of energetic fillings were assessed by analyzing the effects of the acoustic gap (differences between interrupted and uninterrupted targets) and their modulations by the segregation parameter. In this gap-based analysis ($N = 14$) the same statistical methods were applied as for the behavioral data (see above).

Second, the effects of illusory fillings were determined independently of energetic fillings, based on listeners' continuity ratings of the same interrupted targets. This illusion-based analysis involved data from trials in which listeners had rated the same stimulus as 'continuous' or 'discontinuous'. Trials were first randomly rejected from the rating condition that occurred more frequently so that both rating conditions involved the same number of trials, separately for each segregation condition and for each listener. Significance of illusory filling effects was assessed with paired t-tests, using data from listeners who switched their illusory ratings most frequently (zero-octave condition: $N = 11$; threshold condition: $N = 13$). Main effects were tested using a mixed-model 2-way ANOVA ($N = 10$) as described above, except that segregation and ratings were treated as fixed factors (each including two levels).

All EEG analyses were applied to listeners' mean component cluster data. All average power estimates forwarded to parametric statistical testing were well-fitted with normal distributions as assessed by Lilliefors tests (for details see Kiebel *et al.*, 2005).

3 Results

3.1 Behavioral results

Analysis of continuity rating data revealed that stronger energetic fillings (induced by smaller spectral segregations) evoked stronger illusory fillings of the interrupted

target ($F_{2, 26} = 165.6, p < 10^{-15}$; Fig. 2, black circles). The opposite effect was found for the uninterrupted target: True continuities became obscured when the target was masked ($F_{2, 26} = 16.4, p < 10^{-5}$; gray circles). Strong energetic fillings thus markedly reduced listeners' ability to hear the difference between illusory and non-illusory targets during noise interruptions, whereas weak energetic fillings revealed the gap (noise \times gap interaction, $F_{2, 12} = 383.9, p < 10^{-11}$). For the individual illusion thresholds (average estimated threshold: 0.2 ± 0.1 octaves) and also for the zero-octave condition, ratings of the interrupted target frequently switched between illusory continuity and true discontinuity as expected, indicating bistable illusory fillings.

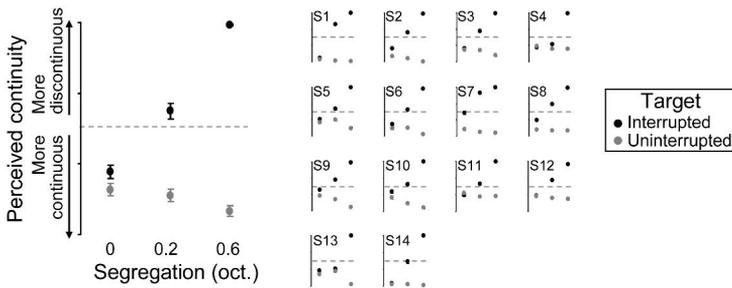


Figure 2. Behavioral results. Average continuity ratings (\pm SE) of all stimuli obtained during EEG, averaged across listeners (left) and for all individual listeners (right). Spectral segregation of the target and the noise disrupted continuity illusions of interrupted targets (black) and disclosed true continuities of uninterrupted targets (gray). For two ambiguous segregation conditions (0 and 0.2 octaves), interrupted targets evoked bistable continuity percepts (dashed lines), facilitating EEG analysis of percept switches. oct., octaves.

3.2 Effects of energetic fillings on EEG power spectra

Analysis of average component power spectra (Fig. 3A) revealed robust modulations in EEG oscillations in all stimulus conditions: Theta-band (3–7 Hz) power largely increased during the noise interval (Fig. 3A, small rectangles; all $t_{13} = 3.6\text{--}7.3, p = 0.004\text{--}10^{-6}$) and beta-band (13–27 Hz) power mainly decreased during the post-noise interval (large rectangles; all $t_{13} = -2.4\text{--}-3.6, p = 0.04\text{--}0.004$), relative to the average power in the 500-ms pre-noise interval (baseline). These modulations indicated general task-induced synchronization and desynchronization, respectively, which were subsequently analyzed for condition effects.

The following results are summarized in Figure 3 and Figure 6A: Statistical comparisons of interrupted and uninterrupted targets revealed that theta-band power increased significantly more during the noise when the target contained a gap ($p < 0.05$; Fig. 3B, red marks). Averaged across the noise interval, this effect was most significant when the target band was largely segregated from the noise spectrum (0.6-octave condition). Theta-band power was also modulated by the spectral segregation between the target and the noise. This effect was significant for interrupted targets, and a similar trend was observed for uninterrupted targets (Fig. 3C, left). The modulations indicated an effect of the strength of the energetic

filling of the gap, and more generally of the spectral segregation between the target and the noise, on the observed theta-band synchronization.

Beta-band power was found significantly lower when the target contained a gap, compared to targets without gap, especially in the post- noise interval ($p < 0.05$; Fig. 3B, red marks). Averaged across this interval, a main effect of the gap was evident, and no effect of spectral segregation was observed (Fig. 3C, right). Therefore, the gap in the target enhanced the observed beta-band desynchronization, irrespective of the strength of energetic filling.

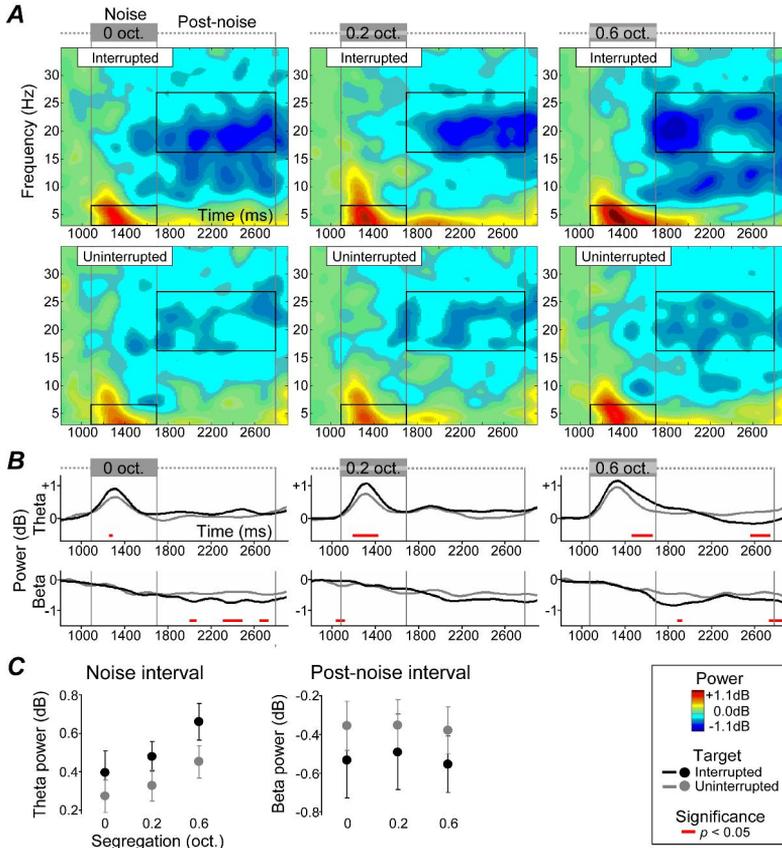


Figure 3. Results from gap-based analysis of average component power spectra. **A**, Power spectra induced by the interrupted target (upper row) and the uninterrupted target (lower row) for three levels of spectral segregation, relative to the pre-noise interval. The gap was accompanied by increased theta-band synchronization (red) during the noise (small rectangles), and by increased beta-band desynchronization (blue) during and after the noise (large rectangles). oct., spectral segregation in octaves; vertical lines, noise onsets and offsets. **B**, Spectral power time series for interrupted targets (black) and uninterrupted targets (gray), illustrating the timing of these effects (red marks) for each level of spectral segregation. **C**, Average theta-band power (left) and beta-band power (right) in the noise interval and in the post-noise interval, respectively. The gap enhanced theta-band synchronization especially when the gap was not energetically filled (0.6-octave condition). The gap further enhanced beta-band desynchronization after the noise. These results indicate sensory influences of the acoustic gap on ongoing EEG oscillations in theta and beta bands.

3.3 Effects of illusory fillings on EEG power spectra

The results so far show that oscillatory EEG processes were modulated by the strength of the energetic filling of the gap. These modulations also reflected influences of illusory fillings since these correlated highly with energetic fillings (see behavioral results; Fig. 2). To disentangle perceptual and acoustic influences, we further investigated the effects of illusory fillings independently of those of energetic fillings.

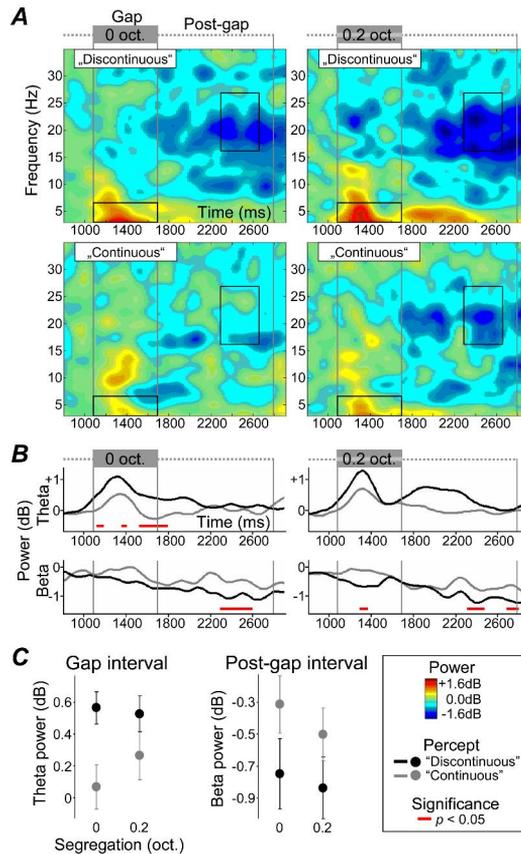


Figure 4. Results from illusion-based analysis of average component power spectra. **A**, Power spectra induced by the true discontinuity percept (upper row) and the continuity illusion (lower row) of the same interrupted target for two levels of energetic filling (columns), relative to the pre-gap interval. Gap perception was accompanied by increased theta-band synchronization (red) during the gap (small rectangles) and by increased beta-band desynchronization (blue) after the gap (large rectangles), respectively. **B**, Spectral power time series for true discontinuity percepts (black) and continuity illusions (gray), illustrating the timing of these percept effects (red marks) for each level of energetic filling. **C**, Average theta-band power (left) and beta-band power (right) in the gap interval and in the post-gap interval, respectively. Gap perception enhanced theta-band synchronization and beta-band desynchronization. These perceptual effects were independent of the energy in the gap, thus indicating non-sensory influences on theta and beta oscillations.

Results from this illusion-based analysis are summarized in Figure 4 and Figure 6B: Statistical analysis of average component power spectra (Fig. 4A) revealed that theta-band power was significantly more increased when listeners reported hearing the gap, compared to continuity illusions ($p < 0.05$; Fig. 4B, red marks). Averaged across the gap interval (Fig. 4A, small rectangles), this effect of illusory filling was highly significant, especially when the gap was fully filled with acoustic energy (Fig. 4C, left). Beta-band power decreased significantly more when listeners reported hearing the gap, especially around 600–900 ms post-gap ($p < 0.05$; Fig. 4B, red marks). Averaged across this interval (Fig. 4A, large rectangles), a strong effect of illusory fillings was evident for both levels of energetic filling (Fig. 4C, right). Therefore, the perception of the gap in the target enhanced theta-band synchronization during the gap, and also beta-band desynchronization after the gap. These non-sensory modulations were highly consistent with the modulations induced by acoustic gaps (Fig. 3A).

3.4 Effects of energetic fillings on event-related potentials

Inspection of average component waveforms revealed prominent deflections locked to the noise followed by more sustained deflections after the noise in all stimulus conditions (Fig. 5A). Statistical comparisons of interrupted and uninterrupted targets indicated transient effects of the gap during and after the noise (paired t-tests, $p < 0.01$; Fig. 5A, red marks). The earliest and most significant effects were observed in a large negative peak around 100 ms after noise onset when the target band was largely segregated from the noise spectrum (0.6-octave condition: $p < 0.004$). The average latency of this ERP (133 ± 3 ms across listeners) was compatible with that of N1, a potential that has been associated with gap detection.

Statistical analyses showed that the amplitude of N1 to noise onsets ($N1_{on}$) was enhanced when targets contained a gap and also when target and noise spectra were largely segregated (Fig. 5B left; see Fig. 6C for details). These modulations were highly similar to those observed in the theta band (Fig. 3C, left), indicating effects of the strength of the energetic filling of the gap, and more generally of the spectral segregation between the target and the noise.

Despite its small amplitude, the N1 to noise offsets ($N1_{off}$, average peak latency: 123 ± 4 ms) differed significantly from baseline (all $t_{13} = -3.0$ – -8.0 , $p = 0.009$ – 10^{-6}), except for one condition (0.6-octave condition, uninterrupted targets). Statistical analysis yielded a slightly different pattern than that observed for $N1_{on}$ (Fig. 5B right; see Fig. 6C for details): Effects of spectral segregation were smaller for interrupted targets, and reversed trends were observed for interrupted targets. $N1_{off}$ was further modulated by the onset of the interrupted target after the gap (*i.e.* by the target re-onset) especially when the gap was largely segregated from the noise, as revealed by comparing interrupted and uninterrupted targets. Therefore, $N1_{off}$ was not generally affected by the spectral segregation of the target and the noise, but mainly by the strength of the energetic filling. Compared to $N1_{on}$, this pattern more closely resembled listeners' continuity ratings ($N1_{on}$: Pearson's $R = -0.2$, $p < 0.03$; $N1_{off}$: Pearson's $R = -0.4$, $p < 0.002$; see behavioral results, Fig. 2).

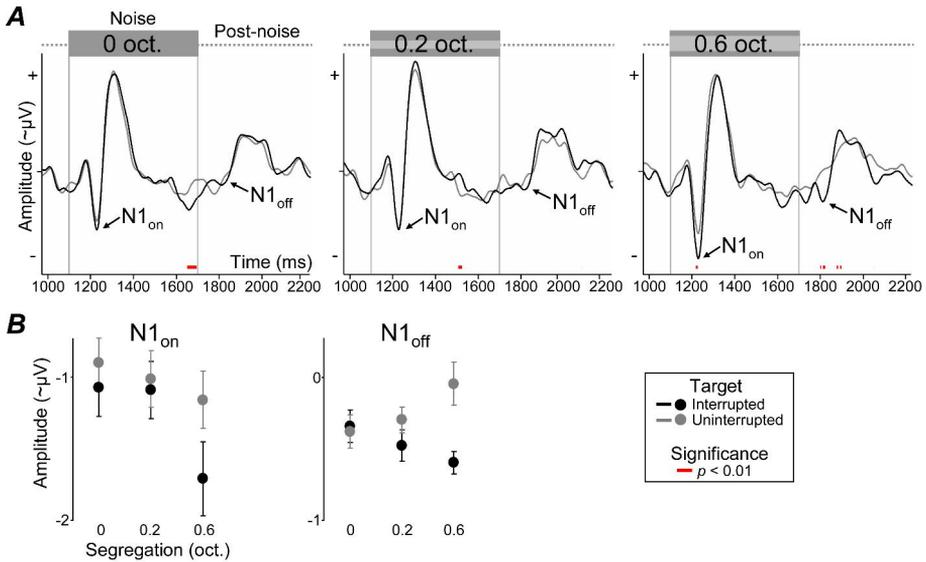


Figure 5. Results from gap-based analysis of average component waveforms. **A**, ERPs evoked by the interrupted target (black) and the uninterrupted target (gray), during and after the noise for three levels of spectral segregation. N1 was evoked by the onset and offset of the noise (see $N1_{on}$ and $N1_{off}$ labels). The earliest effects of the gap (red marks) appeared when the gap was not energetically filled (0.6-octave condition). oct., spectral segregation in octaves; vertical lines, noise onsets and offsets; all waveforms plotted on the same scale proportional to μV . **B**, Average amplitude (\pm SE) of $N1_{on}$ (left) and $N1_{off}$ (right) for all stimulus conditions. $N1_{on}$ amplitude was significantly modulated by the gap and by the spectral segregation of the target and the noise, consistent with the observed modulations in the theta band (Fig. 3C, left). $N1_{off}$ showed a slightly different pattern which correlated more with listeners' perceptual ratings (Fig. 2).

4 Discussion

A robust pattern of neural oscillations was observed in all conditions: noisy interruptions in targets induced event-related synchronization in the theta rhythm during the noise, followed by sustained desynchronization in the beta rhythm after the noise. Strikingly, it was found that this pattern is further modulated by non-sensory processes related to the continuity illusion. Similar modulations were also induced by sensory processes related to the sound energy in the interruptions. Finally, it was observed that the sound energy in the interruptions also affects ERP components 130 ms after the onset and offset of these interruptions. This pattern of results suggests that top-down factors such as auditory attention modulated the salience of sensory gaps and, therefore, determined the continuity illusion.

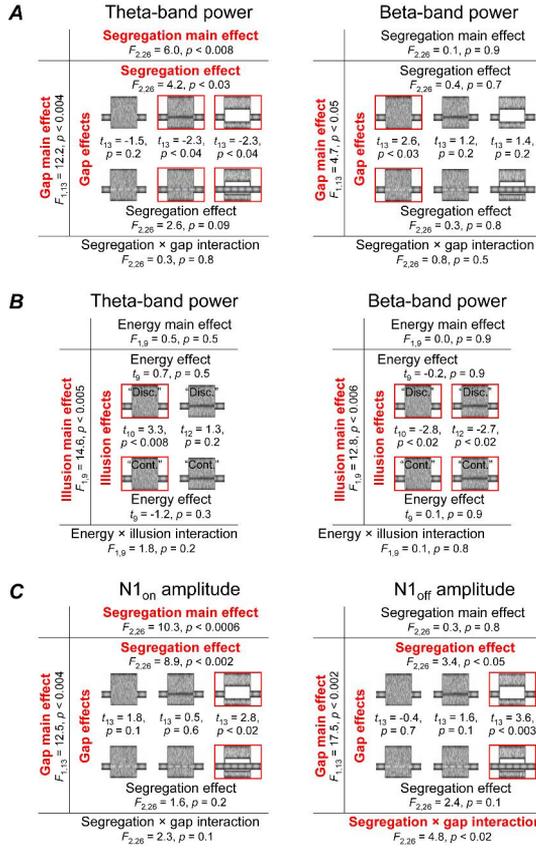


Figure 6. Summary of statistical results. **A**, Results from gap-based analysis of theta-band power during the noise (left) and beta-band power after the noise (right). **B**, Results from illusion-based analysis of theta-band power during the gap (left) and beta-band power during the post-gap interval (right). **C**, Results from gap-based analysis of N1 amplitude to noise onset (left) and noise offset (right). Truncated spectrograms represent stimulus conditions. “Disc.” and “Cont.” labels indicate associated discontinuity and continuity percepts (B). Types of effect are labeled, and statistically significant effects are highlighted ($p < 0.05$).

4.1 Theta synchronization as an index for the salience of sensory edges

In the literature, theta-band synchronization has been associated with the encoding of new sensory input and with increased task demands (for review, see Klimesch, 1999). Theta synchronization may occur during passive listening (Kolev *et al.*, 2001) and can be enhanced by non-sensory factors, for example, when listeners are attentive and perform an active auditory discrimination task (Cacace & McFarland, 2003; McFarland & Cacace, 2004; Molnar *et al.* 2008). Given these previous findings, the theta synchronization that we observed during the noise may reflect neural processes related to 1) the sensory encoding of the noise and 2) the increased listening demand associated with the noise.

Our results that this synchronization was further increased when the target contained a gap and when the target was spectrally segregated from the noise, indicate that sensory processes related to the salience of the spectrotemporal gap, or to the detection of acoustic edges (Nakajima *et al.*, 2000; Fishbach *et al.*, 2001; Herdener *et al.*, 2007), were involved. Crucially, we found that this synchronization was also increased when listeners actually heard the gap, compared to the continuity illusion. This non-sensory effect indicates that top-down factors such as attention may have modulated either the sensory processing of acoustic edges or, more generally, the perceptual salience of these edges during the noise. This interpretation is partially supported by previous electrophysiology findings, which demonstrated that spontaneous theta oscillations influence baseline excitability and auditory-evoked responses of primary auditory cortical neurons in monkeys (Lakatos *et al.*, 2005). Additionally, our interpretation is consistent with the general notion that modulations in oscillations may reflect top-down control related to the perceptual interpretation of sensory features (for review, see Engel *et al.*, 2001).

4.2 Beta desynchronization as an index for the continuity illusion

Beta-band desynchronization has been observed in active auditory discrimination tasks (Cacace & McFarland, 2003; McFarland & Cacace, 2004) and may index neural processes related to response selection and motor preparation (Pfurtscheller & Lopes da Silva, 1999; Kaiser *et al.*, 2001, 2007; Alegre *et al.*, 2003a, 2003b, 2004, 2006; Maekinen *et al.*, 2004). Furthermore, beta desynchronization may be larger and occur earlier when listeners make easy auditory discriminations, compared to difficult discriminations (Kaiser *et al.*, 2007).

Given these previous findings, the beta desynchronization after the noise that we observed can be interpreted as response selection processes related to the continuity rating task. We found that this desynchronization was increased when listeners heard the gap during the noise, which may indicate that ratings of discontinuity percepts were easier than ratings of continuity percepts. This was most apparent for targets comprising fully unmasked gaps: These stimuli induced the largest beta desynchronization, and listeners reported them in a questionnaire as most easy to judge. Selecting an appropriate behavioral response required our listeners to mentally match their sensory input with individually-defined perceptual criteria. Such schema-based matching processes may generally occur after the continuation of the post-gap target has been evaluated (see Introduction). We observed small percept-related modulations in the beta band already during the gaps, which may be related to the fact that targets were known to continue after the gap and thus required no evaluation. However, these effects became stable only around 600 ms post-gap. It can be concluded that by that time, listeners had selected a response and, therefore, had illusorily filled the gap.

4.3 The auditory-evoked N1 as an index for acoustic edge detection

Our result that N1 was evoked by the onsets and offsets of unmasked gaps is consistent with gap detection studies that used shorter gaps without coinciding noise (Joutsiniemi *et al.*, 1989; Pantev *et al.*, 1996; Rupp *et al.*, 2002; Heinrich *et al.*, 2004; Michalewski *et al.*, 2005; Pratt *et al.*, 2005; Lister *et al.*, 2007). We observed that under masking conditions, the N1 evoked by gaps was substantially reduced and did not differ significantly from the N1 associated with truly uninterrupted targets. This result reveals that N1 is sensitive to the sensory energy in the target band, irrespective of which sound (target or noise) induced this energy. The result that more spectrally segregated sounds evoke larger N1 components is in line with findings from previous studies on the continuity illusion (Micheyl *et al.*, 2003), auditory streaming (Gutschalk *et al.*, 2005; Snyder *et al.*, 2006), or masked-speech intelligibility (Martin *et al.*, 1997, 1999, 2005). We observed that the spectrotemporal gap exhibited similar effects on N1 and on the concurrent theta synchronization, consistent with previous findings that the two measures may be coupled (Fuentemilla *et al.*, 2006; Grau *et al.*, 2007). This similarity corroborates our previous interpretation that the increased theta synchronization may reflect the detection of acoustic edges.

We also noticed that the effects of spectral segregation on N1 were reversed for truly uninterrupted targets at noise offsets. While the N1 to noise onset reflected more the spectral relation between the target and the noise, we found that N1 to noise offset correlated more with listeners' perception of the target band. It is thus conceivable that listeners may have perceived the gap in the target after its sensory segregation from the noise. This idea is also supported by results from an MMN study which showed that the integration of short sequential tones within a stream may occur after these sounds have been segregated from other interleaved tones (Sussman, 2005).

4.4 Potential auditory cortical EEG sources

The EEG activities that we investigated displayed mainly central-medial scalp topographies. Even though we did not model the underlying EEG sources, there are two indications that auditory cortical regions may have been involved: First, theta synchronization has been observed at frontal scalp locations (for review see Klimesch, 1999) and also above temporal regions (Cacace & McFarland, 2003; Grau *et al.*, 2007). Furthermore, N1 has been proposed to originate in the AC (Nääätänen & Picton, 1987; Woods, 1995). Given these previous findings, the strongly time-locked theta synchronization and N1 that we observed possibly reflected auditory cortical processes. The observed beta desynchronization, however, more likely indicated activity in sensory-motor cortical sources (for review see Pfurtscheller & Lopes da Silva, 1999). Second, the observed effects were similar to the modulations in AC that we had previously observed using fMRI (Riecke *et al.*, 2007). The EEG activities observed here may be localized in similar regions, since we used the same stimuli and task, and partially the same participants, as in our previous study. The central-medial scalp topographies of these activities likely reflected synchronous bilateral

auditory cortical processes which were merged into a single central component due to their temporal dependencies (Makeig *et al.*, 1997).

4.5 Conclusions

We found that EEG oscillations that are thought to index auditory attention are modulated by acoustic and perceptual factors that may determine the salience of acoustic edges during continuity illusions. We observed similar acoustic effects for concurrent event-related potentials that are known to index gap detection. Finally, we noticed that oscillations associated with decision-making are modulated mainly by perceptual factors after the sensory gaps have been processed.

In summary, our findings set temporal constraints on the acoustic and perceptual analysis of continuity illusions: Sensory processes related to the encoding and detection of acoustic edges may emerge around 130 ms after gap onset and may further evolve during the gap. Perceptual processes related to the formation of the continuity illusion may be active during and after the gap, and may interact with sensory processes involved in the encoding of acoustic edges. These processes could coordinate the grouping of the target fragments into illusorily unified objects which may be fully constructed by about 600 ms after the gaps.

Acknowledgements

We thank Hanna Renvall and Alexander Gutschalk for useful comments on the study, Anke Walter for help with the data acquisition, Fabrizio Esposito for advice on the analysis, and Claudia Schreiner for constructive comments on the manuscript.

References

- Alegre M, Gurtubay IG, Labarga A, Iriarte J, Malanda A, Artieda J. (2003a). Alpha and beta oscillatory changes during stimulus-induced movement paradigms: effect of stimulus predictability. *Neuroreport*, 14, 381–385.
- Alegre M, Labarga A, Gurtubay IG, Iriarte J, Malanda A, Artieda J. (2003b). Movement-related changes in cortical oscillatory activity in ballistic, sustained and negative movements. *Exp Brain Res*, 148, 17–25.
- Alegre M, Gurtubay IG, Labarga A, Iriarte J, Valencia M, Artieda J. (2004). Frontal and central oscillatory changes related to different aspects of the motor process: a study in go/no-go paradigms. *Exp Brain Res*, 159, 14–22.
- Alegre M, Imirizaldu L, Valencia M, Iriarte J, Arcocha J, Artieda J. (2006). Alpha and beta changes in cortical oscillatory activity in a go/no go randomly-delayed-response choice reaction time paradigm. *Clin Neurophysiol*, 117, 16–25.
- Bell AJ, Sejnowski TJ. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput*, 7, 1129–1159.
- Braaten RF, Leary JC. (1999). Temporal induction of missing birdsong segments in European starlings. *Psychol Sci*, 10, 162–166.
- Bregman AS. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Cacace AT, McFarland DJ. (2003). Spectral dynamics of electroencephalographic activity during auditory information processing. *Hear Res*, 176, 25–41.
- Ciocca V, Bregman AS. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Percept Psychophys*, 42, 476–484.
- Contreras-Vidal JL, Kerick SE. (2004). Independent component analysis of dynamic brain responses during visuomotor adaptation. *Neuroimage*, 21, 936–945.
- Delgutte B. (1990). Physiological mechanisms of psychophysical masking: observations from auditory-nerve fibers. *J Acoust Soc Am*, 87, 791–809.
- Delorme A, Makeig S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, 134, 9–21.
- Delorme A, Sejnowski T, Makeig S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, 34, 1443–1449.
- Engel AK, Singer W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci*, 5, 16–25.
- Fechner G. (1860). *Elements of Psychophysics*. New York: Holt, Rinehart, and Winston.
- Fishbach A, Nelken I, Yeshurun Y. (2001). Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. *J Neurophysiol*, 85, 2303–2323.
- Fletcher H. (1940). Auditory patterns. *Rev Modern Physics*, 12, 47–65.
- Fuentemilla L, Marco-Pallarés J, Grau C. (2006). Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *Neuroimage*, 30, 909–916.
- Grau C, Fuentemilla L, Marco-Pallarés J. (2007). Functional neural dynamics underlying auditory event-related N1 and N1 suppression response. *Neuroimage*, 36, 522–531.
- Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci*, 25, 5382–5388.
- Heinrich A, Alain C, Schneider B. (2004). Within- and between-channel gap detection in the human auditory cortex. *Neuroreport*, 15, 2051–2056.
- Heinrich A, Carlyon RP, Davis MH, Johnsrude IS. (2008). Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. *J Cogn Neurosci*, 20, 1737–1752.
- Herdener M, Esposito F, Di Salle F, Lehmann C, Bach DR, Scheffler K, Seifritz E. (2007). BOLD correlates of edge detection in human auditory cortex. *Neuroimage*, 36, 194–201.
- Houtgast T. (1972). Psychophysical evidence for lateral inhibition in hearing. *J Acoust Soc Am*, 51, 1885–1894.

- Joutsiniemi SL, Hari R, Vilkmann V. (1989). Cerebral magnetic responses to noise bursts and pauses of different durations. *Audiology*, 28, 325–333.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. (2000a). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol*, 111, 1745–1758.
- Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ. (2000b). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37, 163–178.
- Jung TP, Makeig S, Westerfield W, Townsend J, Courchesne E, Sejnowski TJ. (2001). Analysis and visualization of single-trial event-related potentials. *Hum Brain Mapp*, 14, 166–185.
- Kaiser J, Birbaumer N, Lutzenberger W. (2001). Event-related beta desynchronization indicates timing of response selection in a delayed-response paradigm in humans. *Neurosci Lett*, 312, 149–152.
- Kaiser J, Lennert T, Lutzenberger W. (2007). Dynamics of oscillatory activity during auditory decision making. *Cereb Cortex*, 17, 2258–2267.
- Kiebel SJ, Tallon-Baudry C, Friston KJ. (2005). Parametric analysis of oscillatory activity as measured with EEG/MEG. *Hum Brain Mapp*, 26, 170–177.
- Klimesch W. (1999). EEG alpha and theta rhythmicities reflect cognitive and memory performance: a review and analysis. *Brain Res Rev*, 29, 169–195.
- Kolev V, Rosso OA, Yordanova J. (2001). A transient dominance of theta ERP component characterizes passive auditory processing: evidence from a developmental study. *Neuroreport*, 13, 2791–2796.
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol*, 94, 1904–1911.
- Lee TW, Girolami M, Bell AJ, Sejnowski TJ. (2000). A unifying information-theoretic framework for independent component analysis. *Comput Math Appl*, 39, 1–21.
- Lister JJ, Maxfield ND, Pitt GJ. (2007). Cortical evoked response to gaps in noise: within-channel and across-channel conditions. *Ear Hear*, 28, 862–878.
- Mäkinen VT, May PJ, Tiitinen H. (2004). Human auditory event-related processes in the time-frequency plane. *Neuroreport*, 15, 1767–1771.
- Makeig S, Bell AJ, Jung TP, Sejnowski TJ. (1996). Independent component analysis of electroencephalographic data. *Adv Neural Inf Process Syst*, 8, 145–151.
- Makeig S, Jung TP, Bell AJ, Ghahremani D, Sejnowski TJ. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc Natl Acad Sci USA*, 94, 10979–10984.
- Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ. (2002). Dynamic brain sources of visual evoked responses. *Science*, 295, 690–694.
- Makeig S, Delorme A, Westerfield M, Jung TP, Townsend J, Courchesne E, Sejnowski TJ. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol*, 2, e176.
- Martin BA, Sigal A, Kurtzberg D, Stapells DR. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds/ba/and/da. *J Acoust Soc Am*, 101, 1585–1599.
- Martin BA, Kurtzberg D, Stapells DR. (1999). The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/and/da. *J Speech Lang Hear Res*, 42, 271–286.
- Martin BA, Stapells DR. (2005). Effects of low-pass noise masking on auditory event-related potentials to speech. *Ear Hear*, 26, 195–213.
- McFarland DJ, Cacace AT. (2004). Separating stimulus-locked and unlocked components of the auditory event-related potential. *Hear Res*, 193, 111–120.
- Michalewski HJ, Starr A, Nguyen TT, Kong Y-Y, Zeng F-G. (2005). Auditory temporal processes in normal-hearing individuals and in patients with auditory neuropathy. *Clin Neurophysiol*, 116, 669–680.
- Micheyl C, Carlyon, RP, Shtyrov Y, Hauk O, Dodson T, Pullvermüller F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *J Cogn Neurosci*, 15, 747–758.
- Miller GA, Licklider JCR. (1950). The intelligibility of interrupted speech. *J Acoust Soc Am*, 22, 167–173.

- Miller CT, Dibble E, Hauser MD. (2001). Amodal completion of acoustic signals by a nonhuman primate. *Nat Neurosci*, 4, 783–784.
- Molnár M, Csuhaj R, Gaál ZA, Czigler B, Ulbert I, Boha R, Kondákor I. (2008). Spectral characteristics and linear-nonlinear synchronization changes of different EEG frequency bands during the CNV. *Psychophysiology*, 45, 412–419.
- Näätänen R, Picton TW. (1987). The NI wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24, 375–425.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends Neurosci*, 24, 283–288.
- Nakajima Y, Sasaki T, Kanafuka K, Miyamoto A, Remijn G, ten Hoopen G. (2000). Illusory recouplings of onsets and terminations of glide tone components. *Percept Psychophys*, 62, 1413–1425.
- Onton J, Delorme A, Makeig S. (2005). Frontal midline EEG dynamics during working memory. *Neuroimage*, 27, 341–356.
- Onton J, Westerfield M, Townsend J, Makeig S. (2006). Imaging human EEG dynamics using independent component analysis. *Neurosci Biobehav Rev*, 30, 808–822.
- Pantev C, Eulitz C, Hampson S, Ross B, Roberts LE. (1996). The auditory evoked "off" response: sources and comparison with the "on" and the "sustained" responses. *Ear Hear*, 17, 255–265.
- Patterson RD, Allerhand MH, Giguere C. (1995). Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform. *J Acoust Soc Am*, 98, 1890–1894.
- Petkov CI, O'Connor KN, Sutter ML. (2003). Illusory sound perception in macaque monkeys. *J Neurosci*, 23, 9155–9161.
- Petkov CI, O'Connor KN, Sutter ML. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54, 153–165.
- Pfurtscheller G, Lopes da Silva FH. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol*, 110, 1842–1857.
- Pratt H, Bleich N, Mittelman N. (2005). The composite N1 component to gaps in noise. *Clin Neurophysiol*, 116, 2648–2663.
- Recanzone GH, Sutter ML. (2008). The biological basis of audition. *Annu Rev Psychol*, 59, 119–142.
- Repp BH. (1992). Perceptual restoration of a 'missing' speech sound: auditory induction or illusion? *Percept Psychophys*, 51, 14–32.
- Riecke L, Van Opstal J, Goebel R, Formisano E. (2007). Hearing Illusory Sounds in Noise: Sensory-perceptual Transformations in Primary Auditory Cortex. *J Neurosci*, 27, 12684–12689.
- Riecke L, Van Opstal AJ, Formisano E. (2008). The auditory continuity illusion: A parametric investigation and filter model. *Percept Psychophys*, 70, 1–12.
- Rupp A, Gutschalk A, Hack S, Scherg M. (2002). Temporal resolution of the human primary auditory cortex in gap detection. *Neuroreport*, 13, 2203–2207.
- Snyder JS, Alain C, Picton TW (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci*, 18, 1–13.
- Sugita Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *Neuroreport*, 8, 1155–1159.
- Sussman ES. (2005). Integration and segregation in auditory scene analysis. *J Acoust Soc Am*, 117, 1285–1298.
- Warren RM. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.
- Warren RM, Obusek CJ, Ackroff JM. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149–1151.
- Warren RM, Bashford JA, Healy EW, Brubaker BS. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Percept Psychophys*, 55, 313–322.
- Warren RM. (1999). *Auditory Perception: A new Analysis and Synthesis*. Cambridge, UK: Cambridge University Press.
- Woods DL. (1995). The component structure of the NI wave of the human auditory evoked potential. *Electroencephalogr Clin Neurophysiol*, 44, S102–109.
- Wrightson JM, Warren RM. (1981). Incomplete auditory induction of tones alternated with noise: Effects occurring below the pulsation threshold. *J Acoust Soc Am*, 69, S105–106.

Supplemental results

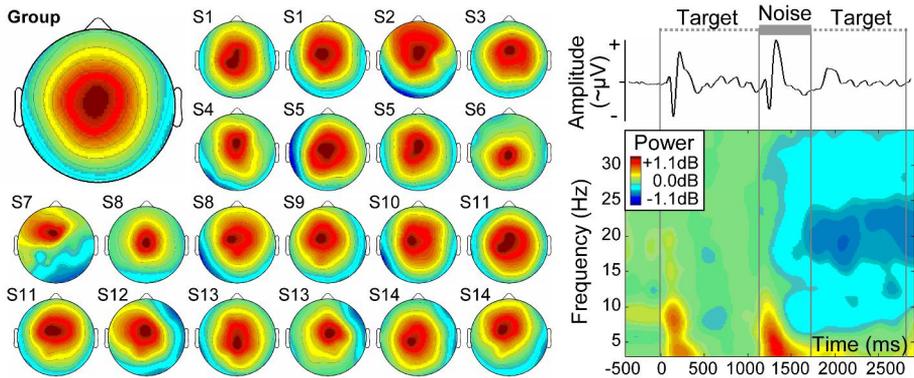


Figure S1. Anatomical-functional characteristics of clustered components of interest. Average component cluster scalp topography (enlarged) and individual component scalp topographies from all participants ($N = 14$), indicating source activities projected to central-medial scalp regions. The average cluster power spectrum (see time-frequency plot) revealed significant theta-band (3–7 Hz) synchronization (red) locked to the onsets of the stimulus and the noise, and sustained beta-band (13–27 Hz) desynchronization (blue) during and mainly after the noise, relative to the 500-ms pre-noise interval. The average cluster ERP waveform exhibited prominent onset-locked peaks, consistent with the theta-band synchronizations.

FMRI-based localization of neural representations of illusory voice continuity in auditory cortex

Abstract

When a portion of a sound is replaced by masking noise the interrupted sound may be perceptually restored, thus giving rise to a continuity illusion. Functional magnetic resonance imaging (fMRI) has indicated that tone continuity illusions are processed in primary auditory cortex (PAC), whereas vowel illusions have been associated with activity in vowel-sensitive cortical regions. This suggests that the neural substrates of the continuity illusion may differ depending on the complexity of the restored sound. It remains unclear, however, whether neural processes that restore interrupted voice sounds operate on voice percepts or frequency components constituting these vowels. In this fMRI study we report preliminary results showing that the masking of gaps in voice sounds is reflected in activity in superior temporal gyrus (STG), a region associated with voice processing. Activity in PAC was found to parallel listeners' reports of perceived continuity of identical voice stimuli, consistent with previous results on tones. Together with the findings on tone illusions these preliminary data suggest that voice-sensitive regions in STG could play a role in integrating interrupted voice sounds with noise while PAC may facilitate the restoration of these voices by filling in missing frequency components.

Based on: Riecke L, Walter A, Sorger B, Formisano E. Illusory filling of noise-interrupted voices in auditory cortex. *In preparation.*

1 Introduction

In natural environments, meaningful auditory perception requires the tracking of sounds in the presence of other possibly louder sounds. The auditory continuity illusion (Miller & Licklider, 1950) is a striking example for how the auditory system constructs sounds of interest even under such noisy conditions. The continuity illusion describes the phenomenon of perceiving an interrupted sound as apparently continuous, provided that noise masks the interruption (Houtgast, 1972; Warren *et al.*, 1972). For example, when a phoneme in a spoken sentence is replaced by a cough or noise the interrupted speech may sound more continuous and intelligible (Warren, 1970; Powers & Wilcox, 1977; Bashford *et al.*, 1992). Extensive psychoacoustic research has demonstrated continuity illusions also for target sounds of lower complexity, including tones, sweeps, and melodies (for reviews, see Bregman, 1990; Warren, 1999), suggesting general constructive mechanisms. The illusion is thought to emerge from perceptual processes that integrate the fragmented target sound with the interrupting masker based on learned schemas (Ciocca & Bregman, 1987), which may facilitate smooth hearing of fragmented sounds in noisy scenes.

Recent research on the underlying neural mechanisms has provided converging evidence that the auditory cortex may play an important role in the construction of continuity illusions. Electrophysiological recordings in primary auditory cortex (PAC) of monkeys have shown suppressed activity of neurons that respond to gaps in tones may be suppressed when these gaps are masked by noise (Petkov *et al.*, 2007). A subset of neurons that were more sensitive to uninterrupted tones showed sustained activity during the masked gaps. Therefore, neurons in monkey PAC respond to interrupted tones as if they were continuous, provided that these interruptions are masked. In humans, functional magnetic resonance imaging (fMRI) revealed that hemodynamic activity in PAC scales with the masking of gaps in tones by noise (Riecke *et al.*, 2007). This activity increased when the gaps were gradually unmasked, indicating neural processes related to the detection of gaps in tones. The activity further paralleled changes in listeners' continuity percepts of identical stimuli, suggesting non-sensory modulations of cortical representations of tones. In sum, these studies have shown that tone continuity illusions are represented already in the PAC. Given the tonotopic layout of sound frequency representations in PAC (e.g. Formisano *et al.*, 2003; Petkov *et al.*, 2006), the results put forward a neural model in which early cortical processing stages restore interruptions in individual frequency bands.

In a recent fMRI study on vowel continuity illusions in humans, Heinrich and colleagues (2008) used spectrally complex sounds (synthetic vowels) in which portions of individual formant-like frequency components were replaced by loud or soft noise. When noise masks the spectral gaps in such sounds, the fragmented formants can be restored, thus facilitating the integration of the fragmented formant with other intact formants to speech-like vowel percepts (Carlyon *et al.*, 2002a, 2002b). It was found that the speech-like vowel percepts evoke less activity in PAC

as opposed to vowel sounds in which these gaps were unmasked, consistent with the previous results on tones. The opposite effect was observed in vowel-preferring regions in posterior middle temporal gyrus (MTG) which were more activated by vowel sounds with masked formant gaps. The authors concluded that vowel continuity illusions are represented in vowel-preferring regions in MTG, suggesting that cortical processes that restore interrupted formants operate prior to this stage.

A problem with this interpretation is that there is no direct evidence showing that restoration of individual formants (component-based restoration) occurs prior to the emergence of vowel continuity illusions. Such illusions could also arise from restoration of partially masked vowels (object-based restoration) that were previously formed by integrating intact and noise-interrupted formants. This alternative interpretation is compatible with a recent psychoacoustic study demonstrating that the restoration of interrupted frequency components occurs after their integration with other simultaneous components (Darwin, 2005). The study revealed that the continuity illusion of a complex of interrupted harmonics requires the masking of the gaps in all harmonics. Unmasking of interrupted harmonics belonging to a different complex does not affect this illusion, which implies that continuity illusions apply to already integrated harmonics.

While component-based restoration would be expected to activate individual frequency representations in PAC, the results by Heinrich and colleagues (2008) are more suggestive of vowel-based restoration processes in vowel-sensitive cortical regions. However, the previous results do not allow definite conclusions in this regard as listeners' reports of vowel continuity illusions were not obtained during the fMRI measurements, so a direct link between the perceived sound restoration and physiology could not be established.

In the present fMRI study we measure blood oxygenation level dependent (BOLD) responses in auditory cortex to natural voice sounds interrupted by masking noise and we identify the regions in which activity levels co-vary with subjective reports of voice continuity. The masking of frequency components of the voice sounds was varied across four parametric levels, and listeners rated the overall perceived continuity of voices on a four-point scale while functional images were collected. To dissociate brain regions related to voice continuity illusions from those related to veridical continuity percepts we presented physically interrupted and uninterrupted voices. The methods used here are similar to those used in our previous fMRI study on tone continuity illusions (Riecke *et al.*, 2007).

2 Materials and methods

2.1 Participants

Seven human volunteers (mean age: 27 years, four women) with normal hearing abilities participated after providing informed consent. The local ethics committee approved the experimental protocol.

2.2 Stimuli, design, and task

Vowel utterances (/a/ vocalizations) produced by different speakers (two males, two females) were recorded in a sound-attenuated chamber and used as targets (duration: 2800 ms, Fig. 1A). The fundamental frequencies of the four voices (f_0 : 100, 148, 224, 231 Hz) were determined by global peaks in the different sound spectra. Broadband Gaussian-noise bursts (duration: 600 ms) were used as maskers. A spectral notch was inserted in noise bursts using a band-stop filter centered on f_0 . To vary the masking level, the width of the spectral notch was parameterized across four levels from zero to eight octaves, including two individually-defined intermediate widths (0.4 and 1.2 octaves on average; Fig. 1B). Note that the eight-octave notch essentially removed the noise from the stimulus. The different maskers were matched for their overall spectral power and superimposed on the spectrotemporal center of targets on a logarithmic-linear scale. The onsets and offsets of the gaps and maskers in the interrupted target stimuli were synchronized. All onsets and offsets were linearly ramped with 25-ms rise-fall times. Stimuli were passed through an eight-octave band-filter centered on f_0 . Uninterrupted target stimuli were spectrotemporally matched, except that no gaps were inserted in targets. All stimuli were matched for overall spectral power and sampled with 16 bits at 44.1 kHz in Matlab 7.0.1 (The MathWorks Inc., Natick, MA, USA). Stimulus parameter settings are summarized in Figure 1A.

Before imaging, listeners were trained in the MR scanner to attend to targets and to rate their overall perceived continuity on a four-point scale (labeled with “most likely continuous”, “probably continuous”, “probably discontinuous”, “most likely discontinuous”). In the fMRI experiments, blocks of stimulation alternating with stimulation-free baseline blocks of same duration (20 s) were presented in pseudorandomized order. Experimental conditions (Fig. 1B) were presented during stimulation blocks comprising all four voices in randomized order. Voice intervals were alternated with visually cued 2200-ms response intervals during which listeners performed ratings. Stimuli were presented diotically via Commander XG headphones (Resonance Technology Inc., Northridge, CA, USA) at approximately 70 dB sound pressure level (SPL).

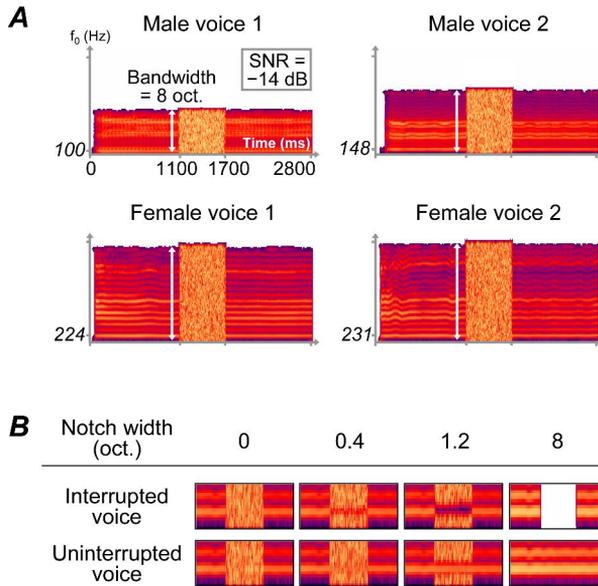


Figure 1. Stimuli and experimental design. **A**, Auditory stimuli comprised natural voice sounds (produced by four different speakers) and interrupting noise, filtered with an eight-octave bandpass centered on the different speakers' fundamental frequencies (f_0 , italic numbers). oct., octaves; SNR, relative spectrum level of voice and noise in decibel (dB). **B**, Stimulus conditions included interrupted and uninterrupted voices (exemplified for female voice 2 [see panel A], see enlarged spectrogram excerpts). Local noise energy levels were parameterized to control the masking of f_0 and adjacent frequency components. Voices comprising masked gaps were expected to evoke continuity illusions. Stimuli were matched for their SNR and bandwidth (white arrows) on logarithmic scales.

2.3 Imaging, data preprocessing, and statistical analysis

Images were collected with a Siemens Allegra 3-Tesla MRI system (Siemens Medical Systems, Erlangen, Germany). BOLD signal changes were measured with a head coil using a gradient echo planar imaging (EPI) sequence (time to echo: 30 ms, acquisition time: 2200 ms, repetition time: 5000 ms, field of view: $256 \times 256 \text{ mm}^2$, matrix size: 128×128 , slice thickness: 2 mm). A clustered volume EPI technique was applied that permitted presentation of auditory stimuli in silence between subsequent volume acquisitions (e.g. Jäncke *et al.*, 2002; Van Atteveldt *et al.*, 2004). During each experiment, 134 volumes were collected, each comprising 27 axial contiguous slices centered on the Sylvian fissure and covering the entire auditory cortex. In total, 28 functional runs were obtained (four runs for each participant). Structural T1-weighted volumes optimized for gray-white matter contrast were obtained using an MPRAGE pulse sequence (voxel resolution: $1 \times 1 \times 1 \text{ mm}^3$).

MRI data were preprocessed and analyzed using Brain Voyager QX 1.9 (Brain Innovation, Maastricht, The Netherlands). For group analyses, unsmoothed functional data and cortical surface-based methods were used that permit improved localization of auditory activation foci compared to Talairach space-based methods (Desai *et*

al., 2005). Statistical analyses involved a general linear model as previously described (Riecke *et al.*, 2007): The model included partially uncorrelated (orthogonalized) predictors for voice gaps, masking changes, and mean continuity ratings (determined by averaging all ratings within each block). For one listener, several ratings were not logged and therefore interpolated based on the grand mean of logged ratings. In the stimulus-based analysis of the fMRI data, the masking parameterizations for interrupted and uninterrupted voices were compared to identify vertices exhibiting significant masker \times gap interactions. In the subsequent percept-based analysis, the orthogonalized perceived continuity ratings of interrupted and uninterrupted voices were compared to determine vertices showing rating \times gap interactions. The resulting random-effect maps were thresholded at $t_6 = 2.5$ (uncorrected $p < 0.05$). To reduce inflated type-I error probabilities caused by multiple comparisons, the thresholded maps were further submitted to a 10 mm^2 cluster size criterion. For further details on the methods, see our previous study (Riecke *et al.*, 2007).

3 Results

3.1 Behavioral results

Behavioral data analysis revealed that the perceived continuity of interrupted voices increased significantly with masking levels. The opposite effect was found for uninterrupted voices (masker \times gap interaction: $F_{3,4} = 459.1$, $p < 10^{-5}$, Fig. 2A), in line with our previous results on tone continuity illusions (Riecke *et al.*, 2007, 2008). Consistent effects were evident for all individual voices (Fig. 2B).

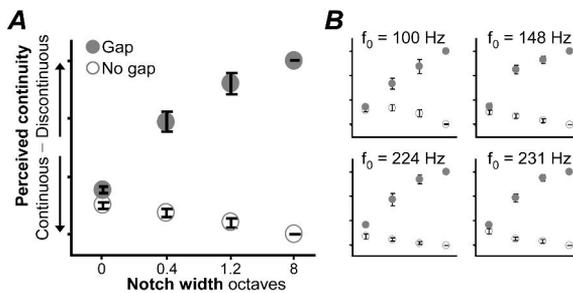


Figure 2. Behavioral results. **A**, Average continuity ratings (mean \pm standard error [SE]) across seven listeners of interrupted voices (solid circles) and uninterrupted voices (open circles) for different noise maskers. Wide notches in the noise unmasked the gaps in interrupted voices and disrupted the continuity illusion. The opposite effect was found for uninterrupted voices, which were perceived as more continuous at lower masking levels. **B**, Same as panel A, but for two individual male (upper row) and female voices (lower row). The perceptual effects of the noise (A) were speaker-independent. f_0 , fundamental frequency.

3.2 Imaging results

Interrupted and uninterrupted voices evoked widespread activity in bilateral auditory cortex relative to baseline (Fig. 3A), the peak of which was located in lateral transverse gyrus (Heschl's gyrus [HG]) (Fig. 3A, red arrows). The parametric manipulation of the masker in these regions was analyzed by contrasting the parametric notch predictors for interrupted versus uninterrupted voices. This analysis revealed that the correlation between activity levels in posterior portions of superior temporal gyrus (STG) and voice masking levels was significantly determined by whether voices were interrupted or uninterrupted (masker \times gap interaction; Fig. 3B, blue regions; $t_6 = 2.5$, uncorrected $p < 0.05$, cluster threshold: 10 mm^2). This effect was also observed in the right transverse temporal sulcus (Heschl's sulcus) and in middle and lateral portions of the left HG. The most significant clusters were identified in bilateral posterior STG and defined as regions of interest (ROIs; Fig. 3B, blue arrows). Compared to analogously defined ROIs in our tone illusion study (Riecke *et al.*, 2007), these ROIs were situated more dorso-caudally (Euclidean distances in Talairach space: 17.7 mm and 14.5 mm for right and left hemisphere, respectively). Further analysis of the patterns of the masker \times gap interactions in these ROIs revealed that for interrupted voices, parametric increases in masking evoked significant decreases in brain activity (Fig. 3C; ROI analysis in right and left hemisphere: $t_6 = 2.3$ and 3.5, $p = 0.02$ and 0.0004 respectively), whereas no such effects were observed for uninterrupted voices (ROI analysis in right and left hemisphere: $t_6 = -1.5$ and -0.5, $p = 0.1$ and 0.6 respectively). These stimulus-based results suggest that posterior STG plays a role in the detection of changes in the masking of gaps in f_0 .

To investigate which regions were involved in changes in listeners' actual percepts alone (*i.e.* without concomitant masking changes) another analysis was performed based on listeners' rating data that had been obtained during fMRI. This percept-based analysis revealed that activity in middle portions of bilateral HG correlated significantly more with changes in perceived continuity of interrupted voices than with changes in continuity perception of uninterrupted voices (rating \times gap interaction; Fig. 3B, green regions; $t_6 = 2.5$, uncorrected $p < 0.05$, cluster threshold: 10 mm^2). The most significant clusters were defined as ROIs (Fig. 3B, green arrows) and found proximate to analogously defined ROIs in our tone illusion study (Euclidean distances in Talairach space: 2.4 mm and 9.9 mm for right and left hemisphere, respectively). ROI analysis in the left hemisphere showed significant effects of continuity ratings on BOLD responses to interrupted voices but not to uninterrupted voices (Fig 3C; $t_6 = 1.9$ and -1.1, $p = 0.05$ and 0.3 respectively). Comparable trends were observed in the right hemisphere (Fig 3C; $t_6 = 1.7$ and -0.9, $p = 0.08$ and 0.4 respectively). As these BOLD response changes in bilateral HG were not related to stimulus changes but to changes in ratings, they likely reflected spontaneous changes in stimulus perception and interpretation.

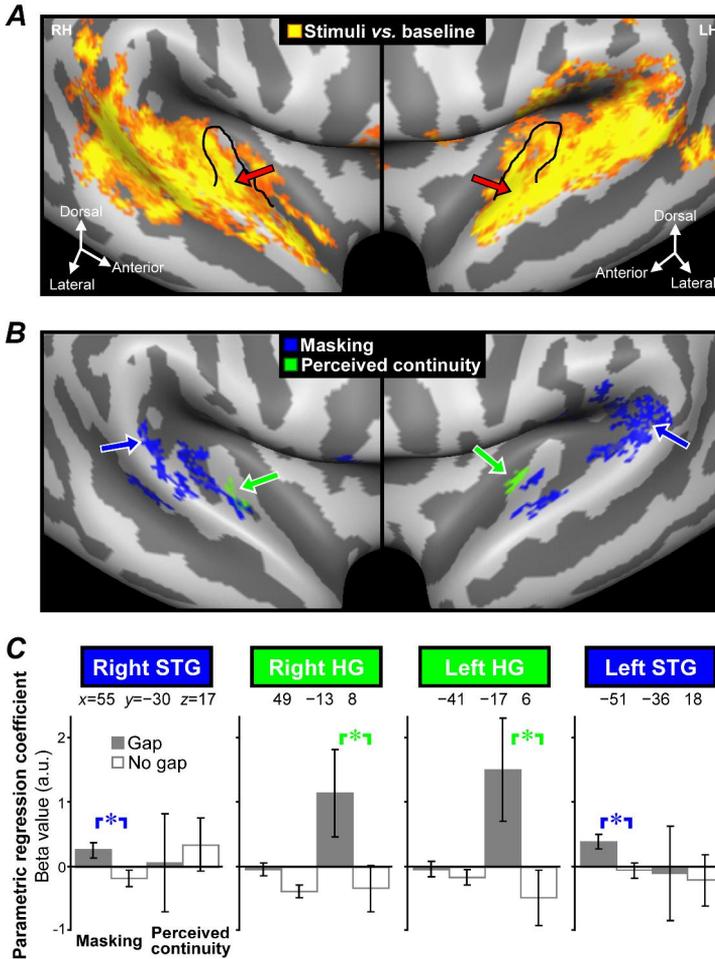


Figure 3. Functional neuroimaging results. **A**, Average activation maps projected onto the inflated average cortical surface of the right hemisphere (RH) and left hemisphere (LH) after cortical alignment across seven listeners. Voice stimuli evoked widespread activity in auditory cortical regions compared to baseline. Arrows indicate locations of peak activity. Black lines delineate the location of Heschl's gyrus (HG). **B**, Bilateral activity in posterior portions of superior temporal gyrus (STG) correlated more with the masking of interrupted than uninterrupted voices (blue regions). Bilateral activity in HG correlated more with the perceived continuity of interrupted than uninterrupted voices (green regions). Random-effect maps were thresholded at $t_0 = 2.5$ (uncorrected $p < 0.05$) and active cortical patches smaller than 10 mm^2 excluded to reduce false-positive probabilities. **C**, Parametric regression coefficients (mean \pm SE across listeners) associated with masking of f_0 and perceived continuity of interrupted voices (solid rectangles) and uninterrupted voices (open rectangles) in four peak regions. Positive values indicate parametric decreases in BOLD response to parametric increases in masking or continuity, and vice versa for negative values. Increases in masking decreased activity to interrupted but not to uninterrupted voices in bilateral posterior STG (panel B, blue arrows). In bilateral HG (panel B, green arrows), perceived continuity of only interrupted voices explained additional variance in activity. Asterisks indicate significance at $p < 0.05$. Average Talairach coordinates are given in units of mm for each region of interest. a.u., arbitrary units.

4 Discussion

The finding that only masking of interrupted voice sounds parametrically affects BOLD responses in posterior STG, especially in the left hemisphere, suggests that neural activity in this region does not reflect masking *per se* but the presence of acoustic energy in f_0 during interruptions. Since it is the energy in f_0 that determines the continuity illusion the observed activity may also reflect differences in the perception of gaps in voices. Regions in posterior STG, particularly in the left hemisphere, have been commonly associated with the processing of voice sounds (Belin *et al.*, 2000, 2002; for reviews, see Scott & Wise 2004; Scott, 2005). These regions also show enhanced sensitivity to onsets and offsets in speech sounds (Harms *et al.*, 2005) and to voice onset times in vowel sounds, especially in the left hemisphere (Jäncke *et al.*, 2002). It is thus conceivable that voice-sensitive regions in posterior STG increase their activity when utterances are perceived that are separated by a gap. This interpretation is consistent with previous fMRI evidence of increased activity in planum temporale (PT) in response to silent interruptions in non-speech sounds (Mustovic *et al.*, 2003; Herdener *et al.*, 2007), and is compatible with the general notion that PT is involved in the analysis of acoustically complex stimuli (for review see Griffiths & Warren, 2002).

We observed comparable effects of gaps in voices in specific regions of HG which have been suggested to constitute PAC (Rivier & Clarke, 1997; Morosan *et al.*, 2001; Rademacher *et al.*, 2001). The observed effects also fit well with previous fMRI results on the continuity illusion (Riecke *et al.*, 2007; Heinrich *et al.*, 2008). The modulations found in PAC were weaker than those in the regions that may be voice-sensitive, which implies that auditory cortical processes related to gap perception operate mainly on representations of objects rather than on representations of individual frequency components.

Strikingly we found that subjective reports of voice continuity correlated with activity in PAC, consistent with our previous fMRI results on tone continuity illusions (Riecke *et al.*, 2007). Since these effects were unrelated to sensory stimulus changes, they likely reflected perceptual influences of non-sensory factors such as attention. Attention may trigger mental processes that match sensory stimulus representations with perceptual expectancies or schemas (Bregman, 1990; Alain & Bernstein, 2008). Schemas are crucial for the perceptual interpretation of distorted speech signals (Repp *et al.*, 1992) because they affect neural processes related to restoration, as evidenced by recent electroencephalography results (Sivonen *et al.*, 2007). Here we identified effects related to voice restoration in a region that is also involved in the restoration of interrupted tones (Petkov *et al.*, 2007; Riecke *et al.*, 2007), corroborating the notion that activity in PAC reflects perceptual rather than acoustic stimulus properties (Nelken, 2004). More importantly our finding indicates that continuity illusions of voices rely on the restoration of interrupted frequency representations (component-based restoration) rather than voice representations (object-based restoration).

The absence of restoration-related effects in voice-sensitive regions suggests that interrupted frequency representations are not fed forward to these regions after their

restoration in PAC. The most plausible explanation for the observed pattern thus may be that individual frequency components are integrated before their restoration (Darwin, 2005) and represented as interrupted voice percepts in voice-sensitive regions, as indicated by the voice gap-related effects in posterior STG. Subsequent restoration of interrupted voice percepts may be achieved remotely by primary auditory cortical processes that complete individual frequency representations, explaining the illusion-related effects in PAC.

Our results differ from those by Heinrich and colleagues (2008) in several aspects which may be related to differences in stimulus material and analyses: Regarding stimulus differences, we used natural voice stimuli comprising temporal gaps all of which evoked clearly audible voice percepts. Heinrich *et al.* employed uninterrupted synthetic vowel stimuli with spectral gaps; these stimuli evoked vowel-like percepts only when the gaps were masked. Regarding analysis differences, we identified a region in posterior STG sensitive to voice gap percepts based on comparing energy-matched stimuli evoking interrupted versus uninterrupted voice percepts. Heinrich *et al.* identified a region in MTG sensitive to vowel-like percepts based on contrasting energy-unmatched stimuli evoking vowel-like versus vowel-unlike percepts. Furthermore, Heinrich *et al.* did not collect listeners' continuity ratings during fMRI which may explain why they could not detect illusion-related modulations in PAC in their analyses. In sum, regions in posterior STG may represent gaps in voices, whereas more ventral regions in MTG may indicate sound energy levels and vowel perception, consistent with the previous notion that the two regions serve different functional purposes (Wise *et al.*, 2001).

A limitation of the present study is the relatively low statistical significance of the fMRI results. This may be related to the small sample size and also to the conservative analyses that were applied to unsmoothed, relatively high spatial-resolution fMRI data. Since a consistent pattern of activity peaks was observed in HG and STG in both hemispheres, explanations in terms of multiple comparisons seem unlikely.

Despite these cautionary remarks, we tentatively conclude that voice-sensitive cortical regions integrate voice sounds with noise and analyze the resulting voice percept representations for offsets and onsets. Based on the matching of these representations with perceptual schemas, early auditory cortical regions may restore interrupted frequency components. We speculate that this restoration process could involve cortical feedback signals that selectively enhance masked frequency representations in PAC to fit a cortical stimulus representation with a matching schema. Non-auditory cortical regions could then read out and interpret the distributed representation of the voice percept from auditory cortex (Formisano *et al.*, 2008) in order to select an appropriate behavioral response. We are currently collecting further data to corroborate these preliminary conclusions.

Acknowledgement

We thank Claudia Schreiner for constructive comments on the script.

References

- Alain C, Bernstein LJ. (2008). From sounds to meaning: the role of attention during auditory scene analysis. *Curr Opin Otolaryngol Head Neck Surg*, 16, 485–489.
- Bashford JA Jr, Riener KR, Warren RM. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Percept Psychophys*, 51, 211–217.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312.
- Belin P, Zatorre RJ, Ahad P. (2002). Human temporal lobe response to vocal sounds. *Cogn Brain Res*, 13, 17–26.
- Bregman AS. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Carlyon RP, Deeks J, Norris D, Butterfield S. (2002a). The continuity illusion and vowel identification. *Acta Acustica United with Acustica*, 88, 408–415.
- Carlyon RP, Micheyl C, Deeks JM, Moore BCJ. (2002b). FM phase and the continuity illusion. *J Acoust Soc Am*, 111, 2468.
- Ciocca V, Bregman AS. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Percept Psychophys*, 42, 476–484.
- Darwin CJ. (2005). Simultaneous grouping and auditory continuity. *Percept Psychophys*, 67, 1384–1390.
- Desai R, Liebenthal E, Possing ET, Waldron E, Binder JR. (2005). Volumetric vs. surface-based alignment for localization of auditory cortex activation. *Neuroimage*, 26, 1019–1029.
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40, 859–869.
- Formisano E, De Martino F, Bonte M, Goebel R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, 322, 970–973.
- Griffiths TD, Warren JD. (2002). The planum temporale as a computational hub. *Trends Neurosci*, 25, 348–353.
- Harms MP, Guinan JJ Jr, Sigalovsky IS, Melcher JR. (2005). Short-term sound temporal envelope characteristics determine multisecond time-patterns of activity in human auditory cortex as shown by fMRI. *J Neurophysiol*, 95, 210–222.
- Heinrich A, Carlyon RP, Davis MH, Johnsrude IS. (2008). Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. *J Cogn Neurosci*, 20, 1737–1752.
- Herdener M, Esposito F, Di Salle F, Lehmann C, Bach DR, Scheffler K, Seifritz E. (2007). BOLD correlates of edge detection in human auditory cortex. *Neuroimage*, 36, 194–201.
- Houtgast T. (1972). Psychophysical evidence for lateral inhibition in hearing. *J Acoust Soc Am*, 51, 1885–1894.
- Jäncke L, Wüstenberg T, Scheich H, Heinze HJ. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, 15, 733–746.
- Miller GA, Licklider JCR. (1950). The intelligibility of interrupted speech. *J Acoust Soc Am*, 22, 167–173.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage*, 13, 684–701.
- Mustovic H, Scheffler K, Di Salle F, Esposito F, Neuhoff JG, Hennig J, Seifritz E. (2003). Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage*, 20, 429–434.
- Nelken I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol*, 14, 474–480.
- Petkov CI, Kayser C, Augath M, Logothetis NK. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol*, 4, e215.
- Petkov CI, O'Connor KN, Sutter ML. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54, 153–165.

Chapter 5

- Powers GL, Wilcox JC. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *J Acoust Soc Am*, 61, 195–199.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage*, 13, 669–683.
- Repp BH. (1992). Perceptual restoration of a “missing” speech sound: auditory induction or illusion? *Percept Psychophys*, 51, 14–32.
- Riecke L, Van Opstal AJ, Goebel R, Formisano E. (2007). Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *J Neurosci*, 27, 12684–12689.
- Riecke L, Van Opstal AJ, Formisano E. (2008). The auditory continuity illusion: A parametric investigation and filter model. *Percept Psychophys*, 70, 1–12.
- Rivier F, Clarke S. (1997) Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex. *Neuroimage*, 6, 288–304.
- Scott SK, Wise RJS. (2004). The functional neuroanatomy of prelexical processing of speech. *Cognition*, 92, 13–45.
- Scott SK. (2005). Auditory processing – speech, space and auditory objects. *Curr Opin Neurobiol*, 15, 197–201.
- Sivonen P, Maess B, Lattner S, Friederici AD. (2006). Phonemic restoration in a sentence context: evidence from early and late ERP effects. *Brain Res*, 1121, 177–189.
- Van Atteveldt N, Formisano E, Goebel R, Blomert L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43, 271–282.
- Warren RM. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.
- Warren RM, Obusek CJ, Ackroff JM. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149–1151.
- Warren RM. (1999). *Auditory Perception: A new Analysis and Synthesis*. Cambridge, UK: Cambridge University Press.
- Wise RJS, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. (2001). Separate neural subsystems within Wernicke’s area. *Brain*, 124, 83–95.

Identification of contextual determinants of illusory continuity

Abstract

The human auditory system is efficient at restoring sounds of interest. In noisy environments, for example, an interrupted target sound may be illusorily heard as continuing smoothly when a loud noise masks the interruptions. In quiet environments, however, sudden interruptions might signal important events. In that case, restoration of the target sound would be disadvantageous. Achieving useful perceptual stability may require the restoration mechanism to adapt its output to current perceptual demands, a hypothesis which has not yet been fully evaluated. In this study, we investigated whether auditory restoration depends on preceding auditory scenes, and we report evidence that restoration adapts to the perceived continuity of target sounds and to the loudness of interrupting sounds. In the first experiment, listeners adapted to illusory and non-illusory tone sweeps (targets) and interrupting noise, and we observed that the perceived continuity of the target and the loudness of the interrupting noise influenced the extent of subsequent restorations. A second experiment revealed that these adaptation effects were unrelated to the adapted spectra, indicating that non-sensory representations of the perceived auditory scene were involved. We argue that auditory restoration is a dynamic illusory phenomenon which recalibrates continuity hearing to different acoustic environments.

1 Introduction

A major challenge for the auditory system is to ensure stability of relevant sound objects in the presence of environmental noise. This stability is facilitated by perceptual filling mechanisms that may restore noisy interruptions in a foreground sound, thereby creating a continuity illusion of the foreground. In a loud scene, for example, an interrupted voice can be illusorily heard as continuing through noise (Miller & Licklider, 1950), which may help to restore the actual speech signal and improve its intelligibility (Warren, 1970; Powers & Wilcox, 1977; Bashford *et al.*, 1992).

Since its discovery (Miller & Licklider, 1950) the continuity illusion has been investigated extensively under several names such as pulsation threshold (Houtgast, 1972), temporal induction (Warren *et al.*, 1972), contextual concatenation (Warren, 1984), amodal completion (Miller *et al.*, 2001), and illusory filling (Petkov *et al.*, 2003; for reviews, see Bregman, 1990; Warren, 1999). Early research revealed that the continuity illusion depends on the masking of the gaps in the interrupted sound (Houtgast, 1972; Warren *et al.*, 1972), or on the absence of sensory evidence for these gaps (Warren *et al.*, 1972; Dannenbring, 1976; Bregman & Dannenbring, 1977). Another determining factor is the similarity of the sound fragments that surround a noisy interruption. For example, when the fragments of an interrupted sweep have the same frequency trajectory or when they are proximate to each other in frequency and time, they are more likely to be grouped and to produce a continuity illusion (Ciocca & Bregman, 1987). Thus, the relevance of the fragment that follows the interruption implies that the continuity illusion depends on the acoustic context (Warren, 1983; Ciocca & Bregman, 1987). Furthermore, it has been observed that the continuity illusion may fade out or fade in during long noise interruptions (Wrightson & Warren, 1981; Warren *et al.*, 1994). This apparent partial extension of the fragmented sound during the interruption indicates that the illusion is not an all-or-none phenomenon, but a perceptual continuum (Bregman, 1990).

More recent research on the continuity illusion has extended the previous findings and stressed the relevance of acoustic onsets and offsets (edges). When a sweep intersects with a silent interruption of a longer sweep, the gap may be misattributed to the shorter sweep while the longer sweep may appear illusorily continuous (Nakajima *et al.*, 2000). Furthermore, two spectrally segregated sweeps that partly overlap in time may be illusorily perceived as a single continuous sweep, accompanied by an additional illusory tone during the overlap (Remijn *et al.*, 2001; Remijn & Nakajima, 2005). This effect is observed irrespective of the presence or absence of masking noise, and persists even when the spectral gap is wider than one critical band (Fletcher, 1940), indicating the involvement of non-peripheral mechanisms. These findings are consistent with the previous results (see previous paragraph) and support a model of auditory grouping proposed by Nakajima and colleagues (2000). According to this model, the illusory continuity of a sound may emerge from the perceptual binding of spectrotemporally proximate edges of different sounds.

An aspect of the continuity illusion which has received not much attention so far is the stability of the illusion across different acoustic environments. This aspect is relevant as restoration of a given sound may be desired under certain circumstances, but not under other circumstances. For example, in quiet environments, an abrupt interference by a loud sound from the background might reflect a meaningful event and thus it should not be smoothed into a sound of interest in the actual foreground. To achieve appropriate levels of perceptual stability, the restoration mechanism would need to adapt its output to current environmental demands.

Auditory adaptation phenomena illustrating how hearing depends on recent acoustic input have been well described in the literature. For example, the same sound may be perceived as substantially softer when a louder sound is presented shortly before (Marks, 1994). Such contrastive aftereffects have been commonly interpreted in terms of adaptation or habituation, which can be understood as reductions in behavioral or neurophysiologic responses caused by prior or ongoing stimulation. Aftereffects may occur for various low-level stimulus properties and in various sensory modalities as evidenced by previous research. Exposure to sounds was shown to induce temporary changes in the ability to discriminate intensities (Zeng *et al.*, 1991; Zeng & Turner, 1992; Carlyon & Beveridge, 1993; Plack *et al.*, 1995; Zeng & Shannon, 1995; Plack, 1996; Oberfeld, 2007, 2008) or to detect target sounds in noise (Penner, 1974; Kidd & Feth, 1982; Viemeister & Bacon, 1982; Wright *et al.*, 1993). Furthermore, amplitude- or frequency-modulations (AMs or FM, respectively) of preceding sounds were found to temporarily increase detection thresholds for AMs or FM of subsequent sounds (Kay & Matthews, 1972; Green & Kay, 1973, 1974; Regan & Tansley, 1979; Tansley & Suffield, 1983; Moody *et al.*, 1984; Wojtczak & Viemeister, 2003, 2005).

These and other (Rosenblith *et al.*, 1947; Zwicker, 1964) auditory aftereffects may reflect a ubiquitous mechanism that enables the auditory system to adapt to current probabilities in the acoustic environment. Further findings of adaptation to non-illusory higher-level sound properties such as sound source location (Frissen *et al.*, 2003, 2005; Phillips & Hall, 2005), phonemic category (Eimas & Corbit, 1973; Cooper, 1974; Diehl *et al.*, 1978; Samuel & Newport, 1979; Simon & Studdert-Kennedy, 1978; Ohde & Sharf, 1979; Sawusch & Jusczyk, 1981; Landahl & Blumstein, 1982) or voice gender (Schweinberger *et al.*, 2008) have extended this hypothesis. Therefore, hearing might depend not only on the acoustic input but also on its perceptual interpretation. Such a perceptual adaptation mechanism could have ecological value because it potentially improves the ability to discriminate rare sound objects in the perceived auditory scene.

To assess whether this adaptation also applies to auditory restoration, we investigated whether the continuity of tone sweeps (targets) or the masking potential of interrupting noise influence subsequent restorations of fragmented target sounds. Listeners were presented with series of schematic auditory scenes, consisting of sweeps, noise, or both (restoration condition), and rated their continuity. An ambiguous scene (probe) created a bistable restoration condition which was presented after different series of unambiguous scenes (adaptors). In the first

experiment, listeners were adapted to illusory continuous, truly continuous, or discontinuous sweeps, and to loud or soft noise. We obtained evidence that auditory restoration adapts to the perceived continuity of the sweeps and also to the loudness of the interrupting noise. In a second experiment, we investigated whether this adaptation is specific to the spectra of the adapting sounds. Listeners adapted to sweeps or noise whose spectra did not overlap with the portion of the spectrum required for illusory restoration. We found that auditory restoration adapts even when the adapted and restored spectra are incongruent, and we observed that the masking potential of the noise has little impact. The results indicate that auditory restoration depends on the perceived continuity and the loudness of preceding sounds, irrespective of their frequency content. We argue that these aftereffects can be explained by adaptation to abstract, non-sensory representations of the auditory scene.

2 Materials and methods

2.1 Participants

26 volunteers (age: 24 ± 3 years, mean \pm standard deviation [SD]) with normal hearing abilities, mainly students from Maastricht University, participated in the study after providing informed consent. Two different groups of twelve listeners participated in experiment 1 and 2, respectively, and two other listeners participated in both experiments. Participants were uninformed about the background of the study, except for two (one of the authors and one research assistant). The local ethical committee approved the procedure.

2.2 Stimuli

An auditory scene was simulated by stimuli of 5000-ms duration (Fig. 1A) consisting of a tone sweep, white noise, or both (Fig. 1B,C). The tone's frequency was logarithmically increased from 1 to 3 kHz and its amplitude was pulsed at 2 Hz, resulting in an ascending sweep which was repeatedly interrupted by silence. For experiment 1 (Fig. 1B), the noise was band-passed from 0.9 to 3.6 kHz (two octaves, 3-dB cutoff frequencies) so that the noise covered the sweep's spectrum. To enable the probing of restoration in these spectra, the noise was inserted in the silent gaps of the sweep such that the onsets and offsets of noise and gaps were synchronized. All onsets and offsets were linearly ramped with 10-ms rise-fall times. For experiment 2 (Fig. 1C), some sweep and noise spectra were modified such that they did not overlap with the probe spectrum described above. The sweep's frequency was logarithmically decreased from 0.8 to 0.3 kHz, resulting in a pitch modulation which was reversed relative to that in experiment 1. The noise was band-passed from 0.45 to 7.2 kHz (four octaves) and band-stopped (notched) from 0.9 to 3.6 kHz (two octaves). Stimuli were sampled at 44.1 kHz with 16 bit resolution using Matlab 7.0.1 (The MathWorks Inc., Natick, MA, USA). Stimuli were presented diotically at maximal 80 dB sound pressure level (SPL) using Presentation 9.30

software (Neurobehavioral Systems, Inc., Albany, CA, USA), a Creative Sound Blaster Audigy 2ZS sound card (Creative Technology, Ltd., Singapore), and a Sennheiser HMD 25-1 headset (Sennheiser electronic, Wedemark, Germany).

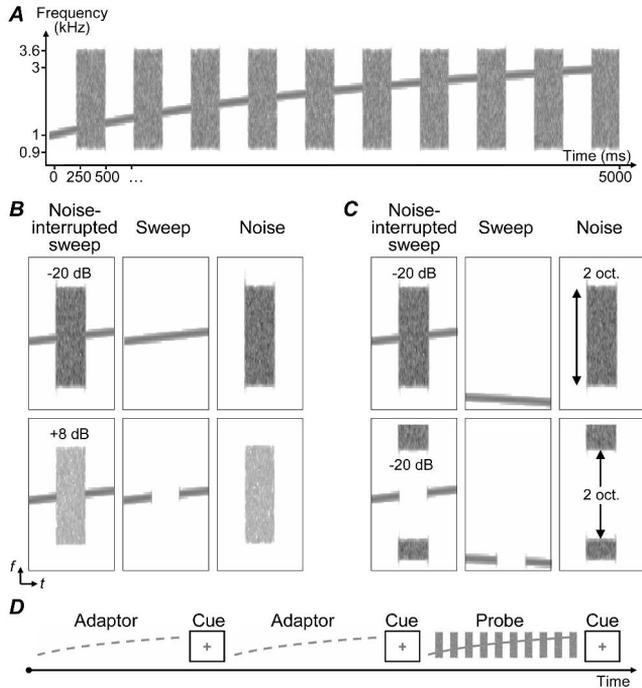


Figure 1. Stimulus and experimental design. **A**, Sound spectrogram and acoustic settings of a schematic auditory scene consisting of a discontinuous sweep repeatedly interrupted by noise. The interrupted target sound tends to be perceptually restored when the noise masks the interruptions, yielding a continuity illusion. Noise of an individually defined intermediate loudness was used to create an ambiguous masking condition (probe) under which restoration becomes bistable. **B**, Truncated spectrograms of the six adaptation stimuli (adaptors) presented in experiment 1. Sweeps of differently heard continuity and noise of different masking potential were presented either simultaneously (left column) or separately (middle and right col.). For experiment 2 (**C**), the spectra of some adaptors were modified such that they were incongruent with the probe spectrum (**A**). dB-values indicate SNR; oct., octaves. **D**, Adaptation, probe, and response intervals (the latter denoted by a visual cue) within a single trial sketched for a “Sweep” condition in experiment 1. Listeners rated stimulus continuity on a four-point scale.

2.3 Design and task

For each experiment, three pairs of adaptors were designed to test listeners for adaptation to three different aspects of the simulated auditory scenes (Fig. 1B,C). In the “Noise-interrupted sweep” conditions, the interrupted ascending sweep was alternated with noise. For experiment 1 (Fig. 1B), the overall noise level was varied (signal-to-noise ratio [SNR]: -20 or 8 dB, respectively) and the noise and sweep spectra overlapped. For experiment 2 (Fig. 1C), the overall noise level was held

constant at the higher value (SNR: -20 dB), and the noise spectrum was either congruent or incongruent with the sweep spectrum. Thus, the noise was eligible to either mask the sweep or not in each experiment and, therefore, these stimuli were expected to adapt listeners to illusory continuity or true discontinuity of the target sound (see Introduction). In the “Sweep” conditions, the sweep was either uninterrupted (*i.e.* the sweep’s amplitude was not pulsed) or interrupted by silence, and no noise was present. The sweep’s spectrum and pitch modulation were either congruent (exp. 1) or incongruent (exp. 2) with those of the probe. These stimuli were designed to evoke adaptation to different target sounds that were physically continuous or discontinuous, in the absence of noise. The sweep level remained constant throughout both experiments. The “Noise” conditions were identical to the “Noise-interrupted sweep” conditions, except that no sweep was presented. These stimuli were expected to adapt listeners to noise of high or low masking potential, either at covarying (exp. 1) or fixed (exp. 2) overall noise level, in the absence of the sweep.

Participants were seated in a sound-attenuated chamber and performed a forced-choice task. They were given written instructions to rate each stimulus’ overall continuity by pressing a button on a four-point scale (labeled as „> 75% continuous”, “> 50% continuous”, “> 50% discontinuous”, “> 75% discontinuous”). For the noise-interrupted sweep conditions, participants were instructed to attend to the sweeps and to ignore the noise, rendering these sounds the perceived foreground and background, respectively (Thurlow, 1957). Response intervals were indicated by a visual cross turning green at the stimulus offset. Before the main experiments, the participants’ ability to hear continuity illusions and perform the task was assessed in twelve training trials. Individual continuity illusion thresholds were estimated based on the method of limits (Fechner, 1860): Multiple series of noise-interrupted ascending sweeps were presented in which the SNR was either gradually decreased from +1.5 dB, or increased from -20 dB, in average steps of 1.5 dB. Thresholds were estimated as the average SNR at which listeners’ ratings switched from non-illusory discontinuity to illusory continuity, or vice versa. These thresholds were offset by an average step of -1.5 dB and implemented as intermediate noise levels to define individual, putative ambiguous, probe stimuli (Fig. 1A). These stimuli were expected to evoke bistable hearing of the target’s continuity, as observed during informal pilot experiments.

In the main experiments, each trial was comprised of an adaptation interval during which one of the six unambiguous adaptors (Fig. 1B,C) was presented twice, followed by a test interval during which the ambiguous probe was presented (Fig. 1D). The inter-stimulus interval (ISI) was fixed at 100 ms. Each of the six conditions was presented three times in randomized blocks of 4.5-min duration. Five individually randomized blocks were presented in total and listeners were allowed to take breaks in between. One listener’s data revealed bistable adaptor ratings during two successive blocks in experiment 1. These data likely reflected a fluctuation in vigilance and were discarded, which did not affect the overall results (Fig. 3A, subject S12).

2.4 Statistical analysis

For each experiment, each listener's continuity rating data were averaged across the 15 trials presented for each condition. The average ratings were statistically analyzed using a general linear model and a two-way analysis of variance (ANOVA) for repeated measures in SPSS 12.0.1 (SPSS inc., Chicago, ILL, USA). The conditions "Noise-interrupted sweep", "Sweep", and "Noise" (Fig. 1B,C) were included in the model as a single within-subject factor with three levels (labeled as adaptor type). The experimental manipulations of the different adaptor types were included in another within-subject factor with two levels (labeled as adaptor level). Interactions between these two factors were tested with *F*-tests. For testing simple effects, pairwise comparisons between individual conditions were analyzed using two-tailed paired samples *t*-tests. Differences between experiments were analyzed with two-tailed independent samples *t*-tests. Aftereffects were computed by subtracting the average probe ratings at each adaptor level separately for each adaptor type. The relative contributions of the separated adaptor types (sweeps, noise) to the aftereffects of the simultaneous adaptor type (noise-interrupted sweeps) were estimated using a linear regression model. The model's fit was assessed with an ANOVA and an *F*-test. Significance of the estimated regression coefficients was assessed with two-tailed one-sample *t*-tests. For all group analyses, inflated type-I error probabilities caused by multiple comparisons were corrected for using Bonferroni's method.

3 Results

3.1 Threshold

The average continuity illusion threshold was estimated as -6.6 ± 2.7 dB (mean \pm SD) across listeners. Thresholds estimated from descending and ascending SNR continua differed significantly (-4.7 and -8.5 dB, respectively; $t_{25} = 3.88$, $p < .001$). Listeners' reported ability to detect true discontinuities thus improved after being presented repeated continuity illusions, and vice versa, consistent with the notion of continuity adaptation.

3.2 Adaptor continuity ratings

Group analysis of listeners' adaptor continuity ratings revealed consistent results for experiment 1 and 2 (Fig. 2A,B). The midpoint of the continuity rating scale was considered to be the point at which auditory restoration became bistable (Fig. 2, large tick marks). As expected, all adaptors were rated significantly different from this point (exp. 1: all $|t_{13}| > 8.50$, all $p < 10^{-6}$; exp. 2: all $|t_{13}| > 8.02$, all $p < 10^{-6}$). Noise-interrupted sweeps were rated illusorily continuous when the noise masked the gaps in the sweep (exp. 1: $t_{13} = 18.12$, $p < 10^{-10}$; exp. 2: $t_{13} = 14.23$, $p < 10^{-9}$; Fig. 2A,B, col. 1,2), consistent with our previous results (Riecke *et al.*, 2008). Truly interrupted and uninterrupted sweeps alone were rated as discontinuous and continuous,

respectively (exp. 1: $t_{13} = 32.44, p < 10^{-14}$; exp. 2: $t_{13} = 19.51, p < 10^{-11}$; Fig. 2A,B, col. 3,4). Noise alone was rated as discontinuous, irrespective of its overall level (exp. 1: $t_{13} = -1.42, p = .18$) and masking potential (exp. 2: $t_{13} = -0.46, p = .66$; Fig. 2A,B, col. 5,6). Therefore, these stimuli adapted listeners to hearing strong sweep continuity or discontinuity in the presence or absence of masking or unmasking noise.

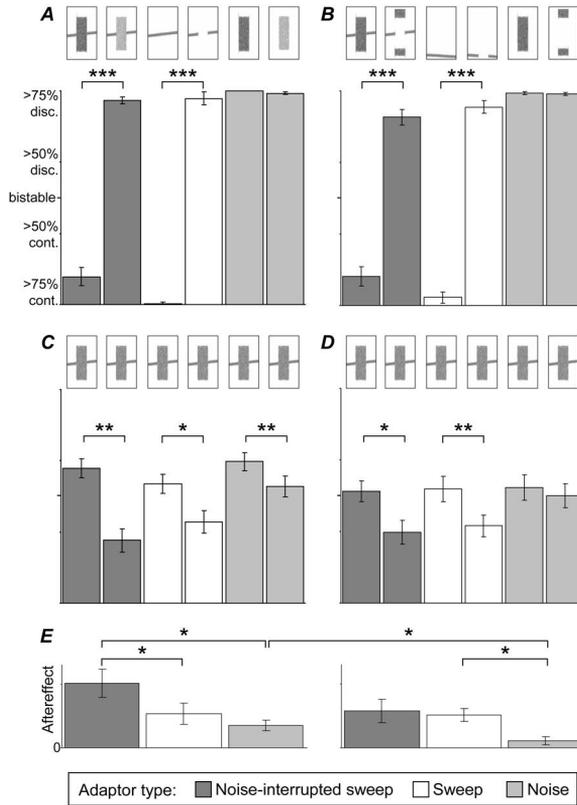


Figure 2. Results from group analysis for experiment 1 and 2. **A, B,** Mean continuity ratings of all adaptors (indicated by the truncated spectrograms) \pm standard error (SE) across listeners ($N = 14$) for experiment 1 (A) and 2 (B). The adaptors evoked unambiguous continuity ratings. **C, D,** Mean continuity ratings of the probe \pm SE across listeners after hearing twice the adaptors depicted above (A, B). The true or illusory continuity of preceding sweeps exhibited aftereffects on continuity ratings of the same subsequent ambiguous target, indicating adaptation to continuity (B, C, col. 1–4). Preceding loud noise supported discontinuity ratings of the ambiguous target, indicating adaptation to noise level (C, col. 5,6). **E,** Mean differential continuity ratings of the probe \pm SE across listeners representing contrastive aftereffects of the three adaptor types in experiment 1 (left) and 2 (right). Adaptation to continuity (col. 2) affected adapted and non-adapted portions of the spectrum. The aftereffects of the noise (col. 3) indicate adaptation to loudness rather than to masking potential. The aftereffects of illusory sweeps (col. 1) were mainly explained by the aftereffects of non-illusory sweeps (col. 2), indicating adaptation to restoration. All ratings are plotted on the same four-point scales. disc., discontinuous; cont. continuous; middle tick marks, presumed point of bistability; asterisks, significance levels (0.05, 0.005, or 0.0005) corrected for multiple comparisons.

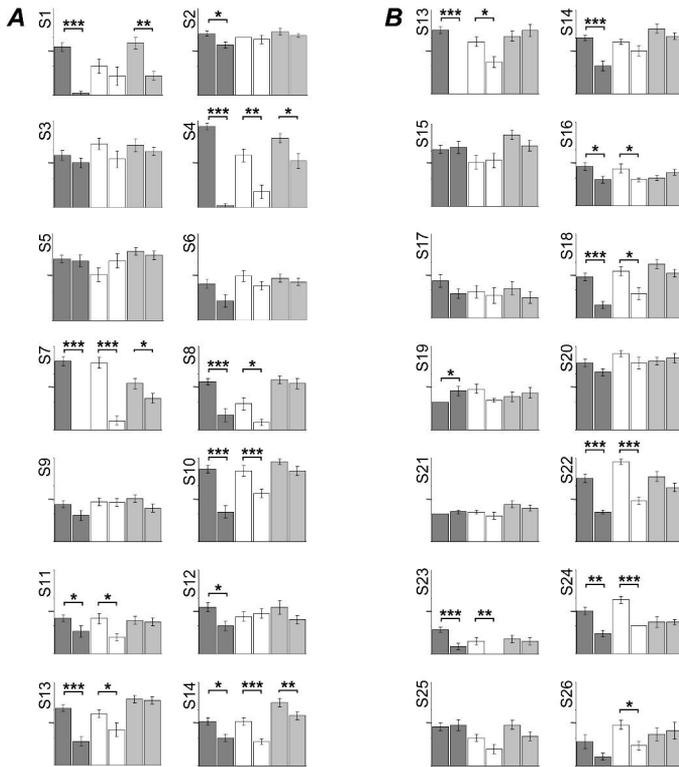


Figure 3. Results from single-subject analysis for experiment 1 and 2. **A, B,** Mean continuity ratings of the probe \pm SE across trial repetitions for all listeners in experiment 1 (**A**) and 2 (**B**). 15 out of 26 listeners showed consistent significant aftereffects of perceived sweep continuity (col. 1,2). Note that two listeners (S13, S14) participated in both experiments. For details, see Figure 2.

3.3 Probe continuity ratings

Group analysis of listeners' continuity ratings of the same ambiguous probe (Fig. 3A,B) revealed similar results for experiment 1 and 2 (Fig. 2C,D). The probe was rated as significantly different from the presumed point of bistability when the preceding adaptors comprised a sweep and were perceived as discontinuous (noise-interrupted sweeps: $t_{13} = -3.02$, $p < .01$ [exp. 1] and $t_{13} = -3.05$, $p < .009$ [exp. 2]; interrupted sweeps: $t_{13} = -2.23$, $p < .04$ [exp. 1] and $t_{13} = -2.78$, $p < .02$ [exp. 2]; Fig. 2C,D, col. 1–4). Regarding differences between the individual adaptation conditions, the type and level of the adaptors exhibited an interaction (exp. 1: $F_{2,12} = 8.16$, $p < .006$; exp. 2: $F_{2,12} = 4.87$, $p < .03$), and the adaptor level exhibited simple effects for noise-interrupted sweeps (exp. 1: $t_{13} = -4.63$, $p < .0005$; exp. 2: $t_{13} = -3.18$, $p < .007$; Fig. 2C,D, col. 1,2) and for sweeps alone (exp. 1: $t_{13} = -3.23$, $p < .007$; exp. 2: $t_{13} = -5.03$, $p < .0003$; Fig. 2C,D, col. 3,4). Notably, the effect of the adaptor level on the adaptor ratings (Fig. 2A,B) versus the subsequent probe ratings

(Fig. 2C,D) was reversed: Adaptation to discontinuous sweep percepts biased listeners to rate the probe as more illusorily continuous, and vice versa for adaptation to continuous sweep percepts (Fig. 2C,D, col. 1–4). Therefore, prior presentation of these stimuli induced contrastive aftereffects on how listeners rated the same ambiguous probe, consistent with the notion of adaptation (see Introduction). For the noise alone, the adaptor level exhibited simple effects on (exp. 1: $t_{13} = -4.20$, $p < .001$) or induced trends in (exp. 2: $t_{13} = -1.74$, $p = .11$; Fig. 2C,D, col. 5,6) the probe ratings, partially consistent with the aftereffects observed for the noise-interrupted sweeps. Adaptation to loud masking noise biased listeners to rate the ambiguous probe as more discontinuous than adaptation to soft unmasking noise (Fig. 2C, col. 5,6), whereas adaptation to masking or unmasking fixed-level noise did not affect these ratings significantly (Fig. 2D, col. 5,6).

3.4 Comparison of different aftereffects

To assess the effects of varying the adaptor spectra, the aftereffects of each adaptor type were compared across the two experiments (Fig. 2E). This analysis revealed that spectral changes in the sweep adaptor did not affect the aftereffects of sweep continuity ($t_{26} = 0.13$, $p = .90$), indicating that these aftereffects did not depend on which part of the spectrum was adapted. The aftereffects of the noise's masking potential were significantly stronger when the overall noise level covaried with the masking potential, compared to when the overall level remained fixed ($t_{26} = 2.27$, $p < .03$). A similar trend was observed when the noise interrupted the sweep ($t_{26} = 1.54$, $p = .13$). Therefore, the aftereffects of noise depended on the overall noise level rather than on its masking potential.

Aftereffects were generally stronger for noise-interrupted sweeps than for noise (exp. 1: $t_{13} = 3.68$, $p < .003$; exp. 2: $t_{13} = 2.32$, $p < .04$) or congruent sweeps alone (exp. 1: $t_{13} = 3.58$, $p < .003$). Linear regression was used to unravel the relation between the different aftereffects (Fig. 2E, col. 1–3). For experiment 1, the aftereffects of noise-interrupted sweeps were largely explained by a weighted sum of the individual aftereffects (sweeps: $t = 4.70$, $p < .001$; noise: $t = 2.92$, $p < .01$; model fit: $R^2 = .79$, $F_{2, 11} = 20.74$, $p < .0002$), whereas for experiment 2 only the aftereffects of sweeps were explanatory ($R^2 = .34$, $F_{1, 12} = 6.20$, $p < .03$). Therefore, the aftereffects of noise-interrupted sweeps were more related to the perceived continuity of the sweeps than to the acoustic properties of the noise. The aftereffects of sweeps and noise were not significantly correlated (exp. 1: Pearson's $R = .28$, $p = .17$; exp. 2: $R = -.06$, $p = .42$), suggesting that two adaptation mechanisms were differentially involved.

4 Discussion

We investigated whether the continuity of simple target sounds and the masking potential of noise influence illusory restorations of subsequent interrupted target sounds. Our results show that preceding sweeps and noise both determine the extent of illusory restoration. First, we observed that interrupted sweeps increased

the degree of subsequent restorations, whereas uninterrupted sweeps made restoration less likely. Second, loud noise reduced the extent of subsequent restorations, whereas soft noise increased their likelihood. These aftereffects are indicative of adaptation to continuity and noise level, respectively. We further investigated whether these two factors affect restorations of sounds that are spectrally incongruent with the adapting sounds. We found that continuity adaptation influenced the degree of restorations of non-adapted parts of the spectrum, and that the masking potential of interrupting noise had little impact. In summary, these results show that restoration adapts to the continuity and loudness of preceding sounds irrespectively of their frequency content.

Our conclusions are corroborated by two further observations. First, adaptation to restoration conditions (simultaneously presented sweeps and noise) revealed that continuity illusions reduced the extent of subsequent restorations, while true discontinuity percepts made restoration more likely. These aftereffects of preceding continuity illusions were largely explained by the aftereffects of true continuity percepts, indicating that listeners adapted to the restored rather than physical continuity. Second, the aftereffects of preceding continuity illusions further scaled with the loudness of the interrupting noise, suggesting that two adaptation mechanisms were differentially involved.

4.1 Previous studies on auditory adaptation

Recent studies on non-illusory phenomena have consistently reported that auditory adaptation applies to higher-level sound properties. It has been shown that listeners' judgments of the identity of a phoneme, the gender of a voice, or the spatial location of a tone depend on prior exposure to the respective sound categories (see Introduction). It was commonly observed that the adaptive effects were influenced slightly by preceding sound frequencies. Tones, for example, were reported to affect the localization of subsequent tones, even when the consecutive tone frequencies differed by two or four octaves (Frissen *et al.*, 2003, 2005). Similarly, gender categorizations of voices may be determined by preceding voices, but not by preceding tones of different frequencies (Schweinberger *et al.*, 2008). These results suggested that changes in auditory categorization may reflect adaptive changes in abstract, non-sensory sound representations. Our results consistently show that aftereffects on restoration are almost unaffected by the adapted sound frequencies, supporting the previous notion that non-sensory sound representations may have adapted.

There is also evidence that preceding sounds may influence the loudness of subsequent sounds (Marks, 1988; Mapes-Riordan & Yost, 1999; Scharf *et al.*, 2002). Adaptation to loud tones may cause loudness reductions of 10 dB or more in subsequent softer tones (Marks, 1993; Ariei & Marks, 2003a; Nieder *et al.*, 2003; Oberfeld, 2007), especially for adaptor levels of 60–80 dB (Mapes-Riordan & Yost, 1999) and for tones of similar frequency (Marks & Warner, 1991; Marks, 1994; Wagner & Scharf, 2006). Loudness reductions may emerge about 200 ms after the adapting sound and may persist for several seconds (Ariei & Marks, 2003b; Ariei *et*

al., 2005; Wagner & Scharf, 2006; Nieder *et al.*, 2007). These values are compatible with the sound levels and time scales in our study, and therefore loudness adaptation may account for the noise level aftereffects that we observed. Specifically, the loud noise could have reduced the loudness of subsequent noise, thereby enhancing the relative audibility of the partially masked gaps, and suppressing the restoration of the target.

4.2 Contextual factors influencing auditory restoration

A common view on auditory restoration states that the putative underlying mechanisms analyze incoming sounds for their relative spectrum levels, *i.e.* for whether the interrupting sound masks the foreground gaps. Depending on the gaps' relative audibility, the gaps may or may not be filled by the interrupting sound, which yields a foreground percept that can be categorized as either illusorily continuous or truly discontinuous (for review, see Bregman, 1990; Warren, 1999).

Our data demonstrate that the extent of restoration of an interrupted target sound depends on preceding continuity illusions and other factors which generally influence continuity perception, *i.e.* true target continuity and noise levels. The absence of frequency selectivity that we observed raises doubts on the possibility that adaptation modulated the sensory input to the restoration mechanisms. This finding, together with the observed adaptation to perceived continuity, supports the notion that the perceptual output of auditory restoration was affected by the preceding sounds. In other words, adaptation probably did not influence the processing of low-level acoustic properties, but the integration of the target and noise spectra, and the categorization of the resulting percept. Modulations of restoration may induce adaptive changes in the audibility of the continuity illusion, whereas modulations of categorization may cause adaptive shifts in an implicit continuity criterion.

Even though these factors cannot be fully disentangled from our data, explanations by continuity criterion shifts alone seem unlikely. First, larger criterion shifts would be expected after stronger continuity ratings, producing stronger bias towards discontinuity ratings. Our results are inconsistent with this prediction since truly continuous sweeps induced significantly weaker bias than illusory continuous sweeps, even though the former were rated significantly more continuous. Second, similar criterion shifts and response biases would be expected after similarly rated adaptors. This prediction is also incompatible with our data given the different biases induced by the indifferently rated noise adaptors. These aftereffects more likely reflected loudness adaptation (see previous paragraph) which has been shown to influence perceived sound levels rather than categorization criteria (Arieh & Marks, 2003a). Finally, debriefings indicated that our listeners were unaware of being repeatedly presented with the same ambiguous sound, and our impression was that adaptation influenced the ability to perceptually integrate the noise with the target.

In summary, these considerations suggest that listeners did not rate continuity simply based on previous button presses. It is possible that preceding sounds induced adaptive changes in restoration and thereby shifted listeners' continuity criteria, or vice versa. Preceding sounds may have altered schemas of auditory

continuity, thereby recalibrating listeners' perceptual interpretation of the ambiguous illusory sound (Bregman, 1990). Further experiments including appropriate control conditions and analysis in terms of signal detection may help disentangle these possibilities.

4.3 Potential neural mechanisms for perceived continuity adaptation

Previous research has shown that auditory restoration may involve activity in primary auditory cortical neurons which typically respond to gaps in sounds. The responses of such neurons may be suppressed when a noise facilitates restoration (Petkov *et al.*, 2007; for review see Recanzone & Sutter, 2008) and may further scale with changes in the illusory continuity of an interrupted sound (Riecke *et al.*, 2007). Other neurons in primary auditory cortex were shown to adapt for several seconds (Ulanovsky *et al.*, 2003, 2004; Bartlett & Wang, 2005; Werner-Reiss *et al.*, 2006) and alter their responses to consecutive tones, even when the adapted tone frequency differed by one octave (Bartlett & Wang, 2005).

It may be possible that neurons in primary auditory cortex change their responses also after continuity adaptation, thereby modulating the detectability of subsequent gaps. Furthermore, their output could be gated by top-down factors, such as by adaptive schemas of continuity, which bias the perceptual interpretation of ambiguous gaps. At present, these potential mechanisms remain speculative and neural correlates of the observed adaptation effects are yet to be determined. The present study provides a useful basis for controlling continuity illusions without confounding stimulus changes, ideally suited for neurophysiologic investigations of auditory restoration.

4.4 Conclusions

Auditory restoration seems to be a flexible phenomenon which adapts to foreground continuities and background levels in the perceived auditory scene. Ten seconds of adaptation are sufficient to affect restorations within the following five seconds. We propose that the tracking of sounds of interest during noisy interruptions may improve after hearing discontinuities. Potentially relevant interruptions, however, may become more salient after hearing ongoing foregrounds or loud noise. These enhancing effects may help optimizing continuity perception and facilitate useful hearing in different acoustic environments.

Acknowledgements

We are grateful to Gerard van Breukelen for help with the statistical analysis, and to Peter De Weerd, and Bettina Sorger for useful comments on an earlier version of the script.

References

- Arieh Y, Marks LE. (2003a). Time course of loudness recalibration: Implications for loudness enhancement. *J Acoust Soc Am*, 114, 1550–1556.
- Arieh Y, Marks LE. (2003b). Recalibrating the auditory system: A speed-accuracy analysis of intensity perception. *J Exp Psychol Hum Percept Perform*, 29, 523–536.
- Arieh Y, Kelly K, Marks LE. (2005). Tracking the time to recovery after induced loudness reduction. *J Acoust Soc Am*, 117, 3381–3384.
- Barlett EL, Wang X. (2005). Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol*, 94, 83–104.
- Bashford JA Jr, Riener KR, Warren RM. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Percept Psychophys*, 51, 211–217.
- Bregman AS, Dannenbring GL. (1977). Auditory continuity and amplitude edges. *Can J Psychol*, 31, 151–159.
- Bregman AS. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge.
- Carlyon RP, Beveridge HA. (1993). Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise. *J Acoust Soc Am*, 93, 2886–2895.
- Ciocca V, Bregman A S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Percept Psychophys*, 42, 476–484.
- Cooper WE. (1974). Adaptation of phonetic feature analyzers for place of articulation. *J Acoust Soc Am*, 56, 617–627.
- Dannenbring GL. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Can J Psychol*, 30, 99–114.
- Diehl RL, Elman JL, McCusker SB. (1978). Contrast effects on stop consonant identification. *J Exp Psychol Hum Percept Perform*, 4, 599–609.
- Eimas PD, Corbit JD. (1973). Selective adaptation of linguistic feature detectors. *Cognit Psychol*, 4, 99–109.
- Fechner G. (1860). *Elements of Psychophysics*. Holt, Rinehart, and Winston, New York.
- Fletcher H. (1940). Auditory patterns. *Rev Modern Physics*, 12, 47–65.
- Frissen I, Vroomen J, de Gelder B, Bertelson P. (2003). The aftereffects of ventriloquism: are they sound-frequency specific? *Acta Psychol*, 113, 315–327.
- Frissen I, Vroomen J, de Gelder B, Bertelson P. (2005). The aftereffects of ventriloquism: generalization across sound-frequencies. *Acta Psychol*, 118, 93–100.
- Green GGR, Kay RH. (1973). The adequate stimuli for channels in the human auditory pathways concerned with the modulation present in frequency-modulated tones. *J Physiol*, 234, 50–52.
- Green GGR, Kay RH. (1974). Channels in the human auditory pathways concerned with the waveform of the modulation present in amplitude- and frequency-modulated tones. *J Physiol*, 241, 29–30.
- Houtgast T. (1972). Psychophysical evidence for lateral inhibition in hearing. *J Acoust Soc Am*, 51, 1885–1894.
- Kay RH, Matthews DR. (1972). On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. *J Physiol*, 225, 657–677.
- Kidd G Jr, Feth LL. (1982). Effects of masker duration in pure-tone forward masking. *J Acoust Soc Am*, 72, 1384–1386.
- Landahl KL, Blumstein SE. (1982). Acoustic invariance and the perception of place of articulation: a selective adaptation study. *J Acoust Soc Am*, 71, 1234–1241.
- Mapes-Riordan D, Yost WA. (1999). Loudness recalibration as a function of level. *J Acoust Soc Am*, 106, 3506–3511.
- Marks LE. (1988). Magnitude estimation and sensory matching. *Percept Psychophys*, 43 511–525.
- Marks LE, Warner E. (1991). Slippery context effect and critical bands. *J Exp Psychol Hum Percept Perform*, 17, 986–996.

- Marks LE. (1993). Contextual processing of multidimensional and unidimensional auditory stimuli. *J Exp Psychol Hum Percept Perform*, 19, 227–249.
- Marks LE. (1994). 'Recalibrating' the auditory system: The perception of loudness. *J Exp Psychol Hum Percept Perform*, 20, 382–396.
- Miller GA, Licklider JCR. (1950). The intelligibility of interrupted speech. *J Acoust Soc Am*, 22, 167–173.
- Miller CT, Dibble E, Hauser MD. (2001). Amodal completion of acoustic signals by a nonhuman primate. *Nat Neurosci*, 4, 783–784.
- Moody DB, Cole D, Davidson LM, Stebbins WC. (1984). Evidence for a reappraisal of the psychophysical selective adaptation paradigm. *J Acoust Soc Am*, 76, 1076–1079.
- Nakajima Y, Sasaki T, Kanafuka K, Miyamoto A, Remijn G, ten Hoopen G. (2000). Illusory recouplings of onsets and terminations of glide tone components. *Percept Psychophys*, 62, 1413–1425.
- Nieder B, Buus S, Florentine M, Scharf B. (2003). Interactions between test-and inducer-tone durations in induced loudness reduction. *J Acoust Soc Am*, 114, 2846–2855.
- Nieder B, Buus S, Cazals Y, Scharf B. (2007). Loudness reduction induced by a contralateral tone. *J Acoust Soc Am*, 122, 35–37.
- Oberfeld D. (2007). Loudness changes induced by a proximal sound: loudness enhancement, loudness recalibration, or both? *J Acoust Soc Am*, 121, 2137–2148.
- Oberfeld D. (2008). The mid-difference hump in forward-masked intensity discrimination. *J Acoust Soc Am*, 123, 1571–1581.
- Ohde RN., Sharf DJ. (1979). Relationship between adaptation and the percept and transformations of stop consonant voicing: effects of the number of repetitions and intensity of adaptors. *J Acoust Soc Am*, 66, 30–45.
- Penner MJ. (1974). Effect of masker duration and masker level on forward and backward masking. *J Acoust Soc Am*, 56, 179–182.
- Petkov CI, O'Connor KN, Sutter ML. (2003). Illusory sound perception in macaque monkeys. *J Neurosci*, 23, 9155–9161.
- Petkov CI, O'Connor KN, Sutter ML. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54, 153–165.
- Phillips DP, Hall SE. (2005). Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level. *Hear Res*, 202, 188–199.
- Plack CJ, Carlyon RP, Viemeister NF. (1995). Intensity discrimination under forward and backward masking: role of referential coding. *J Acoust Soc Am*, 97, 1141–1149.
- Plack CJ. (1996). Loudness enhancement and intensity discrimination under forward and backward masking. *J Acoust Soc Am*, 100, 1024–1030.
- Powers GL, Wilcox JC. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *J Acoust Soc Am*, 61, 195–199.
- Recanzone GH, Sutter ML. (2008). The biological basis of audition. *Annu Rev Psychol*, 59, 119–142.
- Regan D, Tansley B. (1979). Selective adaptation to frequency-modulated tones: Evidence for an information-processing channel selective to frequency changes. *J Acoust Soc Am*, 65, 1249–1257.
- Remijn GB, Nakajima Y, ten Hoopen G. (2001). Continuity perception in stimulus patterns consisting of two partly overlapping frequency glides. *J Music Percept Cogn*, 7, 77–91.
- Remijn GB, Nakajima Y. (2005). The perceptual integration of auditory stimulus edges: An illusory short tone in stimulus patterns consisting of two partly overlapping glides. *J Exp Psychol Hum Percept Perform*, 31, 183–192.
- Riecke L, Van Opstal AJ, Formisano E. (2008). The auditory continuity illusion: A parametric investigation and filter model. *Percept Psychophys*, 70, 1–12.
- Riecke L, Van Opstal AJ, Goebel R, Formisano E. (2007). Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *J Neurosci*, 27, 12684–12689.
- Rosenblith WA, Miller GA, Egan JP, Hirsh IJ, Thomas GJ. (1947). An auditory afterimage? *Science*, 106, 333–335.
- Samuel AG, Newport EL. (1979). Adaptation of speech by nonspeech: evidence for complex acoustic cue detectors. *J Exp Psychol Hum Percept Perform*, 5, 563–578.

- Sawusch JR, Jusczyk P. (1981). Adaptation and contrast in the perception of voicing. *J Exp Psychol Hum Percept Perform*, 7, 408–421.
- Scharf B, Buus S, Nieder B. (2002). Loudness enhancement: induced loudness reduction in disguise? *J Acoust Soc Am*, 112, 807–810.
- Schweinberger SR, Casper C, Hauthal N, Kaufmann JM, Kawahara H, Kloth N, Robertson DM, Simpson AP, Zäske R. (2008). Auditory adaptation in voice perception. *Curr Biol*, 18, 684–688.
- Simon HJ, Studdert-Kennedy M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. *J Acoust Soc Am*, 64, 1338–1357.
- Tansley BW, Suffield JB. (1983). Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. *J Acoust Soc Am*, 74, 765–775.
- Thurlow W. (1957). An auditory figure-ground effect. *Am J Psychol*, 70, 653–654.
- Ulanovsky N, Las L, Nelken I. (2003). Processing of low-probability sounds by cortical neurons. *Nat Neurosci*, 6, 391–398.
- Ulanovsky N, Las L, Farkas D, Nelken I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci*, 24, 10440–10453.
- Viemeister NF, Bacon SP. (1982). Forward masking by enhanced components in harmonic complexes. *J Acoust Soc Am*, 71, 1502–1507.
- Wagner E, Scharf B. (2006). Induced loudness reduction as a function of exposure time and signal frequency. *J Acoust Soc Am*, 119, 1012–1020.
- Warren RM. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.
- Warren RM, Obusek CJ, Ackroff JM. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149–1151.
- Warren RM. (1983). Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception. *J Audio Eng Soc*, 31, 623–629.
- Warren RM. (1984). Perceptual restoration of obliterated sounds. *Psychol Bull*, 96, 371–383.
- Warren RM, Bashford JA, Healy EW, Brubaker BS. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Percept Psychophys*, 55, 313–322.
- Warren RM. (1999). *Auditory Perception: A new Analysis and Synthesis*. Cambridge University Press, Cambridge.
- Werner-Reiss U, Porter KK, Underhill AM, Groh JM. (2006). Long lasting attenuation by prior sounds in auditory cortex of awake primates. *Exp Brain Res*, 168, 272–276.
- Wojtczak M, Viemeister NF. (2003). Suprathreshold effects of adaptation produced by amplitude modulation. *J Acoust Soc Am*, 114, 991–997.
- Wojtczak M, Viemeister NF. (2005). Forward masking of amplitude modulation: Basic characteristics. *J Acoust Soc Am*, 118, 3198–3210.
- Wright BA, McFadden D, Champlin CA. (1993). Adaptation of suppression as an explanation of enhancement effects. *J Acoust Soc Am*, 94, 72–82.
- Wrightson JM, Warren RM. (1981). Incomplete auditory induction of tones alternated with noise: Effects occurring below the pulsation threshold. *J Acoust Soc Am*, 69, S105–106.
- Zeng FG, Turner CW. (1992). Intensity discrimination in forward masking. *J Acoust Soc Am*, 92, 782–787.
- Zeng FG, Turner CW, Relkin EM. (1991). Recovery from prior stimulation. II: Effects upon intensity discrimination. *Hear Res*, 55, 223–230.
- Zeng FG, Shannon RV. (1995). Possible origins of the non-monotonic intensity discrimination function in forward masking. *Hear Res*, 82, 216–224.
- Zwicker E. (1964). 'Negative afterimage' in hearing. *J Acoust Soc Am*, 36, 2413–2415.

Chapter 7

General discussion

The studies presented in this thesis addressed the question of how the brain constructs smooth and meaningful sound percepts from noisy acoustic stimuli. The underlying mechanisms were investigated by means of the continuity illusion, a phenomenon that facilitates the perceptual restoration of noise-interrupted sounds. While the obtained results have been discussed in detail in the respective chapters, the present chapter aims at collectively interpreting the findings in a more general and speculative context. Specifically, bottom-up and top-down influences on the continuity illusion are discussed, and a neural model for the continuity illusion is proposed based on the present and previous evidence.

1 Bottom-up and top-down influences on the continuity illusion

The neurophysiologic and psychoacoustic studies presented in this thesis generally indicate that continuity illusions depend not only on acoustic stimuli but also on cognitive factors. The most robust finding was that continuity illusions are associated with decreased auditory cortical activities compared to gap percepts (Riecke *et al.*, 2007; chapter four & five). Despite its consistency with previous findings (Micheyl *et al.*, 2003; Petkov *et al.*, 2007), this result may be surprising given that the evoked illusions required perceptual fillings of spectrotemporal gaps spanning several semitones and hundreds of milliseconds. One may expect that such illusions ‘fill in’ neural activities and therefore induce increased metabolic demands. Contrary to this expectation, the present data collectively indicate that neural activities in auditory cortex (AC) increase with the number of acoustic edges in the sensory input. Furthermore, these activities can be modulated by listeners’ interpretation of these edges (Riecke *et al.*, 2007; chapter four). This suggests that attention played a role in the perceptual analysis of the sensory input.

1.1 The role of acoustic edges for the continuity illusion

Acoustic edges can be understood as energy transients along the time-frequency axes in the sound spectrogram. They are produced by spectrotemporal gaps in a sound and mark the boundaries of auditory events (Fishbach *et al.*, 2001). The present results show that the continuity illusion depends on the integration of auditory events comprising proximate edges (Riecke *et al.*, 2008), consistent with previous results (Ciocca & Bregman, 1987; Remijn *et al.*, 2001; Remijn & Nakajima, 2005). Previous neurophysiologic studies have reported that neurons at several sites along the auditory pathway respond vigorously to sound gaps (for reviews, see Phillips *et al.*, 2002; Recanzone & Sutter, 2008). These and other findings have led to proposals of cognitive models for edge-based perceptual grouping (Bregman, 1990; Nakajima *et al.*, 2000) and neural models for edge detection (Fishbach *et al.*, 2001; Shamma, 2001), all of which consider acoustic edges as crucial for the formation of sound objects. As previously stated, the present results indicate increased metabolic demands in AC associated with the processing of acoustic edges in the sensory input. Together with the previous results this implies that neural processing in AC primarily reflects the disentanglement of different sound objects in the auditory scene.

1.2 The role of attention for the continuity illusion

Previous research has suggested that selective attention plays an important role in auditory grouping phenomena such as auditory streaming. It has been proposed that attention may enhance the segregation of concurrent sound streams (for reviews see Carlyon, 2004; Alain & Bernstein, 2008). The present data on continuity illusions generally support this view: The result that illusions evoke decreased auditory cortical activities compared to gap percepts, for example, suggests that the grouping of interrupted sounds with noise, as represented in AC, reflects a relatively effortless mode of stimulus interpretation. This notion is further corroborated by previous EEG findings showing that continuity illusions may occur even when listeners are not attentive (Micheyl *et al.*, 2003). In line with the previous considerations (see section 1.1. this chapter) this 'default' mode of auditory organization may imply that during continuity illusions only little processing resources are allocated to neural representations of acoustic edges.

Nevertheless, the continuity illusion may also be sensitive to volitional control as suggested by the results presented in chapter three to six. It was observed that subjective reports of the perceived continuity of the same sounds are paralleled by hemodynamic activities in auditory cortical regions that respond strongly to acoustic gaps in tones (Riecke *et al.*, 2007; chapter five). Moreover, spontaneous switches from reports of true discontinuity to continuity illusions of the same stimuli were found to decrease EEG activities related to acoustic gaps (chapter four). Finally, contextual factors were found to influence continuity illusions of the same stimuli (Riecke *et al.*, 2009). All these changes in continuity judgments were unrelated to concurrent sensory changes and, therefore, they could be ascribed to non-sensory changes such as fluctuations in listeners' attention. Such fluctuations possibly reflect changes in listeners' perceptual expectancies or schemas which are crucial for the continuity illusion (Bregman, 1990). It is thus conceivable that attention modulates the mechanisms that are responsible for the continuity illusion.

Attention indeed plays an important role in the detection of temporal and spectral sound gaps as shown by previous studies using non-illusory stimuli. Specifically, selective attention to temporal gaps may 'sharpen' the receptive fields of auditory cortical neurons responding to these gaps (Fritz *et al.*, 2007). Furthermore, selective attention can enhance the sensory segregation between tones and notched noise, as indexed by modulations in event-related potentials originating from AC (Okamoto *et al.*, 2007). Therefore, the present and previous results suggest that attention may modulate the neural representation of acoustic edges in AC and thus the perception of different sound objects.

Conclusively it appears that sound processing in the AC has a bias towards integrating spectrotemporally proximate sound fragments into smooth and possibly illusory sound percepts. Mental efforts to attend to gaps may enhance acoustic details of auditory stimuli by sharpening neural filters in AC involved in the integration process (Fritz *et al.*, 2007; see also Riecke *et al.*, 2008). This 'zooming in' may improve the detection of acoustic edges and thus the identification of veridical sound object boundaries.

2 Potential neural mechanisms for the continuity illusion

The present findings provide new evidence for the neural basis of the continuity illusion. According to Bregman's (1990) classical view on auditory scene analysis, the continuity illusion emerges from two successive processing stages. The first stage may include early automatic processes that extract sensory features from acoustic input. The second stage may involve voluntary processes that match the extracted sensory evidence with perceptual expectancies or schemas. Since its proposal this cognitive model has received considerable interest among auditory researchers (for a recent review see Ciocca, 2008), but yet an account for neural mechanisms underlying continuity illusions is missing. The present section attempts to integrate the present and previous neurophysiologic findings with Bregman's framework. A central aspect of the neural model proposed here (Fig. 1) is the dynamic processing of sensory gaps in stimulus representations in AC.

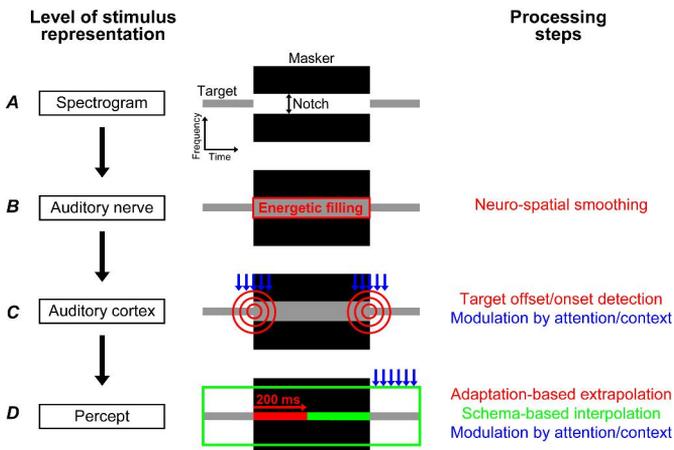


Figure 1. Sketch of a neural model for continuity illusions. **A**, An interruption in a target sound (gray line) is partially masked by noise (black rectangle) comprising a spectral notch (white rectangle). **B**, Peripheral auditory filters fill in energy in the omitted critical target band (gray rectangle), smearing the sensory stimulus representation. **C**, Auditory cortical neurons analyze the filled target band representation for offsets and onsets. The sensitivity of these neural edge detectors (red circles) is modulated by attention and the acoustic context (blue arrows). **D**, When no target offset is detected, neural adaptation facilitates buffering of the fading target sound for about 200 ms (red arrow), which is experienced as a temporary perceptual extension of the target into the masker (red line). Auditory cortical neurons comprising large complex spectrotemporal receptive fields (green rectangle) match the fading target representation with learned perceptual patterns after the gap, which may lead to perceptual interpolation of the interrupted target sound (green line). These neurons are modulated by attention and the acoustic context (blue arrows).

2.1 Detection of sensory edges in cortical sound representations

When a gap in a target sound is partially masked by noise (Fig. 1A), the target and the noise may stimulate the same fibers in the auditory nerve (Delgutte, 1990). This may evoke uninterrupted excitations in peripheral neurons which 'smear' the sensory

stimulus representation (Warren, 1970; Warren *et al.*, 1972; for review, see Warren, 1999; Fig. 1B, red rectangle). The first cortical processing stage for the continuity illusion could be implemented by neurons in AC that respond strongly to energy transients, so-called neural 'onset/offset detectors' (for reviews see Phillips *et al.*, 2002; Recanzone & Sutter, 2008). These onset/offset detectors may analyze the sensory target representation for the presence of acoustic edges in order to determine boundaries of sound objects (Fishbach *et al.*, 2001; Herdener *et al.*, 2007; Fig. 1C, red circles). The energy changes induced by the onset of masking noise may excite large neural populations of onset detectors (Petkov *et al.*, 2007) that subsequently synchronize their output. These synchronized neural activities could give rise to large electric field potentials within 130 ms (chapter four), signaling the onset of a new concurrent object (Alain *et al.*, 2001).

When the gap is unmasked, these neural responses are accompanied by additional responses of neurons detecting the concurrent target offset (Petkov *et al.*, 2007; chapter four). Such offset detectors could operate on target sound representations in brain regions sensitive for particular perceptual properties. For tones, for example, neurons in primary AC may be involved while for more complex sounds, such as voices, neurons in voice-encoding regions may be recruited (Riecke *et al.*, 2007; Heinrich *et al.*, 2008; chapter five). The sensitivity of the offset detectors in primary AC may be modulated by top-down signals from higher cortical regions involved in allocating attention to the sound gap (see section 1.1. this chapter) or possibly by contextual factors (Riecke *et al.*, 2009; Fig. 1C, blue arrows). These modulations could be communicated via neural oscillations in the theta range (chapter four) and induce adaptive receptive field changes in auditory cortical neurons representing the gap (Fritz *et al.*, 2007).

When the gap is masked, however, an offset is not evident in the cortical representation of the target sound. Consequently, auditory neurons may fail to detect the gap and suppress their offset-responses (Petkov *et al.*, 2007; Riecke *et al.*, 2007; chapter four, five). The interrupted sound may then be represented as uninterrupted such that the sustained energy of the noise in the target band can be attributed to the target sound. In this case the target can be perceptually extrapolated into the noise (Bregman, 1990; Fig. 1D, red line). This short-lived extrapolation may occur pre-attentively (Micheyl *et al.*, 2003) and could be mediated by neurons in AC sensitive to sustained sensory input (Petkov *et al.*, 2007), or by mechanisms of neural adaptation (Ulanovsky *et al.*, 2003, 2004) or forward suppression (Fishman *et al.*, 2001, 2004; Wehr & Zador, 2005; for review see Micheyl *et al.*, 2007). These presumably automatic mechanisms may sample the abstract acoustic regularities in the cortical stimulus representation in intervals of 200 ms and facilitate the detection of unexpected acoustic patterns (Näätänen *et al.*, 2001). During gaps longer than 200 ms, the buffered target representation may be gradually 'overwritten' by more recent sensory input (*i.e.* by the ongoing noise) which may be accompanied by the fading of the target percept (Wrightson & Warren, 1981; Warren *et al.*, 1994).

2.2 Matching of cortical sound feature representations with stored perceptual schemas

While the first cortical processing stage for continuity illusions may be fast and automatic, the second stage may be slower and more sensitive to cognitive control. At this stage, the target sound may be perceptually interpolated based on neural processes that match the faded target representation with stored schemas of auditory organization (Bregman, 1990; see also Griffiths & Warren, 2002). Successful matching requires that the spectral trajectory of the post-masker target continues that of the pre-masker target consistent with learned auditory patterns or perceptual expectancies (Ciocca & Bregman, 1987). The matching process may thus occur after the gap, provided that neurons in AC fail to detect the target re-onset after the gap. This process requires a perceptual buffer of relatively large capacity (Cowan, 1984) that could be implemented by auditory cortical filters interpolating sensory input across large, frequency-specific spectrotemporal windows (Husain *et al.*, 2005; Riecke *et al.*, 2008; Fig. 1D, green rectangle). The matching process may be modulated by top-down factors such as auditory attention (Bregman, 1990), or possibly by bottom-up factors such as the auditory context (Riecke *et al.*, 2009; Fig. 1D, blue arrows). Successful matching may activate primary auditory cortical representations of the interrupted target band and result in the perceptual restoration of the overall target (Riecke *et al.*, 2007; chapter five; Fig. 1D, green line). Finally, the matched output may be transmitted to cortical regions involved in decision-making or sensory-motor transformation via neural oscillations in the beta range (chapter four).

2.3 Conclusions

The proposed neural model explains how different processing stages in the AC could analyze fragmented noisy input for continuities. While the first processing stage relies on neural edge detectors that adapt their output to current attention demands, the second stage involves neural filters that integrate sensory edges across complex, possibly sound category-specific, time-frequency windows. The neural model extends the prevailing cognitive model on continuity illusions (Bregman, 1990) and generates new testable predictions which can be readily addressed in future research. For example, conditions in which participants attend to gaps, or imagine a target sound in the same illusory stimulus, would be expected to differentially activate auditory cortical regions involved in gap processing. Moreover, the proposed framework could inform future developments of neural network models (e.g. Husain *et al.*, 2005) or artificial hearing devices. To support smooth hearing in noisy environments, for example, algorithms implemented in such devices should not filter out noise in acoustic input but exploit it for restoration purposes. Feedback loops from previous acoustic input should be included to adapt the algorithms' sensitivity to different levels of acoustic energy in sound gaps. The illusory filter proposed in chapter two could serve as a basic processing unit in this regard.

References

- Alain C, Arnott SR, Picton TW. (2001). Bottom-Up and Top-Down Influences on Auditory Scene Analysis: Evidence From Event-Related Brain Potentials. *J Exp Psychol Hum Percept Perform*, 27, 1072–1089.
- Alain C, Bernstein LJ. (2008). From sounds to meaning: the role of attention during auditory scene analysis. *Curr Opin Otolaryngol Head Neck Surg*, 16, 485–489.
- Bregman AS. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge.
- Carlyon RP. (2004). How the brain separates sounds. *Trends Cogn Sci*, 8, 465–471.
- Ciocca V, Bregman AS. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Percept Psychophys*, 42, 476–484.
- Ciocca V. (2008). The auditory organization of complex sounds. *Front Biosci*, 13, 148–169.
- Cowan N. (1984). On short and long auditory stores. *Psychol Bull*, 96, 341–370.
- Delgutte B. (1990). Physiological mechanisms of psychophysical masking: observations from auditory-nerve fibers. *J Acoust Soc Am*, 87, 791–809.
- Fishbach A, Nelken I, Yeshurun Y. (2001). Auditory edge detection: A neural model for physiological and psychophysical responses to amplitude transients. *J Neurophysiol*, 85, 2303–2323.
- Fishman YI, Reser DH, Arezzo JC, Steinschneider M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res*, 151, 167–187.
- Fishman YI, Arezzo JC, Steinschneider M. (2004). Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am*, 116, 1656–1670.
- Fritz JB, Elhilali M, David SV, Shamma S. (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear Res*, 229, 186–203.
- Heinrich A, Carlyon RP, Davis MH, Johnsrude IS. (2008). Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. *J Cogn Neurosci*, 20, 1737–1752.
- Herdener M, Esposito F, Di Salle F, Lehmann C, Bach DR, Scheffler K, Seifritz E. (2007). BOLD correlates of edge detection in human auditory cortex. *Neuroimage*, 36, 194–201.
- Husain FT, Lozito TP, Ulloa A, Horwitz B. (2005). Investigating the neural basis of the auditory continuity illusion. *J Cogn Neurosci*, 17, 1275–1292.
- Micheyl C, Carlyon RP, Shtyrov Y, Hauk O, Dodson T, Pullvermüller F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *J Cogn Neurosci*, 15, 747–758.
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends Neurosci*, 24, 283–288.
- Nakajima Y, Sasaki T, Kanafuka K, Miyamoto A, Remijn G, ten Hoopen G. (2000). Illusory recouplings of onsets and terminations of glide tone components. *Percept Psychophys*, 62, 1413–1425.
- Okamoto H, Stracke H, Wolters CH, Schmael F, Pantev C. (2007). Attention improves population-level frequency tuning in human auditory cortex. *J Neurosci*, 27, 10383–10390.
- Petkov CI, O'Connor KN, Sutter ML. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54, 153–165.
- Phillips DP, Hall SE, Boehnke SE. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hear Res*, 167, 192–205.
- Recanzone GH, Sutter ML. (2008). The biological basis of audition. *Annu Rev Psychol*, 59, 119–142.
- Remijn GB, Nakajima Y, ten Hoopen G. (2001). Continuity perception in stimulus patterns consisting of two partly overlapping frequency glides. *J Music Percept Cogn*, 7, 77–91.
- Remijn GB, Nakajima Y. (2005). The perceptual integration of auditory stimulus edges: An illusory short tone in stimulus patterns consisting of two partly overlapping glides. *J Exp Psychol Hum Percept Perform*, 3, 183–192.
- Riecke L, Van Opstal AJ, Goebel R, Formisano E. (2007). Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *J Neurosci*, 27, 12684–12689.
- Riecke L, Van Opstal AJ, Formisano E. (2008). The auditory continuity illusion: A parametric investigation and filter model. *Percept Psychophys*, 70, 1–12.

Chapter 7

- Riecke L, Mendelsohn D, Schreiner C, Formisano E. (2009). The continuity illusion adapts to the auditory scene. *Hear Res*, 247, 71–77.
- Shamma S. (2001). On the role of space and time in auditory processing. *Trends Cogn Sci*, 5, 340–348.
- Ulanovsky N, Las L, Nelken I. (2003). Processing of low-probability sounds by cortical neurons. *Nat Neurosci*, 6, 391–398.
- Ulanovsky N, Las L, Farkas D, Nelken I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci*, 24, 10440–10453.
- Warren RM. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.
- Warren RM, Obusek CJ, Ackroff JM. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149–1151.
- Warren RM, Bashford JA, Healy EW, Brubaker BS. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Percept Psychophys*, 55, 313–322.
- Warren RM. (1999). *Auditory Perception: A new Analysis and Synthesis*. Cambridge, UK: Cambridge University Press.
- Wehr M, Zador AM. (2005). Synaptic mechanisms of forward suppression in rat auditory cortex. *Neuron*, 47, 437–445.
- Wrightson JM, Warren RM. (1981). Incomplete auditory induction of tones alternated with noise: Effects occurring below the pulsation threshold. *J Acoust Soc Am*, 69, S105–106.

Summary

The aim of this thesis is to broaden our understanding of the ability to track sounds in noisy environments. The general question addressed in the different chapters is: How does the auditory system in the brain construct smooth and meaningful sound percepts from noisy acoustic stimuli? The mechanisms underlying this phenomenon were investigated by means of the continuity illusion in which an interrupted sound is illusorily perceived as continuing through noise. The continuity illusion is a graded perceptual state which requires that acoustic energy of the noise masks the interruption.

A psychoacoustic study presented in *chapter two* investigated how the different acoustic parameters of noise influence the continuity illusion. The results revealed that continuity illusions of interrupted tones tend to be stronger when the noise 1) has shorter duration and 2) is a stronger masker. These effects on perceived continuity partially depended on the frequency of the tone, and were not evident for physically uninterrupted tones. Based on estimated thresholds for the continuity illusion, a neural filter model was adapted to describe the interaction of the different masking parameters that affect the continuity illusion. The estimated widths of the proposed 'illusory' filters suggested the involvement of constructive perceptual mechanisms that may operate on the output of peripheral masking mechanisms, especially at high sound frequencies. These findings set constraints on the spectral resolution of the mechanisms underlying the continuity illusion, and provide a stimulus set that can be readily applied for investigation of neural correlates.

A functional magnetic resonance imaging (fMRI) study presented in *chapter three* aimed at localizing the proposed constructive neural mechanisms. Using stimuli that were fine-tuned in the previous study, it was found that the masking of gaps in interrupted tones by noise is accompanied by decreased activities in auditory cortical regions. Since no such effect of masking was observed for physically uninterrupted tones, the observed decreases could be ascribed to the filling of gaps by acoustic energy. Importantly, listeners' subjective experience of continuity illusions was paralleled by activities in primary auditory cortex. These findings show that 1) the amount of acoustic energy in tone gaps is processed in early auditory cortex and 2) the subjective experience of continuity illusions of tones is represented in primary auditory areas.

An electroencephalography (EEG) study presented in *chapter four* investigated the relative neural timing of the acoustic-perceptual processes localized in the previous fMRI study. Using the same stimuli and task as before, it was found that 1) increased acoustic energy in tone gaps decreases EEG activities about 130 ms after gap onset and 2) continuity illusions reduce oscillatory EEG activities during and after the gaps. The results indicated that cognitive factors such as attention may influence the sensory processing of acoustic energy during the interruption. These findings set

constraints on the timing of cortical processes involved in the acoustical analysis and perceptual interpretation of noise-interrupted sounds.

A preliminary fMRI study presented in *chapter five* aimed at localizing neural mechanisms for continuity illusions of complex natural sounds. Using voice sounds as stimuli and the same methods as in the previous fMRI study, it was observed that the masking of interrupted voices by noise evokes decreased activities in auditory cortical regions that may be sensitive to speech sounds. Listeners' reports of voice continuity illusions were reflected by activities in primary auditory cortex, consistent with the previous findings on tone continuity illusions. These initial results suggest that 1) the amount of acoustic energy in voice gaps is processed in speech-sensitive cortical regions and 2) voice continuity illusions recruit primary auditory cortical regions that process continuity illusions of the frequency components constituting the voice sound.

A psychoacoustic study presented in *chapter six* investigated whether the continuity illusion can be influenced by contextual factors. The results revealed that continuity illusions of the same ambiguous tone glide are modulated by the loudness and the perceived continuity of preceding sounds. These context effects were unrelated to the sound spectrum of the preceding sounds, indicating non-sensory adaptive influences. The findings revealed that the mechanisms underlying continuity illusions are non-static which may serve to recalibrate continuity hearing to different acoustic environments.

In summary, the findings obtained in these studies show that the construction of smooth sound percepts from noisy acoustic stimuli involves not only static acoustic mechanisms, but also active and adaptive perceptual mechanisms in the brain. Primary auditory cortical regions play a dominant role in this regard, since they process the sound energy in noisy interruptions that is crucial for the continuity illusion. These early sensory-perceptual processes are further modulated by listeners' attention and likely also by the sensory context. Based on the accumulating evidence obtained in these and other studies, a neural model for the continuity illusion is proposed in *chapter seven*. For future research, an implementation of the proposed mechanisms in artificial hearing devices may be envisioned, wherein the illusory filter proposed in chapter two could serve as a basic processing unit.

Samenvatting

Deze dissertatie tracht de kennis te verruimen over het mechanisme dat ons in staat stelt geluiden waar te nemen die verstoord zijn door akoestische ruis. De hoofdvraagstelling is: Hoe is het auditieve systeem in staat een betekenis te geven aan een geluid dat verstoord wordt door verschillende variaties van akoestische ruis? Het mechanisme dat aan dit fenomeen ten grondslag ligt werd onderzocht met behulp van de continuïteitsillusie. Bij deze illusie wordt een geluid, ondanks een onderbreking door akoestische ruis, als een continu ervaren. Deze illusie is een perceptuele ervaring die geleidelijk gebeurt, en die vereist dat de onderbreking gemaskeerd wordt door de akoestische energie van ruis.

In *hoofdstuk twee* wordt een psychoakoestische studie gepresenteerd waarin onderzocht werd op welke manier de verschillende akoestische parameters van akoestische ruis de continuïteitsillusie beïnvloeden. De resultaten onthullen dat continuïteitsillusies van tonen sterker zijn wanneer de akoestische ruis 1) een korte duur heeft en 2) een sterk maskerende eigenschap heeft. De waargenomen continuïteit hing gedeeltelijk af van de frequentie van de toon. Het effect van de ruis was echter afwezig als de gepresenteerde tonen fysisch continue waren. Gebaseerd op de geschatte drempelwaarde van de illusie werd een neuronaal filtermodel voorgesteld. Dit model beschrijft hoe de continuïteitsillusie beïnvloed wordt door de verschillende parameters van akoestische ruis en hun onderlinge interacties. De geschatte breedte van de voorgestelde 'illusionaire' filters suggereert dat een constructief proces verantwoordelijk is dat zich baseert op de output van de perifere maskeringmechanismen. Dit effect is groter bij geluiden die hogere frequenties bevatten. Deze bevindingen bepalen de spectrale resolutie van de mechanismen die aan de continuïteitsillusie ten grondslag liggen. Hiernaast verschaffen ze akoestische stimuli voor onderzoek naar de neuronale correlaten van de continuïteitsillusie.

De functionele magnetische resonantie tomografie (fMRT) studie die in *hoofdstuk drie* gepresenteerd wordt was gericht op het lokaliseren van de neuronale mechanismen van de continuïteitsillusie. Er werd gebruik gemaakt van de akoestische stimuli die in de voorafgaande studie geoptimaliseerd werden. Een conclusie van deze studie is dat een onderbreking van een toon, gemaskeerd met ruis, tot een afname leidt van activiteit in de auditieve cortex. Omdat deze maskeringeffecten niet optraden indien de toon fysisch continue was, konden de geobserveerde effecten worden toegeschreven aan de opvulling van de interruptie met akoestische energie. Daarnaast werd de relatie beschreven tussen de subjectieve ervaring van de continuïteitsillusie en activiteitsniveau van de primaire auditieve cortex. Deze bevindingen laten zien dat 1) de akoestische energie tijdens de onderbreking van een toon al in de primaire auditieve gebieden verwerkt wordt en 2) de subjectieve ervaring van de continuïteitsillusie van een toon in de primaire auditieve gebieden gerepresenteerd wordt.

De electroencefalografie (EEG) studie die in *hoofdstuk vier* gepresenteerd wordt onderzocht de neuronale timing van de akoestische en perceptuele processen die in de fMRT studie gelokaliseerd werden. Door gebruik te maken van de taak en stimuli van de fMRT studie, werden de volgende bevindingen gedaan: 1) De verhoogde akoestische energie tijdens de onderbreking van een toon leidt tot een verminderde EEG activiteit 130 ms na het begin van de onderbreking en 2) de continuïteitsillusie onderdrukt de oscillaties in het EEG signaal gedurende en na deze onderbreking. Deze resultaten tonen aan dat cognitieve factoren zoals aandacht de sensorische verwerking van de akoestische energie beïnvloeden, vooral tijdens de onderbreking van het geluid. Daarnaast bieden de resultaten informatie over de timing van de corticale processen die betrokken zijn bij de akoestische analyse en de perceptuele interpretatie van geluiden die onderbroken worden door ruis.

De voorbereidende fMRT studie die in *hoofdstuk vijf* gepresenteerd wordt had het doel om de neuronale mechanismen voor continuïteitsillusies van natuurlijke, complexere geluiden te lokaliseren. Er werd gebruik gemaakt van stemgeluiden in combinatie met de methoden van de eerdere fMRT studie. Een bevinding van deze studie was dat het maskeren van de onderbroken stemmen door akoestische ruis de activiteit verminderde in auditieve corticale gebieden waarvan vermoed wordt dat ze betrokken zijn bij de verwerking van intacte stemmen. De subjectieve ervaring van de continuïteitsillusie van een stemgeluid correspondeerde met de activiteit van de primaire auditieve cortex. Deze bevinding was in overeenstemming met de resultaten van de eerdere fMRT studie waarin illusies van tonen onderzocht werden. Deze voorbereidende resultaten suggereren dat 1) specifieke auditieve hersengebieden die gevoelig zijn voor stemgeluiden ook de akoestische energie tijdens de interruptie van een stem verwerken en 2) continuïteitsillusies van stemmen in primaire auditieve gebieden verwerkt worden die ook betrokken zijn bij het invullen van alleenstaande frequentie componenten.

Een psychoakoestische studie die in *hoofdstuk zes* gepresenteerd wordt onderzocht in welke mate de continuïteitsillusie van contextuele factoren afhangt. De resultaten onthullen dat illusies van de dezelfde stimulus gemoduleerd kunnen worden door de geluidsssterkte en de waargenomen continuïteit van voorafgaande geluiden. Deze context effecten waren van niet-sensorische aard omdat deze niet aan het geluidsspectrum van de voorafgaande geluiden gerelateerd waren. Deze resultaten tonen dat het mechanisme dat aan de continuïteitsillusie ten grondslag ligt, van niet-statische aard is. Deze flexibiliteit is waarschijnlijk nuttig bij het aanpassen van geluidswaarneming aan verschillende akoestische omgevingen.

Samengevat laten de bevindingen van deze studies zien dat de continue geluidswaarneming van een verstoorde akoestische input niet enkel is gebaseerd op statische akoestische mechanismen in het brein, maar ook op actieve en adaptieve perceptuele mechanismen. Primaire auditieve gebieden spelen daarbij een belangrijke rol omdat ze de geluidsenergie waarop de continuïteit illusie gebaseerd is verwerken tijdens een onderbreking van een geluid. Deze vroege sensorisch - perceptuele processen worden verder beïnvloedt door aandacht en mogelijk ook door de sensorische context. Op basis van de bevindingen van deze en andere

studies wordt in *hoofdstuk zeven* een neuronaal model voor de continuïteitsillusie voorgesteld. Toekomstig onderzoek kan zich richten op de implementatie van de voorgestelde mechanismen in kunstmatige hoorapparaten. Daarbij zou de filter die in hoofdstuk twee voorgesteld werd als een basale verwerkingseenheid kunnen gebruikt worden.

Acknowledgements

Writing this book would not have been possible without the help and the friendship of several people whom I would like to thank in this section. First of all, I am grateful to my supervisors Elia Formisano and Rainer Goebel. Elia, your passion for scientific research evoked my interest in this field already long before you gave me the opportunity of working on this project with you. You taught me a vast number of theoretical, practical, and formal aspects that turned out to be essential for my scientific career, including the development of my own ideas and research interests. Thank you for sharing your expertise and for giving me such an inspiring and rewarding time, also after office hours. Rainer, you strengthened my belief in my work and you opened new paths of thinking for me. You introduced me to your creative views on neural filtering and to several other exciting topics such as artificial intelligence and meditation. Thank you for your optimism and for your generosity that clearly went beyond this project.

I further would like to thank a number of colleagues who helped me with setting up and conducting the studies presented in this book: John, you shared your sharp ideas on study design and you formed the basis for our 'illusory' filter model. Peter, your critical comments taught me how beneficial it can be to stay patient and to be very precise when designing and interpreting studies. Milene, you introduced me to the realm of EEG and you gave me invaluable advice on study design, data acquisition, and analysis. Gerard, your expertise and critical advice on statistics were essential for applying proper analyses to my data. Armin, your professionalism and patience often helped me to understand and solve my problems with fMRI data analyses. Anke, you assisted me a lot with EEG measurements and your enthusiastic efforts contributed greatly to our voice study. Claudia, you helped me with conducting our adaptation study with impressive accuracy and reliability, and you gave me excellent comments on this book. Danny, your clever ideas strongly contributed to the study design and the writings presented in this book; I'm sorry that you left Maastricht. Nick, you supported me with segmenting MR images when I was close to going mad from this task. Aline, Michelle, Christianne, and Claudia (Simons), I appreciate your assistance with my EEG measurements. I also acknowledge the efforts of all the volunteers who participated in my experiments and the technical support from the IT and Instrumentation staffs of the Faculty of Psychology and Neuroscience. In particular, I would like to thank Jordi Bruijnzeels, Sven Gijzen, and Johan Gielissen for helping me out with several technical issues.

Annemie and Riny, you took care of countless formalities that I had to arrange in the context of this project – without your support I would have been lost many times. I appreciate your professionalism and your friendliness which constitute the foundation of the great functioning of the Department of Cognitive Neuroscience. In this regard I also would like to thank all actual and past members of the Department (especially

those from the Audition and Vision groups) whom I did not mention yet; you all contributed in some or other way to an open and highly stimulating working environment.

Some colleagues also became part of my life outside the office and made me feel comfortable living in Maastricht. Henk and Berni, thank you for the relaxed musical evenings at your place. I am looking forward to continuing our long conversations about science, religion, and punk music. Joel, your suggestions helped me a lot with organizing my thesis and promotion; and Judith, you prepared delicious dinners and you gave me the right haircut. I am grateful to both of you for your helpfulness and for being such pleasant minds. Vera, we have been working together now for eight years and I feel sorry that you will leave Maastricht soon. Thank you for sharing this time and for being a faithful and understanding friend. Federico, my (former) roomie, thank you for going together through the ups and downs of PhD student life and for arranging many tasty events. I really enjoyed our evenings at home, especially when Giancarlo, Fabrizio, and Tommaso were also there. Noel, thank you for all your help (especially with the Dutch summary of this book) and for being such a relaxed roomie at work. Our guitar jams and our mutual respect for misbehaviour in office, in hotel rooms, and in many other places mean something special to me. Nico, Martin, Emanuel, and Tobias, my old posse from Aachen Rock City, thank you for the long nights that we spent with live music, barbecue, beverages, and more or less meaningful discussions about science or anything. You guys are awesome friends and you kept my mind in balance, especially when I felt overloaded by work.

I am particularly grateful for the support from my family. Margarete and Rolf, my dear parents, without you nothing would have been possible. I deeply appreciate your care, trust, and patience. Katja and Bernd, thank you for believing in me as a scientist and for being the sweetest and most caring older sister and brother-in-law (respectively) that I can imagine. Betty, my love, you were there whenever I needed you and your inexhaustible helpfulness was essential for finishing this book. I am thankful for your love and trust and everything you gave me, especially during the times where I was occupied with writing. I feel deeply honored being your partner and I am looking forward to us moving on to new paths.

Curriculum vitae

Lars Riecke was born on February 5th 1976 in Osnabrück, Germany. In 1995 he graduated from high school at the Kardinal-von-Galen Gymnasium in Mettingen. In 1996 he enrolled at the Faculty of Engineering at Rheinisch-Westfälische Universität Aachen. He obtained his bachelor's degree (Vordiplom) in Electrical Engineering three years later. In 2000 he registered at the Faculty of Psychology at Maastricht University, where he also worked as a teaching assistant. From 2003–2004 he studied at the Cognitive Science department at University of California San Diego. In 2004 he obtained his master's degree in Psychology (*cum laude*) at Maastricht University with a specialization in Cognitive Neuroscience. He worked on his PhD project at the Cognitive Neuroscience department at Maastricht University from 2005–2009, where he currently continues his works as a postdoctoral fellow.